REGRESSION MODELLING OF MORTALITY

A thesis submitted for the degree of

MASTER OF STATISTICS

at The Australian National University

Timothy Sean HIGGINS

February 2002

DECLARATION OF AUTHENTICITY

This thesis is my own work and all sources used have been acknowledged.

Timothy Sean Higgins

February 2002

ACKNOWLEDGEMENTS

Above all, I am grateful for the help and support of my supervisor, Professor Christopher Heathcote, during the preparation of this thesis. He not only offered invaluable suggestions regarding structure and content, but also provided the initial encouragement to pursue this topic. I am particularly thankful for his encouragement to continue and complete the thesis after taking an extended period of time off.

I am also grateful to Professor Terry O'Neill and Dr Steven Stern, for sympathising with my outside work commitments, and for allowing me numerous extensions in order to complete the thesis.

I would also like to thank Dr Ewa Tabeau and the Netherlands Interdisciplinary Demographic Institute (NIDI) for providing the mortality data used in the thesis.

Finally, I want to thank my wife Salli for putting up with my occasional bad moods during the more stressful days, and for allowing me to hog the computer.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SUMMARY

The populations of developed countries are ageing due to a decrease in fertility rates and mortality rates. In addition, the developed world is faced with the imminent retirement of a large cohort of individuals born post-WWII known as the 'baby boomers'. As the ratio of the dependent population to the labour force rises, pressure will increase on the solvency of social security and health care systems, and the dynamics of society will change. The future demographic structure of the elderly population will be dependent to a large degree on future mortality rates. Accurate forecasts require accurate models for the processes of ageing and death.

The thesis develops the regression approach of Heathcote and Higgins (2001a, 2001b). A regression model for mortality is developed that is founded on the assumptions of independent lives within and between cohorts, and identically distributed risk of death for individuals within a cohort. Under these assumptions, it can be shown that the errors for a regression model of a mortality surface are asymptotically normally distributed with zero means and a fixed variance that depends on the number of survivors. An advantage of the regression approach is that it is founded on the natural pattern of human mortality along cohorts, unlike traditional cross-sectional techniques of modelling mortality.

A comprehensive analysis of historical trends in Dutch mortality is undertaken to aid in the determination of appropriate regression models for the Netherlands. Four models are produced using weighted least squares, two for each sex; years 1890 to 1938, ages 50 to 90 as one model; years 1946 to 1991, ages 50 to 90 as another. The covariates selected include a polynomial in age, a linear term in year, and interactions between age and year, among others.

Extra-binomial variation is identified, the source of which is speculated upon in the context of selection and debilitation resulting from heterogeneity within cohorts. A particular advantage of the formulation of the stochastic structure is that modifications to the assumption of identical distributions within cohorts may allow heterogeneity of hazard rates to be considered in the future. In this instance, however, the source of the extra-

binomial variance cannot be ascertained due to the lack of disaggregated data, and a mixed model with random effects is developed to account for the unobserved covariates that are responsible for the extra-binomial variation. The extra variance is modelled as an exponential function of age and year, and maximum likelihood methods are used to estimate new parameters for both the mean and the additional variance component. The information matrix from the maximum likelihood fit allows standard errors to be produced, thereby enabling confidence and prediction intervals to be placed on the fitted surface.

A number of interesting features are apparent from analysis of the derivatives of the Dutch male and female fitted surfaces. In particular, the age at which the rate of increase in female mortality peaks appears to be shifting to higher ages in a linear fashion over time. Exploration of the residuals and the derivatives from the male model suggest that a strong cohort effect may be present. A disadvantage of the parsimonious regression models developed is that cohort effects are not explicitly modeled. For comparison with the regression models, fixed effects models (based on the techniques of Wilmoth (1990) and Lee and Carter (1992)) are developed where cohort effects are modelled explicitly.

Forecasts to 2030 are produced by extrapolation of the regression models. It is shown how expert opinions may be incorporated by altering the parameter values. The regression models are modified based on observations of historic trends and alternative projections are presented. Cohort life expectancies and conditional probabilities of survival are generated. Finally, it is shown that relatively small uncertainty in mortality projections can have a dramatic influence on the future structure of the population in the Netherlands, providing further justification for the importance of mortality modelling and forecasting.

1 Introduction

1.1 Motivation for the Study of Mortality

In many ways, the future of humanity is determined by the demographic make-up of its population. Consumption of resources, the environmental impact of human actions, technological progress, and the general standard of living are all affected by demographic changes. In the developed world, fertility rates have been steadily decreasing along with mortality rates. This has led to an increase in the dependency ratio of most developed countries, being the ratio of the population aged less than 15 and greater than or equal to 65 to the population aged between 15 and 64:

"fewer people of productive age will increasingly have to provide for an expanding number of elderly dependants" (World Health Organisation,1997).

While this begs the question of what constitutes a "productive age" given improvements in general health, it is clear that an ageing population potentially implies major social change.

Due to a dramatic temporary increase in birth rates following World War II, much of the developed world is experiencing the ageing of a large cohort of individuals known as the "baby-boomers". The fact that this cohort is quickly approaching retirement age has implications to the short-term and long-term maintenance of social support systems, including the funding of social security schemes, private and state insurance schemes, health care insurance and provision, and the changing demand for goods and services (for a discussion of the financial implications of the ageing population see, for example, Denton and Spencer,1996a; Manton et al., 1993; Jeanneret,1983; Rice,1983; Thorburn,1999).

More than any other demographic variable, mortality should have a significant influence on policy decisions for the next fifty years. Although low fertility rates need to be addressed in order to guarantee the existence of future generations of tax paying citizens, the baby-boomers will start to retire within the next ten years, causing a dramatic change in the dependency ratio. The extent of the change will be determined to a large degree by the state

of elderly mortality. For example, it has been shown that reductions in mortality rates at ages 60 to 80, have a greater influence on reducing life expectancy at birth than at younger ages: *"...there is considerable potential for saving years of life and increasing life expectancy by reducing mortality in old age, more than is generally realized"* (Vaupel, 1986).

Implications for social security and public health care depend also on morbidity. If the average age to disability increases along with life expectancy, then effective costs to social systems may not rise substantially. Whether compression of morbidity will occur is an important question, though it is not discussed in this thesis (for a discussion of projections and analysis of morbidity see Smith, 1993; Mathers, 1990; Van Den Berg Jeths et al., 2001)

In order to assess the impact of mortality on the future demographic structure, there is a growing need to produce accurate forecasts of mortality rates in developed countries. It is logical that the most accurate forecasts require an understanding of the underlying process, including the factors that affect the process. Models of mortality ultimately aim to capture the underlying process, though in addition to this goal they have a range of important functions. In addition to aiding forecasts, models may allow us to better understand the biological basis of our own mortality, including illuminating the debate on the existence of a limit to the human lifespan. They may help uncover epidemiological relationships between risk factors thus leading to improvements in medical technology or public knowledge of health risks. They can facilitate comparison of mortality between countries. They can aid in making inferences from incomplete data and they can be used to construct life tables for use in insurance calculations.

The study of mortality and its processes is very much a multidisciplinary task. Demographers, economists, statisticians, actuaries, biologists, geneticists, physicists and geriatricians, have contributed to the field. For the remainder of this chapter, we review the general pattern of historic mortality in developed countries, theories of ageing and methods of modelling mortality, and finally forecasting techniques. Throughout, justification for and references to the regression approach taken in this thesis are given.

1.2 Twentieth Century Mortality in the Developed World.

Mortality in developed countries during the 20th century has seen a steady and dramatic decrease. The microbiology revolution lead to the creation of vaccines. Pasteurisation reduced contagion and dairy herds were tuberculin tested (Smith, 1993). An increase in living standards was brought on by better education, improved food-supplies, and public health improvements (Gage, 1994; Lopez, 1983). These factors, coupled with the introduction of antibiotics, lead to a decline in infectious and parasitic diseases. As a result, death in the later half of the 20th century and the early 21st century has become a more predictable event, where the major causes of death are due to degenerative or chronic conditions such as cancer and heart disease (Lopez,1983:85).

Another feature of mortality in developed countries during the 20th century has been the changing sex differential. Lopez (1983:86-88) shows that between 1900 and the late 1970s there was a continual increase in the differential between males and females for most developed countries, with a rapid divergence after the middle of the century (called the Scandinavian pattern by Lopez). Since World War II, with the emergence of the degenerative diseases as the leading causes of death, the longevity gains of females have consistently exceeded those of males, though the differential has narrowed more recently in many developed countries (see for example, Van Hoorn and De Beer, 2001).

A major task in demographic circles is finding the cause of the sex differential: is it due to genetic and biological differences affecting metabolism, ageing and susceptibility to particular chronic conditions; or is it due to sociological factors, such as diet and lifestyle ? It is generally agreed that a combination of genetic and environmental factors are responsible. Biology certainly plays a significant role given that in most regions of the world, irrespective of dietary habits, female mortality is lower than male mortality (Tickle,1997; Lopez,1983). Lopez (1983) contains a wealth of information on the pattern and possible causes of the mortality differential including discussions on nutrition, lifestyle, and biological differences.

The incidence of mortality from cardiovascular disease, cancer and motor vehicle accidents has been higher in males. Degenerative respiratory diseases beyond age 65 are also responsible for the increasing differential post-WWII (Lopez,1983:116). It is speculated (see, for example, Ruzicka and Lopez,1983:486), that the changing roles of women in the developed world, which sees them participating more equally to men in the workforce, may lead to the adoption of some of the unhealthy habits typically associated with men, such as poorer nutrition, and increased stress from work leading to alcohol and cigarette consumption. On the other hand, it is possible that lifestyle factors may continue to more adversely affect males than females, and that biological factors may advance female longevity faster than that of males.

The age pattern of mortality over the last century has varied by country, however, there are a number of common features: Mortality at birth is relatively high, though has dropped considerably over the century due to improvements in medical care; the highest probability of survival is around ages 8 to 12; mortality increases rapidly to age 18, where there often is a 'hump' associated with automobile accidents and suicide (particularly among males); and there is an exponential increase in mortality from age 30 associated with senescence. (For a comprehensive description of historic mortality patterns see Jain (1994), Lilienfeld (1976) and Tuljapurkar and Boe (1998)). At the top of the age range an interesting phenomenon has been noticed whereby the rate of increase in mortality appears to level out and then decrease, diverting from the exponential pattern. This has been studied in detail by Coale (1996), Horiuchi and Coale (1990), and Horiuchi and Wilmoth (1998). Tuljapurkar and Boe (1998) list the recent literature reporting decelerations in the rate of mortality increase at high ages, and state: "*...the nature of mortality at ages over 80 is now moving to center stage in discussions of future mortality change...*". Causes of this phenomenon have been speculated on, and are discussed in detail in the next section.

In Chapter 3 historic Dutch male and female data are decomposed by age, period and cohort, and trends are examined which follow the general pattern of decreasing mortality and changing sex differential that have been discussed here.

1.3 Mortality Models and Theories of Ageing

As mortality rates have stabilised over the past century, modelling of mortality has become more practical. The ultimate goal of most researchers is to explain the age pattern in mortality, though some concentrate more on description than explanation. Demographers and actuaries have traditionally focussed on modelling empirical mortality using parametric models of the risk of mortality without incorporating biological or other assumptions. Evolutionary biologists have primarily attempted to explain the mechanisms underlying senescence in individuals without reference to empirical population data (Carnes et al., 1996).

The approach developed in this thesis is primarily focused on development of a parametric model over age and time; however, the model is founded on the process of ageing from an individual and cohort perspective.

1.3.1 Biological Theories of Ageing

The development of any parametric models should consider current theories regarding senescence, particularly since the choice of function may be dictated by opinions on the existence of a limit to the human lifespan. Theories on the age pattern in mortality include the evolutionary theory of senescence and hypotheses about genetic processes that determine ageing. (Tuljapurkar and Boe, 1998). A critical aspect of the debate is whether or not a biological limit exists to human life expectancy (see for example, Manton, 1993; Warner, 1993).

In the evolutionary model species are assumed to have a pre-determined upper age limit. The argument is based on the belief that natural selection operates to maintain the ability of a species to reproduce, and that the integrity of the organism will naturally diminish after the reproductive cycle is complete. (Olshansky and Carnes, 1997). Under this hypothesis, if existing diseases are cured, new or infrequent diseases or disorders would emerge to restrict the lifespan. In other words, although current mortality may be reduced in the future, a limit would exist based on a biological law of diminishing returns. Despite advocating the

evolutionary theory of senescence, many researchers recognise that biologically based limits to life expectancy may be manipulated if medical technology, including genetic engineering, affect enough of the population (see, for example Olshansky and Carnes,1997; Warner, 1993; Smith,1993).

Those who advocate the hypothesis about genetic processes believe that the primary causes of senescence are due to predisposing risk factors. If the risk factors are reduced, whether this be through genetic engineering or other means, life expectancy will increase.

The evidence regarding a possible limit on human life is inconclusive. Coale (1996) examined mortality patterns for countries with high expectation of life, in particular, Sweden, Japan, Spain and Austria, and found that age patterns for females have become very similar. He suggests that mortality may not continue to decline once female life-expectancy reaches between 84 and 85 years. In contrast, Vaupel et al.(1994) review the trends in mortality for the oldest populations in 27 countries and they find no evidence of convergence in death rates over time. They present additional evidence to suggest that improvements in mortality for the very old may continue.

1.3.2 Parametric Models of Mortality

The hazard function, which is expressed mathematically in Chapter 2, is the instantaneous probability of death for an individual. Developing a parametric hazard model involves regressing the individual risk of dying on a set of explanatory variables. The most influential parametric mortality model in the literature is that proposed by Benjamin Gompertz. Gompertz (1825) recognised that a geometrical pattern in age captured the behaviour of human mortality for large portions of the life table. He proposed the function:

$$\mu(x) = \exp(\theta x)$$

where $\mu(x)$ is the hazard rate at age x, and θ is an unknown parameter. The close fit of the Gompertz function to empirical data suggested that a "law of mortality" may exist to explain the age patterns at death for individuals and populations. (Willekins (2001) shows

that the Gompertz function is a type of extreme value distribution, and argues that the application of extreme value theory may uncover new theories of ageing.)

In 1867 Makeham noticed that the Gompertz equation failed to capture the behaviour of mortality at higher ages and added a constant term in order correct for this deficiency. Makeham developed a theory of partial forces of mortality, where he argued that mortality schedules for specific diseases fit the Gompertz law more closely than that of aggregate mortality (Olshansky and Carnes, 1997).

The Gompertz curve was further modified by Perks, Beard and others. In 1932 Perks proposed modifications to the Gompertz to allow the curve to more closely approximate the slower rate of increase in mortality at older ages. In 1963 Beard developed a model based on the recognition that the changing slope may be due to heterogeneity in the mortality risks (Manton et al, 1993:24,25). The implication of heterogeneity to modelling mortality is discussed later in this chapter.

The parametric models discussed so far are constrained to modelling the portion of the mortality curve associated with senescence. A number of parametric models have been developed to capture the entire life-span. Since mortality varies in a non-linear fashion with age, particularly during childhood and adolescence, such models are necessarily more complex. In 1872 Thiele proposed a model of the force of mortality as a function of three components, one relating to childhood, middle life, and adult ages. Wittstein in 1883 proposed a law modelling the initial rate of mortality as a function of two components, representing childhood and adult age (Hartmann,1987). More recently Heligman and Pollard (1980) proposed an eight parameter model that captures the shape of the mortality curve over the entire age range. Like the Gompertz, their model falters at the old ages. Another model that is amenable to regression techniques is the Siler competing hazard model, which emphasises causes of death (Gage,1994).

Tabeau (2001) provides a thorough review of parametric mortality models ranging from De Moivre in 1725 to a model proposed by Rogers and Little in 1993. Among parametric

models are polynomial functions, which are commonly used in modelling due to their ability to approximate most functions.

A distinguishing feature of parametric hazard models is the assumption concerning the shape of the risk function and how it changes over time. Rather than modelling the hazard function directly, one could model a function of the hazard such as the initial or central rates of mortality which are easier to estimate from empirical data. Possible shapes include the logit or proportional odds model (see, for example, Salhi et al., 1995). A constraint in parametric models is the a priori specifications of the hazard function. To reduce model mispecifications, in Chapter 3 we explore the Dutch data in detail and show how certain polynomials and other parametric models, including the Perks, provide close fits to the empirical Dutch mortality along individual periods. In Chapter 4, we develop a parametric logit model for Dutch mortality that includes polynomials in age as covariates.

When a suitable parametric surface is lacking, non-parametric techniques can be used to model the force of mortality: "*...sometimes it is more realistic to suggest something about the qualitative features of the unknown function than to determine it.*" (Stefanyuk,1994:54).

1.3.3 Heterogeneity and Stochastic Process Models

Although modelling the processes of mortality should be founded on biological theories, modelling the age-pattern of mortality using parametric models of the hazard function is usually undertaken without regard for biological assumptions. Most parametric models do not attempt to model the mechanisms of ageing and, therefore, are of little practical use outside of testing simple statistical hypotheses about the importance of certain covariates (Yashin,2001).

A common criticism of traditional parametric models is that they assume that the population's risk of mortality is homogeneous, when in reality, individual risk of mortality will not only vary due to genetic pre-disposition, but will change over time due to the impact of environmental factors (Yashin et al.,1985; Yashin, 2001). It has long been recognised that aggregate population mortality reflects the compositional dynamics within

cohorts, and that individual age-specific rates may differ substantially from aggregate rates (In 1963 Beard developed a model that accounted for heterogeneity based on this recognition (Manton et al, 1993:24,25)). Different aggregate patterns arise due to selection from varying probabilities of death for individuals within a cohort: when the more frail members of a cohort perish, the more robust individuals survive thus leading to a reduction in cohort mortality rates. In particular, even though subcohorts may follow a Gompertz distribution, if each subcohort possesses a different level of frailty, then the mortality trajectory of the whole may be substantially non-Gompertzian.

The non-Gompertzian shape at the top of the mortality curve apparent in empirical data from developed countries may reflect selection due to cohort heterogeneity to some extent. However, it may also reflect actual individual mortality patterns, where human interventions involving medical technology have reduced the biologically determined rates for the oldest parts of the mortality curve. On the other hand, perhaps mortality is intrinsically non-Gompertzian, and the decrease in the mortality slope at high ages has a, as yet, undiscovered physiological explanation ? Support for this view comes from Carey (1997), who examined medfly mortality with the purpose of developing insights into human mortality. He found that the leveling-off of mortality in medfly populations is not an artifact of changes in cohort composition, which implies that the reduction in the rate of increase in mortality at high ages may have a biological interpretation, rather than being the outcome of selection from cohort composition.

Manton et al. (1986), Vaupel et al.(1979) and Vaupel and Yashin (1985), explore how the observed rate of death for an entire cohort of individuals relates to the probability of death for each of the individuals within the cohort. An implication of heterogeneous mortality risk is that population life tables based on cohort mortality may give an inflated estimate of an individual's life expectancy. Ignoring variability in frailty may result in biased estimates (Vaupel et al.1979; Vaupel and Yashin, 1985). Vaupel and Yashin give the additional warning (1985): "*...because patterns at the individual level may be simpler than composite population patterns, both theoretical and empirical research may be unnecessarily complicated by failure to recognize the effects of heterogeneity*".

The recognition of the impact of heterogeneity on aggregate mortality and the dynamic nature of human experience, has lead to a recent push to incorporate biological mechanisms and stochastic processes into mortality models (see Yashin et al., 1985; Yashin, 2001; Vaupel et al., 1988). Yashin (2001) reviews modern mortality models that are based on stochastic differential equations, and explains their origins in the context of empirical mortality patterns and the physiological theories of ageing and deterioration. He strongly advocates the development of complex models: "*... models of changing frailty, and in particular the stochastic process models of mortality, allow an explicit description of the physiological mechanisms of aging* " (Yashin, 2001:261).

A major restriction in the development of such models is that individual differences in the mortality process are often due to unobserved variables, and that, "*...one must be concerned with [the unobserved variables] effects in order to uncover the underlying relationship between the force of mortality and the variables of interest.*" (Yashin et al. 1985:169).

Unobserved covariates can be included in statistical models through the introduction of random effects (Yashin, 2001). The impact of unobserved covariates may be inferred to some extent from analysis of the variance of the model under the assumption of homogeneity. In particular, under the assumption of homogeneity of risk within a cohort, it is simple to show (see Chapter 2) that the various measures of mortality, including the number of deaths, are binomially distributed. Pollard (1970) analysed Australian data and found that the estimated variance exceeded the expected binomial variance, particularly for older ages. He acknowledged that the specific causes of the excess variation cannot be determined unless additional covariates are available. The approach adopted for the models developed in this thesis involves inclusion of random effects for unobserved variables. In Chapter 2 we develop the theory behind a mixed regression model with random effects, and in Chapter 4 we develop the mixed regression model for the Dutch data. A deficiency with this approach is that the unobserved covariates may distort the observed relationship of fixed effects within the model. However, in the absence of knowing the origins of extra vaiability, it is appropriate to include random effects rather than trying to incorporate

alternative distributional assumptions for mortality risk which may or may not reflect reality.

Ideally, extra sources of variation, such as frailty, should be determined as a function of observable variables, be they behavioural, genetic or social (Thiltges et al., 1995:25). However, a basic question is whether frailty is a purely 'unobserved' characteristic, or whether it can be related to some observable characteristics (Tuljapurkar and Boe, 1998). Manton et al. (1986) focus on individual characteristics and relate mortality to individual risk factors. Salhi et al. (1995) consider covariates such as marital status, occupation, education, and place of birth, among others, and assume that an individual's risk of death is affected by a set of conditions acting in concert rather than individually. Dynamic stochastic models are ideally required to capture the complexity of causative factors and allow for heterogeneous risk, however, models of this sort are in their infancy and have been difficult to develop due to the lack of adequate longitudinal data. As stated by Willekins (2001), and acknowledged by many in the field (see Tabeau, 2001:27):

"*Although process models are advocated as being superior [to trend models], the knowledge base that is required to model the causal mechanisms implicit in process models remains inadequate*" (Willekins,2001:122).

1.3.4 Incorporating Cohort Effects

Modelling mortality is further confounded by the problem of separation of age, period and cohort effects. Since age is equal to period (year) minus birth cohort, the parameters for a model that contain all three terms cannot be uniquely determined. The importance of cohort effects in mortality analysis was first introduced by Andvord in an attempt to explain the changing pattern in deaths from tuberculosis. The methods were elaborated by Frost who noted a possible causative link between infection early in life and high rates of mortality in older life (Lilienfeld, 1976:100-104). That extraordinary events can influence the long-term survival pattern of a cohort is well-known (Ryder, 1980). Since individuals within a cohort have different frailties, an event such as an epidemic or war may result in debilitation of certain subpopulations, leading to selection which may cause a decline in the aggregate

mortality rates for those surviving the event. On the other hand, a debilitating event may lower the constitution of survivors, thus increasing long-term mortality rates. Stochastic differential equations have been developed to clarify the interaction between selection, debilitation, ageing and recuperation. Vaupel et al. (1988) show that debilitating events result in selection, which alters the population heterogeneity, thereby influencing the impact of the debilitating event. In Chapter 6 patterns in Dutch male mortality as a consequence of the Spanish Influenza epidemic of 1918 are discussed in the context of the observations of Vaupel et al. (1988).

Cohorts are crucial in the development of mortality models. Development of models or life tables has often relied on period data, despite that consecutive survival experience follows cohorts. When selecting models for a mortality surface, it is therefore, important to consider the natural progression of an individual's mortality. The cohort approach has been impractical in the past because of the extended length of time necessary to collect data from a single cohort (Ryder, 1980).

A type of model called Age-Period-Cohort (APC) exists based on time series techniques where age, period and cohort effects are considered simultaneously. A linear APC model can be written as a model for log-rates in which age, period and cohort effects combine additively (Tabeau, 2001). Linear dependence among age, period and cohort means that the effects cannot be solved uniquely unless constraints are placed on the model. Wilmoth (1990) modifies the linear APC model by incorporating interactions between age and year. He overcomes the identification problem by only fitting age and year terms and then looks for cohort effects within the residuals. For comparison with the regression models developed in Chapter 4, in Chapter 6 we apply Wilmoth's approach to the Dutch data. Tabeau (2001) mentions another method of modelling known as the non-overlapping cohort approach that relies on doubly classified data; deaths are classified according to birth cohort as well as year of death and age. Although this approach was not pursued in this thesis, it could be undertaken for the Dutch data, and is a topic worthy of future exploration. Time series methods are applied to mortality modelling for forecasting purposes, and so are discussed in the next section.

The fundamental question in all theories of mortality is: "*...to what extent is mortality 'intrinsic' to an individual, and to what extent is it a response to 'environmental' factors?*" (Tuljapurkar and Boe, 1998).

Since the longevity of an individual depends on both genetics and environment, a theory that concentrates on only one aspect will be insufficient. Vaupel et al. (1994:805) conclude: "*The current evidence... suggests a new paradigm of aging that recognizes diverse and often highly plastic aging processes that can be influenced by health interventions, behavioral changes, and environmental improvements and that depend on genetic differences both between and within species*".

The confounding of cohort and period effects, coupled with cohort heterogeneity leading to selection, and the complex and unknown interactions between genetics and the environment, means that separating exogenous and endogenous causes of death is not practical. Dynamic stochastic models of mortality are necessary to accommodate the complexities but detailed longitudinal data, which are currently lacking, are required to formulate and test models (Tabeau et al., 2001). The best environments for such studies would be controlled environments, such as for certain animal populations, where the effects of medical intervention and extrinsic causes of death are removed.

In Chapter 4 a parametric model is chosen and fit to the Dutch male and female mortality data. In the absence of an established theoretical basis upon which to generate a parametric model, we have taken the approach of choosing a function that best fits the empirical data. Rather than advocating a particular law of mortality, we take the view that it is better to follow the data, rather than trying to fit a preconceived function. Unlike many parametric models of mortality, the regression model developed is founded in a clear and transparent way on a process of individual mortality. Although the process underlying the regression model is a simplified view of reality, it could be modified in future work in order to accommodate realistic cohort assumptions such as variable frailty among sub-groups.

1.4 Forecasting Mortality

Despite the recognition of the need for dynamic stochastic models for explaining the process of ageing in human populations, most forecasting procedures are often simple extrapolations from parametric models that do not take biological processes or complexities such as heterogeneous risk into account. Quoting Manton et al. (1993:25): "*The failure to deal with individual aging trajectories, and their cohort mix, makes it difficult to use epidemiological and biomedical evidence on the impact of health changes on the organism in forecasts.*"

As stated in the previous section, this failure is due to the fact that relatively little progress has been made in the development of stochastic process models for mortality due to a lack of empirical longitudinal data. Nevertheless, the ageing population and the strains on social security systems mean that plausible forecasts are necessary, and a variety of methods have been developed and used that are summarised in this section. For a comprehensive review of forecasting models for mortality, see Tabeau (2001), Land (1986), and Tuljapurkar and Boe (1998).

1.4.1 Extrapolating from Fitted Curves

Objections to curve fitting techniques for forecasting purpose are discussed by Alho (1990:523), who notes that: "*... several curves may fit the past data very well but give widely different forecasts.*"

When generating a model for forecasting, a balance must be reached between parsimony and the level of disaggregated detail to use in projections (Tuljapurkar and Boe, 1998). Keyfitz (1982) advocates that the simpler the curve, the more realistic the forecast. When choosing a model, the structure should enable forecasting to proceed: "*Forecasting supposes that there are typical directions of movement through time, and the formulation of the mathematical functions or the model tables must be such that the progression is readily described and projected.*" (Keyfitz,1982:350).

The development of regression models in Chapter 4 follows the principle of parsimony, and a form is chosen such that the model can be used for predictions which are developed in Chapter 7.

Olshansky and Carnes (1997:25), who advocate the evolutionary theory of senescence, believe that extrapolations from parametric models or projections from time series based models may produce implausible forecasts: "*... some demographic models of human senescence produce mortality schedules that appear to be inconsistent with predictions about human mortality from evolutionary theory*".

Their argument only holds if the basic rate of senescence remains fixed. Altering the rate of senescence, however, may indeed be one of the main consequences of the genetic revolution. Whether or not one subscribes to the evolutionary theory of senescence, a general criticism of extrapolation from a parametric curve is that no matter how close the fit is to empirical data within a particular domain, it does not provide a guide to accuracy outside that domain. Therefore, determining the extent of uncertainty from a parametric forecast is a subjective enterprise. Criticisms like those by Olshansky and Carnes can be addressed by setting a particular target level for future mortality rates (or life expectancy) based on expert opinion, and then interpolating to fit an appropriate forecast.

Pollard (1987) gives a comprehensive summary of methods of projecting age-specific mortality rates, including projection by extrapolation of age-specific rates; projection with reference to standard tables; projection by reference to "laws of mortality" (such as the Gompertz or Makeham); and projection by cause of death, among others. Perhaps the most intuitive and easiest to follow method of projecting mortality is to extrapolate age-specific mortality rates from historic trends. Age-specific rates of decline, rather than an average rate of decline across all ages, are favoured since at the present, changes at older ages predominantly influence future life expectancy (Vaupel,1986), and since there is significant historical variability in the rates of change of mortality across ages. Extrapolations also suffer from the subjectivity associated with selecting an appropriate period from which to extrapolate. Tuljapurkar and Boe (1998) review forecasts based on extrapolations and conclude that the most recent patterns in mortality decline do not always provide the best

basis for short or long term projections. Age-specific mortality rates are generated in Chapter 7 for Dutch males and females. Forecasts using these methods are deterministic; they do not assume a stochastic structure for the errors, or recognise the expected propagation of errors as time increases due to uncertainty in the process.

1.4.2 Time Series Methods

Lee and Carter (1992) developed a model that includes a time factor modulated by an age-response as well as a separate age effect. The time factor is treated as a stochastic process, and time series techniques are used to model the factor over time, thus allowing stochastic projection intervals. Although the model performs well on tests of data, and as a result, has been adopted by many forecasters, it doesn't take into account the natural emergence of mortality along cohorts, and as such is not modifiable to accommodate theories of ageing or heterogeneity. Like other techniques based on extrapolation, projections follow from the subjective decision on the length of the historic period used to fit the time series model. In addition, projections are dependent on the form of the time series model selected.

As stated by McNown in a comment on Lee and Carter's paper:"*...forecasts of mortality identical to Lee and Carter's will be produced by directly projecting each age-specific mortality rate at its own historical rate of exponential decline*" (McNown,1992). This is the case in the particular example of Lee and Carter's projection of US mortality since the preferred time series model that fits the data is linear. In our view, the main advantage of the Lee and Carter approach is its ability to generate stochastic prediction intervals based on time series methods. For comparison with, and as an alternative to the regression model, the Lee-Carter model is briefly discussed and modifications are suggested to accommodate cohort effects in Chapters 6 and 7.

Another method that utilises time-series techniques is that of McNown and Rogers (1989). They apply Heligman-Pollard models to individual periods, and then model and forecast the parameters as a multivariate time series. An appeal of this approach is that by parameterisation, they attempt to incorporate an age theory of mortality into the forecasting model. Bell and Monsell (1991) use time series methods to forecast age-specific mortality

rates by principal components. As with the Lee and Carter method both methods do not take the natural cohort structure of mortality into account.

1.4.3 Projecting by Cause of Death

Since many trends in aggregate mortality can be explained by the incidence of specific diseases, it follows that both descriptions of historic mortality and forecasts may be improved by considering cause of death. Improved forecasts are predicated, however, on the assumption that sufficient data and epidemiological knowledge are available that suggests the existence of a stable predictive relationship (Tuljapurkar and Boe, 1998). The task is complicated by misclassification of cause of death data, and dependence between causes of death (competing risks). Alho and Spencer (1990a) review the evidence and conclude that the major contribution of forecasting mortality by cause has been the ability to easily incorporate expert opinion, rather than improved data analysis or forecasts. A potential extension of the regression models developed in this thesis is to cause of death data, with the inclusion of correlations between surfaces in order to accommodate competing risks. In addition to cause of death there is evidence for correlations between mortality and sex, marital status, education and other socioeconomic variables, and racial and ethnic differences (Tuljapurkar and Boe, 1998).

1.4.4 Uncertainty in Mortality Forecasts

Tuljapurkar and Boe (1998) summarise the debate on uncertainty in mortality forecasts, and argue for an explicit probabilistic (stochastic) approach, such as in Lee and Carter (1992) and Alho (1990) (see also Tuljapurkar and Lee, 2000). Uncertainty is embedded in a stochastic model through the choice of parameters and the error structure, which propagate in an additive way over time. Historic analysis serves an important purpose in that it enables quantification of uncertainty in mortality, however, the level of uncertainty is subjectively determined since forecasts are based on the selection of a particular period of historic data. For example, Lee and Carter (1992) use an indicator variable to account for the Influenza epidemic of 1918. Although this reduces the uncertainty in their time series

forecast, it can be argued that such events should be included in the generation of the error structure given that the outbreak of future epidemics cannot be predicted.

Formal methods of forecasting that rely heavily on an analysis of the past are often criticised for an over-reliance on historic patterns (Tulkjapurkar and Boe,1998), and an absence of expert judgement regarding the future. In all forecasts, and in particular those by cause of death, the impact of changing risk factors and public health practice must be considered. This begs the question of how does one predict these changes? Declines in mortality brought on by shifts in lifestyle may be predictable to some extent. For example, excess male mortality noted in many developed countries has been linked, to some extent, to consumption of cigarettes (Coale, 1996). When younger female cohorts that consist of heavy smokers reach old age, mortality rates may increase. Nusselder et al. (1997) attempt to explain the increase in Dutch mortality for the oldest old in the context of causes of death and suggest that a cohort effect for lung cancer among elderly men may be a factor. Van Den Berg Jeths et al. (2001) review epidemiological approaches to forecasting mortality. Van Genugten et al. (2001) analyse Dutch data and demonstrate substantial differences in forecasts when trend extrapolation of cause-specific mortality is compared with epidemiological projections based on risk factors.

In the presence of epidemiological and expert judgement regarding the future, the variance of stochastic forecasts based on historic data should be modified accordingly. Alho (1992) (see also Alho and Spencer (1985)) develops a formal model based on mixed estimation in linear regression that allows the incorporation of expert opinion in forecasting. For an example of the incorporation of expert opinion into mortality forecasts by cause of death using a different approach see Bulatoa and Stephens (1992). With regards to the state of future mortality, current expert opinion is optimistic in light of advances in biotechnology and the mapping of the human genome. Professor Burton of the Victorian Anti-Cancer Council, was recently quoted as saying:"*...We will make significant inroads... early in the next century through the early detection, through prevention of fatal cancers, like lung cancer and through an absolutely fabulous array of new treatments which are coming to us from the molecular genetic revolution*" (Jacobsen, 1998).

1.4.5 Impact of Uncertainty on Future Population

Besides forecast variability due to future uncertainty, forecasts are further hindered by uncertainty in current mortality estimates due to the absence of reliable data, particularly at old ages. Bennett and Olshansky (1996) show that there is an argument for adjustments to US mortality rates over age 95 and possibly between ages 65 and 94. They argue that even small adjustments can have a dramatic consequence to the long-term aged population structure.

An implication of the ageing population and increasing dependency ratio is a substantial strain on the social security systems of developed countries. One way of reducing the long-term financing problem faced by social security is to increase the eligibility age for social security benefits (see, for example, Burtless,1998). In light of increasing life expectancy, the need for re-assessment of eligibility ages is crucial: "*An increase in life spans, when the normal retirement age remains unchanged, is equivalent to a sizable increase in lifetime Social Security benefits* " (Burtless,1988).

In Chapter 7 we consider what is meant by 'old age' from the standpoint of social security eligibility. We use a series of criteria, developed by Denton and Spencer (1996b), to estimate an age today that corresponds to the pension eligibility age of 65 that was introduced in 1957 with the establishment of the General Old Age Benefit Act in the Netherlands (van Oorschot,1998).

Lee and Tuljapurkar (1998) examine the impact of projected mortality decline on the US Social Security System. They show that mortality forecasts used by the Social Security Administration (SSA) have much higher long term mortality rates and a lower range of uncertainty than the levels suggested by an examination of past mortality trends. Australian population projections undertaken by the Australian Bureau of Statistics have also been criticised for not incorporating realistic allowance for uncertainty in forecasted mortality rates (Higgins, 1998).

The fact that uncertainty in mortality projections primarily affects estimates of the older sections of the population is attested by a study of US forecasts. Olshansky (1988) shows that much of the increase between population forecasts over a ten year period was attributable to uncertainty in mortality assumptions for over 65 year olds. Olshansky (1988:482) noted that projections in the early 1960s assumed that mortality declines for older-aged groups would be minimal. This was based on the premise that observed period life-expectancy at birth, which was 70 years, was close to the upper-bound of the human lifespan, an assumption that turned out to be incorrect.

Conservative projections and limited allowance for uncertainty have led to the opinion that: '*...there is a significant likelihood that the future elderly and oldest-old populations will be larger than anticipated in current federal projections.*' (Manton et al., 1993:3).

In Chapter 7 mortality forecasts are generated from the regression models of Chapter 4. Following the approach of Heathcote and Higgins (2001b) it is shown how external opinions can be used to modify parameters of the regression models in order to generate alternative forecasts. We apply various mortality forecasts to the Dutch population and show how uncertainty in future rates can lead to dramatic structural changes in population composition, particularly for the elderly.

2 Theory

2.1 Introduction

With x denoting age, t period, and $t-x$ being the corresponding birth cohort (all measured in years) we can formulate a regression model for the Dutch mortality surface, as

$$Y_{x,t-x} = \delta_{x,t-x} + \varepsilon_{x,t-x}$$

where $Y_{x,t-x} = \log\left(\frac{\tilde{q}_{x,t-x}}{1-\tilde{q}_{x,t-x}}\right)$ are the observed log(odds), $\delta_{x,t-x} = \log\left(\frac{q_{x,t-x}}{1-q_{x,t-x}}\right)$ is the smooth population surface being modelled, and $\varepsilon_{x,t-x}$ are the random errors. The age specific rate of mortality within one year for a cohort born at $t-x$ is $q_{x,t-x}$ and it is estimated by $\tilde{q}_{x,t-x}$.

When the $q_{x,t-x}$ are defined along cohorts by the life table function $q_{x,t-x} = 1 - \frac{l_{x+1,t-x}}{l_{x,t-x}}$, where $l_{x,t-x}$ is the number of survivors aged x at year t, it is shown that under the assumptions of independence between errors for different cohorts, and independent but identically distributed lives within a cohort, the errors $\varepsilon_{x,t-x}$ over the mortality surface are, for large populations, asymptotically independently and normally distributed with zero mean and variances given by:

$$\sigma^2_{bin}(t,x) = \frac{1}{l_{x,t-x} q_{x,t-x}\left(1-q_{x,t-x}\right)}$$

Since the normal distribution is uniquely defined by its mean and covariance matrix, weighted least squares or maximum likelihood methods can be used to estimate parameters for a smooth surface, where the weights are chosen as the inverse of the binomial variance.

A common feature of real data is that of overdispersion due to the presence of random variables. In particular, the failure of the assumption of identically distributed lives within a cohort may lead to heteroscedastic error terms. A solution is to modify the model by including random covariates $\eta_{x,t-x}$, thus resulting in a mixed model:

$$Y_{x,t-x} = \delta_{x,t-x} + \eta_{x,t-x} + \varepsilon_{x,t-x}$$

Extra-binomial variance can then be attributed to the variance component of the random effects. If we assume zero mean for the random effects, and zero covariance between random effects, and between fixed and random effects, then the total variance for the model is simply the sum of the binomial variance and the extra variance component. If we model the extra variance as a function of the covariates of the mean model, then maximum likelihood techniques can be used to simultaneously estimate parameters for both the mean model and the variance component for the random effects. The details and assumptions underlying the construction and estimation of the mixed model are covered in this chapter.

Following from the assumption of asymptotic independence and normality, statistical methods are used to determine standard errors for the overall estimates of the fitted log(odds), as well as for conditional probabilities of survival and life expectancies.

2.2 The Hazard Function

We first define some notation: For an expression $A_{x,t-x}$, the subscript $x,t-x$ refers to age x and cohort $t-x$. An identical expression, in terms of age x and year t is given by $A(t,x)$.

In order to determine the distribution of the $\varepsilon_{x,t-x}$, it is first necessary to provide some underlying theory on the probabilistic structure of mortality. Much of the theory below is well established and can be found in Chiang (1968, 1984) as well as most demographic texts with a statistical emphasis.

Let X be a continuous random variable that equals the exact age at death or the exact lifespan of an individual. The probability that the individual will die prior to age x is given by the distribution function

$$F_X(x) = PROB(X \leq x)$$

The hazard function is defined as the probability that an individual experiences death in a small time interval given that the individual has survived up to the beginning of the interval. In other words, it is the instantaneous rate of mortality, or the probability that an individual alive at age x will die in the interval $(x, x+dx)$. If we call the hazard function $\lambda_X(x)$, then the probability of dying in the age interval $(x, x+dx)$ is,

$$\lambda_X(x)dx = PROB(x \leq X \leq x+dx | x < X) + o(dx) = \frac{F_X'(x)}{1-F_X(x)} + o(dx)$$

The probability that an individual alive at age 0 will survive to age x is,

$${}_xp_0 = 1 - F_X(x) = \exp\left\{-\int_0^x \lambda(u)du\right\} \qquad (2.2.1)$$

where the conditional probability of survival to age $(x+a)$ given survival to age x is defined as,

$${}_ap_x = \Pr(T > x+a | T > x) \qquad (2.2.2)$$

Equation (2.2.1) assumes that our hazard varies with age in the same way for all individuals (ie. all cohorts). We could modify this to incorporate different hazards for different cohorts. This situation would give a more realistic approximation of the behaviour of human mortality. For an individual born in the cohort $t-x$, the probability that the individual survives to age x is

$$ {}_{x}p_{0,t-x} = \exp\left\{-\int_{0}^{x} \lambda_{t-x}(u)du\right\} \tag{2.2.3} $$

where $\lambda_{t-x}(x)$ is the hazard for the $t-x$ cohort.

We can generalise the above and write the probability that an individual from cohort $t-x$ survives to x+1 given survival to age x,

$$ {}_{1}p_{x,t-x} = p_{x,t-x} = \exp\left\{-\int_{x}^{x+1} \lambda_{t-x}(u)du\right\} \tag{2.2.4} $$

Note, the subscript 1 is often left off the symbol ${}_{1}p_{x,t-x}$. This expression implies that we have a distinct hazard function for each cohort. In other words, the force of mortality varies with age for all individuals in a particular cohort, or that the individuals within a cohort make up a homogeneous group. The assumption of homogeneity (Assumption (ii) in the next section) does not take into account the possibility of differing frailties within a cohort due to disability or other sources of heterogeneity. The assumption of homogeneity may lead to overdispersion in general, and extra-binomial variation in the Dutch data in particular.

2.3 Mortality and the Binomial Distribution

We can think of the mortality of the Netherlands as a sample from a larger hypothetical population, and that observed Dutch mortality values vary randomly about unknown population values. For example, if $\tilde{l}_{x,t-x}$ is a random variable for the number of Dutch survivors then it will vary around the unknown population quantity $l_{x,t-x}$ with a particular variance, which, as will shown below, is based on the binomial distribution. There is a strong argument, however, for assuming that the unknown population quantities are random variables themselves (Pollard,1970). For example, as discussed in Chapter 1, survivors in a

particular cohort may have a mixture of frailties, which ultimately influences the aggregate shape of the mortality curve. In Section 2.6 we consider the impact of population quantities as random variables; for the moment, however, we assume that population quantities are fixed, and we introduce a model for mortality based on the classical assumptions regarding the distribution of lives within a cohort:

Assumptions

(i) Cohorts are statistically independent with regard to mortality. In other words, errors from a regression model of mortality are independent for different cohorts.

(ii) individuals within a cohort have lifespans that are independent of each other, and are homogeneous or identically distributed. That is, they share the same hazard function or force of mortality;

That the number of survivors from a cohort is binomially distributed follows from Assumption (ii). This is shown below.

For the birth cohort $c = t - x$

$$\mathrm{I}_{t-x}^{(j)}(x) = \begin{cases} 1 & \textit{individual } j \textit{ survives to age}(x) \\ 0 & \textit{otherwise} \end{cases}$$

Here $\mathrm{I}_{t-x}^{(j)}(x)$ has a bernouilli distribution with mean and variance

$$E\left(\mathrm{I}_{t-x}^{(j)}(x)\right) = {}_{x}p_{0,t-x} \tag{2.3.1}$$

$$VAR\left(\mathrm{I}_{t-x}^{(j)}(x)\right) = {}_{x}p_{0,t-x}\left(1 - {}_{x}p_{0,t-x}\right) \tag{2.3.2}$$

where ${}_{x}p_{0,t-x}$ is the probability of survival from birth to age x for an individual in cohort $c = t - x$ and is given in equation (2.2.3).

A non-zero covariance exists for an individual within a cohort. Ie. An individual's mortality at age $x+1$ in year $t+1$ is dependent upon the individual's past mortality at age x, year t.

$$COV\left(\mathrm{I}_{t-x}^{(j)}(x+1),\mathrm{I}_{t-x}^{(j)}(x)\right)={}_{x+1}p_{0,t-x}\left(1-{}_{x}p_{0,t-x}\right) \tag{2.3.3}$$

By Assumption (i), for two different individuals i and j for all values of k and m,

$$COV\left(\mathrm{I}_{t-x+m}^{(i)}(x+k),\mathrm{I}_{t-x}^{(j)}(x)\right)=0. \tag{2.3.4}$$

Note that equations (2.3.1), (2.3.2) and (2.3.3) are the moments of $\mathrm{I}_{t-x}^{(j)}(x)$ conditional on a fixed and known ${}_{x}p_{0,t-x}$. This assumes that the population value of ${}_{x}p_{0,t-x}$, which is unknown, is constant. If the population measure ${}_{x}p_{0,t-x}$ is a random variable with an unknown variance due to the failure of the assumption of identical distributions at age x within a cohort, then the unconditional moments of $\mathrm{I}_{t-x}^{(j)}(x)$ have to take into account the uncertainty in ${}_{x}p_{0,t-x}$ (this is discussed in more detail in Section 2.6). For the moment, however, we assume that Assumption (ii) holds.

Define the random variable for the number of lives at exact age x of the cohort born at time $t-x$, $\tilde{l}_{x,t-x}$ as

$$\tilde{l}_{x,t-x}=\sum_{j=1}^{l_{0,t-x}}\mathrm{I}_{t-x}^{(j)}(x)$$

where $l_{0,t-x}$ equals the number of births corresponding to the cohort $t-x$. The observed number of survivors $\tilde{\tilde{l}}_{x,t-x}$ at (t,x) is a realisation of the random variable $\tilde{l}_{x,t-x}$.

It follows from the independence of lives within a cohort, that $\tilde{l}_{x,t-x}$ is binomially distributed with parameters $l_{0,t-x}$ and ${}_{x}p_{0,t-x}$. Therefore, the mean and variance are:

$$E\left(\tilde{l}_{x,t-x}\right)=l_{0,t-x}\left({}_{x}p_{0,t-x}\right) \tag{2.3.5}$$

$$VAR\left(\tilde{l}_{x,t-x}\right)=l_{0,t-x}\left({}_{x}p_{0,t-x}\right)\left(1-{}_{x}p_{0,t-x}\right) \tag{2.3.6}$$

Extending the argument for (2.3.3) shows that the covariance for $\tilde{l}_{x,t-x}$ and $\tilde{l}_{x+k,t-x}$ is:

$$COV\left(\tilde{l}_{x,t-x},\tilde{l}_{x+k,t-x}\right)=l_{0,t-x}\left({}_{x+k}p_{0,t-x}\right)\left(1-{}_{x}p_{0,t-x}\right) \tag{2.3.7}$$

An implication of the failure of Assumption (ii) is that $\tilde{l}_{x,t-x}$ may be the sum of *different* binomial random variables, meaning that ${}_{x}p_{0,t-x}$ would vary within each cohort at age x.

Since $\tilde{l}_{x,t-x}$ is binomially distributed, it follows that ${}_{x}\tilde{p}_{0,t-x}$ is a binomial proportion with realisation ${}_{x}\tilde{\tilde{p}}_{0,t-x}$ and cohort life-table function:

$${}_{x}\tilde{p}_{0,t-x}=\frac{\tilde{l}_{x,t-x}}{l_{0,t-x}}.$$

Similarly, the cohort age-specific probability of survival to age $x+1$ given survival to age x, $\tilde{p}_{x,t-x}$ (see equation (2.2.4)) is a binomial proportion with realisation $\tilde{\tilde{p}}_{x,t-x}$ and cohort life-table function:

$$\tilde{p}_{x,t-x}=\frac{\tilde{l}_{x+1,t-x}}{\tilde{l}_{x,t-x}},$$

mean

$$E\left(\tilde{p}_{x,t-x}\right)=p_{x,t-x}$$

and variance

$$VAR(\tilde{p}_{x,t-x}) = E\left(\frac{1}{\tilde{l}_{x,t-x}}\right) p_{x,t-x}(1-p_{x,t-x})$$

The Weak Law of Large Numbers (WLLN) holds for $\tilde{l}_{x,t-x}$ as $l_{0,t-x} \to \infty$. In other words,

$$\lim_{l_0 \to \infty} \Pr\left[\left|\tilde{l}_{x,t-x} - l_{x,t-x}\right| > \delta\right] = 0 \quad \text{for any } \delta > 0, \text{ and}$$

$$\left(\frac{1}{\tilde{l}_{x,t-x}}\right) \xrightarrow{p} \left(\frac{1}{l_{x,t-x}}\right) \qquad \text{as } l_{0,t-x} \to \infty.$$

See for example, Schervish (1995:397), and also chapter 10 of Chiang (1968).

Therefore, the asymptotic variance of $\tilde{p}_{x,t-x}$ is

$$VAR(\tilde{p}_{x,t-x}) = \frac{p_{x,t-x}(1-p_{x,t-x})}{l_{x,t-x}}$$

The cohort age-specific probability of death by age $x+1$ given survival to age x, is defined as:

$$\tilde{q}_{x,t-x} = 1 - \tilde{p}_{x,t-x} = 1 - \frac{\tilde{l}_{x+1,t-x}}{\tilde{l}_{x,t-x}},$$

with the realisation of this random variable denoted by $\tilde{\tilde{q}}_{x,t-x}$.

2.4 The Form of the Regression Surface

We can formulate a regression model for a mortality surface, as

$$Y_{x,t-x} = \delta_{x,t-x} + \varepsilon_{x,t-x} \tag{2.4.1}$$

where the response $Y_{x,t-x}$ and the fixed component $\delta_{x,t-x}$ are a function $h(.)$ of the initial rate of mortality $q_{x,t-x}$, or

$$Y_{x,t-x} = h(\tilde{q}_{x,t-x}) = h(q_{x,t-x}) + \varepsilon_{x,t-x}$$

The function chosen for modeling the male and female Dutch data is the logistic function:

$$h(q_{x,t-x}) = \log\left(\frac{q_{x,t-x}}{1-q_{x,t-x}}\right)$$

Although there are other possible choices, such as the probit or complementary log-log function, the logistic (or logit) function is more easily interpretable, since it is simply the logarithm of the odds (McCullagh and Nelder,1989:109). In addition, the logarithm is a differentiable, monotonic function of $q_{x,t-x}$, and counteracts the exponential increase in mortality as age increases. The trend in the logarithm of the odds of mortality as age increases appears almost linear from age 50 onwards (see Chapter 3 for a thorough analysis into Dutch mortality patterns). This has obvious advantages when it comes to model selection.

The model can then be written as,

$$\log\left(\frac{\tilde{q}_{x,t-x}}{1-\tilde{q}_{x,t-x}}\right) = \log\left(\frac{q_{x,t-x}}{1-q_{x,t-x}}\right) + \varepsilon_{x,t-x}$$

The statistical properties of $\tilde{l}_{x,t-x}$ have been determined in the previous section under Assumptions (i) and (ii), so we can now consider the stochastic properties of $\varepsilon_{x,t-x} = \log\left(\frac{\tilde{q}_{x,t-x}}{1-\tilde{q}_{x,t-x}}\right) - \log\left(\frac{q_{x,t-x}}{1-q_{x,t-x}}\right)$, which is a function of $\tilde{l}_{x,t-x}$ and $\tilde{l}_{x+1,t-x}$.

In Section 2.5 we present a proof showing that the regression errors $\varepsilon_{x,t-x}$ along the same cohort are asymptotically normally distributed. In addition, the covariance structure for $\varepsilon_{x,t-x}$ and $\varepsilon_{x+k,t-x}$, which is complicated for period probabilities, simplifies to zero for probabilities along a cohort, thereby resulting in asymptotic independence.

2.5 The Asymptotic Distribution of $\varepsilon(t, x)$.

The classical assumption of homogeneity (Assumption (ii)) leads to the binomial distribution for the number of survivors $l_{x,t-x}$ aged x in a cohort. Recall from Chiang (1968) that the Central Limit Theorem and the Weak Law of Large Numbers hold for the binomial frequencies $l_{x,t-x}$.

Result

As $l_{0,t-x} \to \infty$, $\varepsilon_{x,t-x} = \log\left(\frac{\tilde{q}_{x,t-x}}{1-\tilde{q}_{x,t-x}}\right) - \log\left(\frac{q_{x,t-x}}{1-q_{x,t-x}}\right)$ converges in distribution to a normally distributed random variable Z where,

$$E(Z) = 0$$

$$VAR(Z) = \frac{1}{l_{x,t-x}q_{x,t-x}(1-q_{x,t-x})} = \frac{1}{l_{0,t-x}({}_xp_{0,t-x})q_{x,t-x}(1-q_{x,t-x})},$$

and $\varepsilon_{x,t-x}$ and $\varepsilon_{x+k,t-x}$ are asymptotically independent for all values of (t,x) in the Lexis plane.

Proof

For this proof we are referring to only one cohort, therefore, all subscripts referring to cohort $t-x$ has been omitted. For example, $p_{x,t-x} = p_x$

The distribution of the errors from the regression model (2.4.1) can be found by noting the following:

$$\varepsilon_x = \log\left(\frac{\tilde{q}_x}{1-\tilde{q}_x}\right) - \log\left(\frac{q_x}{1-q_x}\right)$$

$$= \log\left(\frac{\tilde{l}_x - \tilde{l}_{x+1}}{\tilde{l}_{x+1}}\right) - \log\left(\frac{l_x - l_{x+1}}{l_{x+1}}\right)$$

$$= \log\left(\frac{\tilde{l}_x - \tilde{l}_{x+1}}{l_x - l_{x+1}}\right) - \log\left(\frac{\tilde{l}_{x+1}}{l_{x+1}}\right)$$

Using $\log\left(\frac{\tilde{l}_{x+1}}{l_{x+1}}\right) = \log\left(1 + \frac{\tilde{l}_{x+1} - l_{x+1}}{l_{x+1}}\right)$

$$= \frac{\tilde{l}_{x+1} - l_{x+1}}{l_{x+1}} + o_p\left(\frac{1}{\sqrt{l_0}}\right), \qquad \text{as } l_0 \longrightarrow \infty,$$

and, $\log\left(\frac{\tilde{l}_x - \tilde{l}_{x+1}}{l_x - l_{x+1}}\right) = \frac{\tilde{l}_x - \tilde{l}_{x+1} - (l_x - l_{x+1})}{l_x - l_{x+1}} + o_p\left(\frac{1}{\sqrt{l_0}}\right), \qquad \text{as } l_0 \longrightarrow \infty,$

we find,

$$\varepsilon_x = \left(\frac{\tilde{l}_x - \tilde{l}_{x+1}}{l_x - l_{x+1}}\right) - \left(\frac{\tilde{l}_{x+1}}{l_{x+1}}\right) + o_p\left(\frac{1}{\sqrt{l_0}}\right)$$

$$= \frac{\tilde{l}_x}{(l_x - l_{x+1})} - \frac{\tilde{l}_{x+1}}{(l_x - l_{x+1})}\frac{l_x}{l_{x+1}} + o_p\left(\frac{1}{\sqrt{l_0}}\right)$$

Now, since $l_x = l_0({}_xp_0)$ and $l_{x+1} = l_0({}_{x+1}p_0)$, this expression reduces to,

$$\varepsilon_x = \frac{1}{l_0}\left[\frac{\tilde{l}_x}{({}_xp_0 - {}_{x+1}p_0)} - \frac{\tilde{l}_{x+1}}{({}_xp_0 - {}_{x+1}p_0)p_x}\right] + o_p\left(\frac{1}{\sqrt{l_0}}\right)$$

$$= \frac{1}{l_0}\left[\frac{\tilde{l}_x - l_x}{{}_xp_0 - {}_{x+1}p_0} - \frac{\tilde{l}_{x+1} - l_{x+1}}{({}_xp_0 - {}_{x+1}p_0)p_x} + \frac{l_x}{{}_xp_0 - {}_{x+1}p_0} - \frac{l_{x+1}}{({}_xp_0 - {}_{x+1}p_0)p_x}\right] + o_p\left(\frac{1}{\sqrt{l_0}}\right)$$

$$= \frac{1}{l_0}\left[\frac{\tilde{l}_x - l_x}{{}_xp_0 - {}_{x+1}p_0} - \frac{\tilde{l}_{x+1} - l_{x+1}}{({}_xp_0 - {}_{x+1}p_0)p_x} + \frac{l_x p_x - l_{x+1}}{({}_xp_0 - {}_{x+1}p_0)p_x}\right] + o_p\left(\frac{1}{\sqrt{l_0}}\right)$$

$$= \frac{1}{l_0}\left[(\tilde{l}_x - l_x)\frac{1}{{}_xp_0 - {}_{x+1}p_0} + (\tilde{l}_{x+1} - l_{x+1})\frac{-1}{({}_xp_0 - {}_{x+1}p_0)p_x}\right] + o_p\left(\frac{1}{\sqrt{l_0}}\right) \qquad (2.5.1)$$

Similarly, we can write

$$\varepsilon_{x+k} = \log\left(\frac{\tilde{q}_{x+k}}{1-\tilde{q}_{x+k}}\right) - \log\left(\frac{q_{x+k}}{1-q_{x+k}}\right)$$

$$= \frac{1}{l_0}\left[(\tilde{l}_{x+k} - l_{x+k})\frac{1}{{}_{x+k}p_0 - {}_{x+k+1}p_0} + (\tilde{l}_{x+k+1} - l_{x+k+1})\frac{-1}{({}_{x+k}p_0 - {}_{x+k+1}p_0)p_{x+k}}\right] + o_p\left(\frac{1}{\sqrt{l_0}}\right)$$

(2.5.2)

Therefore, the compound asymptotic distribution of ε_x and ε_{x+k} depends on the joint distributions of $\tilde{l}_x$, $\tilde{l}_{x+1}$, $\tilde{l}_{x+k}$ and $\tilde{l}_{x+k+1}$.

By the multivariate Central Limit Theorem (Schervish,1995:643; Rao,1973:128) as $l_0 \longrightarrow \infty$, the normed vector of these frequencies is asymtotically normally distributed. Specifically, let

$$\underline{L} = \begin{bmatrix} l_x \\ l_{x+1} \\ l_{x+k} \\ l_{x+k+1} \end{bmatrix} = \begin{bmatrix} l_0({}_xp_0) \\ l_0({}_{x+1}p_0) \\ l_0({}_{x+k}p_0) \\ l_0({}_{x+k+1}p_0) \end{bmatrix}, \qquad \underline{\tilde{L}} = \begin{bmatrix} \tilde{l}_x \\ \tilde{l}_{x+1} \\ \tilde{l}_{x+k} \\ \tilde{l}_{x+k+1} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{l_0} \mathrm{I}_c^{(j)}(x+i) \\ \sum_{j=1}^{l_0} \mathrm{I}_c^{(j)}(x+1+i) \\ \sum_{j=1}^{l_0} \mathrm{I}_c^{(j)}(x+k+i) \\ \sum_{j=1}^{l_0} \mathrm{I}_c^{(j)}(x+k+1+i) \end{bmatrix}$$

We can write the four-by-four covariance matrix for each of the IID vectors as,

$$\underline{C} = \begin{bmatrix} {}_xp_0(1-{}_xp_0) & {}_{x+1}p_0(1-{}_xp_0) & {}_{x+k}p_0(1-{}_xp_0) & {}_{x+k+1}p_0(1-{}_xp_0) \\ {}_{x+1}p_0(1-{}_xp_0) & {}_{x+1}p_0(1-{}_{x+1}p_0) & {}_{x+k}p_0(1-{}_{x+1}p_0) & {}_{x+k+1}p_0(1-{}_{x+1}p_0) \\ {}_{x+k}p_0(1-{}_xp_0) & {}_{x+k}p_0(1-{}_{x+1}p_0) & {}_{x+k}p_0(1-{}_{x+k}p_0) & {}_{x+k+1}p_0(1-{}_{x+k}p_0) \\ {}_{x+k+1}p_0(1-{}_xp_0) & {}_{x+k+1}p_0(1-{}_{x+1}p_0) & {}_{x+k+1}p_0(1-{}_{x+k}p_0) & {}_{x+k+1}p_0(1-p_{0,x+k+1}) \end{bmatrix}$$

Since $\underline{\tilde{L}}$ is the sum of a sequence of l_0 IID random vectors, it follows that the Multivariate Central Limit Theorem applies, and

$$\frac{(\underline{\tilde{L}} - \underline{L})}{\sqrt{l_0}} \xrightarrow{d} N(0, \underline{C}).$$

In order to prove asymptotic normality for ε_x, and asymptotic independence between ε_x and ε_{x+k}, we can use the following condition of the Multivariate Central Limit Theorem (Schervish,1995:643; Rao,1973:128):

If $\underline{\underline{A}}$ is a non-singular matrix with four columns, and if $\underline{\underline{A}}\underline{\tilde{L}}$ is a linear combination of the $\tilde{l}_x$, $\tilde{l}_{x+1}$, $\tilde{l}_{x+k}$ and $\tilde{l}_{x+k+1}$, then:

$$\frac{\underline{\underline{A}}\left(\underline{\tilde{L}}-\underline{L}\right)}{\sqrt{l_0}} \xrightarrow{d} N\left(\underline{0}, \underline{\underline{A}}\underline{\underline{C}}\underline{\underline{A}}'\right) \tag{2.5.3}$$

Referring back to equations (2.5.1) and (2.5.2), we can write $\underline{\underline{A}}$ as:

$$\underline{\underline{A}} = \begin{pmatrix} \frac{1}{{}_xp_0 - {}_{x+1}p_0} & \frac{-1}{\left({}_xp_0 - {}_{x+1}p_0\right)p_x} & 0 & 0 \\ 0 & 0 & \frac{1}{{}_{x+k}p_0 - {}_{x+k+1}p_0} & \frac{-1}{\left({}_{x+k}p_0 - {}_{x+k+1}p_0\right)p_{x+k}} \end{pmatrix}$$

Then,

$$\begin{bmatrix} \varepsilon_x \\ \varepsilon_{x+k} \end{bmatrix} = \begin{bmatrix} \log\left(\frac{\tilde{q}_x}{1-\tilde{q}_x}\right) - \log\left(\frac{q_x}{1-q_x}\right) \\ \log\left(\frac{\tilde{q}_{x+k}}{1-\tilde{q}_{x+k}}\right) - \log\left(\frac{q_{x+k}}{1-q_{x+k}}\right) \end{bmatrix} = \frac{\underline{\underline{A}}}{l_0} \begin{bmatrix} \tilde{l}_x - l_x \\ \tilde{l}_{x+1} - l_{x+1} \\ \tilde{l}_{x+k} - l_{x+k} \\ \tilde{l}_{x+k+1} - l_{x+k+1} \end{bmatrix} + o_p(1)$$

Therefore,

$$\sqrt{l_0}\begin{bmatrix} \varepsilon_x \\ \varepsilon_{x+k} \end{bmatrix} = \frac{\underline{\underline{A}}\left(\underline{\tilde{L}}-\underline{L}\right)}{\sqrt{l_0}} + o_p(1)$$

It follows from the Multivariate Central Limit Theorem (equation (2.5.3)) that,

$$\sqrt{l_0}\begin{bmatrix} \varepsilon_x \\ \varepsilon_{x+k} \end{bmatrix} \xrightarrow{d} N\left(\underline{0}, \underline{\underline{A}}\underline{\underline{C}}\underline{\underline{A}}'\right)$$

All that remains is to evaluate $\underline{\underline{A}}\underline{\underline{C}}\underline{\underline{A}}'$:

$$\underline{AC} = \begin{pmatrix} 0 & -1 & -\frac{{}_{x+k}p_0}{{}_{x+1}p_0} & -\frac{{}_{x+k+1}p_0}{{}_{x+1}p_0} \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Therefore,

$$\underline{ACA'} = \begin{pmatrix} 0 & -1 & -\frac{{}_{x+k}p_0}{{}_{x+1}p_0} & -\frac{{}_{x+k+1}p_0}{{}_{x+1}p_0} \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{{}_{x}p_0 - {}_{x+1}p_0} & 0 \\ \frac{-1}{({}_{x}p_0 - {}_{x+1}p_0)p_x} & 0 \\ 0 & \frac{1}{{}_{x+k}p_0 - {}_{x+k+1}p_0} \\ 0 & \frac{-1}{({}_{x+k}p_0 - {}_{x+k+1}p_0)p_{x+k}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{({}_{x}p_0 - {}_{x+1}p_0)p_x} & -\frac{{}_{x+k}p_0}{{}_{x+1}p_0}\left(\frac{1}{{}_{x+k}p_0 - {}_{x+k+1}p_0}\right) + \frac{{}_{x+k+1}p_0}{{}_{x+1}p_0}\left(\frac{1}{({}_{x+k}p_0 - {}_{x+k+1}p_0)p_{x+k}}\right) \\ 0 & \frac{1}{({}_{x+k}p_0 - {}_{x+k+1}p_0)p_{x+k}} \end{pmatrix}$$

The term in the off-diagonal disappears since $\frac{{}_{x+k+1}p_0}{p_{x+k}} = {}_{x+k}p_0$, resulting in zero covariances.

Therefore,

$$\underline{ACA'} = \begin{pmatrix} \frac{1}{({}_{x}p_0 - {}_{x+1}p_0)p_x} & 0 \\ 0 & \frac{1}{({}_{x+k}p_0 - {}_{x+k+1}p_0)p_{x+k}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{l_0}{l_x q_x (1-q_x)} & 0 \\ 0 & \frac{l_0}{l_{x+k} q_{x+k} (1-q_{x+k})} \end{pmatrix}$$

The diagonal terms are the variances of $\sqrt{l_0}\varepsilon_x$ and $\sqrt{l_0}\varepsilon_{x+k}$

In addition, by Assumption (i), ε_x and ε_{x+k} are independent for different cohorts.

Therefore, ε_x and ε_{x+k} are asymptotically normally and independently distributed, where ε_{x+i} has zero mean and asymptotic variance of $\frac{1}{l_{x+i}q_{x+i}(1-q_{x+i})}$.

This concludes the proof. A briefer account of the asymptotic independence can be found in Chiang (1968:228).

2.6 Model Fitting and Extra-Binomial Variation

Recapping, our model is:

$$Y_{x,t-x} = \delta_{x,t-x} + \varepsilon_{x,t-x} = \log\left(\frac{q_{x,t-x}}{1-q_{x,t-x}}\right) + \varepsilon_{x,t-x}$$

where,

$$VAR(\varepsilon_{x,t-x}) = \sigma^2_{bin}(t,x) = \frac{1}{l_{x,t-x}q_{x,t-x}(1-q_{x,t-x})}$$

For each observation at age x and cohort $t-x$ we can write $\log\left(\frac{q_{x,t-x}}{1-q_{x,t-x}}\right)$ in terms of a linear combination of p covariates $X_1 \ldots X_p$, with parameters $\underline{\beta} = (\beta_1 \ldots \beta_p)$:

$$Y_{x,t-x} = \delta_{x,t-x}(\underline{\beta}) + \varepsilon_{x,t-x} = \sum_{i=1}^{p} \beta_i X_i + \varepsilon_{x,t-x}$$

For n observations, we can write this in vector notation, as

$$\underline{Y} = \underline{\delta}(\underline{\beta}) + \underline{\varepsilon} = \underline{X}\underline{\beta} + \underline{\varepsilon} \qquad (2.6.1)$$

where $\underline{Y}$ is a n x 1 vector of observed log(odds), $\underline{X}$ is a n x p matrix of covariates, $\underline{\beta}$ is a p x 1 vector of parameters to be estimated, and $\underline{\varepsilon}$ is a n x 1 vector of errors with $E(\underline{\varepsilon}) = \underline{0}$ and $VAR(\underline{\varepsilon}) = \underline{V} = diag(\underline{\sigma}^2_{bin})$.

Since the formula is linear in the unknown parameters, and since the asymptotic variance differs for different (t,x), the method of weighted least squares can be used to estimate the parameters. The theory behind weighted least squares is well known and only the main results are covered here (see, for example, Ryan (1997), Weisberg (1985))

Weighted least squares involves minimising $\sum_{(t,x)} w_{x,t-x}\left[Y_{x,t-x} - \delta_{x,t-x}(\underline{\beta})\right]^2$ over the entire surface $L(t,x)$, where each $w_{x,t-x}$ is the inverse of the corresponding binomial variance. This implies that any observation with a high variance will be discounted in determination of the fitted parameters $\hat{\underline{\beta}}$. In matrix notation, the fitted parameters $\hat{\underline{\beta}}$ are estimated from:

$$\hat{\underline{\beta}} = \left(\underline{X'V^{-1}X}\right)^{-1}\underline{X'V^{-1}Y}$$

where $\underline{V}^{-1}$ is the n x n matrix of weights, or the inverse of the covariance matrix which is given, in this instance, by

$$\underline{V}^{-1} = \begin{bmatrix} l_{x,c}q_{x,c}(1-q_{x,c}) & 0 & \cdots & 0 \\ 0 & l_{x+1,c}q_{x+1,c}(1-q_{x+1,c}) & \cdots & 0 \\ \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & l_{x+m,c+j}q_{x+m,c+j}(1-q_{x+m,c+j}) \end{bmatrix}$$

where age ranges from x to $x+m$, and cohort ranges from c to $c+j$.

This is equivalent to $\hat{\underline{\beta}} = \left(\underline{X}^{w\prime} \underline{X}^{w} \right)^{-1} \underline{X}^{w\prime} \underline{Y}^{w}$, where $\underline{X}^{w} = \sqrt{\underline{W}}\,\underline{X}$ and $\underline{Y}^{w} = \sqrt{\underline{W}}\,\underline{Y}$,

where $\underline{W}$ is a n x 1 vector of weights and is equal to $\underline{\sigma}_{bin}^{-2}$.

In spite of the theoretical justification for the weights used, as is more common than not in regression modeling (see, for example, McCullagh and Nelder (1989), Williams (1982), Moore and Tsiatis (1991)), there may be an apparent source of extra variation in the models generated under the Assumptions of Section 2.3.

A means of determining the extent of extra variation is to produce the sum of squares of the standardised residuals. It is well known that the sum of squares of standardised normal variables have an approximate chi-square distribution, $\chi^2(n)$ (see, for example, Rao (1973:181)) . It follows that the standardised residuals from the regression model $\left(\underline{Y} - \underline{X}\hat{\underline{\beta}}\right)' \underline{V}^{-1} \left(\underline{Y} - \underline{X}\hat{\underline{\beta}}\right)$ should compare closely with a $\chi^2(n-p)$ distribution if they are sufficiently standard normal. In particular,

$$E\left\{\left(\underline{Y} - \underline{X}\hat{\underline{\beta}}\right)' \underline{V}^{-1} \left(\underline{Y} - \underline{X}\hat{\underline{\beta}}\right)\right\} = n - p$$

When the standardised residuals exceed $n-p$, a factor can be found to adjust the fit. An estimate of the factor is:

$$\psi = \frac{\left(\underline{Y} - \underline{X}\hat{\underline{\beta}}\right)' \underline{V}^{-1} \left(\underline{Y} - \underline{X}\hat{\underline{\beta}}\right)}{n-p} \qquad (2.6.2)$$

(see, for example, Williams (1982)). Adjusting the fitted model by an aggregate factor has no effect on the parameter estimates; rather it inflates the standard errors of the parameters.

Parameterising a component of the variance and using maximum likelihood provides an alternative approach. By minimising the negative log-likelihood using the Splus function 'ms', we can solve for the new parameter estimates for the mean and variance.

It is plausible that the appearance of extra-binomial variation is due to one or more random effects for which we have no information. In particular, it is likely that in reality the individuals within a cohort are not identically distributed, but vary in underlying frailty due to genetic pre-dispositions. In addition, it is likely that frailties change over time due to the interaction of lifestyle, period effects, and constitution.

We could modify the model by assuming that the population probability of survival for individuals within a cohort is no longer fixed at ${}_xp_{0,t-x}$, but is a random variable that modifies the underlying cohort-specific hazard. In other words, ${}_x\tilde{p}_{0,t-x}$ no longer varies simply due to it being a binomial proportion of random variables, but also contains variability since it represents a mixture of distributions with differing frailties. The moments of ${}_x\tilde{p}_{0,t-x}$ are then,

$$E\left({}_x\tilde{p}_{0,t-x}\right)={}_xp_{0,t-x}$$

$$VAR\left({}_x\tilde{p}_{0,t-x}\right)=g_1\left({}_x\tilde{p}_{0,t-x}\right)$$

$$COV\left({}_x\tilde{p}_{0,t-x},{}_{x+k}\tilde{p}_{0,t-x}\right)=g_2\left({}_x\tilde{p}_{0,t-x},{}_{x+k}\tilde{p}_{0,t-x}\right)$$

where $g_1(.)$ and $g_2(.)$ are unknown functions. In this situation, using the identities $VAR(X)=E\left(VAR\left(X|Y\right)\right)+VAR\left(E\left(X|Y\right)\right)$ and $E(X^2)=VAR(X)+\left(E(X)\right)^2$, the unconditional moments of $\tilde{l}_x$ in cohort $t-x$ become:

$$E\left(\tilde{l}_{x,t-x}\right)=l_{0,t-x}\left({}_xp_{0,t-x}\right)$$

$$VAR\left(\tilde{l}_{x,t-x}\right)=l_{0,t-x}\left({}_xp_{0,t-x}\right)\left(1-{}_xp_{0,t-x}\right)+l_{0,t-x}\left(l_{0,t-x}-1\right)g_1\left({}_xp_{0,t-x}\right)$$

$$COV\left(\tilde{l}_{x,t-x},\tilde{l}_{x+k,t-x}\right)=l_{0,t-x}\left({}_{x+k}p_{0,t-x}\right)\left(1-{}_xp_{0,t-x}\right)+l_{0,t-x}\left(l_{0,t-x}-1\right)g_2\left({}_xp_{0,t-x},{}_{x+k}p_{0,t-x}\right)$$

Note that:

$$VAR\left(\mathrm{I}_{t-x}^{(j)}(x)\right) = E\left(VAR\left(\mathrm{I}_{t-x}^{(j)}(x)\middle|_x \tilde{p}_{0,t-x}\right)\right) + VAR\left(E\left(\mathrm{I}_{t-x}^{(j)}(x)\middle|_x \tilde{p}_{0,t-x}\right)\right)$$

$$= {}_xp_{0,t-x}\left(1 - {}_xp_{0,t-x}\right)$$

which implies that,

$$VAR\left[\sum_j \mathrm{I}_{t-x}^{(j)}(x)\right] = VAR\left[\tilde{l}_{x,t-x}\right] > \sum_j VAR\left[\mathrm{I}_{t-x}^{(j)}(x)\right]$$

This expression shows that the heterogeneity within a cohort arises due to the mixture distribution of ${}_x\tilde{p}_{0,t-x}$.

The analysis above implies that one way of dealing with extra-binomial variance is to consider the function $\kappa_{t-x}(x)$, such that,

$$VAR\left(\tilde{l}_{x,t-x}\right) = l_{0,t-x}\left({}_xp_{0,t-x}\right)\left(1 - {}_xp_{0,t-x}\right) + \kappa_{t-x}(x)$$

$$COV\left(\tilde{l}_{x,t-x}, \tilde{l}_{x+k,t-x}\right) = l_0\left({}_{x+k}p_{0,t-x}\right)\left(1 - {}_xp_{0,t-x}\right) + \kappa_{t-x}(x, x+k)$$

It is desirable to incorporate the extra variance component in such a way that the regression errors are still asymptotically independently and normally distributed. One way of achieving this is to propose a functional form for $\kappa_{t-x}(x)$ such that, asymptotically,

$$COV\left(\varepsilon_{x,t-x}, \varepsilon_{x+k,t-x}\right) = 0,$$

This could be achieved by presenting $\varepsilon_{x,t-x}$ as a function of $\tilde{l}_{x,t-x}$, as was done in Section 2.4, and then finding the covariance as a function of $\kappa_{t-x}(x, x+k)$, $\kappa_{t-x}(x, x+k+1)$, $\kappa_{t-x}(x+1, x+k)$ and $\kappa_{t-x}(x+1, x+k+1)$.

A difficulty with this approach is that of determining a functional form for $\kappa_{t-x}(x)$ that conforms with asymptotic independence. Assuming that such a function can be found, random effects analysis of variance techniques could possibly be applied to estimate the extra variance component.

Despite its theoretical appeal, this approach was not taken here, but may be the subject of future work. It is apparent that such an approach would be more appealing when mortality data are disaggregated by additional variables such as cause of death. The approach assumes that the extra binomial variation is caused by a unknown mixture of survivor functions within each cohort, and as such the mixture distribution can be considered a random effect. In a similar vein, Pollard (1970) considers the variance structure of the number of deaths in a population under different mortality hypotheses, which includes a hypothesis that assumes that the population aged x consists of sub-classes, each with a different random q_x.

The approach used for dealing with overdispersion in this thesis is based on the existence of unknown random effects, however, it does not explicitly assume that these effects are derived from a particular source. Instead the random effects are included in a general manner, and it is assumed that their inclusion does not affect the asymptotic independence and normality of the regression errors.

The argument for inclusion of random effects is covered in more detail in Chapter 4.

If we let the regression model be written as a mixed model with fixed effects $\underline{\delta} = \underline{X}\underline{\beta}$ (n observations) and random effects $\underline{\eta}$ (an n x q matrix, with q levels), then in matrix notation,

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{\eta} + \underline{\varepsilon} \qquad (2.6.3)$$

where the random effects have the properties $E(\underline{\eta}) = \underline{0}$ and $VAR(\underline{\eta}) = diag(\underline{\gamma}^2)$. In addition, the covariance between different random effects $\underline{\eta}_i$ and $\underline{\eta}_h$ where $i \neq h$ is zero, and $\text{cov}(\underline{\eta}_i, \underline{\varepsilon}') = 0$

Under these assumptions,

$$E(\underline{Y}) = \underline{X}\underline{\beta}$$

$$E(\underline{Y} \mid \underline{U} = \underline{\eta}) = \underline{X}\underline{\beta} + \underline{\eta}$$

and, if $\underline{V}^*$ is the covariance matrix for $\underline{Y} - E(\underline{Y})$, then,

$$\underline{V}^* = \underline{V} + diag(\underline{\gamma}^2) = diag(\underline{\sigma}^2_{bin}) + diag(\underline{\gamma}^2) \qquad (2.6.4)$$

The variance components are taken into account when estimating the fixed effects $\underline{\beta}$. If we know $diag(\underline{\gamma}^2)$, WLS could be used to estimate a best (minimum variance), linear, unbiased estimator (BLUE) of $\underline{\beta}$:

$$\underline{\hat{\beta}} = \left(\underline{X}'\left(\underline{V}^*\right)^{-1}\underline{X}\right)^{-1}\underline{X}'\left(\underline{V}^*\right)^{-1}\underline{Y}$$

Unlike the standard approach, where an overall measure of dispersion ψ is calculated and applied to the variance of the estimates (see, for example, Williams (1982)), we assume that the extra variance can be expressed as an exponential function of the covariates of the mean model. A discussion of the selection of the function and covariates for the Dutch regression models is given in detail in Chapter 4.

If the extra variance $VAR(\underline{\eta})$ is expressed as a function g(.) of x and t, with parameters $\underline{\alpha}$, then the total variance of the residuals at (t,x) is,

$$VAR(Y_{x,t-x} - \delta_{x,t-x}(\underline{\beta})) = \sigma^2_{x,t-x} + g(t,x;\underline{\alpha}).$$

Here we make the additional assumption of zero covariance between the estimated parameters of $g(t,x;\underline{\alpha})$ and $\delta_{x,t-x}(\underline{\beta})$.

Under the assumptions for the mixed model, the residuals are asymptotically normal, and the method of maximum likelihood estimation can be applied to estimate the parameters (see for example Rao,1973:353; Schervish,1995:307)

If the residuals follow a normal distribution, then the asymptotic probability density function is:

$$f\left(t,x;\underline{\hat{\alpha}},\underline{\hat{\beta}}\right) = \frac{1}{\sqrt{2\pi\left(\sigma^2_{x,t-x} + g(t,x;\underline{\hat{\alpha}})\right)}} \exp\left(\frac{-\left(Y_{x,t-x} - \delta(t,x;\underline{\hat{\beta}})\right)^2}{2\left(\sigma^2_{x,t-x} + g(t,x;\underline{\hat{\alpha}})\right)}\right)$$

The likelihood is the product of the distributions over the entire surface. The log-likelihood is the sum of the log of the error distribution over all observations:

$$\sum_{(t,x)} \log\left(f\left(t,x;\underline{\hat{\alpha}}\,\underline{\hat{\beta}}\right)\right) \approx -\frac{1}{2}\sum_{(t,x)} \log\left(2\pi\left(\sigma^2_{x,t-x} + g(t,x;\underline{\hat{\alpha}})\right)\right) - \frac{1}{2}\sum_{(t,x)}\left(\frac{\left(Y_{x,t-x} - \delta(t,x;\underline{\hat{\beta}})\right)^2}{\left(\sigma^2_{x,t-x} + g(t,x;\underline{\hat{\alpha}})\right)}\right)$$

Maximising the log-likelihood involves setting to zero the derivatives with respect to each parameter to be estimated, and solving the resulting equations simultaneously. A function exists in S-plus to help with the estimation process. The form of the likelihood used in fitting the mixed model, and the technique used in maximisation, are given in Section 4.4 of Chapter 4, and Appendix A.

If $\hat{\theta}_n$ is the (assumed unique) consistent maximum likelihood estimator of θ, and assuming that the probability density function has continuous second partial derivatives with respect to the parameters, then the asymptotic variance is the inverse of the Fisher Information matrix I(θ) (see, for example, Schervish (1995:421)):

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} N\left(0, \frac{1}{\mathrm{I}(\theta)}\right)$$

The Fisher Information is produced as an output from S-plus when maximum likelihood estimation is undertaken, therefore allowing straightforward production of standard errors for the parameters and fitted values of both $\delta(t,x;\underline{\hat{\beta}})$ and $g(t,x;\underline{\hat{\alpha}})$.

2.7 Inference from the Fitted Model

2.7.1 Prediction Interval for $\hat{q}_{x,t-x}$

For a known vector of predictors $\underline{x}$, the asymptotic variance for the fitted model is

$$VAR\left(\underline{\hat{Y}} \mid \underline{x}\right) = \underline{x}\left(\underline{X}'\left(\underline{V}^*\right)^{-1}\underline{X}\right)^{-1}\underline{x}'$$

where $\underline{V}^* = diag(\underline{\sigma}^2_{bin}) + diag(\underline{\gamma}^2)$. If we use Maximum Likelihood Estimation to extract the parameter estimates, then $\underline{X}'\left(\underline{V}^*\right)^{-1}\underline{X}$ is equal to the component of the Fisher Information matrix corresponding to the parameters β, $I(\underline{\beta})$. Therefore, the asymptotic variance is:

$$VAR\left(\underline{\hat{Y}} \mid \underline{x}\right) = \underline{x}I(\underline{\hat{\beta}})^{-1}\underline{x}'$$

If we consider the problem of prediction of log(odds) for a single value at the point (t,x) on a mortality surface, given by the scalar $\tilde{w}_{x,t-x}$, where

$$\tilde{w}_{x,t-x} = \underline{x}'\underline{\beta} + \eta_{x,t-x} + \varepsilon_{x,t-x},$$

then the expectation and the variance of the predicted value is given by,

$$E\left(\hat{w}_{x,t-x}\right) = \underline{x}'\underline{\hat{\beta}}$$

$$VAR(\hat{w}_{x,t-x}) = \underline{x}' I(\underline{\beta})^{-1} \underline{x} + V^*_{x,t-x}$$

where $V^*_{x,t-x} = \sigma^2_{bin}(t,x) + \gamma^2(t,x)$ is the variance for a single value at the predicted point (t,x).

The variance for the extra variance component $g(t,x;\underline{\hat{\alpha}})$ can also be found from the Information matrix. If $I(\alpha)$ is the component of the Information matrix relating to the parameters α, and if $\underline{x}_\gamma$ is a known vector of predictors, then,

$$VAR(\gamma^2 \mid \underline{x}_\gamma) = \underline{x}_\gamma I(\alpha)^{-1} \underline{x}_\gamma'$$

Estimated standard errors are found by taking the square-root of the estimated variance.

$$se(\hat{w}_{x,t-x}) = \sqrt{\hat{VAR}(\hat{w}_{x,t-x})} = \sqrt{\underline{x} I(\hat{\beta})^{-1} \underline{x}' + \sigma^2_{bin}(t,x) + g(t,x;\underline{\hat{\alpha}})}$$

$$se(\hat{\gamma}^2) = \sqrt{\hat{VAR}(\hat{\gamma}^2)}$$

95% prediction intervals around the mean can be found by:

$$\hat{w}_{x,t-x} \pm 1.96 se(\hat{w}_{x,t-x}) = \hat{w}_{x,t-x} \pm 1.96 \sqrt{\underline{x} I(\hat{\beta})^{-1} \underline{x}' + \sigma^2_{bin}(t,x) + g(t,x;\underline{\hat{\alpha}})} \qquad (2.7.1)$$

For interpretation, it is more useful to express results in terms of $\hat{q}_{x,t-x}$ rather than log(odds).

A prediction interval for $\hat{q}_{x,t-x}$ can be found by first recognizing that an interval for $\frac{\hat{q}_{x,t-x}}{(1-\hat{q}_{x,t-x})}$ can be generated by taking exponents of the end points of the prediction interval around $\hat{w}_{x,t-x}$.

$$0.95 = \Pr\left(\hat{w}_{x,t-x} - 1.96se\left(\hat{w}_{x,t-x}\right) < \log\left(\frac{q_{x,t-x}}{1-q_{x,t-x}} \right) < \hat{w}_{x,t-x} + 1.96se\left(\hat{w}_{x,t-x}\right) \right)$$

$$= \Pr\left(\exp\left(\hat{w}_{x,t-x} - 1.96se\left(\hat{w}_{x,t-x}\right)\right) < \frac{q_{x,t-x}}{1-q_{x,t-x}} < \exp\left(\hat{w}_{x,t-x} + 1.96se\left(\hat{w}_{x,t-x}\right)\right) \right)$$

After performing some algebra, we obtain the prediction interval for $\hat{q}_{x,t-x}$:

$$\left(\frac{\exp\left(\hat{w}_{x,t-x} - 1.96se\left(\hat{w}_{x,t-x}\right)\right)}{1+\exp\left(\hat{w}_{x,t-x} - 1.96se\left(\hat{w}_{x,t-x}\right)\right)}, \frac{\exp\left(\hat{w}_{x,t-x} + 1.96se\left(\hat{w}_{x,t-x}\right)\right)}{1+\exp\left(\hat{w}_{x,t-x} + 1.96se\left(\hat{w}_{x,t-x}\right)\right)} \right) \tag{2.7.2}$$

Note that because of the impact of the exponential, the distance from the lower bound to $\hat{q}_{x,t-x}$ is not the same as the distance from the upper bound to $\hat{q}_{x,t-x}$.

2.7.2 *Conditional Probability of Survival Calculations*

The conditional probability of survival was defined in equation (2.2.2) as:

$${}_a p_x = \Pr\left(X > a + x \middle| X > x\right)$$

Cross sectional conditional probabilities ${}_a^t p_x$ are calculated by the product of the respective survival rates along individual periods (t). When only cohort-specific probabilities of survival are available such that:

$$\tilde{p}_{x,t-x} = \frac{l_{x+1,t-x}}{l_{x,t-x}},$$

rather than period specific probabilities,

$$\tilde{p}_x = \frac{l_{x+1,t-x-1}}{l_{x,t-x}}$$

the cross-sectional conditional probabilities cannot easily be estimated. Regardless, cross-sectional conditional probabilities of survival are of limited interest, since they do not capture the natural behaviour of mortality along cohorts. Cohort-specific conditional probabilities of survival ${}^{c}_{a}p_x$ are found by multiplying the survival rates along a cohort rather than a particular period:

$$ {}_{a}p_{x,t-x} = p_{x,t-x}p_{x+1,t-x}p_{x+2,t-x}\cdots p_{x+a-1,t-x} $$

$$ {}_{a}\tilde{p}_{x,t-x} = \frac{l_{x+1,t-x}}{l_{x,t-x}}\frac{l_{x+2,t-x}}{l_{x+1,t-x}}\frac{l_{x+3,t-x}}{l_{x+2,t-x}}\cdots\frac{l_{x+a,t-x}}{l_{x+a-1,t-x}} = \frac{l_{x+a,t-x}}{l_{x,t-x}} $$

The sample variance for the cohort specific conditional probability of survival can be found by using a Taylor expansion of a function of $p_{x,t-x}p_{x+1,t-x}p_{x+2,t-x}\cdots p_{x+a-1,t-x}$ (see, for example, Kendall and Stuart (1977:247)). Finding the asymptotic distribution of a function of a random variable in this way is called the delta method (Schervish (1995:401-402)).

Consider the function $g(\tilde{p}_1\tilde{p}_2\tilde{p}_3\ldots\tilde{p}_k)$ which we will write as $g(\underline{\tilde{p}})$, where $\underline{\tilde{p}}$ has mean $\underline{p}$. Then the Taylor series expansion is:

$$ g(\underline{\tilde{p}}) = g(\underline{p}) + \sum_{i=1} g_i'(\underline{p})(\tilde{p}_i - p_i) + O(\frac{1}{\sqrt{l_0}}) $$

It follows that,

$$ VAR[g(\underline{\tilde{p}})] = E\left[\left(\sum_{i=1} g_i'(\underline{p})(\tilde{p}_i - p_i)\right)^2\right] + o(\frac{1}{\sqrt{l_0}}) $$

$$ = \sum_{i=1} g_i'(\underline{p})^2 VAR(\tilde{p}_i - p_i) + \sum_{i\neq j=1}\sum g_i'(\underline{p})g_j'(\underline{p})COV(\tilde{p}_i - p_i, \tilde{p}_j - p_j) + o(\frac{1}{\sqrt{l_0}}) $$

It was shown in Section (2.5) that $\varepsilon_{x,t-x}$ and $\varepsilon_{x+k,t-x}$ are asymptotically independent. It is simple to show, using a similar argument, that $\tilde{p}_i - p_i$ and $\tilde{p}_j - p_j$ are asymptotically independent, which implies that $COV(\tilde{p}_i - p_i, \tilde{p}_j - p_j) = 0$.

Therefore,

$$VAR\left({}_a\tilde{p}_{x,t-x}\right) = VAR\left[g(\underline{\tilde{p}})\right] = \sum_{i=1} g_i'(\underline{p})^2 VAR(\tilde{p}_i - p_i) + o(\frac{1}{\sqrt{l_0}})$$

Now, since

$$g_i'(\underline{p}) = \frac{\partial\left({}_a p_{x,t-x}\right)}{\partial p_{i,t-x}} = \frac{\partial\left(p_{x,t-x}\right)\left(p_{x+1,t-x}\right)\left(p_{x+2,t-x}\right)\ldots\left(p_{x+a-1,t-x}\right)}{\partial p_{i,t-x}} = \frac{{}_a p_{x,t-x}}{p_{i,t-x}},$$

$$VAR\left({}_a\tilde{p}_{x,t-x}\right) = \sum_{h=x}^{a-1}\left(\frac{{}_a p_{x,t-x}^2}{p_{h,t-x}^2} VAR\left(\tilde{p}_{h,t-x}\right)\right) + o(\frac{1}{\sqrt{l_0}})$$

$$VAR\left({}_a\hat{p}_{x,t-x}(\underline{\hat{\beta}})\right) = {}_a\hat{p}_{x,t-x}^2(\underline{\hat{\beta}}) \sum_{h=x}^{a-1}\left(\frac{VAR\left(\hat{p}_{h,t-x}(\underline{\hat{\beta}})\right)}{\hat{p}_{h,t-x}^2(\underline{\hat{\beta}})}\right) + o(\frac{1}{\sqrt{l_0}}) \qquad (2.7.3)$$

To find, $VAR\left(\hat{p}_{x,t-x}\right)$, we can apply the delta method to $VAR\left(\log\left(\frac{\hat{q}_{x,t-x}}{1-\hat{q}_{x,t-x}}\right)\right)$ as follows,

$$VAR\left(\hat{p}_{x,t-x}(\underline{\hat{\beta}})\right) = g'\left(\log(\frac{\hat{q}_{x,t-x}}{1-\hat{q}_{x,t-x}})\right)^2 VAR\left(\hat{Y}_x \middle| X; \underline{\hat{\beta}}\right) + o(\frac{1}{\sqrt{l_0}})$$

$$= \left(\hat{q}_{x,t-x}(1-\hat{q}_{x,t-x})\right)^2 VAR\left(\hat{Y}_x \middle| X; \underline{\hat{\beta}}\right) + o(\frac{1}{\sqrt{l_0}})$$

2.7.3 Life Expectancy Calculations

Both the period and cohort expectation of life for age x are calculated in Chapter 7 following Chiang (1968:192; 1984:161) and Pollard et al. (1974:42-43):

$$e_x = \frac{T_x}{l_x} = \frac{\int_x^{\infty} l_{x+a}\,da}{l_x} \approx \frac{\sum_{a=0}^{\infty} L_{x+a}}{l_x} = \frac{\sum_{a=0}^{\infty}\left(\frac{l_{x+a}+l_{x+1+a}}{2}\right)}{l_x} \qquad (2.7.4)$$

We have made the additional assumption here that each person who dies during any year lives half of the year on average.

For cohort life expectancies, T_x equals the total person years lived by the *cohort* from age x and above. L_x equals the number of years lived by the l_x survivors between ages x and x+1. For the cohort expectation of life, the expected life span for a 40 year old in 1950 depends on l_x for 41 year olds in 1951, 42 year olds in 1952, and so on along the cohort.

For period expectations, T_x equals the total person years lived during the period at age x and above. We calculate the expectations by summing the L_x over a *single period* rather than a cohort. For example, the expected life span for a 40 year old in 1950 depends on l_x for all other ages above 40 *in* 1950.

The sample variance for the cohort expectation of life can be found by first expressing the life expectancy in terms of ${}_a p_x$:

$$e_x \approx \frac{\sum_{a=0}^{\infty} L_{x+a}}{l_x} = \frac{1}{2}\sum_{a=0}^{\infty}\left(\frac{l_{x+a}+l_{x+1+a}}{l_x}\right) = \frac{1}{2}\sum_{a=0}^{\infty}\left({}_a p_x + {}_{a+1}p_x\right)$$

$$= \frac{1}{2} + {}_1p_x + {}_2p_x + {}_3p_x + \ldots + {}_w p_x + \frac{1}{2}\,{}_{w+1}p_x$$

$$\approx \frac{1}{2} + \sum_{a=1}^{\infty} {}_a p_x$$

Using the Taylor expansion, and letting $g(\underline{\tilde{p}}) = e_x \approx \frac{1}{2} + \sum_{a=1}^{\infty} {}_a p_x$

$$g_i'(\underline{p}) = \frac{\partial \sum_{a=1}^{\infty} {}_a p_{x,t-x}}{\partial p_{i,t-x}} = \frac{\sum_{a=1}^{\infty} \partial({}_a p_{x,t-x})}{\partial p_{i,t-x}} = \frac{\sum_{a=1}^{\infty} {}_a p_{x,t-x}}{p_{i,t-x}} = \frac{e_x - \frac{1}{2}}{p_{i,t-x}}$$

Now, since $COV(\tilde{p}_i - p_i, \tilde{p}_j - p_j) = 0$,

$$VAR(\hat{e}_x) = \sum_{i=1} g_i'(\underline{p})^2 VAR(\hat{p}_i) + o(\frac{1}{\sqrt{l_0}})$$

$$= \left(\hat{e}_x - \frac{1}{2}\right)^2 \sum_{i=x}^{k} \frac{VAR(\hat{p}_{i,t-x}(\underline{\hat{\beta}}))}{\hat{p}_{i,t-x}} + o(\frac{1}{\sqrt{l_0}}) \qquad (2.7.5)$$

Chiang (1968:211) derives a variance for e_x in terms of ${}_a p_{x,t-x}$, which is significantly more complicated and more difficult to extract in practice than variance estimates that are based on equation (2.7.5).

3 Extracting and Exploring the Data

3.1 Introduction

Dutch deaths and end of year population numbers are provided for the years 1850 to 1991, ages 0 to 90, 95, 99 or 100 depending on the years[1]. The death data are doubly classified, meaning that the data are classified by age (at death), year of death, and year of birth (cohort). Data in this form allows generation of cohort probabilities of death. Before fitting a model to the data, it is necessary to explore the data for trends and significant patterns. The purpose of this is two-fold; firstly, it sheds light on patterns of past mortality; and secondly, it aids in establishing an appropriate parameterisation for model fitting.

In this chapter we group and analyse the data by sex, age, year and cohort. The data range from 1850 to 1991 for all ages is presented in Section 3.3, though detailed analysis is only conducted on ages 50 to 90, years 1890 to 1991 for the remainder of the chapter.

A range of important conclusions arises as a result of the analysis as follows.

When modelling the age-curve of mortality, traditional functions such as the Gompertz and Makeham fail to adequately capture the trend. Alternative functions, such as the Perks or high order polynomials in age, satisfactorily capture the shape of the mortality curve for both males and females, including the derivatives.

Both male and female mortality display periods of extreme volatility in the 19th century due to various epidemics, and two particularly catastrophic events in the 20th century: the Spanish Influenza of 1918, and the Hunger Winter of 1944-45. There is a marked discontinuity in mortality pre- and post-WWII that is an argument for modelling the surfaces separately. Mortality for both males and females displays a non-linear but smooth pattern post-WWII, related in part to the cardiovascular epidemic of the 1950s to 1970s.

---

[1] Data was obtained up to and including age 100 for years 1850-1869 and 1930-1939. For years 1870-1919 the max age was 90; for 1920-1929 and 1940-1949 the max age was 95; for 1950-1991 the max age was 99.

For both sexes, there is a definite pattern in the estimated derivatives over time that highlights the excess mortality due to the cardiovascular epidemic.

Modelling the post-WWII trend in mortality with a linear trend in year is not appropriate due to the mortality 'hump' due presumably to the coronary heart disease epidemic (Poppel et.al, 1994:7) (not to be confused with the well-known 'accident hump' which is a feature of the age distribution). Although the trend in year can be modelled using a high degree polynomial, the parameters are difficult to interpret and projections are nonsensical. Including a normal density function helps to capture the behaviour of male and female log(odds).There is evidence that the center of the mortality hump changes with age, which may imply that a cohort effect is present. This is supported by graphical inspection of the derivatives with respect to year and age. The importance of cohort effects on adult age-patterns of mortality means that cohort variation, as well as age and year effects, need to be considered when generating a model for Dutch mortality.

3.2 Generating the log(odds) from the Data.

Dutch data for males and females was provided by the Netherlands Interdisciplinary Demographic Institute (NIDI), courtesy of Ewa Tabeau. The main characteristics of the data, and how the data was collected is discussed by Ewa Tabeau et al. (1994). The data consisted of files containing:

$\tilde{\tilde{E}}(t,x)$ the realised end of year t, age x population, and

$\tilde{\tilde{d}}(t,x)$ the realised number of deaths in year t, age x.

The data was provided for years 1850 to 1990. $\tilde{\tilde{E}}(t,x)$ only extended to age 90 for the years 1860-1910.

The realised number of survivors at the beginning of year t, age x, is $\tilde{\tilde{l}}(t,x)$. As deaths are doubly classified according to both year and age, we can examine the Age-Period-Cohort (APC) structure in the data.

For the construction of life-tables, we would expect that the number of births for a particular cohort would equal the total number of deaths along the cohort. Here, however, we are dealing with empirical data, and emigration and immigration have changed cohort sizes from year to year. As such, the raw number of deaths along a cohort does not match up with the initial births $l_{t-x}(0)$. Unfortunately we do not have data on the number of Dutch migrants.

We initially derived $q_{x,t-x}$ for the Dutch data along cohorts according to the well-known life-table derivation $q_{x,t-x} = 1 - \frac{l(t+1,x+1)}{l(t,x)} = \frac{l(t,x) - l(t+1,x+1)}{l(t,x)}$ (see, for example, Chiang (1968, 1984)) . This leads to problems since this form assumes that our population is exempt from migration. Since we are dealing with calculating empirical death probabilities from data where migration has been apparent, the following approximation was used,

$q_{x,t-x} = \frac{\#\,deaths}{l(t,x)} = \frac{d_1(t,x) + d_2(t+1,x)}{l(t,x)}$, which is equivalent to $\frac{l(t,x) - l(t+1,x+1)}{l(t,x)}$ where migration has not occurred. The subscripts on the deaths refer to the classification of the deaths according to birth cohort. A subscript of 1 means that the deaths relate to the later cohort; a subscript of 2 refers to the earlier cohort. For example, $d_1(t,x)$ is the number of deaths in year t, aged x, from the birth cohort $t-x$, whereas $d_2(t,x)$ is the number of deaths in year t, aged x, from the birth cohort $t-x-1$. Figure 3.1 illustrates the partitioning of the data in a Lexis diagram.

The number of survivors at time t+1, age x+1, can be expressed in terms of the number of deaths in the previous year, the number of migrants (EM, IM), and the number of survivors at time t, age x:

$$l(t+1,x+1) = l(t,x) - d_1(t,x) - d_2(t+1,x) + IM(t,x) + IM(t+1,x) - EM(t,x) - EM(t+1,x)$$
$$= [l(t,x) + IM(t,x) + IM(t+1,x)] - [d_1(t,x) + d_2(t+1,x) + EM(t,x) + EM(t+1,x)]$$

$l(t,x)$ was approximated by taking the average of two calculations. Ignoring migration, the number of survivors at time t, age x, should be,

$$l(t,x) = E(t,x) + d_1(t,x)$$

With migration,

$$l(t,x) = E(t,x) + d_1(t,x) + EM(t,x) - IM(t,x)$$

Similarly, we could calculate $l(t,x)$ from the previous year:

$$l(t,x) = E(t-1,x-1) - d_2(t,x-1) - EM(t,x-1) + IM(t,x-1)$$

The number of survivors $l(t,x)$ was approximated by taking the average of these two calculations:

$$\frac{E(t,x) + d_1(t,x) + E(t-1,x-1) - d_2(t,x-1) + [EM(t,x) - EM(t,x-1) + IM(t,x-1) - IM(t,x)]}{2}$$

When $[EM(t,x) - EM(t,x-1) + IM(t,x-1) - IM(t,x)] = 0$ (net migration is zero along each cohort), this reduces to :

$$\tilde{\tilde{l}}(t,x) = \frac{\tilde{\tilde{E}}(t,x) + \tilde{\tilde{d}}_1(t,x) + \tilde{\tilde{E}}(t-1,x-1) - \tilde{\tilde{d}}_2(t,x-1)}{2}$$

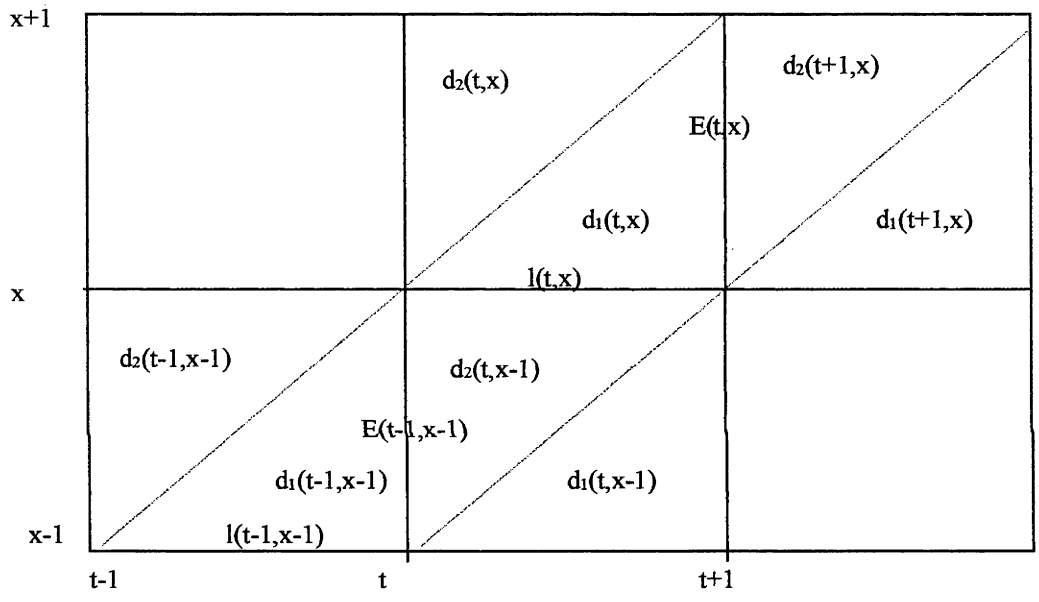
The above empirical form of $q_{x,t-x}$ is believed to be a good non-parametric estimate of the true conditional probability of death along a cohort in the absence of migration. The

distribution of empirical $q_{x,t-x}$ can be approximated by the distribution of the life-table function

$$\tilde{\tilde{q}}_{x,t-x}=1-\left[\tilde{\tilde{\ell}}(t+1,x+1)/\tilde{\tilde{\ell}}(t,x)\right]$$

which has been derived in Chapter 2.

Figure 3.1 Lexis diagram for the generation of $\tilde{\tilde{l}}(t,x)$



Calculation is subsequently made of numerical estimates of the time and age specific probabilities of death and the empirical mortality surface

$$y(t,x)=\ \text{logit}\,\tilde{\tilde{q}}_{x,t-x}=\log\left(\frac{\tilde{\tilde{q}}_{x,t-x}}{1-\tilde{\tilde{q}}_{x,t-x}}\right)=\log(\text{odds of death between } (t,x) \text{ and } (t+1,x+1)$$

referred to as log(odds).

3.3 Analysing log(odds)

As discussed in the previous section, the measure used to study mortality in this report is based on cohorts rather than periods. The main reason cohort measures are used is that they *can* be used. The Dutch data provided is rare in the sense that it is doubly classified and enables extraction of rates along cohorts.

Mortality for individuals naturally emerges along cohorts and so cohort measures enable the study of health processes from the point of view of consecutive human experience. It is widely acknowledged that examination of patterns in mortality using cohort probability measures allows a truer picture of the mortality process, uncovering complexities that may be concealed in the period approach (Ryder,1980; Davies,1987). Quoting Poppel et al:

"... cohorts and not period measures make it possible to study the mechanisms that underlie the regularity in age profiles of mortality." (1994:2).

Life expectancy measures are usually based on periods. Although period life expectancies are informative about the cross-sectional state of mortality for a group, they do not provide an indication of years of expected life in the presence of changing mortality. Cohort life expectancy, although requiring future projections of mortality, provides a more realistic assessment of the expected years of life for an individual. Both period and cohort life expectancies are given in Chapter 7 for the Dutch models generated in Chapters 4 to 6.

The analysis of the data that proceeds is carried out using cohort measures of $q_{x,t-x}$, $\log q_{x,t-x}$ and log(odds). Analysis of the data using period measures of mortality would lead to different conclusions, and would fail to capture many of the cohort effects that are discussed in this chapter.

A perspective plot of log(odds) for Dutch females and males is given in Figure 3.2 and Figure 3.3 respectively.

There are a number of features that immediately stand out in the two perspective plots: there is obvious volatility in log(odds) for the upper ages, particularly between 1850 and 1860 and between 1920 and 1940; and there are apparent periods of excess mortality, notably around 1918/1919 and 1944/1945. Cross-sectional plots have been generated in order to observe patterns in the log(odds).

Figure 3.4 displays the typical features of a human mortality curve:

- high mortality at age 0 followed by a decline reaching a minimum around age 10;
- a relatively steep increase in the late teen years (typically associated with accidents and suicide), though this is only present in the recent past;
- a slow rise between ages 20 and 50, followed by an increasing rate at higher ages.

Mortality appears similar for males and females up to WWII, after which there has been a divergence. The mortality differential between males and females is discussed in some detail in a later part of this chapter. A selection of the mortality curves is presented for comparison in Figure 3.5 and Figure 3.6.

Of note is the dramatic decrease in mortality across all ages over time. The pattern over time is examined in detail in a later part of this chapter.

A feature of the log odds mortality curve at all years is the apparent almost linear trend in age starting around age 50. The subsequent analysis of the data is restricted to ages 50 and above.

Figure 3.2 **Dutch Females. Log(odds). Ages 0 to 100. Years 1850 to 1991.**

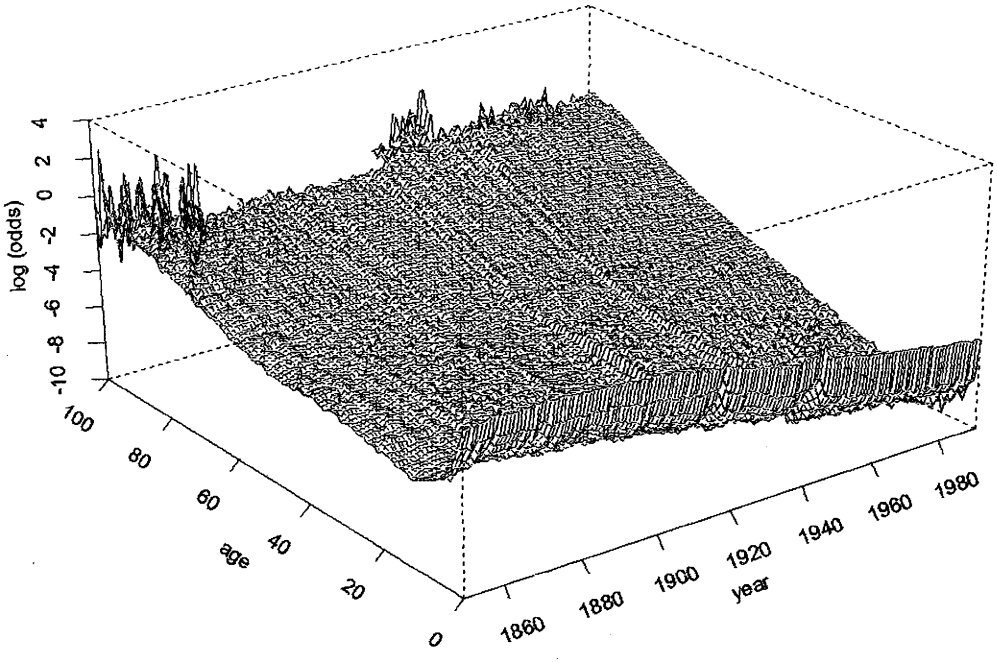


Figure 3.3 **Dutch males. Log(odds). Ages 0 to 100. Years 1850 to 1991.**

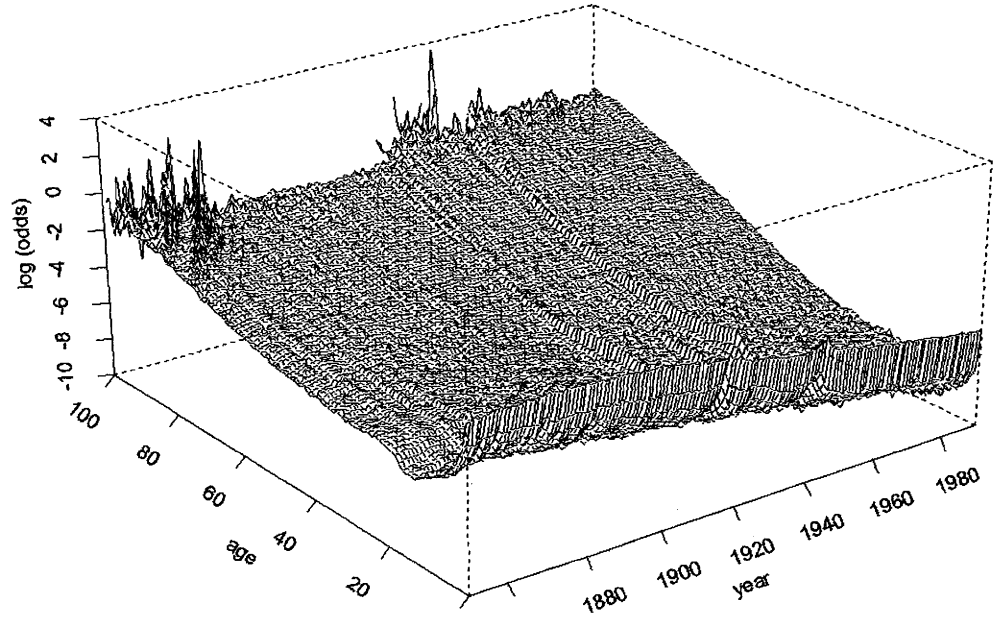


Figure 3.4 **Dutch Males and Females. Cross-section of log(odds) for selected years. (Solid line=females. Dotted line=males. All plots are on the same vertical scale)**

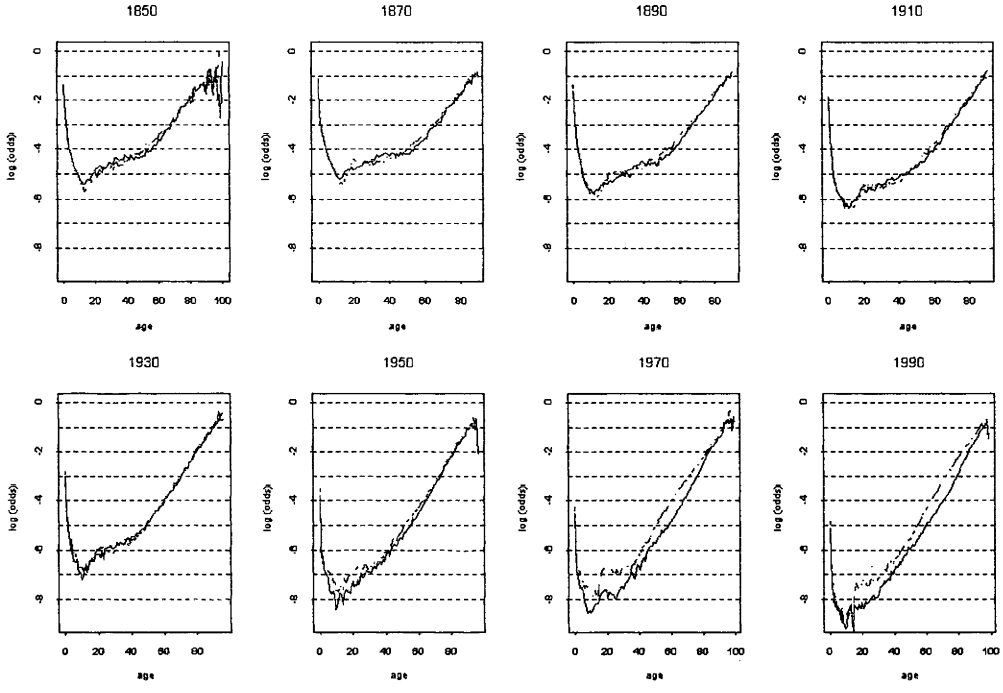


Figure 3.5 **Dutch Females. Cross-section of log(odds) for selected years. Ages 0 to 100.**

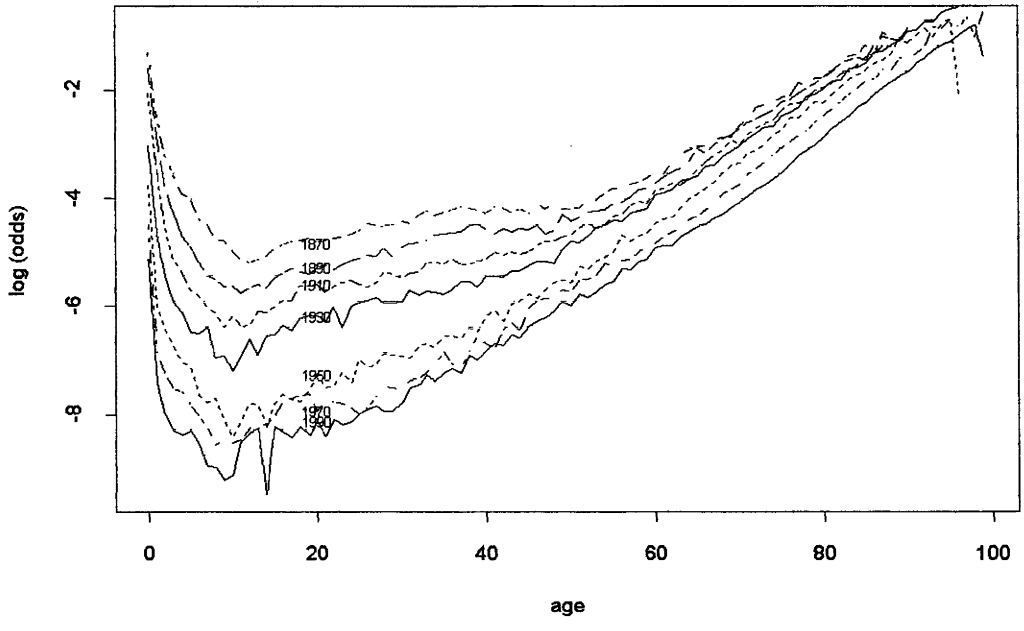
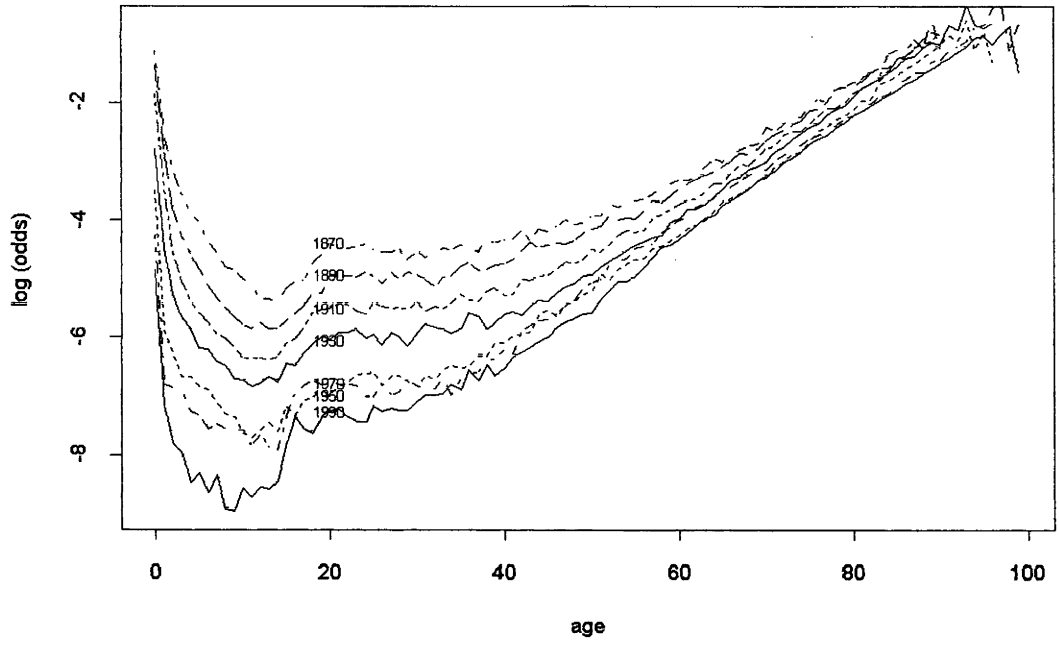


Figure 3.6 **Dutch Males. Cross-section of log(odds) for selected years. Ages 0 to 100.**



3.4 Male and Female log(odds). Ages 50-100.

Figure 3.7 and Figure 3.8 present the female and male log(odds) for ages 50 to 100. The same information is conveyed in contour plots (Figure 3.9 and Figure 3.10) which display the general decrease in mortality rates over time. The figures also display the volatility in log(odds) that is apparent prior to 1946, particularly for ages above 90. The gaps in the plots indicate periods where data are not available.

A feature of both the perspective and contour plots is the periods of excess mortality, particularly that corresponding to the winter of 1944-45.

The fluctuations that are present in the mid to late 1800's were caused by epidemics including cholera in 1853, 1854 and 1855 and measles and typhus also in 1855 (Poppel et al.,1994:5). Following 1871, variability in mortality was reduced, though variability is still quite significant until post WWII.

Due to the excessive volatility in log(odds) for 1850 to 1890, the questionable accuracy of the number of deaths and exposure populations for these years, and the absence of data in the high ages, subsequent analysis of the log(odds) has been constrained to the years 1890 to 1991, ages 50 to 90. This time period encompasses an entire century and so allows a cohort to be tracked through completely.

The male and female log(odds) for 1890 to 1991, ages 50 to 90 are examined in detail in the remainder of this chapter with reference to period, age and cohort. This enables identification of patterns and aids in determining an appropriate parametric structure for the regression modelling performed in the Chapter 4.

Figure 3.7 **Dutch Females. Log(odds). Ages 50 to 100. Years 1850 to 1991.**

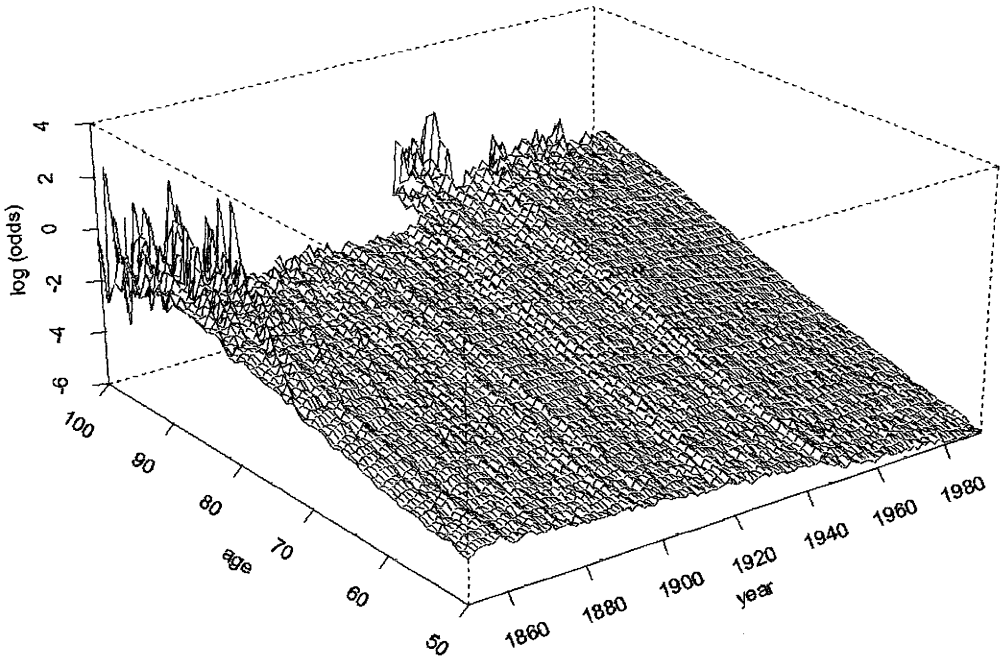


Figure 3.8 **Dutch Males. Log(odds). Ages 50 to 100. Years 1850 to 1991.**

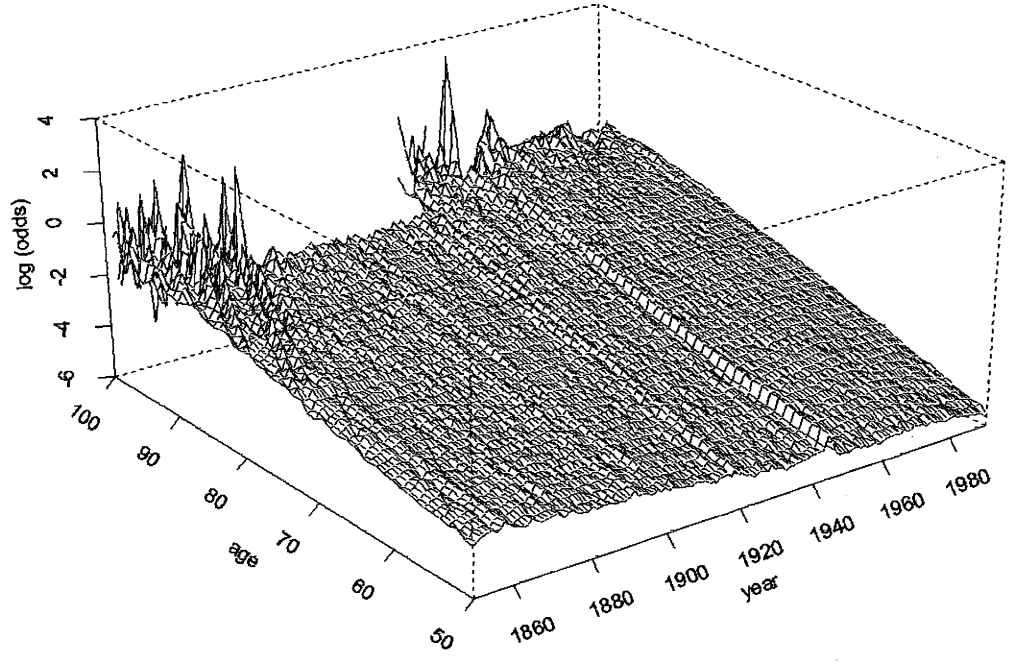


Figure 3.9 **Dutch Females. Contour plot of Log(odds). Ages 50 to 100. Years 1850 to 1991.**

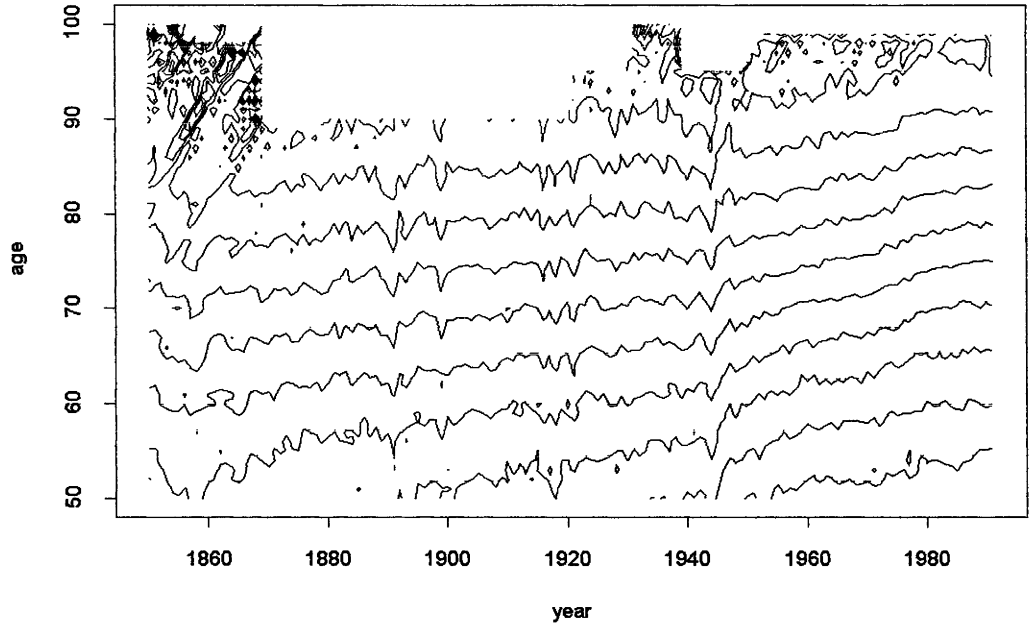
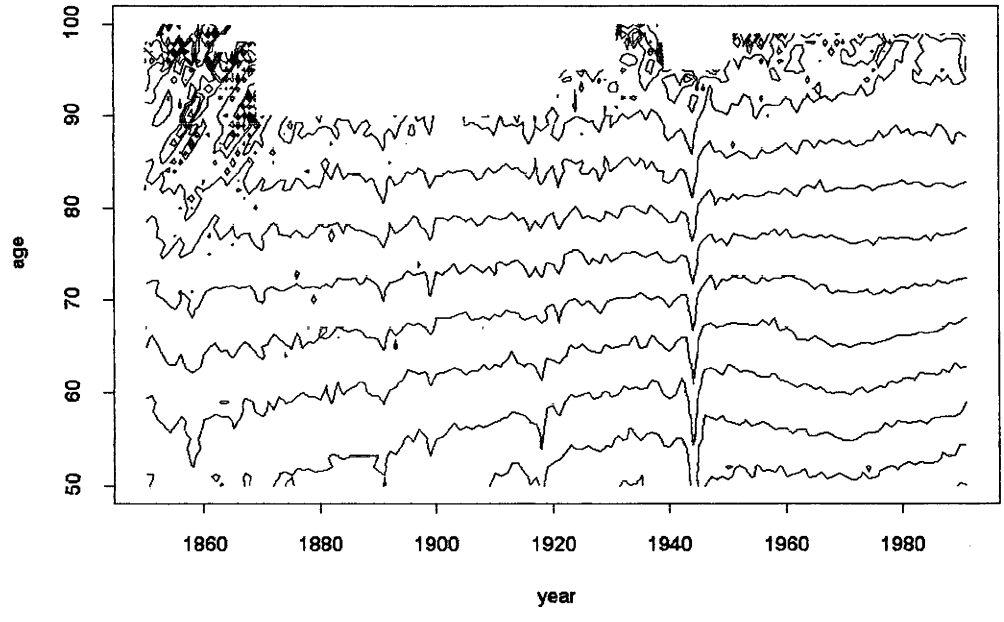


Figure 3.10 **Dutch Males. Contour plot of Log(odds). Ages 50 to 100. Years 1850 to 1991.**



3.5 Analysis of Age Patterns

At first glance, log odds appears linear between ages 50 and 90 over the range of years. Examination of the residuals following a linear fit, however, reveals that there are subtle curvatures at the higher ages that do not conform to linearity. Although this is not so noticeable for earlier periods this century, it has become very apparent in the recent past.

For the sake of consistency with other studies, $\log(q_{x,t-x})$ rather than log(odds) is fitted using a Gompertz and Makeham distribution for the year 1980 to illustrate the failure of the traditional distributions to capture the age pattern of mortality[2].

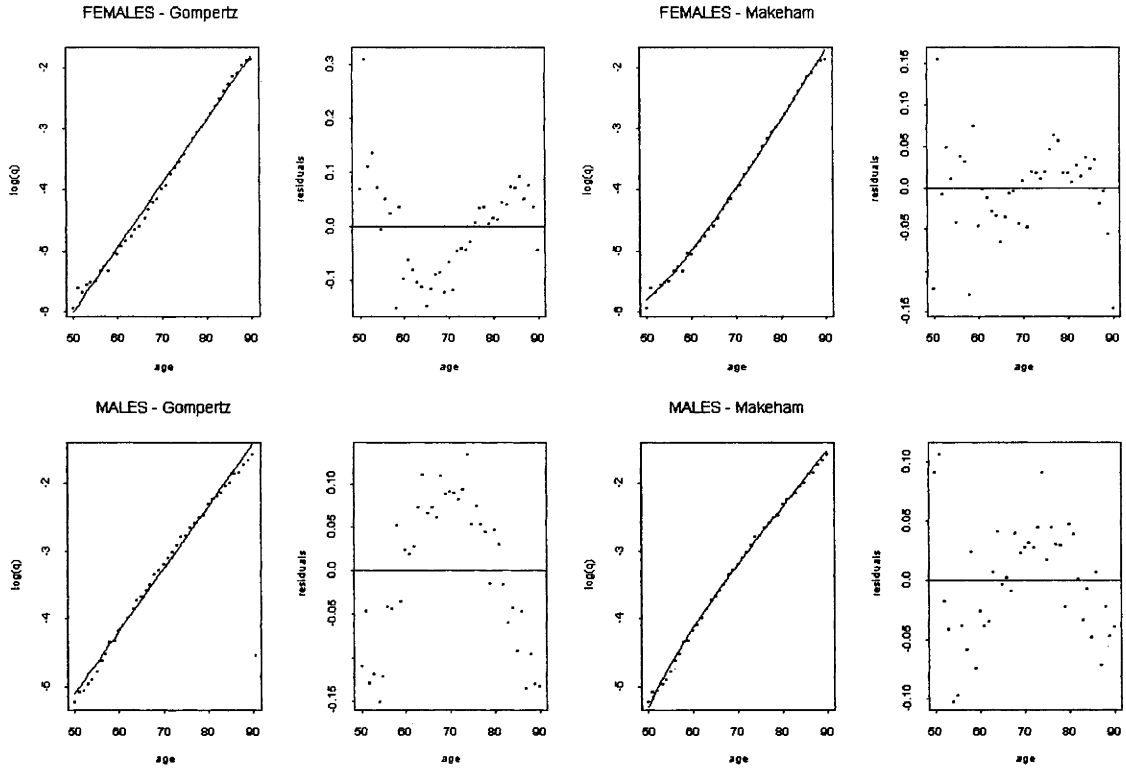
Both the Gompertz and Makeham fail to capture the age pattern of mortality (see Figure 3.11). In particular, for Dutch females both functions tend to under-fit for the older ages, and over-fit between ages 60 and 70. For Dutch males the functions fail to capture the curvature of the mortality rates with age, with both functions and particularly the Gompertz under-fitting between 60 and 80, and over-fitting for the other age ranges

Although the Makeham fit is poor for both sexes, the difference in shape for the male and female fit raises some interesting questions. The rate of change of the male curve seems to be decreasing in general, whereas that of the female curve appears to be increasing. The extent of the difference between the male and female age pattern of mortality and the non-linearity in the data can be revealed by examining the derivatives of the mortality curve. This idea was developed by Horiuchi and Coale who noticed a consistent pattern in the rate of change of female mortality (1990). In particular, they noticed a maximum rate of change followed by a decrease around age 75 in the mortality rates of a number of developed countries. Exploration of the derivatives is taken up in Chapter 5.

---

[2] Rather than modelling the response q, log(q) was regressed on the logarithm of the Gompertz and Makeham expressions using non-linear least squares regression with a Gauss-Newton algorithm. For example, the Gompertz fit was log(q)=log(B)+ux

Figure 3.11 **Dutch Females and Males. Gompertz and Makeham fit to log(q(x)) and residuals for 1980. Ages 50 to 90.**



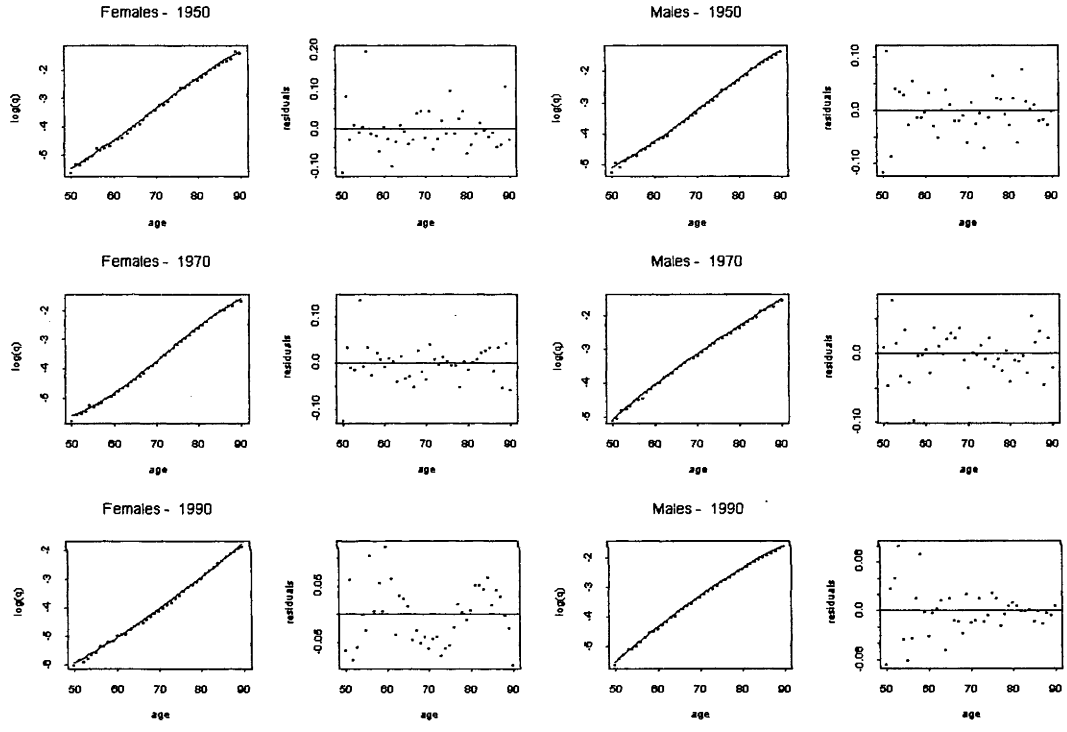
As discussed in Chapter 1, the lack of fit of the standard distributions to aggregate mortality may in part be due to the discrepancy between individual and aggregate mortality in the presence of heterogeneity. In other words, although a particular distribution may not be suitable for the aggregate population, it may in fact still represent individual mortality (Manton et al. (1986); Vaupel et al. (1979), Vaupel and Yashin (1985)). Horiuchi and Coale (1990) show that the assumption of a Makeham distribution for individual mortality over the age range, coupled with gamma distributed frailty, leads to cohort mortality following a Perks distribution. The Perks distribution is:

$$\mu_x = \frac{A + Be^{ux}}{1 + Ce^{ux}}$$

where A, B, C and u are parameters, x is age, and μ_x is the force of mortality at age x (Horiuchi and Coale, 1990).

Although the Perks distribution is applicable to cohorts rather than periods we apply it to periods for illustration of its superiority over the Gompertz and Makeham distributions (see Figure 3.12). As illustrated later in this chapter, the Perks distribution also fits individual cohorts fairly well, though period effects distort the results reducing the interpretability of the parameters.

Figure 3.12 Dutch Females and Males. Perks fit and residuals for selected years. Ages 50 to 90.



The Perks model performs well with the exception of the most recent years for the females. The advantage of the Perks over other distributions is its ability to capture the changing derivative in the mortality curve at older ages. The derivatives emerging from the Perks fit conform closely to the smoothed derivatives, except for the most recent years for females. Horiuchi and Coale (1990) favour the Perks since it has a theoretical foundation based on a plausible mixing distribution. Despite this, there has yet to be any proof that the individual mortality distribution or indeed the frailty distribution that underlie the Perks model are correct. In the absence of justification for the Perks model, an alternative parameterisation

can be sought that captures the age-pattern of mortality, including the curvature of the derivative.

A polynomial in age can be shown to capture the age pattern and derivatives closely over most of the range considered. Third degree or even second degree polynomials appear suitable for the bulk of the mortality surface, whereas inclusion of an additional degree is necessary to capture the fit for females for the most recent years. Modelling the log(odds) or $\log(q_x)$ with a polynomial in age has previously been suggested in the literature. Tabeau (2001) provides a comprehensive review of parameterisation functions for mortality, including polynomial and non-polynomial parameterisations. $\log(q_x)$ as a second degree polynomial in age may be written:

$$\log(q_x) = D + Bx + Cx^2$$

Modelling log(odds) or $\log(q_x)$ by a higher order polynomial in age is simply an extension of this model. Fitting with a higher degree improves the quality of the female fit in general more so than the males, particularly in the recent past.

The analysis undertaken in this section points to a number of conclusions:

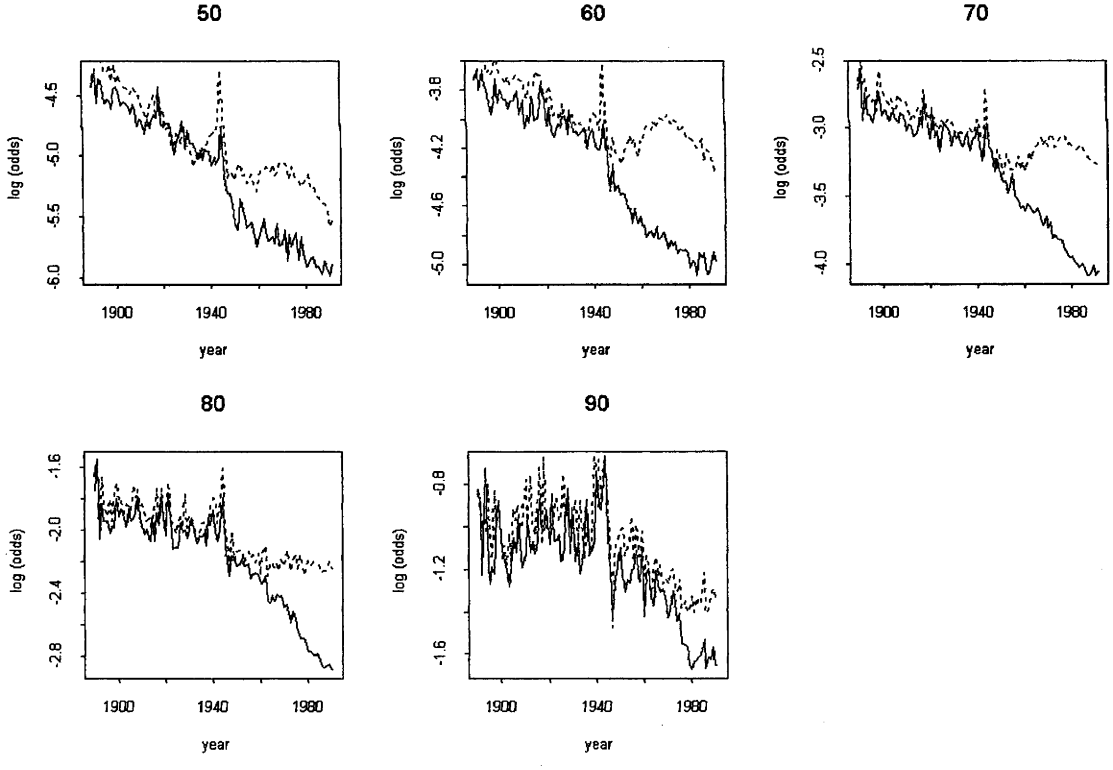
- The age pattern of mortality along different periods differs for males and females;
- traditional functions such as the Gompertz and Makeham do not adequately capture the age-pattern by period, particularly for more recent periods;
- the Perks function, which has a plausible theoretical basis, captures the shape of the mortality curve for both males and females well, including the derivatives, with the exception of the most recent female periods;
- polynomials of degree 2 and 3 capture the mortality curve and derivatives for males, whereas polynomials of degree 4 are needed to adequately model recent female mortality.

The accuracy of fit of the polynomials in age across all periods provides a starting point to parameterisation of the entire mortality surface. This is taken up in Chapter 4.

3.6 Analysis of Time Trends

Figure 3.13 displays the change in log(odds) for males and females between 1890 and 1991 for selected ages. A feature of the bulk of the ages is the similarity between the sexes and the relatively high volatility prior to WWII, and the discontinuity that follows. In particular, there appears to be a sudden drop in mortality rates following 1945 and a change in slope for some ages, particularly for females. The post-WWII drop in mortality was likely caused by a variety of factors, including the introduction of antibiotics, DDT and improvements in public health.

Figure 3.13 **Dutch Females and Males. Cross-section of log(odds) for selected ages. Years 1890 to 1991. (Solid line=females; dotted line=males)**



Males exhibit a 'hump' in mortality between 1950 and 1980 associated with unfavorable trends in cardiovascular disease (Poppel et.al, 1994:7), and it is shown later in this section that a similar, though much smaller hump is present for females.

In addition, there are a number of individual years in the data marked by excess mortality. The peak at 1918 due to the Spanish Influenza is relatively small for ages 50 and above, though is much more apparent for the younger ages[3]. The more noticeable peak is at 1944-45 from the 'Hunger Winter' associated with the blockading of parts of the Netherlands by Germany during WWII.

A clearer indication of the periods of excessive mortality can be obtained by smoothing the log(odds) for each period and then examining the variation along different age profiles. A selection of cross-sections at different ages for the smoothed log(odds) shows what was previously stated: that the two most apparent outliers relate to 1918 and 1944/45. Figure 3.14 also points to excessive mortality at 1891, 1893, 1899, 1916, 1921, 1928 and 1940. In addition, 1947 is marked since it may indicate a period of reduced mortality. Most of these periods show excessive mortality across the older ages but are less apparent for ages 50 and 60. A cross-section plot for females is not reproduced here as it exhibits similar features as Figure 3.14.

The reduced mortality for 1947 may be the direct result of increased death rates in adjoining years due to the Hunger Winter of 1944/45 leading to a selection effect. (see for example, Vaupel et al. (1988) for a discussion of the interaction between debilitation, recuperation and selection).

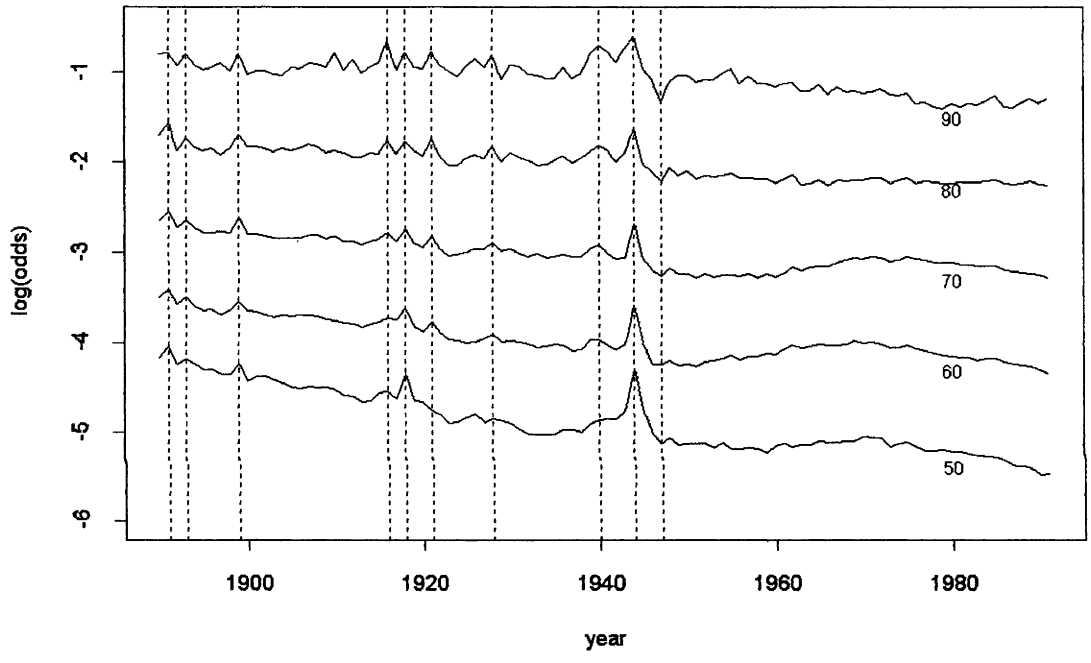
In addition to the high mortality rate at 1940, male log(odds) appear high for subsequent years leading up to 1945 for age 50 as a result of the impact of WWII. For males less than age 50 (particularly ages 15 to 40), WWII had a much more devastating impact. Poppel et.al (1994:5) discuss the crisis of mortality for the Dutch population during World War II,

---

[3] Poppel et al. remark that the population at risk for 1913-1919 is underestimated in the Dutch data, resulting in a slight overestimate of the mortality rates for these years

and note that the mortality rates for 1941-1945 are seriously underestimated as they do not include deportations to Germany. Those deported to Germany and Poland were officially registered as 'removed' from the population and so did not influence the mortality statistics for the Netherlands.

Figure 3.14 **Dutch Males. Cross-sections of period-smoothed log(odds). Years of particularly high or low mortality are indicated by dotted vertical lines.**



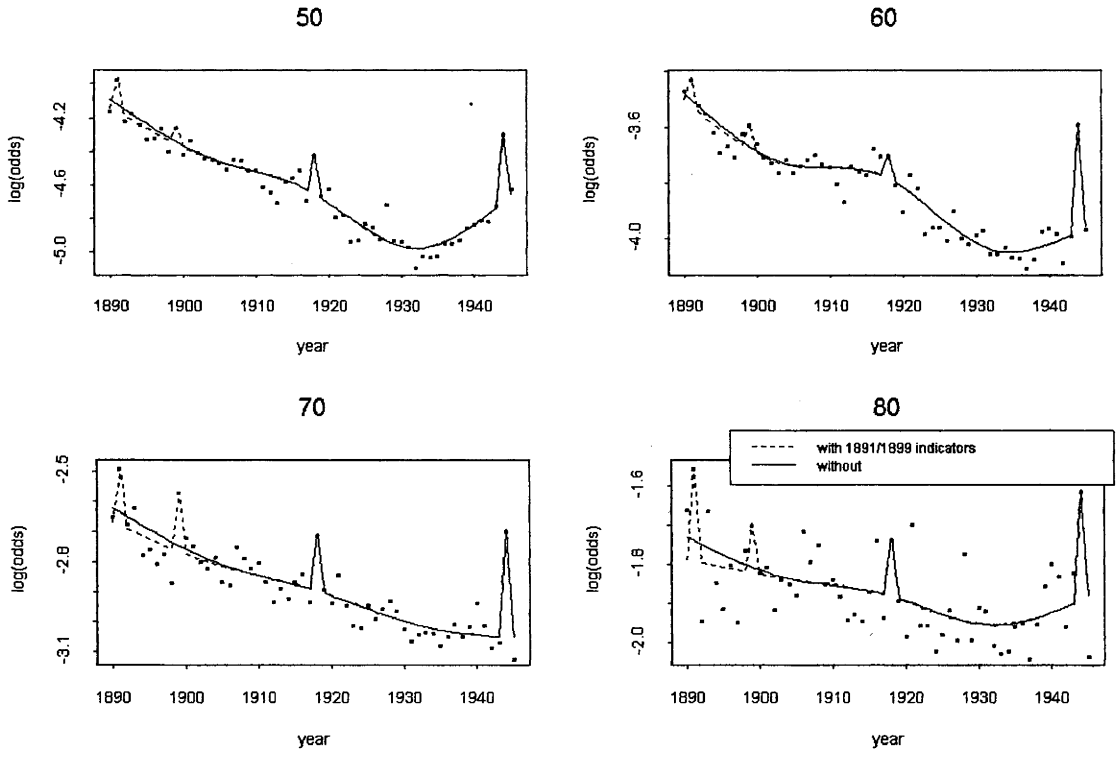
In order to build up a picture of the entire surface, natural splines were fitted to each age between 50 and 90 for the range 1890 to 1991. The discontinuity between pre- and post-WWII mortality means that there is a strong argument for analysing the pre- and post-war periods separately. A spline curve is fitted up to age 1945 and a separate curve is fitted post 1945 to capture the discontinuity.

Indicators are included for those years with particularly high mortality as a consequence of dramatic historical incidents; namely, 1918 and 1944. In addition, it could be argued that an indicator should be included for 1947 due to the apparent drop in mortality and the

plausibility of this being an explainable effect. The inclusion of an indicator for 1947 has an almost negligible effect on the shape of the splines for both males and females and so has been omitted. In the absence of linking extreme events with the periods of excess mortality in the other periods identified in Figure 3.14, it would seem sensible to include these points as part of the smoothed surface rather than giving them separate attention.

The results for males for a selection of ages for 1890 to 1945 are given in Figure 3.15. The effect on the spline fit of including indicators for 1891 and 1899 is displayed by the dotted line. For the sake of brevity, the equivalent plots for females have not been included.

Figure 3.15 **Dutch Males. Cross-section of log(odds) and natural splines (df=4) for selected ages. Years 1890 to 1945.**



When major periods of excess mortality are accounted for, the decline in mortality for females over the period between 1890 and 1945 has been almost linear. Indeed, a linear term in year appears to adequately capture the mortality decline for most ages in this range. Males, on the other hand, particularly between the ages of 50 and 60, have a trend in

mortality that is non-linear. As seen in Figure 3.15, there is an increase in log(odds) from the mid 1930's due to the impact of WWII. This is apparent across most ages between 50 and 90. It is clear that a linear term in year will not sufficiently capture the pattern of male mortality pre-1946 unless numerous indicator variables are inserted for the periods of excess mortality surrounding 1918 and 1944.

Despite the linear trend for females pre-1946 and the appearance of linearity for much of the surface, a linear decline oversimplifies the trend in mortality post-WWII. Although there is a definite decline in log(odds) for all ages between 50 and 90, the rate of decline varies across ages and along years. The empirical log(odds) for each age are smoothed in order to elucidate trends over time. Figure 3.16 displays the log(odds) from 1946 to 1991 with natural splines fitted to a selection of ages for males and females.

Figure 3.16 **Dutch Males and Females. Cross-section of log(odds) and natural splines (df=6) for selected ages. Years 1946 to 1991 (Males = dots; Females = crosses)**

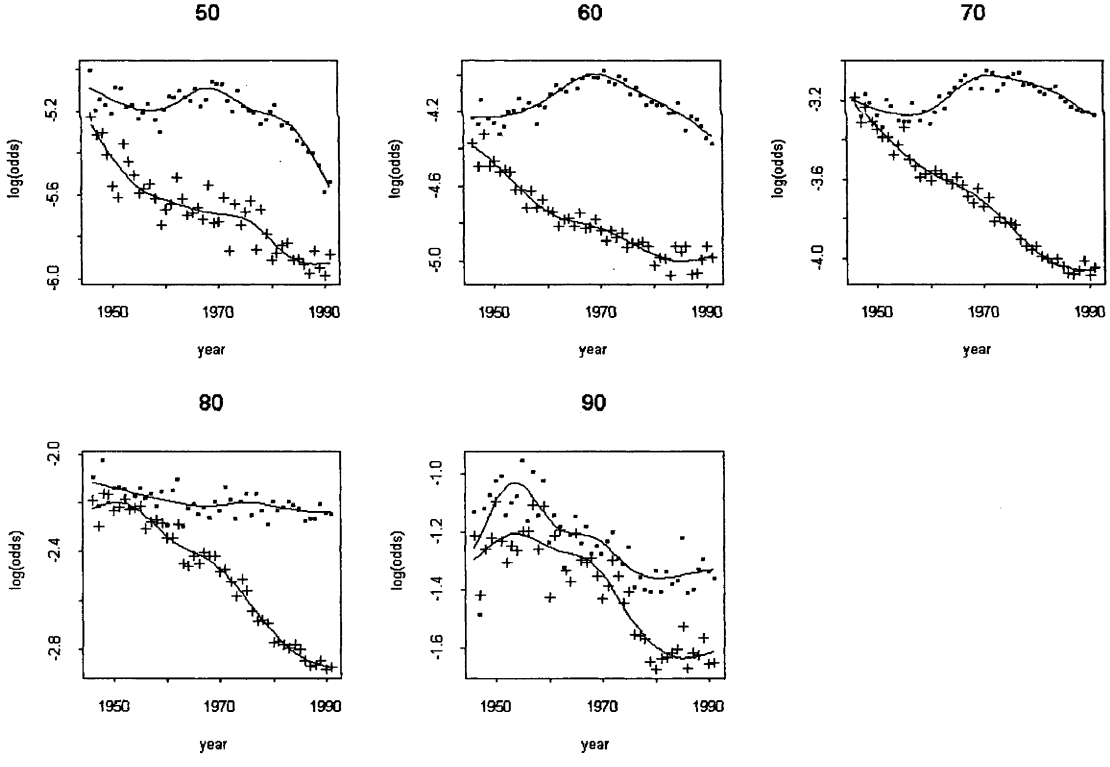


Figure 3.16 reveals some interesting trends. For females at ages 50 to 70, there appears to be a change of slope in log(odds) starting around 1960. Following 1970, mortality begins to fall for these ages, until the mid-seventies, when the rate of change appears to decrease once again. The pattern for the older ages is different, with log(odds) displaying an almost logistic trend. For males, the non-linearity is much more apparent, particularly for the middle ages.

Besides the obvious difference due to the magnitude of the hump in mortality, the differential between the log(odds) for the sexes is startling. A comprehensive analysis of the sex differential requires cause-specific data and knowledge of the social and cultural networks of the Netherlands, and is, therefore, not undertaken here. A number of researchers including Lopez (1983) discuss European and Worldwide trends in the sex differential and can be referred to for interest.

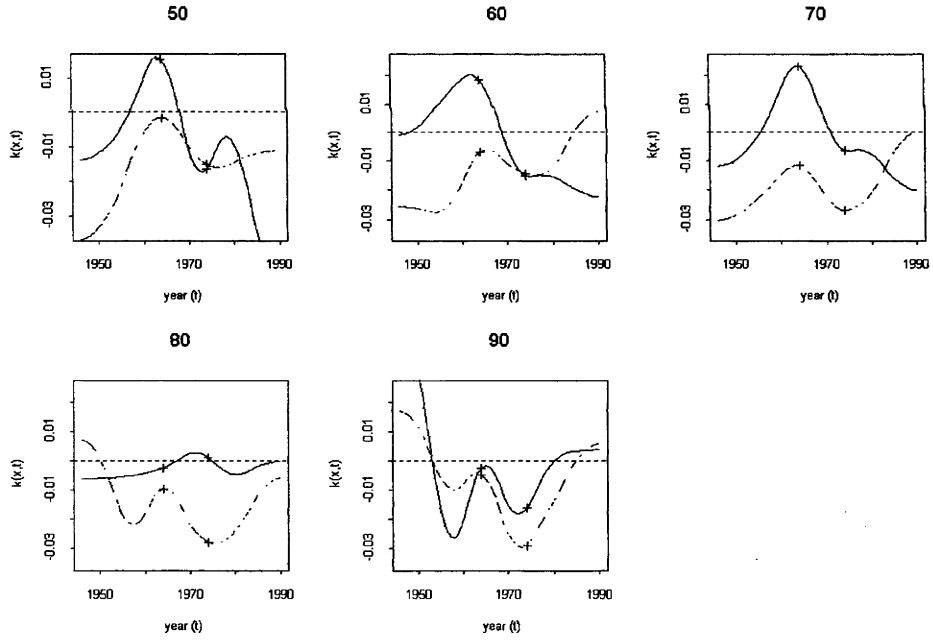
By considering changes in life expectancy, Poppel et al. (1994:9) show that Dutch male and female mortality started to diverge as early as the 1880s, with significant changes in direction at 1898, 1927 and 1975. Much of the early deviation in life expectancy is due to widening male excess mortality in the first year of life. The change in direction in 1975, leading to a narrowing of the differential between ages 50 to 70, is apparent from Figure 3.16, suggesting that the more recent changes in the life expectancy differential are due to changes in the older ages, rather than the younger ones. Indeed, one of Poppel et al's major conclusions is that:

"The recent data suggest that higher age groups participate most in male excess mortality. The significance of infant mortality has completely disappeared... Effects of age 80+ on the total difference between men and women increased to 13% in 1991 and the contribution of this age has never been so high" (1994:13).

The implications of the changing differential to projections of mortality are raised in Chapter 7. In light of the non-linearity that is present for both males and females, derivatives of the log(odds) with respect to year were estimated in order to scrutinise the trend in mortality since 1946. The derivatives were approximated by taking the first

difference between successive smoothed log(odds) along each period, which is equivalent to taking the log(ratios of the odds). The log(odds) were smoothed by using natural splines with six degrees of freedom[4] (see Figure 3.17). The turning points in the approximated derivatives for males and females are almost identical to each other. Both sexes have a maximum derivative near 1964 (corresponding to the first marker in Figure 3.17). For males the maximum derivative for ages 50 to 70 is high and positive, reflecting a steep increase in mortality; for females the corresponding derivatives are close to zero, reflecting a leveling out, rather than a real increase in mortality. The peak mortality for males, corresponding to the top of the 'hump', and identified by the year at which the male derivatives reach zero, is approximately 1968-1970, though there is some evidence that this changes with age.

Figure 3.17 **Dutch Males and Females. Approximate derivatives with respect to year of spline smoothed log(odds) (df=6) for selected ages. Years 1946 to 1991. (Solid line=males; dashed line=females).**



[4] Derivatives approximated from splines with four and eight degrees of freedom were compared to those with six degrees of freedom. It was felt that the derivatives estimated from the smoothed log(odds) with six degrees of freedom sufficiently captured the general shape of the trend.

Both sexes undergo a minimum derivative near 1974 (corresponding to the second marker in Figure 3.17). For ages 50 and 60, the derivatives are almost identical for both sexes, meaning that mortality is decreasing at a similar rate. For ages 70 to 80, the rate of decrease for females is greater than that for males.

For males, following 1974, there is a slight increase in derivative over time, particularly between ages 50 and 60, corresponding to a leveling out in the log(odds). This is followed by a dramatic decrease starting in the early 1980s for 50 to 70 year olds. In contrast, in the decade leading up to 1991, the rate of decrease in female mortality has slowed, although the rate itself may be leveling out.

For 80 to 90 year olds, despite the different magnitudes of mortality, and despite the differing rates of change since WWII, derivatives in the last years are similar between the sexes. The sharp drop in log(odds) for males aged ninety around 1975 may be related to a cohort effect in addition to possible period effects. This is explored in the Section 3.7.

Distinguishing between period and cohort effects in the instance of explaining the 'hump' is not straightforward. The fact that the peak mortality is around 1968-1970 for ages 50 to 75, and since the turning points in the derivatives are at similar years, the 'hump' appears to be predominantly a period effect. However, as will be discussed in Section 3.7, there is evidence that the derivatives also vary according to cohort. In particular, the derivatives appear to become positive soon after the 1885 cohort.

The analysis of the derivatives shows that a linear term in year will not suffice for post-WWII. A polynomial of high enough degree will capture the trend in time, but interpretation of the parameters becomes difficult as the degree increases and higher degree polynomials when projecting are not feasible beyond the range of the fitted model.

An alternative approach is to model the cardiovascular hump, which is particularly present in male mortality, by including a separate non-linear function. The inclusion of a separate function in this instance seems sensible given that the coronary epidemic essentially represents a period of excess mortality that should be treated like an extreme event.

Following Heathcote and Higgins (2001a), a possible function is a normal density estimator with particular mean and variance.

By using an iterative procedure, an optimum mean and standard deviation were selected for the male mortality hump. The process involved fitting a model with linear term in year to the post-1945 data for different choices of mean and standard deviation in the normal density term. The best model used mean 1972 and standard deviation 9, and was selected based on the R-squared value. Using a similar approach for females, the optimum normal density term had mean 1959 and standard deviation of 12. Rather than being used explicitly for the cardiovascular hump as it had been for males, the normal density term for females captures the general non-linear behaviour of the observed log(odds). If a quadratic in year rather than a linear term in year is used in the female model, a normal density term with mean 1968 optimises the female fit.

As seen in Figure 3.18, the normal density is not sufficient to capture many of the non-linearities in mortality, however, it is a means of producing a reasonable fit without relying on higher-order polynomials in year.

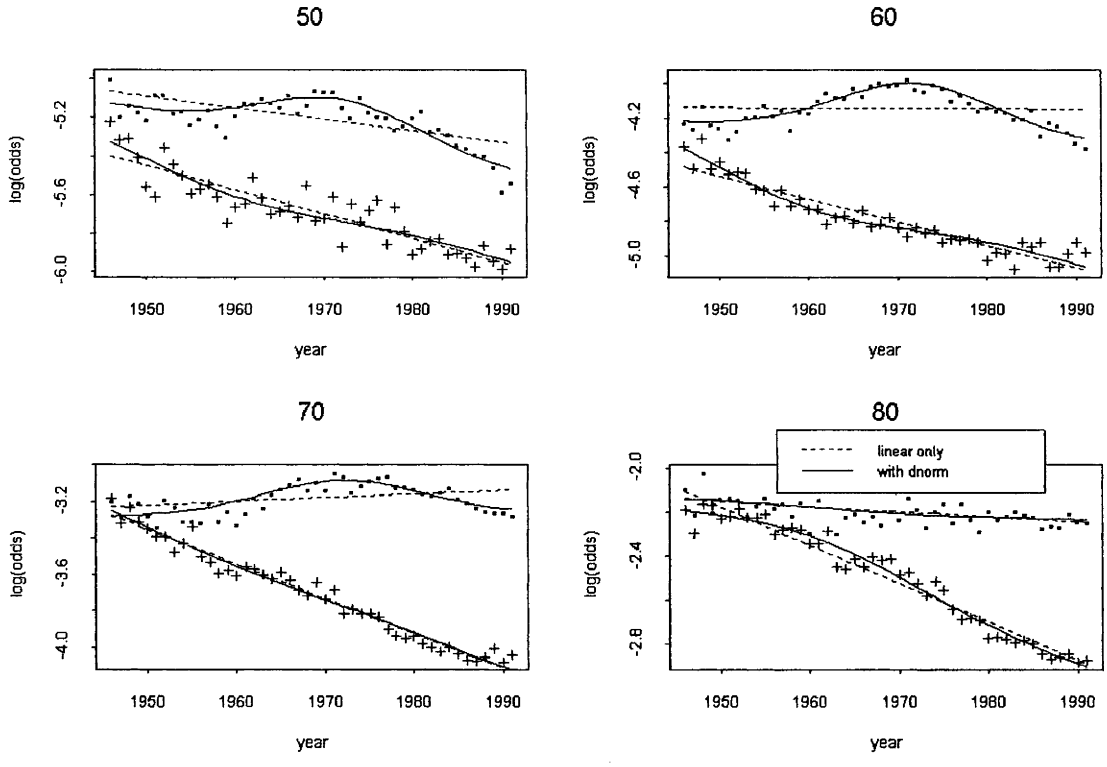
To adequately model the trend post-WWII a more complicated function is required, such as a third degree polynomial in year in addition to the normal density function [5]. Although this fits the data well, the usefulness of such a model is questionable. The inclusion of high degree polynomials discounts the plausibility of using the model for projections, though it can still be useful in producing a smooth surface for descriptive purposes and would enable examination of the variability of the residuals. Another way to deal with the non-linearities in trend is to model the surface by piecewise components.

The approach taken in the Chapter 4 for modelling the mean and variance for male and female log(odds) involves utilising polynomials in age and year and a normal density term as discussed above.

---

[5] A second degree polynomial in year in addition to the normal density function seems adequate for females for most ages

Figure 3.18 **Dutch Males and Females. Cross-section of log(odds) and linear trend in year with and without normal density function. Years 1946 to 1991. MALES (dots): mean=1972, standard deviation=9. FEMALES (crosses): mean=1959, standard deviation=12.**



As a result of the analysis in this section, a number of conclusions are reached:

- Both male and female mortality display periods of extreme volatility in the 19th century due to various epidemics, and two particularly catastrophic events in the 20th century: the Spanish Influenza of 1918, and the Hunger Winter of 1944-45;

- there is a strong argument for modelling pre-and post-WWII mortality surfaces separately due to a dramatic discontinuity in mortality across all ages;

- since WWII, Dutch male and female mortality for 50 to 90 year olds has diverged considerably, though the difference in mortality rates has narrowed in the most recent decade studied;

- both male and female log(odds), and male in particular, underwent an increase in the rate of change. For ages 50 to 70, both male and female log(odds) have a local maximum rate of change near 1964 and a corresponding minimum near 1974. The center of the mortality 'hump' is approximately at 1968, though this appears to change with age;

- Modelling the post-WWII trend in mortality with a linear trend in year is not appropriate due to the mortality 'hump'. Although the trend in year can be modelled using a high degree polynomial in year, the parameters are difficult to interpret and projections are nonsensical. Including a normal density function captures the behaviour of male and female log(odds) to some degree, and reduces the need for a higher order polynomial in year;

3.7 Analysis by Cohort

The analysis to date has involved examination of trends in mortality by period and age. Extraction of cohort effects is confounded by period and age effects, though exploratory analysis can help identify plausible phenomena.

The problem with identification in models with age, period and cohort effects has been discussed in Chapter 1. There is substantial literature on the construction and analysis of Age-Period-Cohort models, and in particular, there are a number of papers that take up the analysis of period versus cohort effects when doubly-classified data are available (see, for example, Tabeau et al. (2001)). Although the Dutch data has been obtained in this form, we leave the comparison of period and cohort derived rates to other researchers.

We proceed in the analysis in a similar vein to the previous two sections; by extracting the data along cohort for males and females and examining the trends.

The Perks model is founded on the age pattern of mortality for an individual over their lifetime and the frailty distribution of a cohort of individuals with the same 'base' mortality. As such, the natural fit for the Perks model is along cohorts, not periods, as had

been carried out in Section 3.6. Fitting the Perks model to cohorts is confounded by period effects that distort the shape of the fit unless specifically allowed for. In the case of extraordinary events such as the Spanish Influenza of 1918 indicator variables can be inserted.

In the event of more subtle period effects, it may be difficult to account for these effects when modelling along cohorts. An example of this is shown in Figure 3.19 which compares four cohorts each separated by five years. Figure 3.19 shows that the mortality for a 60 year old in 1950 is higher than that for a 60 year old in 1960 which in turn is superior to a 60 year old in 1965. If mortality has decreased linearly over time, in the absence of a cohort effect, one would expect that mortality for a 60 year old in 1965 would be lower than the mortality for a 60 year old in 1950.

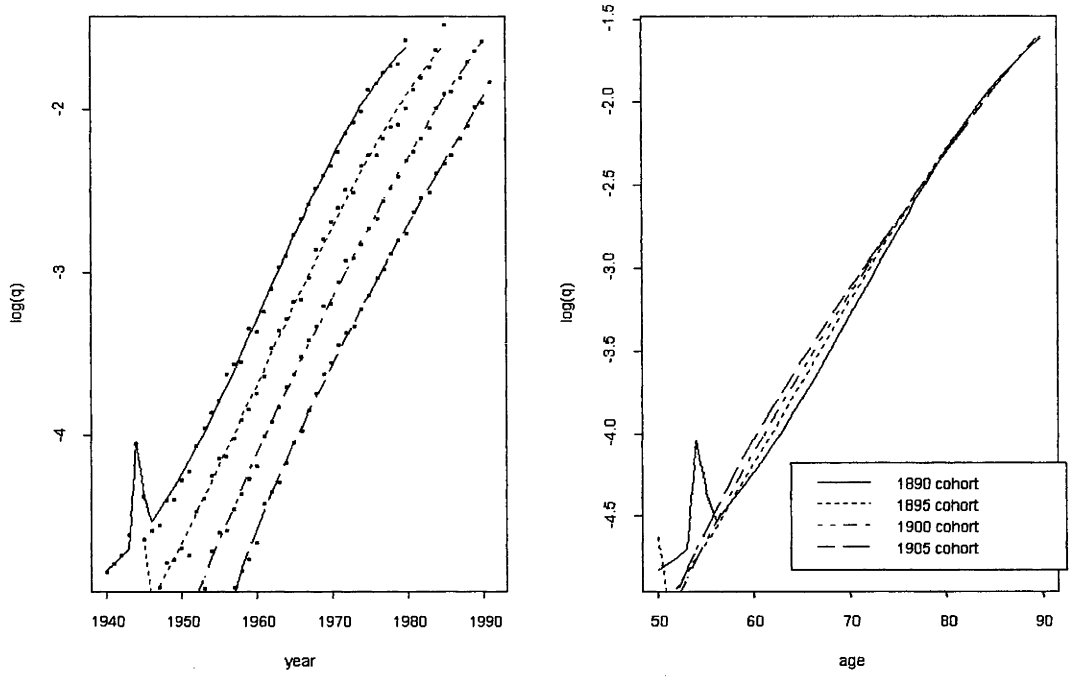
On first inspection, therefore, it could be argued that differences between cohorts are responsible. However, as shown in Section 3.6 male mortality has not decreased linearly over time, but rather a mortality hump with a peak near 1968-1970 dominates from ages 50 to 75. It can be shown that a period of excess mortality can account for the pattern seen in Figure 3.19. In other words, a period effect may get misinterpreted as a cohort effect in the presence of non-linear mortality trends. As a result, different methods are used to try to ascertain whether cohort effects are present in the Dutch data.

The strong relationship between age and mortality, and calendar year and mortality, as attested in Sections 3.5 and 3.6, points to the importance of these variables over cohort in modelling mortality. Tuljapurkar and Boe (1998), in their recent review of the state of affairs of mortality forecasting, state that most researchers have found that mortality is best represented as functions of age and calendar year, and concludes that *"... cohort-effects have played a limited role in the analysis and forecasting of mortality"*.

The discussion to date has not meant to disprove that important cohort effects exist in the Dutch data, rather it has been used to illustrate the difficulties that can arise in distinguishing between cohort and period effects. Cohort effects can still be isolated to some degree and it is argued that an understanding of the patterns is critical in modelling

and forecasting. As will be shown in the remainder of this section and in Chapters 5 and 6, there is strong evidence that a cohort effect transects the male data and is responsible for many of the non-linear mortality patterns apparent in recent years.

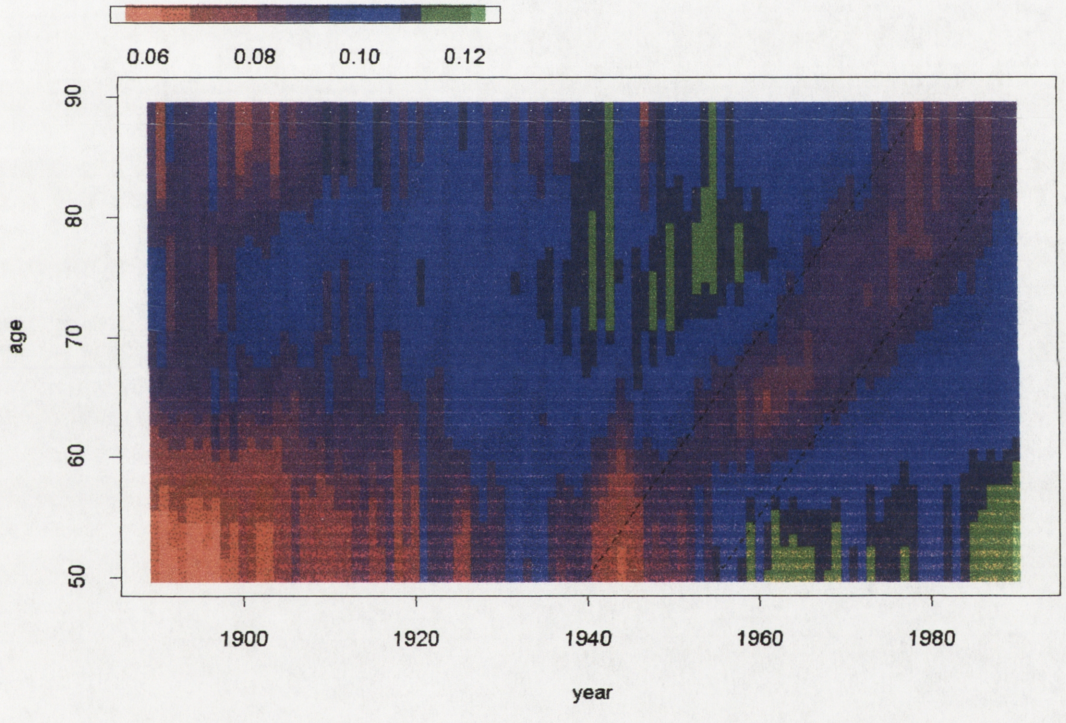
Figure 3.19 Dutch Males. Comparison of four cohorts. Observed log(q) and Perks fit.



Although it is difficult to distinguish period from cohort effects when mortality is decreasing in a non-linear fashion, patterns in the derivatives over individual years may suggest cohort variation. Caselli (1990) searches for cohort effects in French and Italian data by exploring the derivatives along periods. Caselli smoothed the derivatives by a weighted triangular distribution rather than using natural splines as we have done here. A low derivative indicates little change in mortality between subsequent ages. Conversely, a high derivative points to a large increase in mortality. In the event where a pattern in the derivatives permeates along a diagonal of the mortality surface, this may indicate a cohort effect. Figure 3.20 is a contour plot of estimated derivatives for Dutch males.

Since the derivatives in this instance are estimated from subsequent ages along the same period, the pattern is not confounded by period effects related to the mortality hump. Figure 3.20 shows that the change in mortality between cohorts approximately I n the range 1890 to 1905 is relatively low, and in particular, there is a marked change in derivative corresponding approximately to the 1885-1890 cohorts and the 1905-1910 cohorts. This may suggest that the 1885-1890 cohorts have relatively favourable mortality compared to adjacent cohorts.

Figure 3.20 **Dutch Males. Estimated derivatives along individual years based on smoothed log(odds). Ages 50 to 90. Year 1890 to 1991.**



The above analysis has relied on graphical inspection of the mortality rates and derivatives. Wilmoth (1990) distinguishes between age, period and cohort by decomposing the residuals of a fixed effects model. The method has an attractive simplicity to it, though like the graphical methods, involves a degree of judgement. Exploration of the Dutch data using the technique of Wilmoth, and discussion of cohort effects in the context of debilitation and selection is the topic of Chapter 6.

The main conclusions from this brief analysis of Dutch cohort mortality are:

- Cohort effects are difficult to isolate due to confounding with age and period;

- the male mortality hump identified in Section 3.6 has characteristics that identify it as a period effect. However, contour plots of approximated derivatives along year and age suggest that the initial curvature prior to the hump at different ages varies with cohort.

- graphical analysis of the derivatives suggests that a male cohort effect may transect the Dutch data. Male mortality after the age of 60, and in between 60 and 75 in particular, has been more favourable for those individuals who were born around 1885-1895, than for males born in previous years and subsequent years up to 1905-1910.

The importance of cohort effects on adult age-patterns of mortality means that cohort variation, as well as age and year effects, need to be considered when generating a model and projections for Dutch mortality.

Modelling a cohort effect is problematical because of confounding with age and year. Possible cohort effects can be incorporated into a regression model in an indirect way by including high order interactions between terms in age and year. Alternatively, an approach based on separating the year and age effects and decomposing the residuals may offer insight into cohort effects. Modelling the Dutch mortality surface, accounting for year, age and cohort, is the topic of the next three chapters.

4 Modelling Dutch Mortality using Regression – Generating the Model.

4.1 Introduction

Regression techniques can be applied to the problem of modelling a mortality surface, in particular when polynomials in age and year and interactions are the predictors. Under the assumption of independence between cohorts and identical distributions of lifetimes within a cohort, it was shown in Chapter 2 that the residuals from an expected surface are asymptotically normal with a defined binomial variance that is a function of year and age.

Based on the binomial variance, weighted least squares (WLS) is undertaken on parsimonious models of the pre-1939 and post-1945 male and female mortality surfaces. The models incorporate a high degree polynomial in age, a linear term in year, and interactions. For the post-1945 models a normal density function is included to capture the mortality hump associated presumably with cardiovascular disease for the males, and the non-linear decrease in mortality post-WWII for females.

It is shown that extra-binomial variation is present in the residuals from the fitted models. The extra variance is shown to vary according to age and year, implying that there is variability in the mortality surface that cannot be explained by homogeneous binomial variation in survivor numbers. An exponential function is proposed as a model for the extra variance. Maximum likelihood estimation is applied and parameter estimates for the mean and the extra variance of the pre-1939 and post-1945 surfaces are calculated.

The fits can be improved by incorporating higher order interactions between age and year, though at the expense of interpretation. Despite their deficiencies, models that avoid high order interactions are preferred as they form the basis for plausible predictions and allow simple mathematical interpretations of historic mortality patterns.

Final estimates of the proposed models of Section 4.4 are given in Table 4.3, Table 4.4, Table 4.5 and Table 4.6.

4.2 Selecting and Fitting the Regression Models

In Chapter 3, the patterns in log(odds) by age, year and cohort were analysed for both Dutch males and females, ages 50 to 90, years 1890 to 1991. The main conclusions that are relevant to the question of modelling are:

- there is an age-pattern in mortality that can be modelled using the Perks function or a third or fourth degree polynomial in age;
- there are a number of periods of extraordinary mortality that can be captured by indicator variables;
- the trend in log(odds) over the post-WWII years is non-linear, partly due to the presence of a mortality 'hump', and differing rates of mortality decline over time;
- there is a strong discontinuity between the mortality surface pre- and post-WWII;
- a possible cohort effect appears to transect the male mortality surface.

When the errors from a smooth surface have a distribution due to a plausible stochastic assumption, a parametric surface enables inferences to be made about the quality of fit and the distribution of the parameters. Such an approach is only useful if the mean has successfully been modelled. The justification for using regression techniques to model the mortality surface comes from the stochastic structure of the errors. The theory underlying the error distribution is given in Chapter 2.

Given the stochastic basis for the errors, and the resulting advantage of determining parameters, a descriptive parametric surface is developed in this chapter, despite the non-linearities in year that prevent a simple model. The purpose of this is two-fold: to provide a clear picture of mortality across year, age and cohort; and to enable analysis and decomposition of the variability in the surface.

The large degree of non-linearity in the post-WWII mortality trends inhibits fitting a basic model to the surface. If a suitable parametric model can be developed that is not unrealistic when extrapolated beyond the range of the data, then projections may be generated. The topic of projecting mortality is taken up in Chapter 7.

If ordinary least squares is applied in the presence of heteroscedasticity, the estimated coefficients are unbiased but inefficient. The theoretical argument outlined in Chapter 2 suggests that the regression residuals produced by OLS will not be independent and identically distributed. In particular, variation will exist according to the number of survivors and the probability of death at each age and year. In addition, extra-binomial variation may be present that can be modelled as a function of age and year.

The approach taken to modeling mortality in this chapter is similar to that taken by Heathcote and Higgins (2001a). Recapping from Chapter 2, for our mixed model:

$$Y_{x,t-x} = \log\left(\frac{\tilde{q}_{x,t-x}}{1-\tilde{q}_{x,t-x}}\right) = \delta_{x,t-x} + \eta_{x,t-x} + \varepsilon_{x,t-x} \qquad (4.2.1)$$

where, $\eta_{x,t-x}$ is $N(0,\gamma^2(t,x))$, $\varepsilon_{x,t-x}$ is $N(0,\sigma^2_{bin}(t,x))$, and $\sigma^2_{bin}(t,x) = \dfrac{1}{l_{x,t-x}q_{x,t-x}(1-q_{x,t-x})}$ and are assumed independent.

The initial weights used are the inverse of the binomial variance. If there is evidence of extra-binomial variation, $\gamma^2(t,x)$ can be modelled and the fitted surface re-estimated using weights based on the inverse of $\sigma^2_{bin}(t,x)+\gamma^2(t,x)$.

An earlier, slightly different and less satisfactory fit was achieved by using iteratively re-weighted least-squares (IRLS), as done by Heathcote and McDermid (1994) in their regression modelling of French mortality data. At each iteration, odds are estimated from the fitted model and used to determine new weights that are then used in re-estimation of the fit. Although this method is more theoretically sound, the new weights are still estimated based on empirical $l_{x,t-x}$, and so the procedure is not strictly IRLS. The large

sample size means that the weights used in WLS differ only marginally from those produced by the final IRLS fit. Because of the computational simplicity of using WLS, and the fact that differences between the fitted models when using IRLS and WLS are negligible, WLS has been favoured in this analysis. An additional advantage of WLS is seen in Section 4.3, where extra-binomial variance is modeled. Having a fixed value for the explained component of variance greatly simplifies the computation.

Figure 4.1and Figure 4.2 illustrate the relative weights used for males. Similar curves hold for females but are not plotted here. The largest weights occur near age 75 in the most recent years for males (and age 85 for females). These observations are weighted approximately 8% more than 75 year olds in 1980, 16% more than 75 year olds in 1970, and 62% more than 75 year olds in 1960. In comparison to 1890, the most recent 75 year old observations are given nearly three times the weight. In contrast, for 50 year olds, the weights have changed little over the century. Although the number of survivors at this age has risen by nearly 600%, the probability of death has dropped by over 40%, resulting in theoretical variances (and weights) that have relatively small differences.

The weights take into account variation in the uncertainty surrounding estimates of the log(odds) due to survivor numbers and probabilities of survival/death. An examination of the weights uncovers a general pattern of larger weights (smaller variance) for ages between 70 and 80 for the more recent years, where the log(odds) are high and the number of survivors is relatively large compared to earlier years. For the earlier years, the weights are more constant across the age range; the high rate of mortality for the upper ages is balanced out by the small number of survivors at these ages in the early years.

The pattern of weights makes sense intuitively, as we would expect that a larger number of survivors and a larger number of deaths (reflected in the probability of death) should enable more accurate estimation of the actual log(odds) and, therefore, have a smaller variance.

The relationships between log(odds) and the individual variables age, year and cohort have been explored in detail in Chapter 3. In this section, mixed regression models are constructed with polynomials in age and year as predictors.

The general form of the model is:

$$\delta(t,x;\underline{\beta}) = \beta_0 + \sum_{i=1}^{4}\beta_i x^i + \sum_{i=5}^{q}\beta_i t^{i-4} + \sum_{i=1}^{l}\sum_{j=1}^{m}\beta_{q+ij}x^i t^j \qquad (4.2.2)$$

which models $\log\left(\frac{q_{x,t-x}}{1-q_{x,t-x}}\right)$ as a quartic in age x, a polynomial in year t, and interactions between the polynomial terms in age and year. Examination of the pattern in mortality over age suggests the use of a quartic or cubic rather than a quadratic in age.

The advantage of this model is its ability to capture the trends in the data across age and year and the fact that linear regression techniques can be applied when fitting. A disadvantage is the lack of theoretical foundation for the selection of the polynomial in age. As will be shown, interactions between terms in age and year can capture non-linearities in the surface, and in Chapter 6 it is shown that many of the interactions in the Dutch male regression model appear to capture cohort effects.

The fact that interactions can account for cohort variation can be illustrated by a simple example. Consider the model for log(odds) given by

$$\log\left(\frac{q_{x,t-x}}{1-q_{x,t-x}}\right) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 t + \alpha_4 tx + \varepsilon_{x,t-x}$$

where x=age and t=year. Since birth cohort $c = t - x$, we can re-write the model in terms of cohort and age (or cohort and year):

$$\log\left(\frac{q_{x,t-x}}{1-q_{x,t-x}}\right) = \alpha_0 + (\alpha_1 + \alpha_3)x + (\alpha_2 + \alpha_4)x^2 + \alpha_3 c + \alpha_4 cx + \varepsilon_{x,t-x}$$

Interpretation of the parameters and isolation of separate effects becomes very difficult when higher-order interactions between year and age are included.

Figure 4.1 **Dutch Males. Theoretical weights used in WLS fit. Weights = $1/\sigma^2_{bin}(t,x)$**

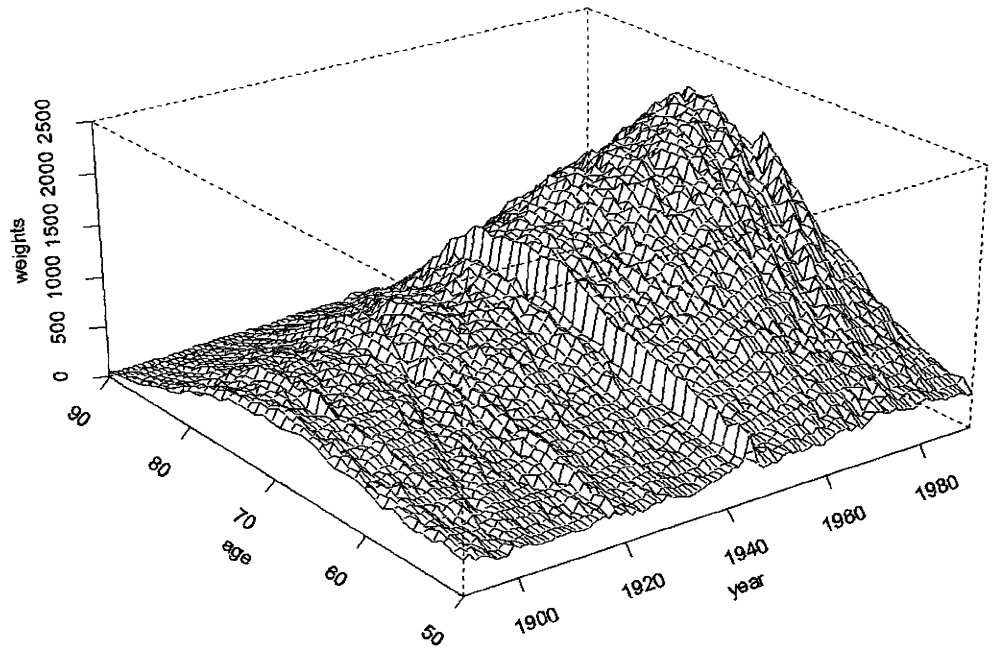
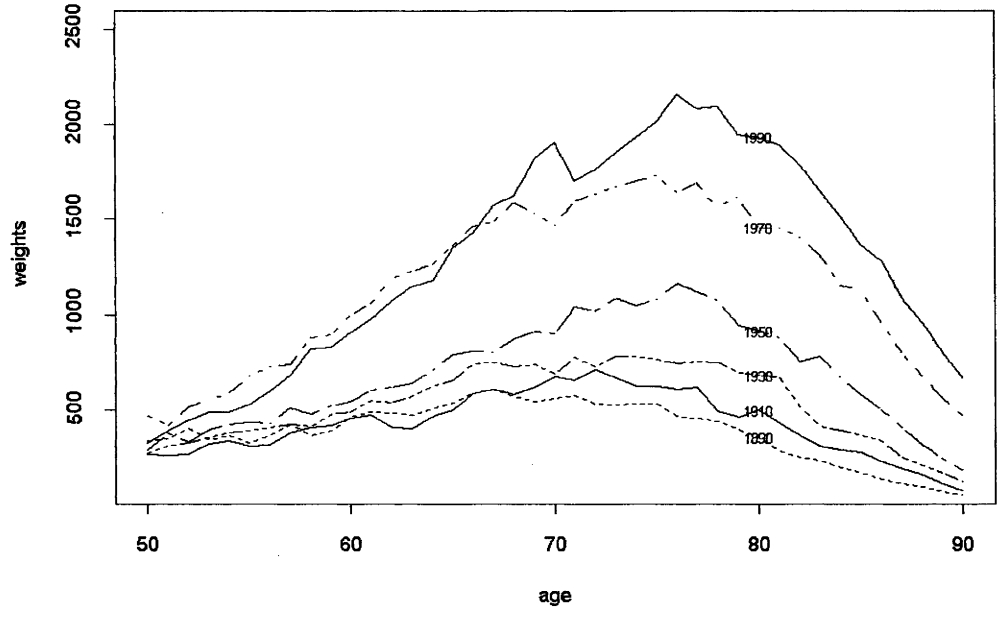


Figure 4.2 **Dutch Males. Theoretical weights for a selection of years. Weights = $1/\sigma^2_{bin}(t,x)$**



Because of the significant discontinuity pre- and post-WWII, two separate fits were constructed. The first fit includes years 1890 to 1938. Years 1939 to 1945 have been excluded from the model since they display excessive mortality for a range of ages as a result of the impact of WWII. The second fit covers the years 1946 to 1991. Interactions between the quartic in age, the indicators, and the polynomial in year are included in the preliminary model.

To account for the period of extraordinary mortality at t=1918 we introduce the indicator (dummy variable) I(1918), where

$$\mathrm{I}(t) = \begin{cases} 1 & if\,(t = year) \\ 0 & otherwise \end{cases}$$

Indicators for the periods of excessive mortality associated with WWII and inclusion of an indicator to capture the post-war change in mortality enables construction of a complete surface over the years 1890 to 1991. This was the approach taken by Heathcote and Higgins (2001a). Analysis of the observed rates indicates that excessive mortality was apparent for the late 1930's as well as 1940-1945. In order to avoid bias in the fit, indicators for each of these years would ideally be required. However, there are few gains in modelling the years of excessive mortality associated with WWII. The results can provide a quantification of the impact of WWII, but they offer no insight into the general trend in mortality over the century.

For this thesis, a different approach to Heathcote and Higgins (2001a) was taken. The pre and post war mortality regimes have been modelled separately. This was decided for a number of reasons. Firstly, it may be reasonable to treat the two periods as independent for much of the age range due to the large discontinuity in mean mortality. Secondly, the volatility in mortality seems to drop off markedly following 1945 suggesting a different variance structure pre and post WWII. It is likely that the main factor underlying the discontinuity and change in variance was the changed medical regime that swept across Europe post WWII which included penicillin as a usable drug by 1943 and the

establishment of public health systems following WWII. Thirdly, it is more realistic to base projections of mortality on post-WWII patterns.

A number of models were fit to the Dutch data with various polynomials in year to compare the impact of higher degrees and interactions. In order to avoid collinearity between the terms in the polynomials, age and year were centered by subtracting the mean of the variables. An alternative would have been to use the S-Plus function poly(). Poly() returns a matrix of orthonormal polynomials, which represents a basis for polynomial regression. We chose to use the centered age and year variables since the resulting fit in both cases is identical, the fitted model based on parameter estimates for the centered variables is easier to replicate, and predictions based on the fitted model don't require the normalisation constants that are needed when poly() is used.

The preliminary models were fit using the S-plus function 'lm'.

4.2.1 Fitting the Initial Models

The models fit to the Dutch male data ranged from inclusion of a linear term in year and its interactions to a quartic in year with its interactions.

The inclusion of high-order interactions greatly increases the number of parameters estimated. For the pre-1939 model with a linear term in year, 14 parameters are estimated whereas 10 are estimated in the absence of interactions between age and year. For the pre-1939 model with a fourth degree polynomial in year, the higher-order interactions increase the number of parameters from 13 to 29.

Increasing the power of year without including interactions has relatively little effect on improving the fit. Indeed, if a polynomial in year to the power of any large number is included, the residuals display the same basic pattern as those from a relatively simple fit. The pattern in the residuals indicates that a possible cohort effect runs through the surface that is not captured by the polynomials in age and year.

Although cohort effects cannot be captured directly in a regression model due to identification problems, they can be indirectly modeled through interactions and through the identity $c = t - x$. As interactions between polynomials in age and year are incorporated, the residuals become less biased. As mentioned in the previous section, a problem with using high-order interactions between age and year, besides that of interpretation of the parameters, is that they do not lead to sensible predictions. As a result, it is tempting to try and constrain the model only to linear terms in year, and account for the non-linearities in another way. One way, suggested in the previous chapter, and used by Heathcote and Higgins (2001a), is to include a normal density function to capture the cardiovascular hump.

The predictor

$$\text{dnorm(t,1972,9)}=\frac{1}{9\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{t-1972}{9}\right)^2\right]$$

is included in the model for post-1945 male mortality. The selection of the mean and variance for the predictor is discussed in Section 3.6 of Chapter 3.

The complete male models using a linear term in year for pre-1939 and post-1945 mortality are given below (eqns 4.2.3 and 4.2.4).

pre-1939 model:

$$\delta(t,x;\underline{\beta}) = \beta_0 + \sum_{i=1}^{4}\beta_i x^i + \beta_5 t + \sum_{i=1}^{4}\beta_{5+i}x^i t + \sum_{i=10}^{13}\beta_i x^{i-10} I(1918) \qquad (4.2.3)$$

post-1945 model:

$$\delta(t,x;\underline{\beta}) = \beta_0 + \sum_{i=1}^{4}\beta_i x^i + \beta_5 t + \sum_{i=1}^{4}\beta_{5+i}x^i t + \sum_{i=10}^{13}\beta_i x^{i-10} dnorm(t,1972,9) \qquad (4.2.4)$$

Dnorm(year,1972,9) gives most weight to the ridge along t=1972 with neighbouring years weighted according to the above normal density. As the shape of the cardiovascular hump varies with age, interactions between the polynomial in age and the normal density ordinate were included in the model.

Equations (4.2.3) and (4.2.4) offer a parsimonious fit that capture many of the major features of the mortality surface and, in avoiding the inclusion of high order terms in year, enable easy interpretation of the parameters.

The linear models in year based on equations (4.2.3) and (4.2.4) were fit to the data using weighted least squares. Step-wise backwards elimination was carried out on the parameter space, where only parameters with a p value less than approximately 5% were retained.

The post-WWII model for females incorporated a normal density term with mean 1959 and standard deviation 12, rather than a mean of 1972 and standard deviation of 9 as for the male model. The normal density term captures the non-linear behaviour in mortality decline post-1945, rather than being used explicitly for the cardiovascular hump as it had been for males. If a fourth degree polynomial in year is included in the model, the optimal normal density term has a mean of 1968, relating explicitly to the cardiovascular hump.

The final models given in Section 4.4, namely equations (4.4.1m), (4.4.2m), (4.4.1f) and (4.4.2f) have the same functional form as those produced by WLS and backwards elimination as described here, so Section 4.4 can be referred to for the final parameterisation of the regression models.

The main difference between the male and female pre-1939 models is the significance of interactions between the polynomial in age and year.

Corresponding fits for males and females based on a fourth-degree polynomial in year and high order interactions between the terms in the fourth-degree polynomial in age and the fourth-degree polynomial in year were produced for comparison with the simpler models.

The fourth-degree polynomial in year and the high order interactions had very little effect on the model R-squared (see Table 4.1).

Table 4.1 R-squared for preliminary models

	Linear term in year	Fourth degree polynomial in year
males		
Pre-1939	99.44%	99.49%
Post-1945	99.80%	99.83%
females		
Pre-1939	99.33%	99.39%
Post-1945	99.79%	99.83%

Although the more complex models provide a closer fit to the observed log(odds), the simpler models are easier to interpret due to the exclusion of higher order terms in year. A disadvantage of the parsimonious models is that they do not fit the data as accurately, and may not capture some of the features of the mortality surface, such as the apparent cohort effect that transects the male surface, which are more closely modelled by including high order interactions. An advantage of the simpler models is that they allow quantification of certain broad features of mortality, such as the rate of change in mortality decline at different ages, and the influence of the mortality hump (in the case of the males). An additional advantage is that suitably modified models may be used for projecting mortality, whereas extrapolating a model that incorporates higher-order terms in year leads to nonsensical projections.

As a result of the relative ease of interpretation, the potential for extrapolation of the parsimonious models, and the small extra explanatory power of the higher-order terms (as evidenced by R-squared), models incorporating higher-order polynomials in year were ultimately rejected. The parsimonious models are examined and developed further in the remainder of this chapter. Numerical results for the final models are presented in Tables (4.3), (4.4), (4.5) and (4.6).

4.2.2 Assessing the Fits

The standard errors for the fitted models are accurate only if the assumptions of independence and identical distributions of lifetimes along cohorts hold. Rather than making inferences based on the parameter estimates and standard errors of the fitted models described in the previous section, the adequacy of the fits are examined in order to test the assumptions.

A reliable way to assess the adequacy of a fitted model is to examine the standardised residuals with respect to the response and predictors (Ryan, 1997). Clusters of residuals and curvilinear patterns can indicate an inappropriate fit. Certain observations may have undue influence on the mean curve and should be weighted down.

A plot of Cook distances for the fitted models based on the linear term in year revealed a number of points of potentially high influence, however all of these, with the exception of the male log(odds) for year 1947, age 90, have a negligible influence on the overall fit. Using the rough guide of a particularly high Cook value when $D>4/n = 0.0021$ (for the post-1945 surface), none of the observations were considered high. Despite this, the most influential point (corresponding to year 1947, age 90) has a relatively large degree of leverage due its position away from the mean. This point was weighted down and the male post-WWII model was refit. The results presented in this chapter correspond to the modified fit.

The residuals by year for the pre-WWII male and female models are biased for parts of the surface due to the extreme variability in mortality from year to year prior to 1939. Unless a large number of indicators is used to capture individual years (besides 1918), the bias will persist. Accuracy of fit is not of paramount importance for the period pre-1939; rather more insight can be gained by trying to quantify the variability in mortality for this period.

An immediate observation of the residuals is that there is evidence of heterogeneity with respect to age and year for some portions of the mortality surface even after taking the binomial weights into account. The weights applied to the pre-WWII male and female

models appear to correct for the heterogeneity between years to some extent, however, too much weight appears to be given to the most recent years at the expense of the period between 1900 and 1915.

The residuals by cohort are unbiased supporting both male and female fitted models. There is definite heterogeneity in the residuals by cohort; the variance decreases for the more recent cohorts, with a marked decrease around the 1880-1890 cohort. The weights derived from the theoretical variance (binomial variance) correct the heterogeneity to a large degree, which implies that much of the difference in variation is due to the relatively low number of survivors distorting the estimation of the rates for early cohorts.

The standardised residuals display a more homogeneous pattern, though there is evidence of under-weighting of the most recent cohorts relative to the remainder of the mortality surface. The remaining heterogeneity in the residuals by cohort may be explained by larger variation in susceptibility to disease or frailties of individuals in early cohorts relative to more recent cohorts. It is plausible that a dramatic increase in medical technology and drugs such as penicillin and antibiotics during the mid 20^{th} century and public health measures were responsible for the drop in heterogeneity for cohorts after 1900. The effect of heterogeneity on population mortality and possible causes of cohort variability are covered in more detail at a later section of this chapter and in Chapters 6 and 7.

An assessment of the adequacy of the variance assumption underlying the structure of the residuals can be achieved by extracting the weighted residual sum of squares (the deviance divided by the residual degrees of freedom) for the fitted model. Under the assumption of independence and identical distribution of mortality along cohorts, this statistic (ψ) should be close to unity (see equation (2.6.2) from Section 2.6, Chapter 2).

Table 4.2 gives the value of ψ for each of the four models introduced above.

Table 4.2 ψ **for the preliminary fitted models**

	Males	Females
<1939	2.18	2.88
>1945	1.93	2.19

For all models the values of ψ are significantly higher than unity. This is not surprising since the models are biased in some regions of the surface, and since the graphical analysis of the residuals shows marked heterogeneity with regard to age, year and cohort. For the models that incorporated a fourth-degree polynomial in year and higher order interaction, the values of ψ are lower (2.02 and 1.60 for males; 2.63 and 1.82 for females), but are still much greater than unity.

The extent of the difference in variability between pre and post-WWII mortality can be quantified to some extent by the higher ψ for the pre-WWII models.

It is important to point out the impact of volatility in mortality rates in a given year and its sensitivity on the variance. As an example, if the indicator that captures the Spanish Influenza (1918) is omitted from the male pre-WWII fit, the value of ψ increases from 2.18 to 2.58. It is, therefore, important to account for periods of extraordinary mortality when exploring the variance of the fitted model. It is difficult to justify fitting indicators to other period effects, however, particularly when many of the fluctuations are not due to extraordinary events.

The fact that the variability is from real fluctuations rather than bias resulting from a poor choice of smooth surface is supported by a complex model using interactions between ten degree polynomials in year and age. When fit to the <1939 male data this produces a variance of 1.85, which is a relatively small reduction from 2.18.

It would seem, therefore, that an extra source of variance is present. Analysis of the residuals, in particular, graphical inspection of the heterogeneity of the residuals as

undertaken above is critical in determining the structure of the extra variance. Further analysis of the heterogeneity and accounting for the extra variability in determining the regression parameters is taken up in the next section.

Despite the heterogeneity present in the model, the fitted models appear unbiased for much of the mortality surface, although there are deficiencies in both male and female fits prior to 1939 due to the volatility in mortality by year. Curvilinearity of the residuals with year for post-1945 females can only be adequately corrected by including non-linear terms or high order polynomials in year in the female model. Although this would improve the fit and form a better model for descriptive purposes, such a model would be of limited use in practice because of its lack of predictive power.

The remainder of this chapter will attempt to quantify the extent and character of the extra variation, and incorporate this into the development of new models.

4.3 Accounting for Extra Binomial Variation

Overdispersion could be attributed to various factors: the presence of outliers, the mis-specification of the model, or variation between the response probabilities. We can deal with this situation in a number of ways. We could plead that the source of this extra variation is due to factors that we have no data for. The best solution in this case may be to weight by an overall dispersion factor, and thus obtain an 'average' estimate of the heterogeneity in the data (see, for example, Williams (1982); Moore and Tsiatis (1991)). Alternatively, we could suppose that this variation is dependent upon birth cohort, year and age effects, and try and model accordingly.

4.3.1 Weighting by Aggregate Dispersion Factor

We first consider weighting by overall dispersion. The theory is given in Section 2.6 of Chapter 2.

With regards to the two male models with the linear term in year, standard errors should be inflated by the square-root of ψ=2.18 and 1.93, or approximately 1.5 and 1.4 respectively.

For the two female models, standard errors should be inflated by the square-root of ψ=2.88 and 2.19, or approximately 1.7 and 1.5. When these factors are applied to the standard errors the fitted models, the significance of the parameters decreases, but most remain significant at the 5% level.

The alternative, and favoured method, is to model the extra-binomial variance as a function of the variables of the mean model. The rationale for this approach and the results are presented in next section.

4.3.2 Modelling the Extra-Binomial Variance

Given that there appears to be a relationship between the heterogeneity from the WLS model and age, year and cohort, there is a strong argument for modelling the extra variance.

The presence of over-dispersion in the Dutch mortality data may be partly caused by heterogeneity within individual cohorts. The assumption of independent and identically distributed lives within a cohort is a simplistic model of reality. Individuals have frailties that depend on numerous factors including genetics. The frailty of an individual changes over time as a result of damage to the immune system brought on by physical stress such as disease, psychological stress, and social habits, such as smoking and diet. Although a 'base' frailty could be thought to exist due to genetic predispositions, actual frailty could be decreased by the adoption of healthy habits and the advancement of medical treatments.

It is not unreasonable to assume that an individual with constant frailty will undergo 'failure' at a rate proportional to the individual's age. The assumption of fixed and identical frailty for individuals in a cohort leads to the presumption of a relatively smooth mortality schedule over the cohort, affected only by period effects. On the other hand, heterogeneous frailties lead to selection in cohorts (Vaupel and Yashin (1985), Vaupel et al. (1988)). As the structure of a cohort changes due to debilitation and selection, the shape of the aggregate mortality curve would be expected to change. In the presence of changing frailty, the volatility in the rates of mortality would be expected to increase.

In these circumstances, one would expect that the extra variation in rates would be highest at the upper ages. The oldest members of a cohort will consist of a relatively small number of individuals, so only a small change in the composition of frailty for this group would have a dramatic effect on the mortality rates. As such, it is hypothesised that the largest extra variation (above that expected from the binomial model) should occur in the oldest ages of the cohorts with the fewest individuals.

In addition, we would expect greater volatility in death rates during periods of excessive mortality. Although the regression models to date have included an indicator to account for the period of excess mortality associated with the Spanish Influenza, other periods of volatility have been smoothed over. One would expect that the variance associated with these regions of the mortality surface will be relatively high.

Rather than correcting for over-dispersion by an 'inflation factor' as proposed in Section 4.3.1, the presence of heterogeneity, and the theoretical justification for heterogeneity due to differing and changing frailties, leads to the proposition that the extra variance should be modeled. In the absence of information pertaining to the composition of frailty within cohorts, such as cause-of-death and incidence of morbidity, we are restricted to model the variance as a function of the covariates in the mortality surface.

The discussions above have provided some justification for expecting a relationship between age, year and cohort with the extra variance. A functional form for the extra variance may be sought by examining the patterns in the variability of the residuals from the mortality surface. Prior to modeling the extra variance it is important to make sure that the fit is unbiased. The bias of the residuals for both the male and female fits is negligible for the most part, with the exception of certain years prior to 1939. As discussed earlier in this chapter, this bias can only be corrected for by including additional indicators, but doing so will distort the 'real variability' in mortality that we are trying to model. The models based on a linear term in year described in Section 4.2 are, therefore, considered to be suitable in order to proceed with constructing a model for the extra variance.

Maximum likelihood estimation is used to estimate the parameters for the extra variance function as well as new estimates for the mean. The theory behind MLE estimation is briefly covered in Section 2.6 of Chapter 2, and the likelihood function used is given in Section 4.4. Maximising the log-likelihood is equivalent to minimising the negative log-likelihood. This was performed by using the S-Plus function 'ms' which was used to minimise the negative log-likelihood over the set of parameters (Venables and Ripley, (1994); Chambers (1993); Chambers and Bates (1993)). The efficiency and convergence of the function was improved by including the derivatives of the negative log-likelihood.

In order to assess the appropriateness of different parameterisations, standard errors for the MLE fit were required. For large samples, MLE estimates are normally distributed with unbiased mean and a covariance matrix given by the inverse of the observed Fisher information matrix, which is equal to the negative of the matrix of second derivatives of the log-likelihood. Since we are minimising the negative log-likelihood, the matrix of second derivatives is equivalent to the observed information matrix for the fitted model.

In order for S-plus to include the information matrix in the output, the second derivatives of the negative log-likelihood were manually included in the algorithm for the fit. An example of the algorithm as applied to the Dutch mortality data is given in Appendix A.

In Section 2.6 of Chapter 2, we introduced the function g(.) of x and t, with parameters $\underline{\alpha}$, which models the extra-binomial variance: $\gamma^2 = g(t,x;\underline{\alpha})$.

An appropriate form for $g(t,x;\underline{\alpha})$ depends on the expected shape of the extra variance with regard to age, year and cohort. Through inspection of the residuals from the fitted models, it is expected that there will be higher variance (less weight) associated with the years 1915-1938, and a marginally flatter variance profile associated with age (in particular, more variance (less weight) for the ages 65 to 80 relative to the other ages).

Examination of the pattern of absolute residuals from a spline fit to the mortality surface suggests that the maximum variation occurs between 1915 and 1950 above age 80, and the

minimum occurs between ages 60 and 80 after 1960. The pattern of binomial variance coincides to a degree with the pattern in the absolute residuals of a spline fit to the mortality surface. In particular, there is more variation at old ages for early years than later years and variance generally drops off as year increases. In addition, there appears to be more variation for the younger ages across all periods. A feature not captured by the explained variance is the high variability between 1915 and 1938 for all ages, particularly 70 and above.

An exponential was used as the base function for the extra variance since this accommodates a wide range of shapes and restricts the extra variance to a positive number. In order to determine a suitable functional form for the extra variance, a range of alternative exponential functions were chosen and MLE was used on the fitted models to determine appropriate parameters for the extra variance.

The models for the extra variance were of the form $\exp\left(\beta_0 + \sum_{i=1} \beta_i Z\right)$, where Z was varied from model to model. The base model considered was

$$g(t,x;\underline{\alpha}) = \exp(\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \alpha_4 t + \alpha_5 t^2 + \alpha_6 tx) \qquad (4.3.1)$$

It is likely that some of the extra variance for the mortality surfaces would not be uniform with age. Interactions between age and year are included in the model to capture potential patterns of non-uniformity. Additional models were considered with interactions $t^2 x$ and $x^2 t$ but the extra parameters were found to be not significant.

The primary purpose of the model $g(t,x;\underline{\alpha})$ is to provide insight into the extra variance. Another purpose of modelling the extra-binomial variance is that it allows us to perform a regression fit that takes into account the variability in the data, thereby producing a superior fit.

4.4 The Fitted Models

MLE was used to estimate new coefficients for the fitted models as well as for the extra variance γ^2. The process used to select the appropriate functional form for the final models was as follows: MLE was performed on the weighted residuals and the parametric surfaces where the weights are given by:

$$1/\left(\sigma^2_{bin}(t,x)+g(t,x;\underline{\alpha})\right),$$

where $\sigma^2_{bin}(t,x)$ is the binomial variance and is treated as fixed, and the parametric surfaces are given by the equations (4.2.3) and (4.2.4) for males and equivalent models for females but with a different normal density ordinate. Standard errors were produced for the parameters of the model. If a parameter for a term in $g(t,x;\underline{\alpha})$ was not significant at the 5% level it was dropped from the model, and the process was repeated with a reduced form for $g(t,x;\underline{\alpha})$.

The distribution function for the $\varepsilon_{x,t-x}$ of (4.2.1) is:

$$f\left(t,x;\hat{\underline{\alpha}},\hat{\underline{\beta}}\right)=\frac{1}{\sqrt{2\pi\left(\sigma^2_{x,t-x}+g(t,x;\hat{\underline{\alpha}})\right)}}\exp\left(\frac{-\left(Y_{x,t-x}-\delta(t,x;\hat{\underline{\beta}})\right)^2}{2\left(\sigma^2_{x,t-x}+g(t,x;\hat{\underline{\alpha}})\right)}\right)$$

The likelihood is the product of the distributions over the entire surface. The log-likelihood is the sum of the log of the error distribution over all observations:

$$\sum_{x=50}^{90}\sum_{t=J1}^{J2}\log f\left(t,x;\hat{\underline{\beta}},\hat{\underline{\alpha}}\right)=-\frac{n}{2}\log 2\pi-\frac{1}{2}\sum_{x=50}^{90}\sum_{t=J1}^{J2}\log\left(\sigma^2(t,x)+g(t,x;\hat{\underline{\alpha}})\right)-\frac{1}{2}\sum_{x=50}^{90}\sum_{t=J1}^{J2}\left(\frac{\left(Y_{x,t-x}-\delta(t,x;\hat{\underline{\beta}})\right)^2}{\sigma^2(t,x)+g(t,x;\hat{\underline{\alpha}}}\right.$$

where J1 and J2 are 1890 and 1938 for the pre-WWII mortality surface, and 1946 and 1991 for the post-WWII surface.

The parameter values, standard errors, and associated t and p values generated using MLE are given in Table 4.3, Table 4.4, Table 4.5, and Table 4.6. The values are untransformed, meaning that they are expressed in terms of age and year, rather than the centered equivalents.

For the pre-1939 surface, the equations for the final male fitted model and the extra variance are:

$$\begin{aligned}\delta(t,x;\underline{\beta}) &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 \\ &\quad + \beta_5 t + \beta_6 xt + \beta_7 tx^2 + \beta_8 tx^3 + \beta_9 tx^4 \\ &\quad + I(1918)(\beta_{10} + \beta_{11} x),\end{aligned} \tag{4.4.1m}$$

$$g(t,x;\underline{\alpha}) = \exp(\alpha_1 + \alpha_2 x + \alpha_3 t)$$

The equivalent pre-1939 female equations are:

$$\begin{aligned}\delta(t,x;\underline{\beta}) &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 \\ &\quad + \beta_5 t + \beta_6 xt \\ &\quad + I(1918)(\beta_7 + \beta_8 x + \beta_9 x^2 + \beta_{10} x^3),\end{aligned} \tag{4.4.1f}$$

$$g(t,x;\underline{\alpha}) = \exp(\alpha_1 + \alpha_2 x + \alpha_3 t + \alpha_4 t^2 + \alpha_5 tx)$$

For the post-1945 surface, the equations for the final male fitted model and the extra variance are:

$$\begin{aligned}\delta(t,x;\underline{\beta}) &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 \\ &\quad + \beta_5 t + \beta_6 xt + \beta_7 tx^2 + \beta_8 tx^3 + \beta_9 tx^4 \\ &\quad + dnorm(t,1972,9)(\beta_{10} + \beta_{11} x + \beta_{12} x^2 + \beta_{13} x^3),\end{aligned} \tag{4.4.2m}$$

$$g(t,x;\underline{\alpha}) = \exp(\alpha_1 + \alpha_2 x + \alpha_3 t + \alpha_4 t^2 + \alpha_5 tx)$$

The equation for the final post-1945 female fitted model is :

$$\begin{aligned}\delta(t,x;\underline{\beta}) = {} & \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 \\ & + \beta_5 t + \beta_6 xt + \beta_7 tx^2 + \beta_8 tx^3 + \beta_9 tx^4 \\ & + dnorm(t,1959,12)(\beta_{10} + \beta_{11} x + \beta_{12} x^2 + \beta_{13} x^3),\end{aligned} \qquad (4.4.2f)$$

$$g(t,x;\underline{\alpha}) = \exp(\alpha_1 + \alpha_2 x + \alpha_3 x^2 + \alpha_4 t + \alpha_5 t^2 + \alpha_6 tx)$$

Estimates of the standard errors involved extracting variances and covariances for the transformed parameters.

When the fit of the linear model based on the original weights and the new weights are compared, the means of the two models are almost identical. However, the models generated with the new weights are more accurate (lower standard errors) as they take as much as possible of the uncertainty in the observed log(odds) into account in determining the best fit.

Table 4.3 Dutch male model (4.4.1m) Parameter values, standard errors, t and p values.

Parameters for the mean	value	standard error	t-value	p-value
Intercept	1.04E+03	1.27E+02	8.15E+00	0.00
x	-5.65E+01	7.28E+00	-7.76E+00	0.00
x^2	1.20E+00	1.56E-01	7.66E+00	0.00
x^3	-1.14E-02	1.49E-03	-7.64E+00	0.00
x^4	4.06E-05	5.31E-06	7.64E+00	0.00
t	-5.43E-01	6.63E-02	-8.18E+00	0.00
xt	2.95E-02	3.80E-03	7.74E+00	0.00
x^2t	-6.23E-04	8.16E-05	-7.64E+00	0.00
x^3t	5.93E-06	7.77E-07	7.64E+00	0.00
x^4t	-2.12E-08	2.78E-09	-7.64E+00	0.00
$I(1918)$	5.80E-01	6.48E-02	8.94E+00	0.00
$I(1918)(x)$	-5.66E-03	9.54E-04	-5.94E+00	0.00
Parameters for the extra variance				
Intercept	4.68E+01	8.08E+00	5.79E+00	0.00
x	3.67E-02	6.46E-03	5.69E+00	0.00
t	-2.90E-02	4.25E-03	-6.81E+00	0.00

Table 4.4 Dutch male model (4.4.2m) Parameter values, standard errors, t and p values.

Parameters for the mean	value	standard error	t-value	p-value
Intercept	-1.09E+03	6.60E+01	-1.65E+01	0.00
x	7.08E+01	3.91E+00	1.81E+01	0.00
x^2	-1.64E+00	8.58E-02	-1.91E+01	0.00
x^3	1.62E-02	8.27E-04	1.96E+01	0.00
x^4	-5.80E-05	2.95E-06	-1.97E+01	0.00
t	5.46E-01	3.35E-02	1.63E+01	0.00
xt	-3.57E-02	1.98E-03	-1.80E+01	0.00
x^2t	8.29E-04	4.35E-05	1.91E+01	0.00
x^3t	-8.20E-06	4.19E-07	-1.96E+01	0.00
x^4t	2.93E-08	1.50E-09	1.96E+01	0.00
$dnorm(year,1972,9)$	-1.97E+02	4.32E+00	-4.56E+01	0.00
$dnorm(year,1972,9)x$	8.87E+00	1.90E-01	4.66E+01	0.00
$dnorm(year,1972,9)x^2$	-1.25E-01	2.76E-03	-4.52E+01	0.00
$dnorm(year,1972,9)x^3$	5.61E-04	1.32E-05	4.26E+01	0.00
Parameters for the extra variance				
Intercept	2.86E+03	4.26E+02	6.73E+00	0.00
x	3.20E+00	2.79E-01	1.15E+01	0.00
t	-3.00E+00	4.31E-01	-6.96E+00	0.00
t^2	7.84E-04	1.09E-04	7.17E+00	0.00
tx	-1.61E-03	1.42E-04	-1.13E+01	0.00

Table 4.5 Dutch female model (4.4.1f) Parameter values, standard errors, t and p values.

Parameters for the mean	value	standard error	t-value	p-value
Intercept	4.50E+01	3.30E+00	1.36E+01	0.00
x	-1.06E+00	1.76E-01	-6.04E+00	0.00
x^2	1.35E-02	3.85E-03	3.52E+00	0.00
x^3	-1.03E-04	3.73E-05	-2.75E+00	0.01
x^4	2.76E-07	1.34E-07	2.05E+00	0.04
t	-2.05E-02	8.08E-04	-2.54E+01	0.00
xt	2.06E-04	1.19E-05	1.74E+01	0.00
$I(1918)$	5.71E+00	1.25E+00	4.57E+00	0.00
$I(1918)(x)$	-2.38E-01	5.37E-02	-4.43E+00	0.00
$I(1918)(x^2)$	3.40E-03	7.67E-04	4.43E+00	0.00
$I(1918)(x^3)$	-1.62E-05	3.65E-06	-4.43E+00	0.00
Parameters for the extra variance				
Intercept	5.10E+03	1.07E+03	4.76E+00	0.00
x	-1.47E+00	7.01E-01	-2.09E+00	0.04
t	-5.25E+00	1.12E+00	-4.68E+00	0.00
t^2	1.35E-03	2.93E-04	4.59E+00	0.00
tx	7.87E-04	3.68E-04	2.14E+00	0.03

Table 4.6 Dutch female model (4.4.2f). Parameter values, standard errors, t and p values.

Parameters for the mean	value	standard error	t-value	p-value
Intercept	1.84E+03	3.45E+02	5.33E+00	0.00
x	-1.11E+02	2.04E+01	-5.46E+00	0.00
x^2	2.51E+00	4.47E-01	5.62E+00	0.00
x^3	-2.46E-02	4.31E-03	-5.70E+00	0.00
x^4	8.77E-05	1.54E-05	5.70E+00	0.00
t	-9.42E-01	1.75E-01	-5.39E+00	0.00
xt	5.69E-02	1.03E-02	5.50E+00	0.00
x^2t	-1.28E-03	2.27E-04	-5.66E+00	0.00
x^3t	1.26E-05	2.19E-06	5.75E+00	0.00
x^4t	-4.49E-08	7.81E-09	-5.75E+00	0.00
$dnorm(year,1959,12)$	2.85E+02	3.00E+01	9.48E+00	0.00
$dnorm(year,1959,12)x$	-1.36E+01	1.33E+00	-1.02E+01	0.00
$dnorm(year,1959,12)x^2$	2.04E-01	1.92E-02	1.06E+01	0.00
$dnorm(year,1959,12)x^3$	-9.71E-04	9.16E-05	-1.06E+01	0.00
Parameters for the extra variance				
Intercept	4.55E+03	1.56E+03	2.92E+00	0.00
x	3.33E+00	9.74E-01	3.42E+00	0.00
x^2	1.72E-03	6.02E-04	2.85E+00	0.00
t	-4.73E+00	1.58E+00	-3.00E+00	0.00
t^2	1.23E-03	4.00E-04	3.08E+00	0.00
tx	-1.81E-03	4.82E-04	-3.75E+00	0.00

Figure 4.3 **Dutch males. Fitted log(odds) based on final models (4.4.1m and 4.4.2m) and observed log(odds) for 1939-1945.**

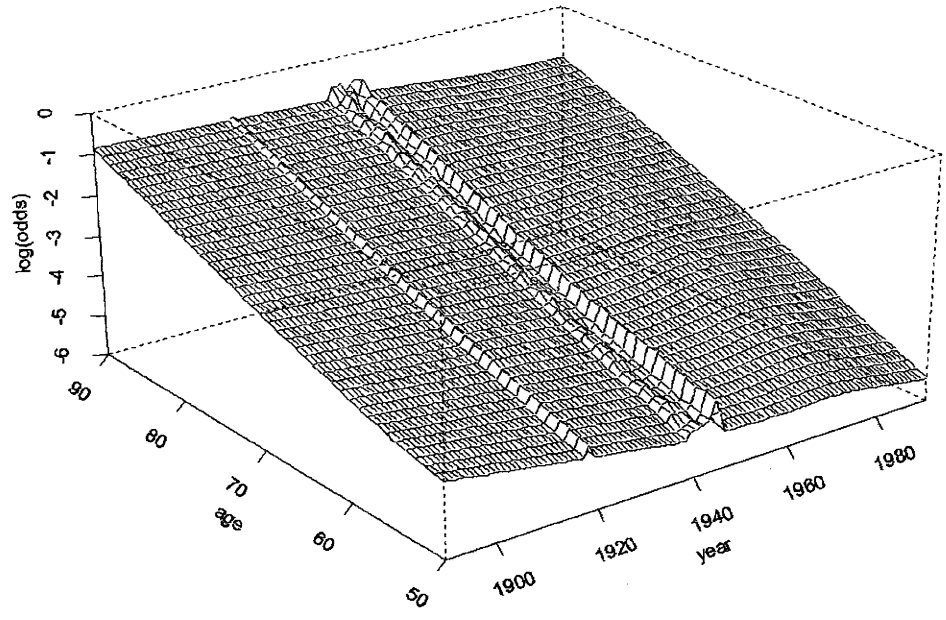


Figure 4.4 **Dutch males. Observed log(odds) for 1890-1991.**

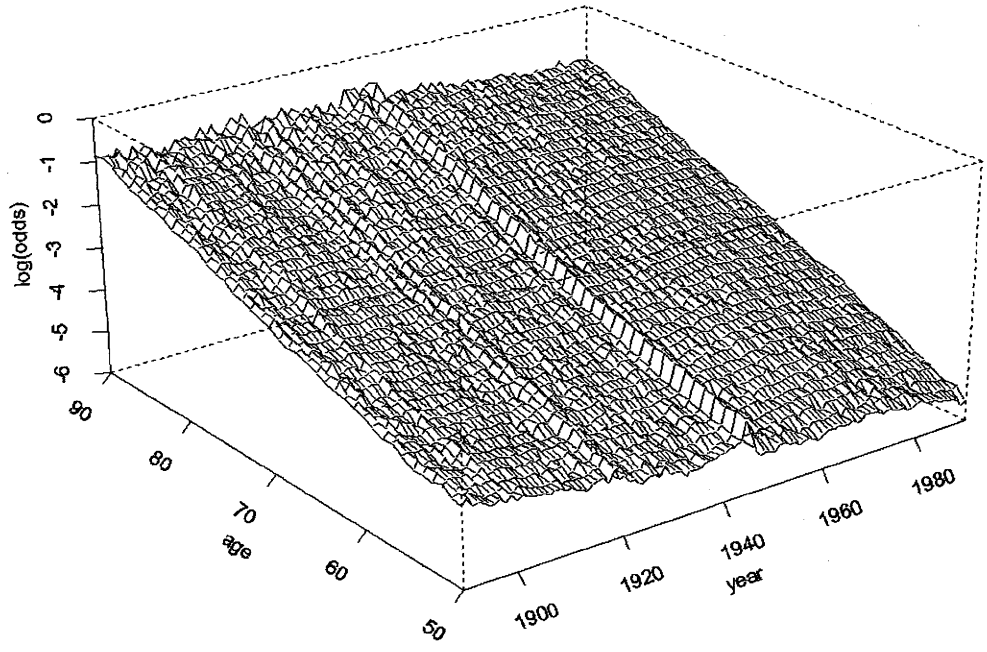


Figure 4.5 **Dutch females. Fitted log(odds) based on final models (4.4.1f and 4.4.2f) and observed log(odds) for 1939-1945.**

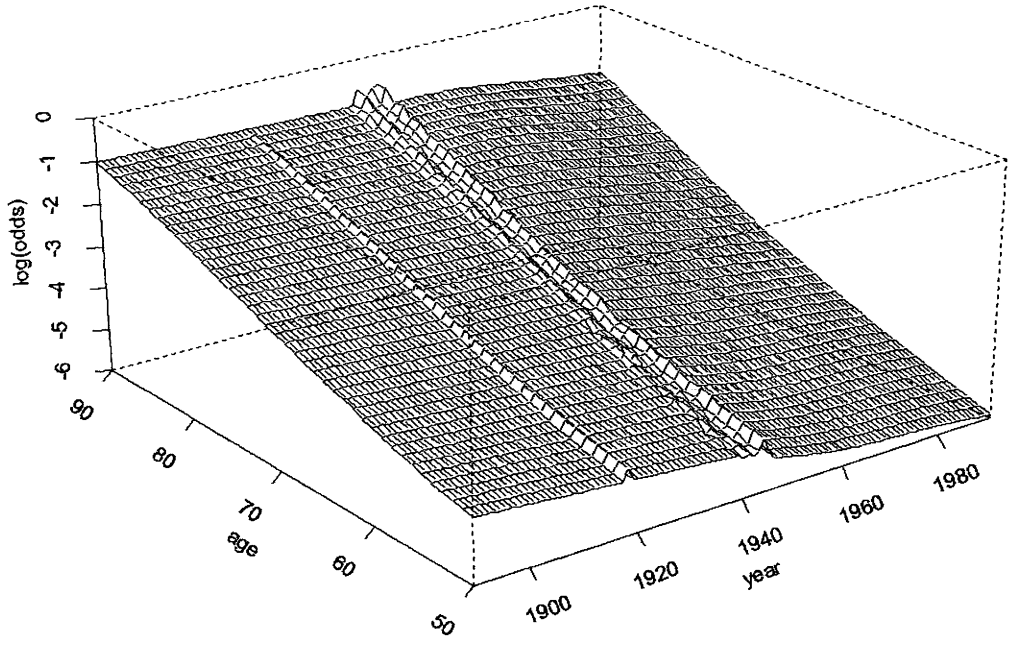


Figure 4.6 **Dutch females. Observed log(odds) for 1890-1991.**

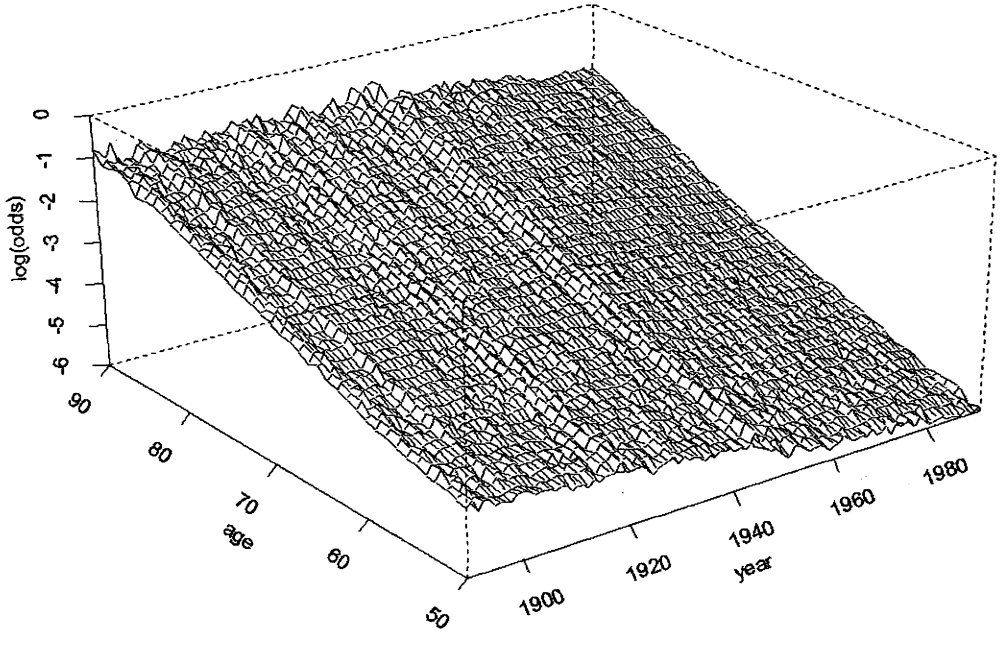


Figure 4.7 **Dutch males. Observed and fitted log(odds) based on final models (4.4.1m and 4.4.2m) (length of vertical scales are identical)**

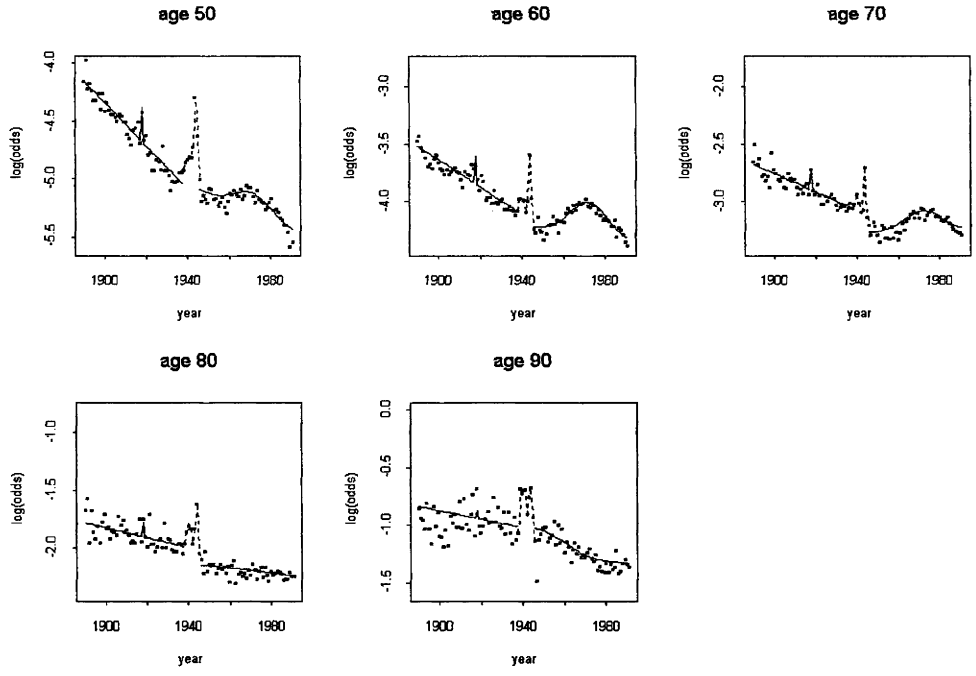
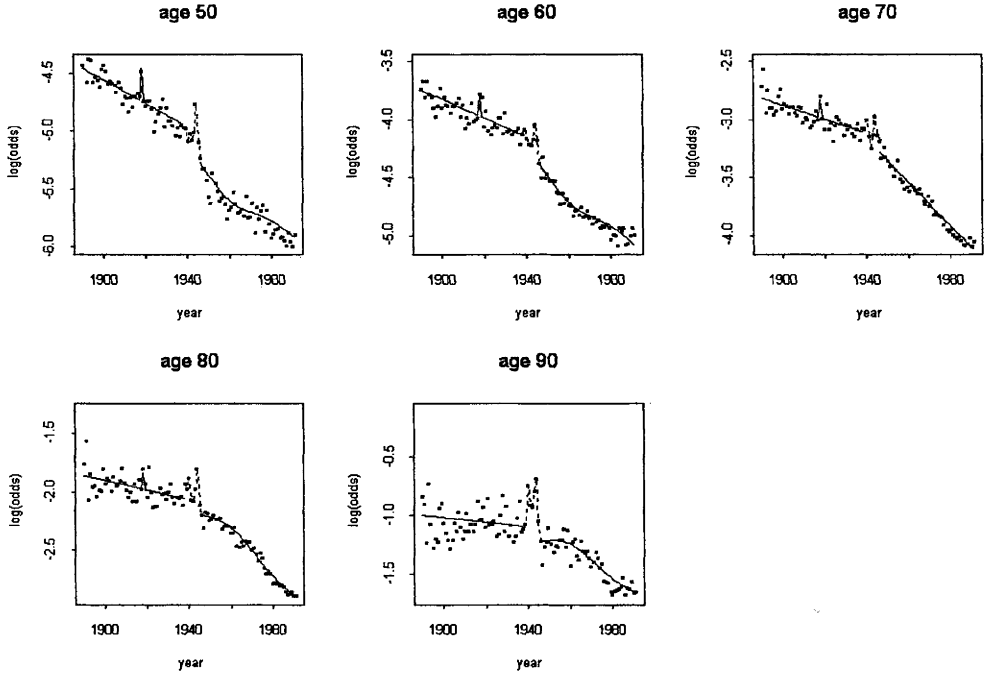


Figure 4.8 **Dutch females. Observed and fitted log(odds) based on final models (4.4.1f and 4.4.2f). (length of vertical scales are identical)**



Of particular interest is the additional component of variance relative to the binomial variance, as well as the total variance emerging from the fitted models. Both sexes display higher mortality variation prior to World War II, with female mortality variation being larger than that of males across the entire age range with the exception of the oldest ages (see Figure 4.9). The reduction in variation following WWII is most marked for men. It is possible that the relatively large reduction for men may be attributed to a decline in infectious and parasitic diseases at the turn of the century benefiting men more rapidly than women (Lopez,1983:90). There is increasing variance for fifty year old females post-WWII, though the sharp increase in the latest decade for this age group is more a function of the relative paucity of the fitted model for this corner of the mortality surface.

Figure 4.9 **Dutch males and females. Total variance for a selection of ages for fitted models (4.4.1m), (4.4.2m), (4.4.1f), (4.4.2f). (Solid line=males; dotted line=females)**

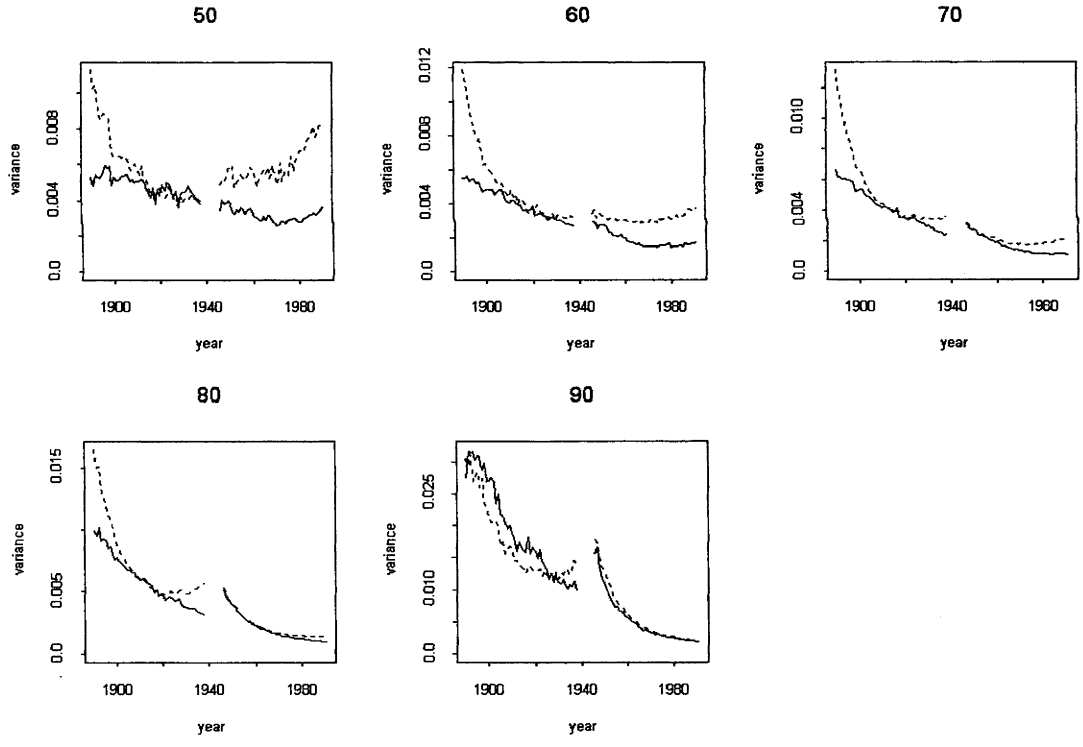


Figure 4.10 and Figure 4.11 display the explained (binomial) variance, the extra variance (γ^2), and the total variance for a selection of years for the final female and male models. Figure 4.12 and Figure 4.13 show the extra variance (γ^2) for females and males

respectively. As expected, for both sexes the highest extra variance is associated with the upper ages of those cohorts with the fewest survivors. The highest percentage increase in variance due to the gamma-squared term corresponds with 70 to 80 year olds. A larger relative variance reduces the weights and produces a more homogenous distribution for the standardised residuals by age.

The excessive volatility in the rates of mortality for the upper ages may be due to heterogeneous frailty distributions within individual cohorts as speculated in Section 4.3.2. In this context, the extra-binomial variation arises from the failure of Assumption (ii) from Section 2.3 in Chapter 2, which assumes that the lifetimes of individuals within a cohort are independent and identically distributed.

The hypothesis of changing frailty may also explain the general pattern of increasing extra variance with age, as changing frailty would have a greater impact as the number of survivors decreases. For recent decades, the extra variance is relatively low. The fact that the variance for ages 60 to 90 is flatter than in earlier decades may be a consequence of the extension of the average life-span, which would lead to less volatility at the older ages. One possible area of further work would be examination of various mixture distributions as a means of attempting to explain the pattern in the extra variance.

Figure 4.10 **Dutch females. Explained variance (dotted line), extra variance (dashed line) and total variance (solid line) for a selection of years for fitted models (4.4.1f, 4.4.2f).**

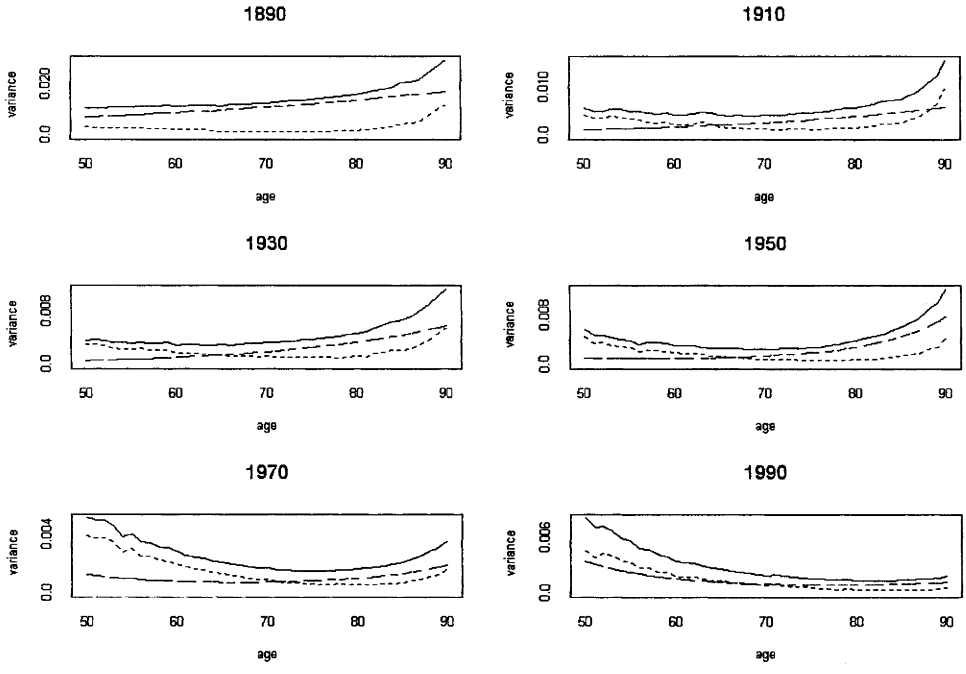


Figure 4.11 **Dutch males. Explained variance (dotted line), extra variance (dashed line) and total variance (solid line) for a selection of years for fitted models (4.4.1m, 4.4.2m).**

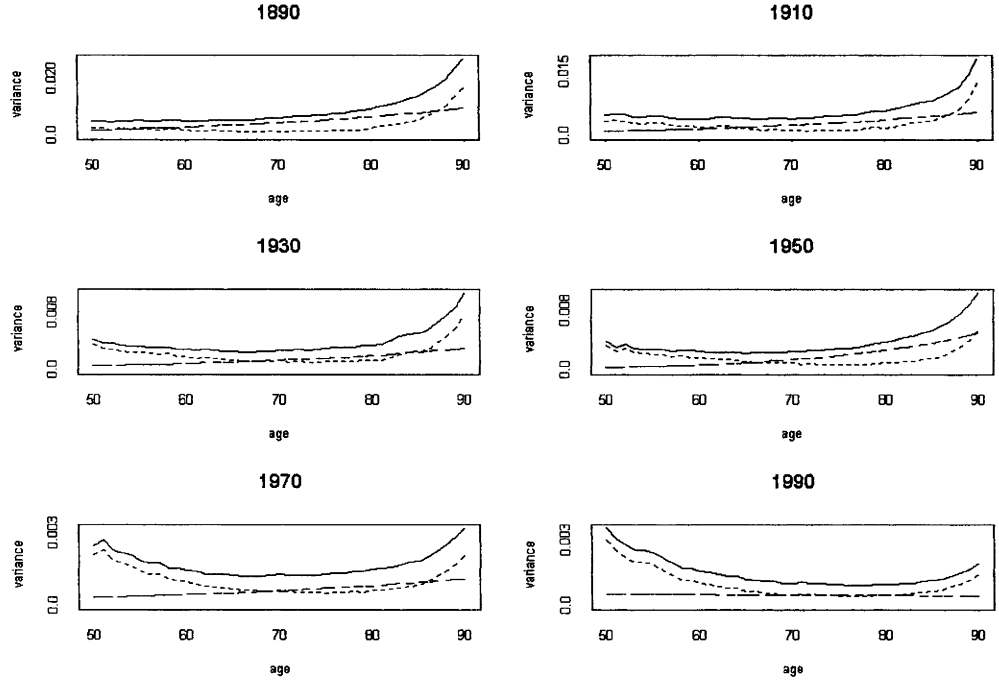


Figure 4.12 **Dutch females.** $g(t,x;\hat{\underline{a}})$ **(extra variance) from fitted models (4.4.1f) and (4.4.2f).**

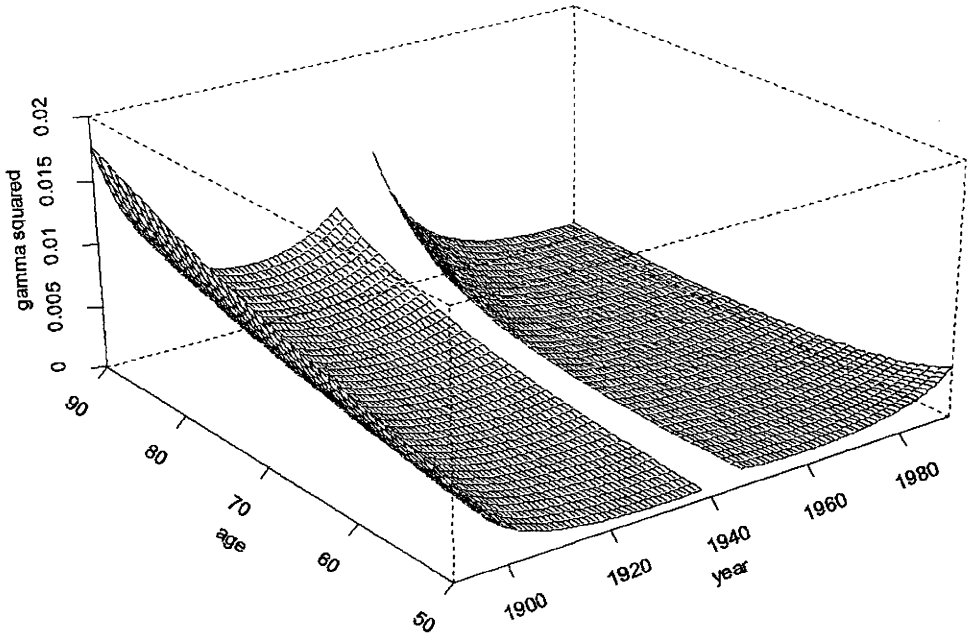
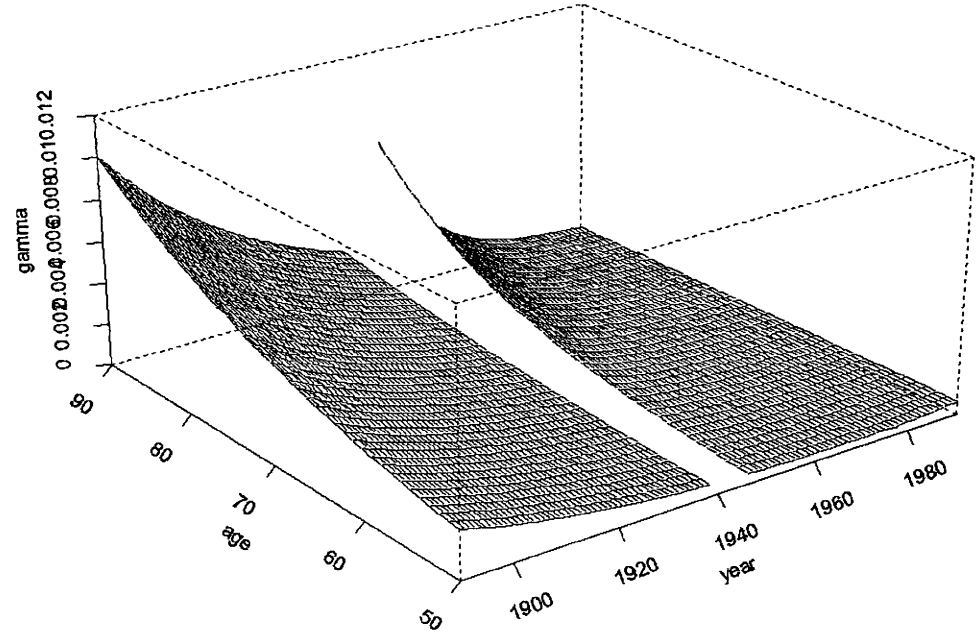


Figure 4.13 **Dutch males.** $g(t,x;\hat{\underline{a}})$ **(extra variance) from fitted models (4.4.1m) and (4.4.2m).**



ψ was estimated (Section 2.6, Chapter 2,) and is given below for the final models.

ψ	<1939	>1945
Males	0.999	1.019
Females	1.005	1.014

For the model of extra variance to adequately capture the dispersion, we would expect ψ to be close to unity. Although this is satisfied for the four models considered, in itself, this does not indicate whether or not the heterogeneity in the residuals has been accounted for. To obtain a graphical assessment of this, the standardised residuals are plotted by age, year and cohort. If the model is appropriate, the distribution of the standardised residuals should be homogeneous over the range of the data with respect to each predictor.

Figure 4.14 and Figure 4.15 display the standardised residuals with respect to age, year and cohort for both the male and female fitted models. In Figure 4.14 residuals for 1947 for ages 89 and 90 have been omitted.

The heterogeneity that appears in the raw residuals with respect to age is reduced when the new weights are incorporated. The pattern of standardized residuals by age for males and females is similar and appear unbiased across all ages. The standardized residuals should have equivalent variances across all ages if the assumption of identical distribution within cohorts applies. The variation at the oldest ages which is present in the unweighted residuals is still present, though reduced when the new weights are taken into account.

Both male and female fits exhibit curvilinearity in the residuals by year. Achieving a closer fit thereby reducing this problem involves including high order terms in year such as the fourth-degree polynomial fits performed earlier in this chapter, or by including additional non-linear functions such as normal density ordinates. For the sake of parsimony, and in the interest of maintaining a linear term in year, the models were not modified despite the pattern in the residuals by year.

Figure 4.14 **Dutch males. Standardized residuals from (4.4.1m) and (4.4.2m).**

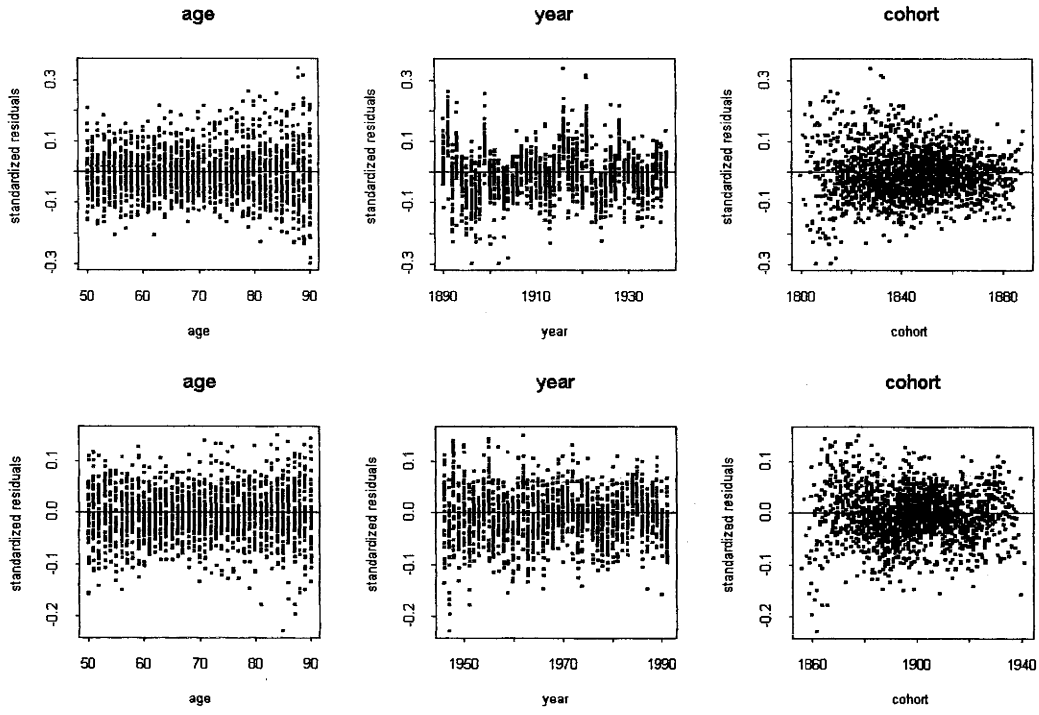
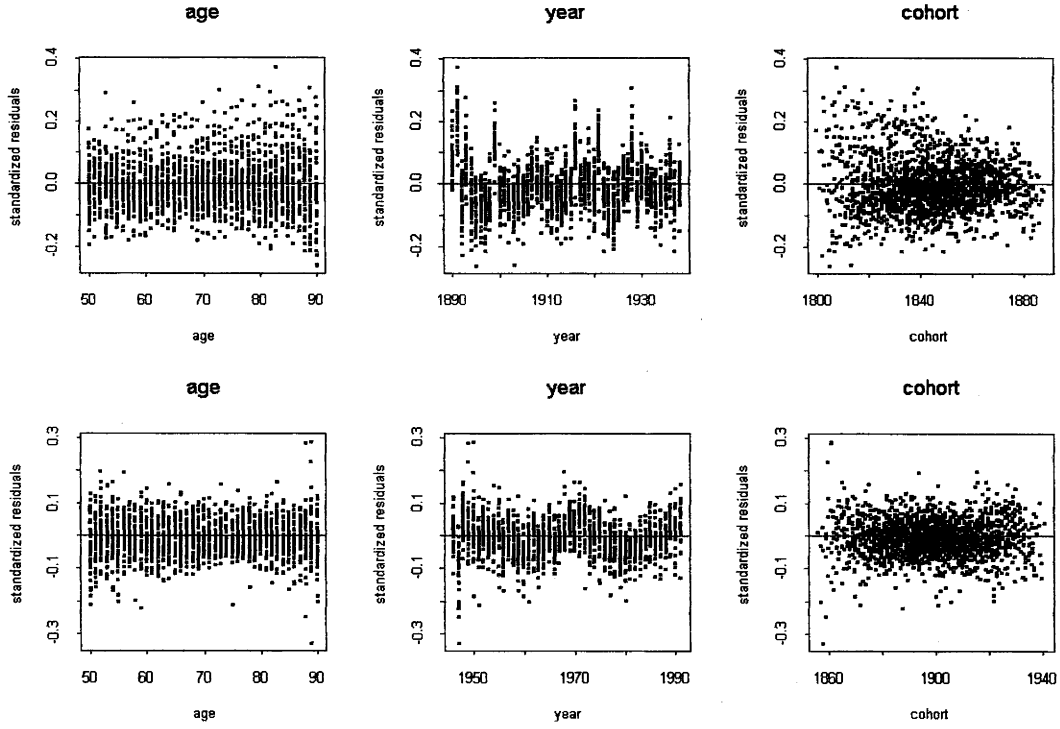
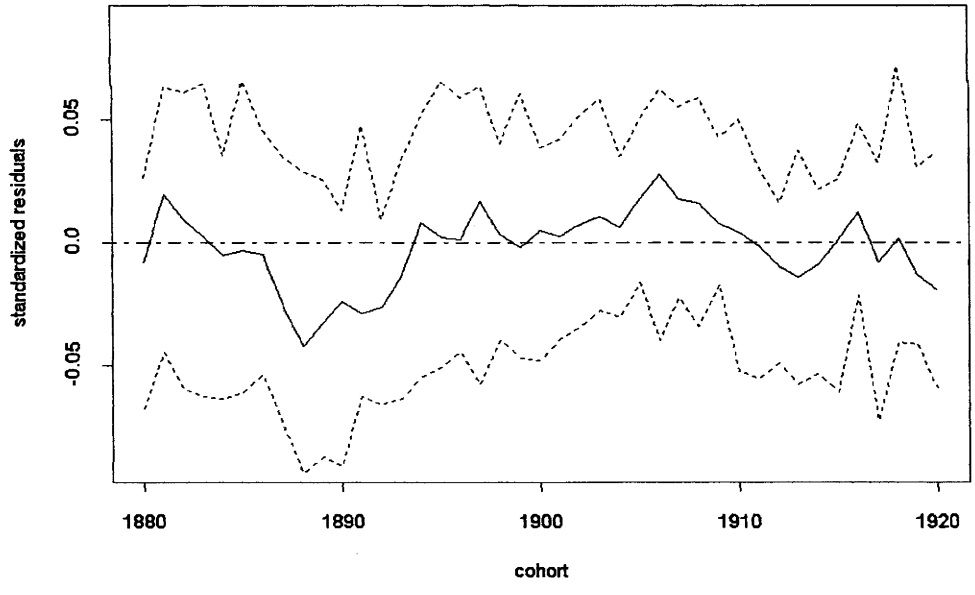


Figure 4.15 **Dutch females. Standardized residuals from (4.4.1f) and (4.4.2f).**



Despite the improvement in the standard errors when MLE techniques are used and the extra variance is modelled, there appears to be a curvilinear pattern in the residuals with respect to cohort for the male model (4.4.2m) that highlights the deficiency in using a simple parametric surface when cohort effects may be present. Figure 4.16 shows the 10%, 50% and 90% quantiles for the standardised residuals by cohort for the male model (4.4.2m). Higher order interactions between year and age are required to reduce this pattern, though it is implausible that such interactions can be justified. It is more plausible that the pattern in the residuals is due to a legitimate cohort effect, and as such, there is an argument for inclusion of an extra term in the model to account for this. An alternative means of modelling mortality that explicitly accounts for cohort variation is the topic of Chapter 6.

Figure 4.16 **Dutch Males. 10%, 50% and 90% quantiles of the standardised residuals by cohort for the fitted model (4.4.2m)**



It should be noted that the technique of estimating extra variance using maximum likelihood methods can be applied to more complicated models that would lead to fitted models with more stable residuals.

4.5 Conclusions

In this chapter, the log(odds) was modeled as a function of age, year and interactions between age and year for both Dutch males and females, aged 50 to 90, years 1890 to 1991. It was shown that:

- periods of extraordinary mortality can be captured by indicator variables;
- a high order polynomial in age and year, and high order interactions are required to capture the non-linearities in the mortality surface;
- weights derived from the assumptions of cohort homogeneity and independence (Assumptions (i) and (ii) from Section 2.3, Chapter 2) can be used to fit a weighted regression model;
- extra-binomial variance due to the failure of the assumption of identical distributions within a cohort means that the standard errors from the basic model are too small, and require adjustment.
- the extra variance can be modelled by an exponential function in age, year and interactions between age and year;
- the form of the extra variance model enables the construction of hypotheses regarding the source of the variation in the context of selection and debilitation.

In the next chapter the fitted models are explored in detail. The fits are expressed in terms of $q_{x,t-x}$ and standard errors are produced for use in prediction intervals. Various features of the male and female fitted surfaces are examined, and comparisons are made between the coefficients. Finally, the derivatives from the fitted surfaces are explored with some interesting conclusions. Conditional probabilities of survival and life expectancies for these models are generated in Chapter 7.

5 Modelling Dutch Mortality using Regression – Interpreting the Results

5.1 Introduction

Standard errors can be produced that allow confidence intervals to be placed on individual parameters, or prediction intervals to be placed on the fitted log(odds). Fitted values and prediction intervals for $q_{x,t-x}$ are given for the fitted models from Chapter 4.

An advantage of the proposed models is that the estimated coefficients allow us to quantify certain aspects of mortality change throughout the period examined. As an example, the impact of the normal density ordinate on male post-WWII mortality, and rate of decline in mortality for the sexes are explored.

First and second differences along cohorts are extracted from the fitted surfaces. An important conclusion of this analysis is that attempts to reveal patterns in the age structure of mortality are confounded by period effects when analysis is undertaken along cohorts. First derivatives generated from the female fitted surface along individual periods take a bell-shaped pattern. Similar patterns were reported by Horiuchi and Coale (1990) for female mortality from a number of countries, and by Horiuchi and Wilmoth (1998) for Swedish and Japanese males and females. An interesting phenomenon that has not been previously reported, is the apparent almost linear shift in the maximum derivative to higher ages as year increases for Dutch females. In the context of selective survival, this may be seen as a slowing down in the pace of ageing, or a shift to higher ages in the boundary between the "old" and the "old old". The maximum derivative appears to be asymptoting near age 81 in recent years which may reflect a limit, or simply a temporary lull.

5.2 Inference

A major advantage of regression techniques is that standard errors can be produced and used in the construction of prediction intervals and confidence intervals for the fitted values. Recapping the theory from Chapter 2, if we let the regression model be written as a mixed model with fixed effects $\delta = \underline{X}\underline{\beta}$ (n observations) and random effects $\underline{\eta}$ (with q levels), then

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{\eta} + \underline{\varepsilon}$$

where $\underline{Y}$ are the observed log(odds), $\underline{\varepsilon}$ are the errors from the fixed effect component and are normally distributed with $E(\underline{\varepsilon}) = \underline{0}$ and $\text{var}(\underline{\varepsilon}) = diag\left(\sigma^2_{bin}(t,x)\right)$, where $\sigma^2_{bin}(t,x)$ is the binomial variance, and the random effects have the properties $E(\underline{\eta}) = \underline{0}$ and $\text{var}(\underline{\eta}) = diag\left(\gamma^2(t,x)\right)$. In addition, the covariance between different random effects $\underline{\eta}_i$ and $\underline{\eta}_h$ where $i \neq h$ is zero, and $\text{cov}(\underline{\eta}_i, \underline{\varepsilon}') = 0$

If we consider the problem of prediction of log(odds) for a single value at the point (t,x) on a mortality surface, given by the scalar $\tilde{w}_{x,t-x}$, where

$$\tilde{w}_{x,t-x} = \underline{x}'\hat{\underline{\beta}} + \eta_{x,t-x} + \varepsilon_{x,t-x},$$

for a known vector of predictors $\underline{x}$ (which includes age, year, and the other variables in the fitted models), then a 95% prediction interval can be found by:

$$\hat{w}_{x,t-x} \pm 1.96\sqrt{\underline{x}'\left(\text{I}(\hat{\underline{\beta}})\right)^{-1}\underline{x} + \sigma^2_{bin}(t,x) + g(t,x;\hat{\underline{\alpha}})}$$

where $\text{I}(\underline{\beta})$ is the component of the Fisher Information matrix relating to $\underline{\beta}$ (see Appendix), and $g(t,x;\hat{\underline{\alpha}})$ is the fitted model for the extra variance component $\gamma^2(t,x)$.

For the sake of interpretation, it is instructive to consider the rate of mortality change in terms of $q_{x,t-x}$ rather than log(odds) since $q_{x,t-x}$ has a clear interpretation. Log(odds) can be converted back to the initial rate of mortality by the transformation

$$q_{x,t-x} = \frac{\exp(\log(odds))}{1+\exp(\log(odds))}$$

A prediction interval for $\hat{q}_{x,t-x}$ is:

$$\left(\frac{\exp\left(\hat{w}_{x,t-x} - 1.96se\left(\hat{w}_{x,t-x}\right)\right)}{1+\exp\left(\hat{w}_{x,t-x} - 1.96se\left(\hat{w}_{x,t-x}\right)\right)}, \frac{\exp\left(\hat{w}_{x,t-x} + 1.96se\left(\hat{w}_{x,t-x}\right)\right)}{1+\exp\left(\hat{w}_{x,t-x} + 1.96se\left(\hat{w}_{x,t-x}\right)\right)} \right),$$

Prediction intervals for the fitted $q_{x,t-x}$ are given in Figure 5.1, Figure 5.2 and Figure 5.3. Included in the figures are 95% confidence intervals around the fitted $\hat{\gamma}^{2}(t,x)$.

Confidence intervals could also be constructed for the fitted $q_{x,t-x}$, but are of limited use since they provide estimates of the uncertainty in the mean value rather than for an individual case. Since population means can be predicted with more precision than individual cases the confidence intervals would be much narrower than the prediction intervals derived above.

Figure 5.1 **Male and Female fitted q(t,x) and prediction intervals 1946-1991. Female fit is below male fit in each case.**

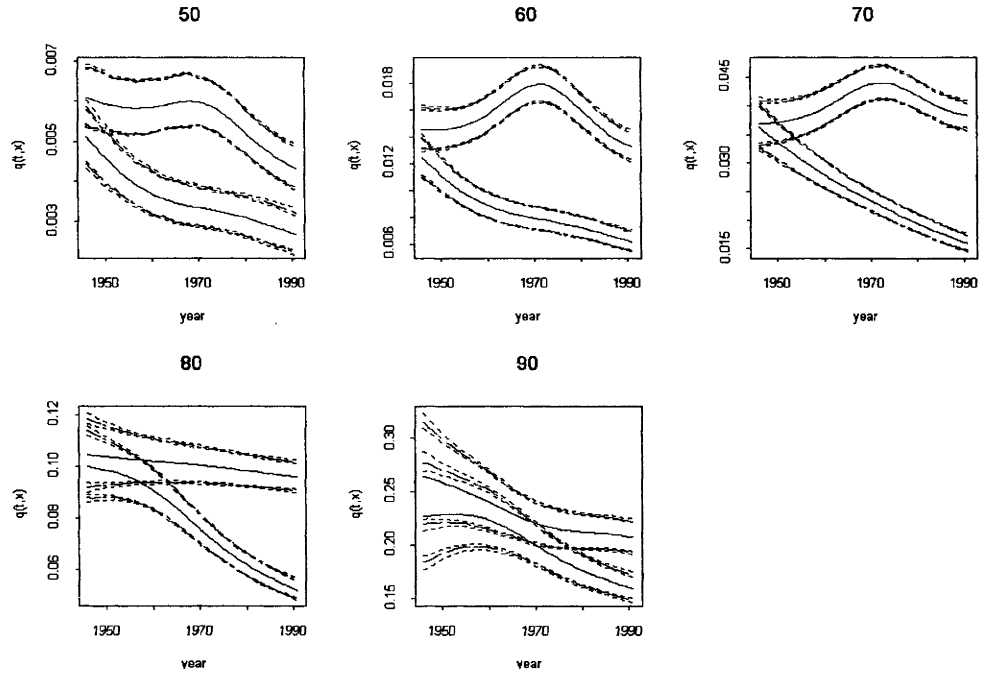


Figure 5.2 **Male observed and fitted q(t,x) and prediction intervals 1890-1991.**

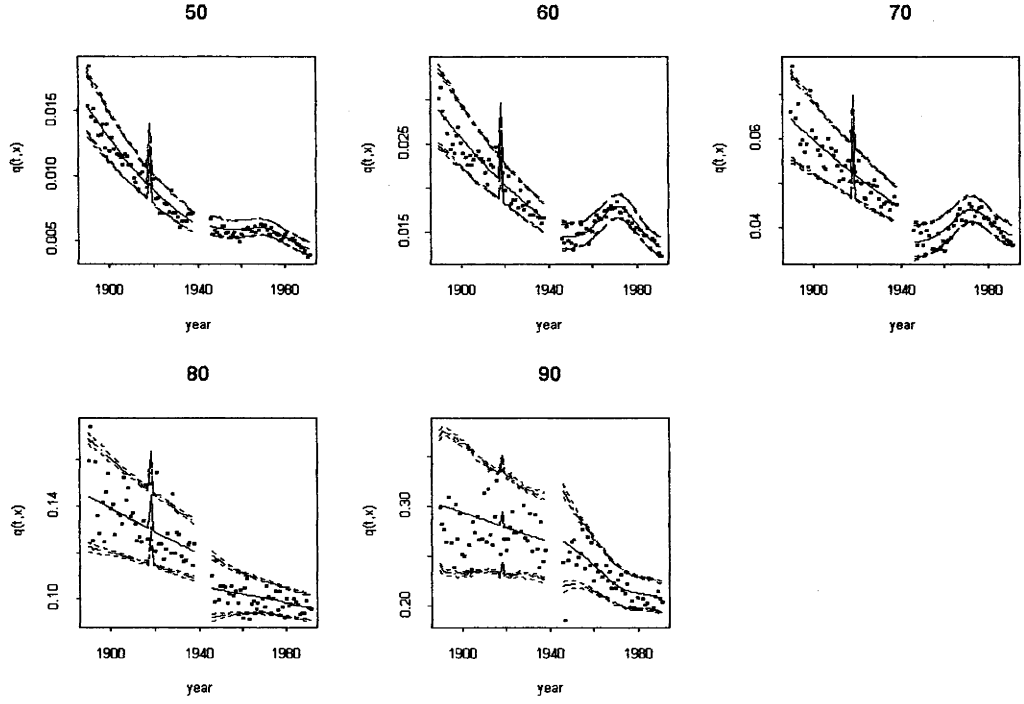
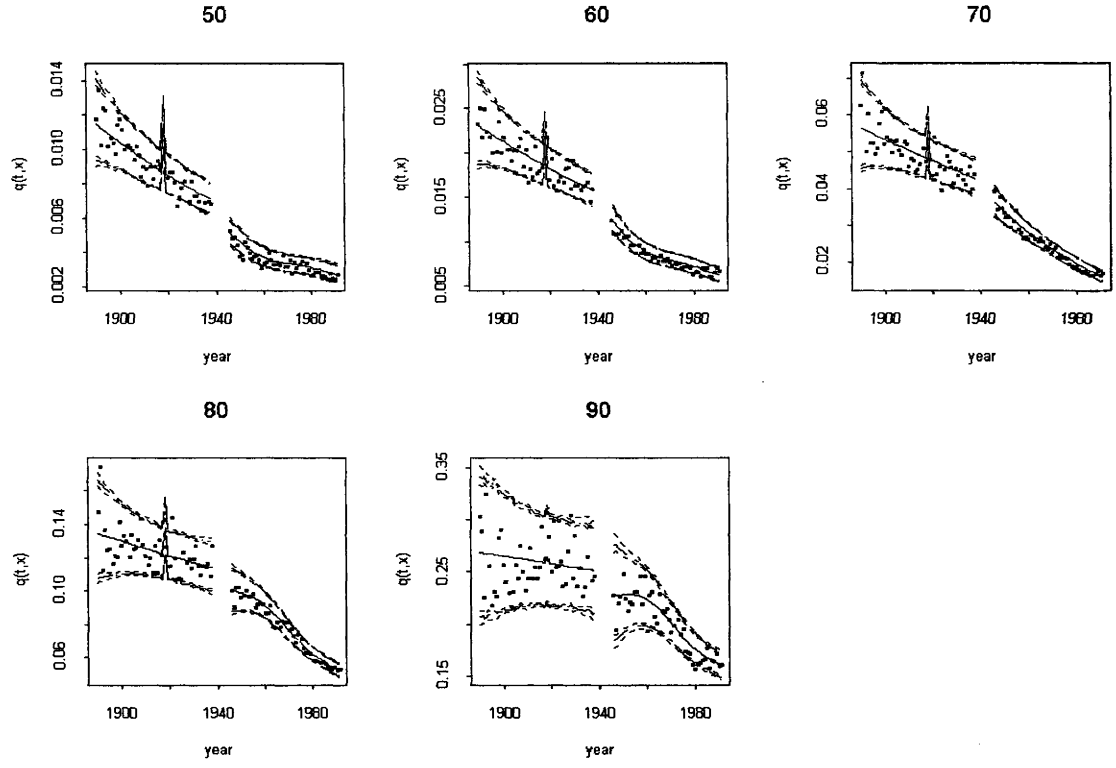


Figure 5.3 Female observed and fitted q(t,x) and prediction intervals 1890-1991.



$\hat{q}(t,x)$ from the fitted models for a selection of ages and years are given in Table 5.1 and Table 5.2. In general, the standard error used in forming the upper bound of a prediction interval around $\hat{q}(t,x)$ will be larger than the corresponding standard error used in the lower bound. This follows from conversion of the lower and upper bounds of the prediction interval surrounding the fitted log(odds) due to exponentiating the end points.

Table 5.1 **Male fitted q(t,x) x 100% (approximate standard errors for use in prediction intervals are given in brackets)**

year age	1890	1900	1910	1920	1930	1950	1960	1970	1980	1990
50	1.54 (0.12)	1.28 (0.10)	1.07 (0.07)	0.89 (0.06)	0.74 (0.05)	0.59 (0.04)	0.58 (0.03)	0.60 (0.03)	0.52 (0.03)	0.44 (0.03)
60	2.89 (0.22)	2.57 (0.18)	2.29 (0.14)	2.04 (0.12)	1.82 (0.10)	1.45 (0.07)	1.57 (0.07)	1.79 (0.07)	1.60 (0.06)	1.34 (0.05)
70	6.42 (0.50)	5.96 (0.41)	5.54 (0.34)	5.14 (0.29)	4.77 (0.24)	3.71 (0.19)	3.96 (0.17)	4.38 (0.15)	4.18 (0.14)	3.83 (0.12)
80	14.40 (1.18)	13.88 (1.03)	13.37 (0.89)	12.88 (0.79)	12.41 (0.68)	10.38 (0.59)	10.21 (0.44)	10.05 (0.36)	9.84 (0.31)	9.62 (0.28)
90	30.14 (3.36)	29.37 (3.36)	28.62 (2.77)	27.87 (2.58)	27.14 (2.19)	25.85 (2.00)	24.03 (1.39)	22.04 (1.00)	21.25 (0.83)	20.78 (0.75)

Table 5.2 **Female fitted q(t,x) x 100% (approximate standard errors for use in prediction intervals are given in brackets)**

year age	1890	1900	1910	1920	1930	1950	1960	1970	1980	1990
50	1.15 (0.12)	1.04 (0.08)	0.94 (0.07)	0.85 (0.06)	0.77 (0.05)	0.46 (0.04)	0.37 (0.03)	0.33 (0.02)	0.31 (0.03)	0.27 (0.03)
60	2.31 (0.25)	2.14 (0.17)	1.97 (0.13)	1.82 (0.11)	1.68 (0.09)	1.11 (0.06)	0.88 (0.05)	0.79 (0.04)	0.72 (0.04)	0.62 (0.04)
70	5.64 (0.61)	5.32 (0.41)	5.03 (0.32)	4.74 (0.27)	4.48 (0.25)	3.37 (0.17)	2.80 (0.12)	2.33 (0.10)	1.95 (0.08)	1.62 (0.07)
80	13.43 (1.49)	12.97 (1.06)	12.53 (0.84)	12.10 (0.74)	11.68 (0.72)	9.86 (0.57)	9.08 (0.40)	7.57 (0.29)	6.18 (0.22)	5.27 (0.19)
90	26.92 (3.45)	26.54 (2.86)	26.17 (2.46)	25.80 (2.20)	25.43 (2.05)	22.84 (1.99)	22.33 (1.40)	20.00 (0.97)	17.59 (0.74)	16.06 (0.61)

As expected from examination of the variance structure of the residuals of the fitted models in the previous chapter, $\hat{q}(t,x)$ for the upper ages and the earliest years have the maximum standard errors, and the lowest standard errors are associated with the most recent years.

The higher variance (and standard errors), in general, for the pre-WWII mortality regime is seen in the figures and tables above. Besides the fall in variance from pre- to post-WWII, there is a fairly steady decline in standard error throughout the 100 years from 1890 to 1990, reflecting a gradual and steady decline in the heterogeneity of mortality over the century.

The parameter estimates for the fitted models can help quantify the extent of mortality change. In order to examine the impact of the mortality hump on the apparent linear decline in male mortality, we can fix age and examine the change in year. For example, for a man aged 60, the model (4.4.2m) reduces to:

$$\delta(\underline{\beta}';t,60) = 0.951 - 0.00266t + 6.44dnorm(t,1972,9)$$

In 1972, $dnorm(1972,1972,9) \cong 0.04433$, so,

$$\delta(\underline{\beta}';1972,60) = -4.00$$

For a seventy-five year old in 1972,

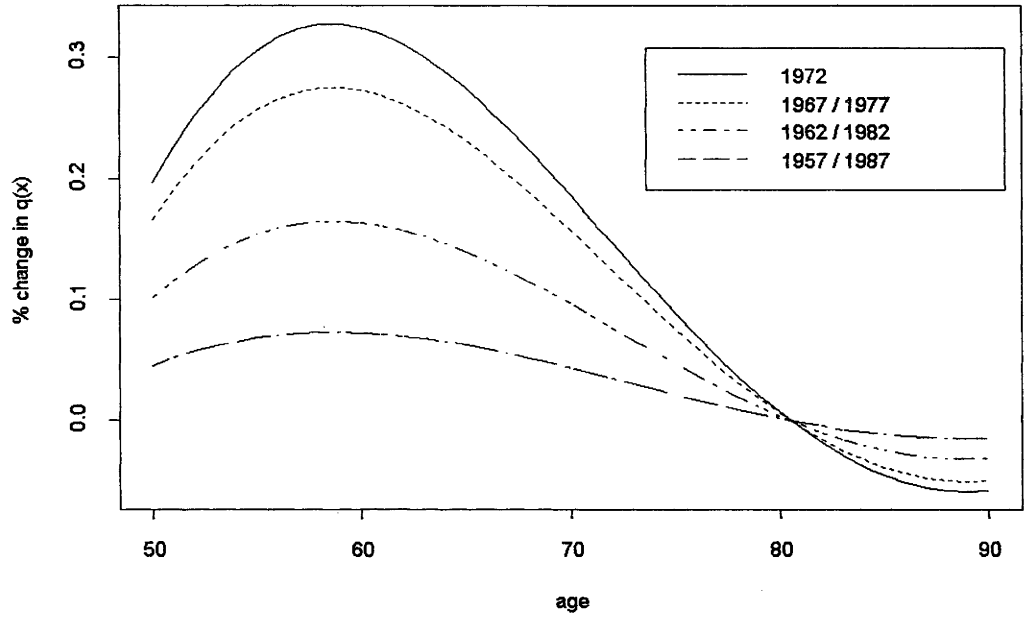
$$\delta(\underline{\beta}';t,75) = -2.43 - 0.000152t + 2.06dnorm(t,1972,9) = -2.64$$

The normal ordinate has a much greater impact for 60 year olds than 75 year olds. For male 60 year olds in 1972, the log(odds) has increased by 0.286 above the mortality level that would have persisted if the hump was not present. In terms of the rate of mortality $q(t,x)$ this equates to a fitted rate of 0.0179 with the mortality hump as opposed to 0.0135 without, which is equivalent to a 32.5% increase. In comparison, the percentage increase for a 75 year old man as a result of the normal density ordinate in 1972 is only 9%.

Figure 5.4 displays the percentage change in the fitted mortality rate as a result of the normal ordinate. The percentage increases represent the change due to inclusion of the normal density ordinate in the male mortality model. An unexpected result of the fit is that the normal ordinate decreases the log(odds) (and, therefore, decreases the mortality rate) at

ages above 80. Figure 5.4 should not be seen to explicitly represent the effect on mortality as a result of CHD, however, in the absence of CHD it is likely that the normal ordinate would not have been required in the model.

Figure 5.4 **Dutch Males. Percentage change in q(t,x) for a selection of years due to inclusion of normal density ordinate**



To illustrate the descriptive advantage of using linear models over non-linear in year, equations (4.4.1m) and (4.4.1f) (pre-WWII) are written out with their parameter values. The indicator capturing excess mortality due to the Spanish Influenza is omitted in the following examples.

Males

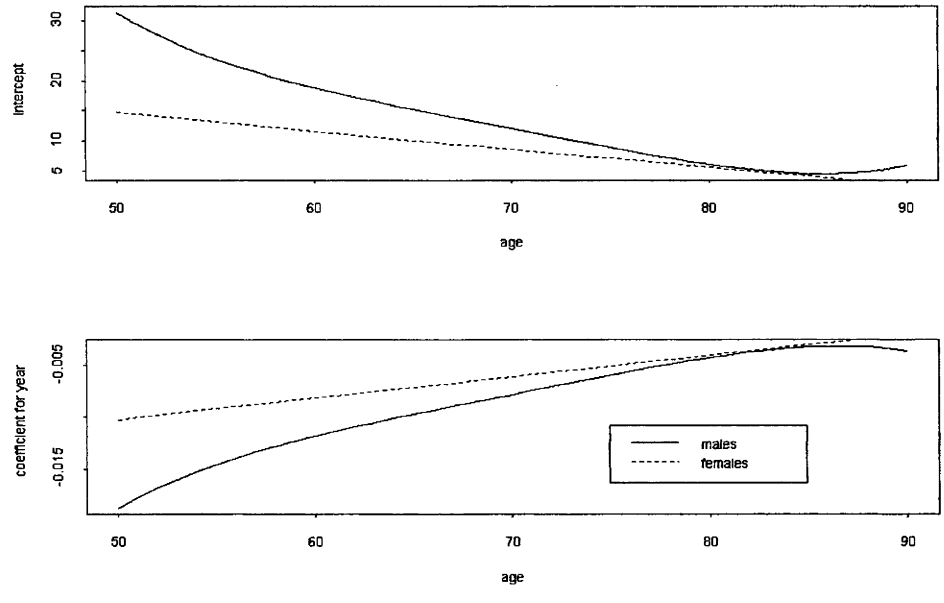
$$\delta(\underline{\beta'};t,x) = 1035.3 - 56.50x + 1.20x^2 - 0.011x^3 - 4e-06x^4 +$$
$$t\left(-0.543 + 0.029x - 0.00062x^2 + 5.93e-06x^3 - 2.12e-08x^4\right)$$

Females

$$\delta(\underline{\beta}';t,x) = 45.0 + 1.06x + 0.0135x^2 - 0.00010x^3 + 2.76e-07x^4 + t(-0.0205 + 2.06e-04x)$$

When age is substituted into the equations, the slope of change in log(odds) with year is revealed. For males aged 50, the coefficient for year reduces to –0.019. For 70, it becomes –0.008, and for 90 it is close to 0. This implies that the rate of decline in log(odds) for 50 year olds was over twice as great as for 70 year olds (on average) between 1890 and 1938. For females aged 50, the coefficient for year reduces to –0.010 and for age 70, it becomes –0.006. Figure 5.5 displays the intercept and coefficient for year for ages 50 to 90 for the models (4.4.1m) and (4.4.1f). Although the rate of decline is much higher for younger women than for older women, the relative difference in mortality decline between the ages and the absolute rate of improvement was greater for males than females pre-WWII. The rate of change in log(odds) from one age to the next can be determined by taking the derivative with respect to age. Exploring the derivatives is the topic of Section 5.3.

Figure 5.5 **Dutch males and females. Intercept and effective coefficient for year for (4.4.1m) and (4.4.1f)**



5.3 Exploring the Derivatives

An advantage of modelling the mortality surface as opposed to individual years is that a surface allows exploration of trends that may only become apparent when a large number of years is considered. In addition to providing checks on the adequacy of the fitted models, patterns in the derivatives raise interesting questions regarding the changing pattern of mortality over time, including the presence of cohort effects (see Section 3.7, Chapter 3).

As mentioned in Section 3.5, Chapter 3, Horiuchi and Coale (1990) proposed the statistic k(x), being the proportional increase or decrease with age in the risk of death at a given age x. Since log(odds) is the statistic of interest in the models proposed in this chapter, the exploration of the derivatives involves log(odds) rather than log(m(x)) used by Horiuchi and Coale. We would expect higher rates of change working with log(odds) than with m(x) since at high ages, increases in m(x) have a proportionately higher impact on odds.

Horiuchi and Coale are restricted in their analysis of derivatives due to the lack of complete data. As a result, they approximate k(x) by the difference in the natural logarithm of the central death rate along individual periods. We do not have this restriction, so we proceed to estimate derivatives along cohorts. Due to volatility in the rates, Horiuchi and Coale (1990) first smooth the data by using moving averages before estimating the derivatives. Since we have a smooth surface in the form of the fitted models (4.4.2m) and (4.4.2f), exact derivatives can be extracted by differentiating the fitted model, or they can be approximated by taking first differences between successive log(odds).

A shortcoming of this approach is that the first differences extracted are dependent on the shape of the smooth surface. Patterns may be revealed that are an artifact of the parameterised surface rather than the real data, and similarly, patterns that may exist in the derivatives of the observed data may be obscured by the smooth surface. A solution is to fit smooth curves to the first differences of the observed data along cohorts. This method was attempted but abandoned due to extreme volatility in the observed differenced log(odds) which restricted the objective fitting of a smooth curve to the differences.

We follow Horiuchi and Coale (1990) and approximate the derivative by the first difference along cohorts of the smooth surface. We define the rate of change of log(odds) for cohort c and age x as:

$$j(c,x) = \frac{d\log(odds(c,x))}{dx}$$

$j(c,x)$ is estimated by the first difference:

$$\hat{j}(c,x) = \log(\hat{o}dds(c,x)) - \log(\hat{o}dds(c,x-1))$$

Figure 5.6 **Dutch females. First and second differences of log(odds) between ages by cohort for female model (4.4.2f).**

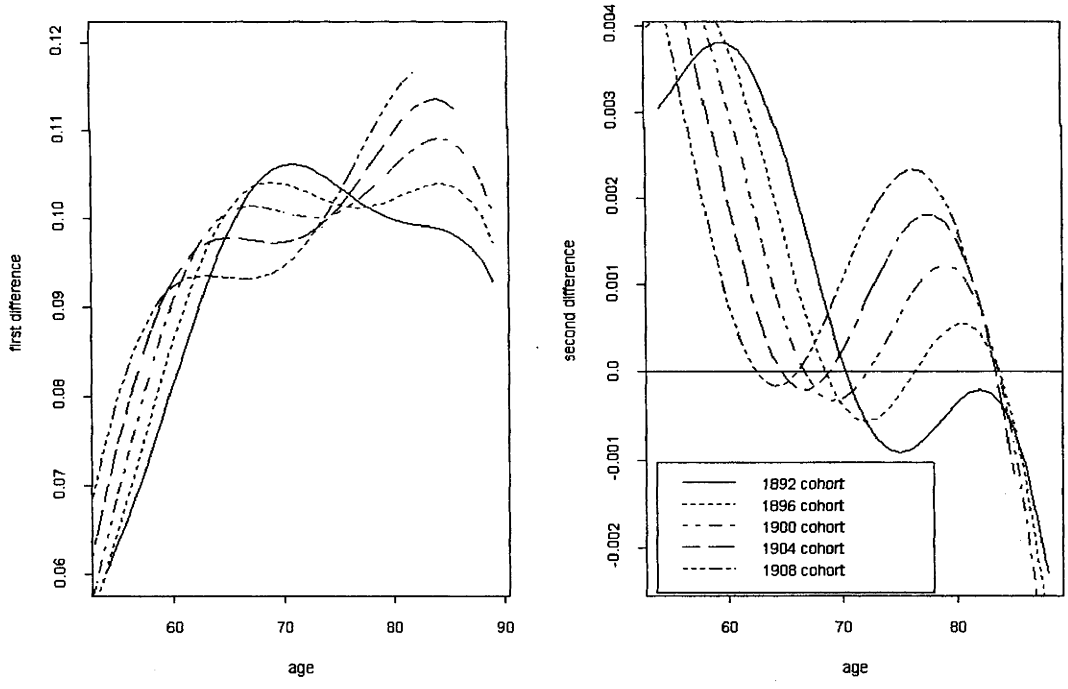
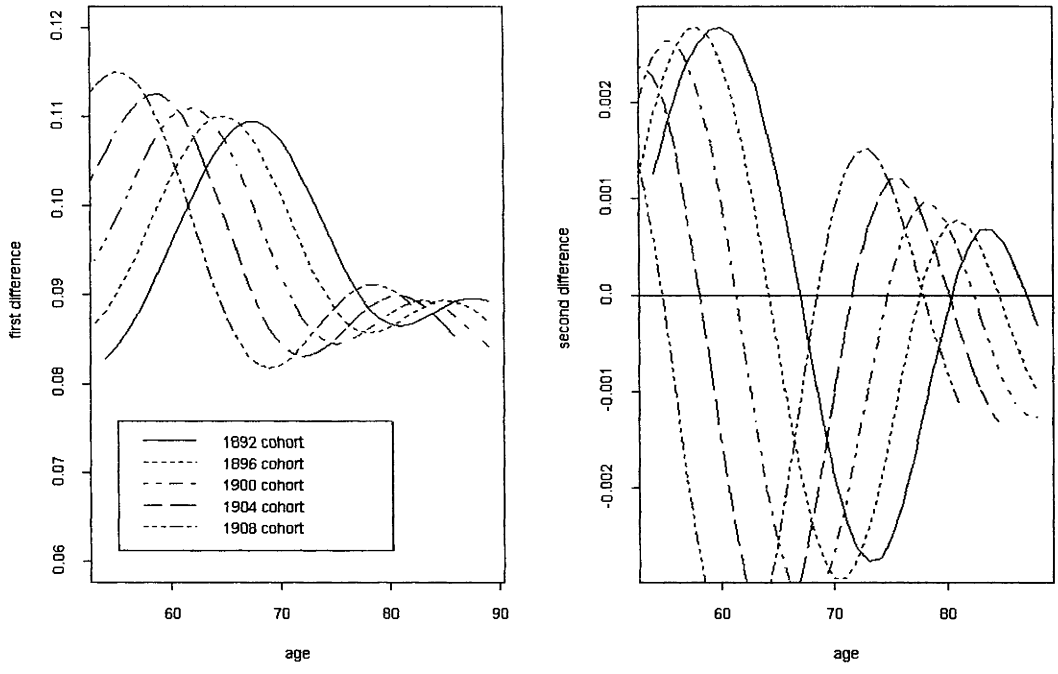


Figure 5.7 **Dutch males. First and second differences of log(odds) over cohorts for male model (4.4.2m).**



The first and second differences were estimated for five cohorts (1892 to 1908) for both the male and female post-1945 fitted models. The results are displayed in Figure 5.6 and Figure 5.7. Both have three sets of turning points (corresponding to where the second differences are zero) for the majority of cohorts examined.

For females, the first turning point in the first differences relates to the part of the mortality curve where the rate of increase in mortality has temporarily peaked. Although the cohorts span sixteen years, the range of the maximum is only seven years. In Section 3.6 of Chapter 3 it was shown that a local maximum derivative with respect to year appears in the mid 1960's, presumably due to the CHD epidemic. The closeness of the maxima in Figure 5.6 above, suggests that the CHD period effect is at least partly responsible for the first turning point in this instance. Table 5.3 lists the age and year at which the turning point in the first differences lies for the female cohorts.

Table 5.3 **Dutch females. Age and year corresponding to turning points in the first differences for log(odds) for a selection of cohorts.**

Cohort	Age at first turning point	Corresponding year	Age at second turning point	Corresponding year
1892	71	1963	-	-
1896	69	1965	76	1972
1900	66	1966	72	1972
1904	64	1968	69	1973
1908	62	1970	66	1974

The second turning point in the first differences relates to the part of the curve where the rate of increase in mortality has leveled off before starting to accelerate yet again (with the exception of the 1892 cohort in this example). For all the female cohorts examined here, this minimum appears to correspond with the period 1972-1974. In Section 3.6 of Chapter 3 it was shown that both males and females undergo a minimum derivative with respect to age near 1974.

The trend in the male differences appears similar to that observed for the females, though the correspondence between the turning points of the first differences and the mortality hump is much stronger, which is to be expected given the extent of the male mortality hump. Table 5.4 lists the age and year at which the turning point lies for each cohort examined. All of the cohorts reveal a maximum derivative corresponding with the late 1950s and early 1960s (when mortality was rising at its quickest rate); and a minimum derivative in the mid 1970s (when the rate of decline in mortality reached a temporary lull following the rapid fall in rates after the mortality hump).

Table 5.4 **Dutch males. Age and year corresponding to turning points for log(odds) for a selection of cohorts.**

Cohort	Age at first turning point	Corresponding year	Age at second turning point	Corresponding year
1892	67	1959	80	1972
1896	64	1960	78	1974
1900	61	1961	75	1975
1904	58	1962	71	1975
1908	55	1963	68	1976

Of note is the negative relationship between age and year at which the first turning point appears for both males and females. The relationship for the male model is quite weak however: the year corresponding to the first turning point varies only four years (from 1959 to 1963) while age varies by twelve years (67 to 55). For females, the negative relationship is stronger, with the year varying by seven years while age varies by nine (see Table 5.3). This is further support for the location and shape of the mortality hump varying with age (first raised in Chapter 3). For the male model this is also confirmed by the significance of the interaction terms between age and the normal density ordinate.

Analysis of Figure 5.6 and Figure 5.7 suggests that the shifting in the estimated derivative as cohort changes is the result of the mortality hump. When effects such as this predominate, they inhibit comparisons between the cohorts through analysis of the derivatives.

The major conclusion of the analysis undertaken in this section to date is that attempts to reveal patterns in the age structure of mortality are confounded by period effects when analysis is undertaken along cohorts.

Due to the presence of strong period effects, we follow Horiuchi and Coale (1990) by exploring derivatives along individual periods. Since the only non-linear component of the mortality surfaces is a function of year, we can express the derivatives with respect to age

for each year as a simple expression by differentiating the parametric surface as below. Here we use the fitted smooth surface. It is monotone and hence the turning points of the derivatives and the differences should have similar properties.

For the female model (4.4.2f),

$$\frac{d\delta(\underline{\beta};t,x)}{dx} = \beta_1' + 2\beta_2' x + 3\beta_3' x^2 + 4\beta_4' x^3 + \beta_6' t + 2\beta_7' tx + 3\beta_8' tx^2 + 4\beta_9' tx^3 + dnorm(t,1959,12)\left(\beta_{11}' + 2\beta_{12}' x + 3\beta_{13}' x^2\right)$$

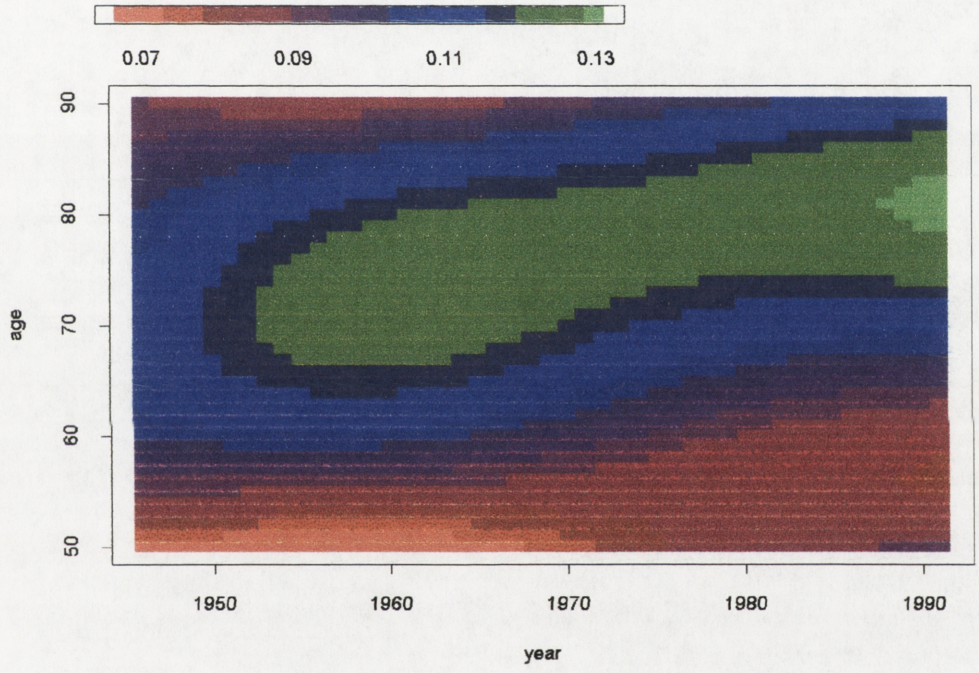
$$\frac{d^2\delta(\underline{\beta};t,x)}{dx^2} = 2\beta_2' + 6\beta_3' x + 12\beta_4' x^2 + 2\beta_7' t + 6\beta_8' tx + 12\beta_9' tx^2 + dnorm(t,1959,12)\left(2\beta_{12}' + 6\beta_{13}' x\right)$$

Setting the second derivative to zero and solving for x with different values of t identifies the turning point for each year. The results of this analysis for females are given in Figure 5.8, which displays the first derivatives for selected years and Figure 5.9, which gives the age corresponding to the maximum of the derivative for each year from 1946 to 1990. As shown here, a polynomial in age also satisfactorily captures the bell-shaped pattern.

The analysis shows that not only does the female derivative display a definite pattern over the age range, but since 1950 the location of the maximum of the derivative appears to be increasing over time (see Figure 5.9). Horiuchi and Coale (1990) note a peak in the female estimated derivatives at around age 75 for different countries, mostly in the late 1960s to the mid 1970s, and a similar pattern is observed by Horiuchi and Wilmoth (1998). The results presented in this section show that the peak for Dutch females appears to increase, almost linearly, from 70 in 1950 to over 80 in 1991. It is worth pointing out that the peak at age 75 corresponds with the late 1960s and early 1970s for the Dutch female data, as it does for most of the years studied by Horiuchi and Coale (1990).

Horiuchi and Coale (1990), and Horiuchi and Wilmoth (1998) speculate on the reasons behind the bell-shaped pattern in the context of individual and aggregate mortality distributions. In particular, if the individual risk of death follows the Makeham equation and if individual frailty is gamma distributed, then the aggregate mortality rates follow the Perks distribution. They show that the bell-shaped derivatives are consistent with the Perks distribution.

Figure 5.8 **Dutch Females. Contour plot of first derivatives of log(odds) along periods from (4.4.2f). Ages 50 to 90. Year 1946 to 1991.**



As shown here, a polynomial in age also satisfactorily captures the bell-shaped pattern. Horiuchi and Coale (1990), and Horiuchi and Wilmoth's (1998) explanation, that large individual variations in mortality are responsible for the shape of the derivatives, is impossible to test with the Dutch data, since cause-of-death and data disaggregated by other variables would ideally be required to test frailty distribution assumptions.

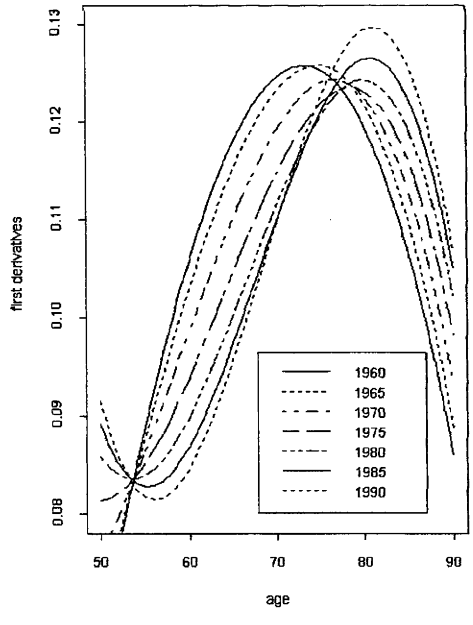
It is, however, worth reflecting on the changing maxima in the context of the ideas of Horiuchi and Coale and Horiuchi and Wilmoth. The combination of individual Makeham

distributions and the Perks distribution can explain the bell-shaped curve as follows (Horiuchi and Coale (1990), and Horiuchi and Wilmoth (1998)): the proportion of mortality risk that is chance-related decreases as a person becomes older, and degeneration-related risk increases proportionately. The increase in the rate of mortality due to degeneration-related risk is countered by a slowing down in the rate of mortality increase due to selection reducing the overall frailty of the surviving population.

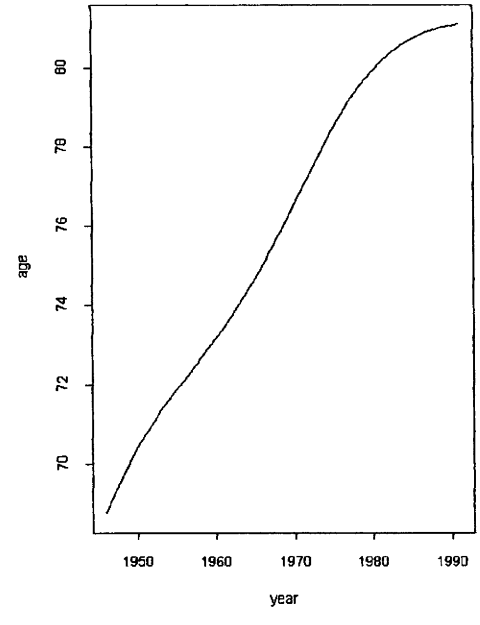
In this context, the changing age at which the maximum occurs for the Dutch females could be seen as a reweighting of the ratio of chance-related to degeneration-related risks over time. In particular, the proportion of mortality risk that is chance-related is increasing over time for the same age, meaning that the dominance of degeneration-related risk is being pushed forward to older ages. Therefore, in the context of selective survival, the Dutch female results point to a slowing down in the pace of ageing.

Figure 5.9 Dutch Females.

First derivatives of log(odds) along periods based on female model (4.4.2f).



Maximum of the derivative by age and year.



Does the maximum of the derivative mark the boundary between the "old" and the "old old" ? As seen in Figure 5.9, the maximum derivative appears to be flattening near age 81. Whether this is a real effect reflecting the approach of a limit or a temporary lull that will be reversed by subsequent mortality data can only be speculated. However, mortality studies based on more recent data provide considerable evidence of a deceleration in the rate of increase of mortality at ages over 85 (Tuljapurkar and Boe,1998). The changing position of the maximum of the derivative over time is further support that simple functions that have constant derivatives over all ages, such as the Gompertz, are not suitable for modelling aggregate Dutch mortality.

In order to ascertain whether the patterns in the first differences and derivatives were an artifact of the function chosen for the mortality surface, high order splines were used to smooth the observed data and first differences were extracted for comparison. The method undertaken in producing the differences in this instance was to fit natural splines through the S-plus function 'ns' with various degrees of freedom, and to calculate:

$$\log(odds_{x,t-x}) - \log(odds_{x-1,t-x}) \qquad \text{for the differences along cohort}$$

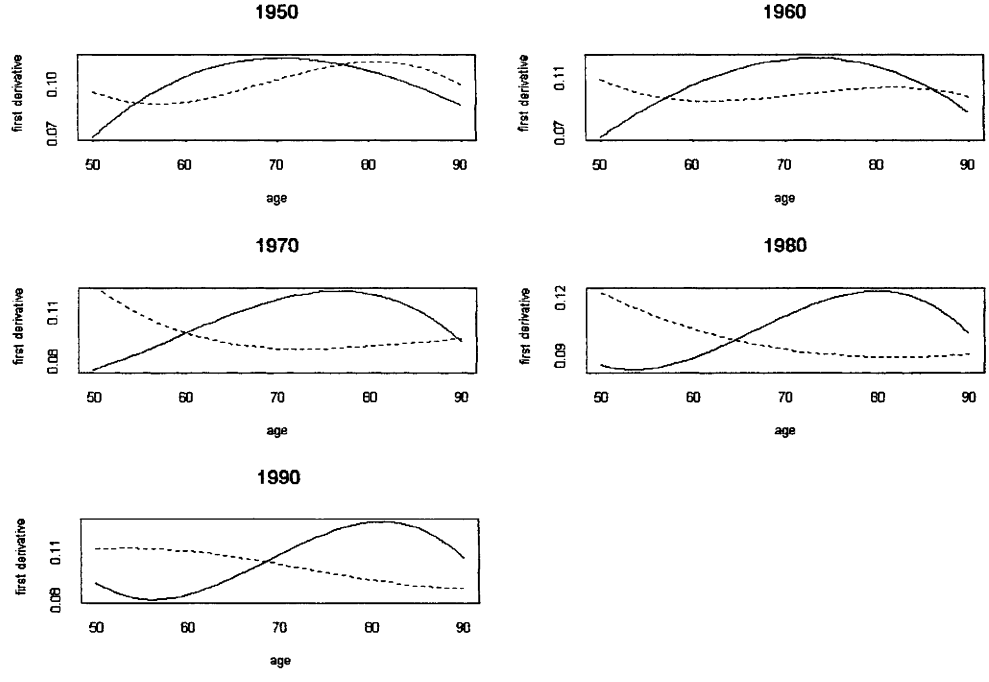
$$\log(odds_{x,t-x}) - \log(odds_{x-1,t-x+1}) \qquad \text{for the differences along period}$$

The patterns in the estimated derivatives were similar to those produced along cohort and period for the most part, thus providing additional support for the fitted models and confirming the validity of the patterns in the derivatives.

The male first derivatives display a dramatically different pattern to the females (see Figure 5.10). Of note is a general decrease in slope for males over recent years, as opposed to a general increase for females. The fluctuating pattern in the derivatives seen for the Dutch males is consistent with the findings of Horiuchi and Coale (1990). However, Horiuchi and Wilmoth (1998) report a similar bell-shaped pattern for Japanese and Swedish male mortality as for females.

Although not apparent from Figure 5.10, as shown in Chapter 3 and Figure 4.16 from the previous chapter, the diagonal pattern in the male derivatives may suggest the presence of a cohort effect. Figure 5.11 displays the first derivatives of the male fit (4.4.2m) and the estimated derivatives (first differences) based on a spline fit to the 1946-1991 surface (reduced version of Figure 3.21). The two dotted lines transecting the plots are the 1890 and 1905 cohorts.

Figure 5.10 **Dutch Males and Females. First derivatives for selected years based on (4.4.2f) and (4.4.2m). Ages 50 to 90. (solid line=females; dotted line=males)**

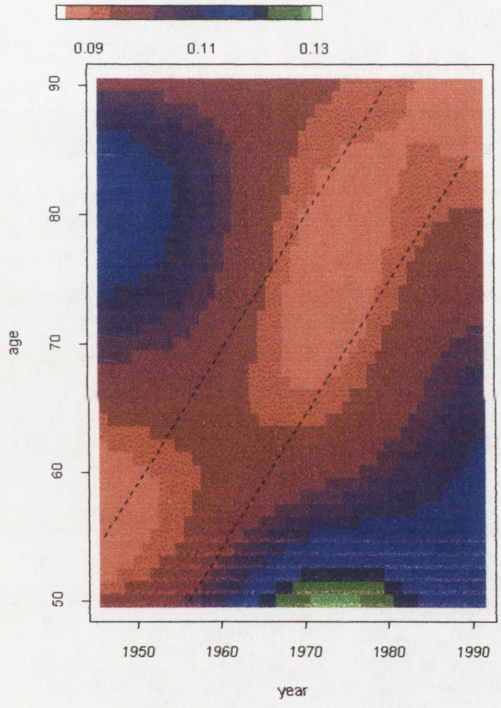


The spline surface contour plot suggests the strong presence of a cohort effect, which is captured to some extent by the fitted male model (4.4.2m) through interactions between polynomials in age and year, and the interactions between the polynomial in age and the normal ordinate centered at 1972. As discussed previously, higher order polynomials in year and interactions between these terms and the polynomial in age would be required to more adequately capture the diagonal pattern observed in the first derivatives. Although modelling by higher order interactions is possible, it is more likely that the true cause of these patterns is due to variation between cohorts.

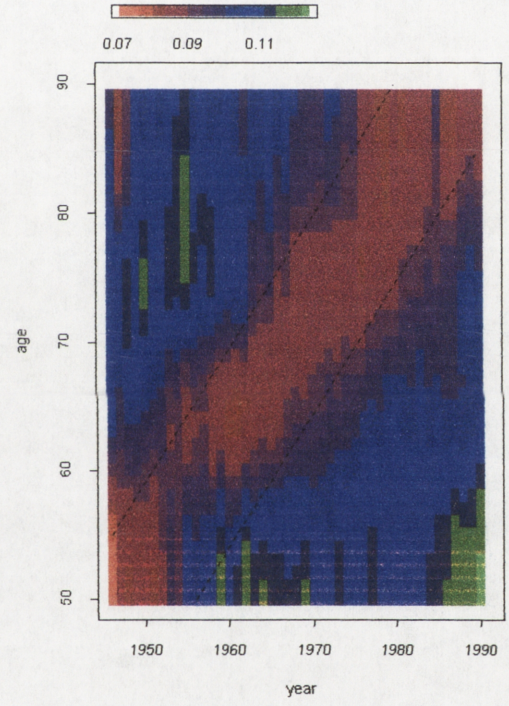
If we take the view that the diagonal pattern in the derivatives is indeed cohort related, then we can generate a new variable θ_c by examination of the residuals by cohort that could be added to the fitted surface. The next chapter takes a similar approach to identification of cohort effects: age and year effects are modelled and the residuals are decomposed. As will be shown in the next chapter, there is a strong argument that the diagonal pattern in the derivatives is indeed due to a cohort effect.

Figure 5.11 Dutch Males.

First derivatives of log(odds) along periods based on male model (4.4.2m). Ages 50 to 90.



Estimated first derivatives along Periods based on spline surface.



5.4 Conclusions

The regression models explored in this chapter using polynomials and interactions result in estimates of parameters with standard errors, thus providing confidence bounds for the mean or prediction intervals for individual values. Extra-binomial weights are incorporated in order to generate realistic standard errors for the rates of mortality. Rates of mortality,

conditional probabilities of survival and period and cohort life expectancies are generated and projected in Chapter 7.

Exploration of the coefficients from the fitted models provide a means of quantifying the change over time in mortality, the coronary epidemic of the 1970s, and other features. Studying the derivatives from the fitted surfaces leads to the identification of some interesting features and raises a number of questions. In particular, a shift in the location of the maximum of the derivative in females along individual periods from 1950 to 1990 has been observed. Is this increase evidence of a slowing down in the pace of ageing, in other words, evidence of increase in lifetime with year? Exploration of derivatives along cohorts offers no additional support to this hypothesis, as changes appear to be masked by period effects.

Despite the advantages of the regression models discussed, a disadvantage of the male model is that it does not explicitly take into account the potential cohort effect running through the surface; rather the effect is implicitly accounted for by interactions between a polynomial in age and year. An alternative means of modelling Dutch male mortality explicitly accounting for cohort effects by using a fixed effects model based on the techniques of Wilmoth (1990) and Lee and Carter (1992) is the topic of the next chapter.

6 Modelling Dutch Mortality using a Fixed Effects Model– Ages 50-90.

6.1 Introduction

In this chapter the approach of Wilmoth (1990) is taken, whereby a cohort effect, if it exists, is assigned the parameter θ_c, where the form of the parameter is determined after fitting any additive terms and the main multiplicative components of the mortality surface. The determination of the number of multiplicative components to fit depends on the results of a singular value decomposition of the residuals after fitting effects for year and age. Lee and Carter (1992) generate a similar model of mortality but do not consider the presence of cohort effects. The model generated by Lee and Carter can be modified along the lines of Wilmoth to include a cohort effect θ_c. An immediate criticism of Wilmoth's approach, which he himself raises, is that the choice of separation of cohort effects from year by age interactions may not be clear. The justification for imposing a cohort effect on the residuals in the case of the Dutch males comes from a visual inspection after fitting the additive components. It is, therefore, a subjective choice by the researcher on whether the pattern observed is indeed cohort related. In essence, Wilmoth's technique contains the same deficiencies as other methods that rely on subjective graphical inspection. It does, however, offer a means of modeling the cohort effect if and when the evidence for such an effect is plausible.

The resulting θ_c for males is given in Figure 6.4, the shape of which points to a strong cohort effect. The shape of θ_c suggests that survivors born between 1885-1895 are relatively more robust than males born in the subsequent cohorts. It is speculated that the Spanish Influenza may have had a long-term debilitating impact on the future mortality of adolescents exposed during 1918, thereby increasing the frailty of these cohorts relative to adjacent cohorts. In contrast, it is plausible that the same outbreak may have lead to a decrease in average frailty for the 1885-1895 cohorts due to removal of the weaker members. Lending weight to the legitimacy of the cohort effect is the closeness of fit that follows when θ_c is applied to the simple additive fit (see equation (6.5.1) and (6.6.2)).

The assumptions from Chapter 2 that underlie the construction of the regression model in Chapter 4 are not applied to the generation of the models in this chapter. Future work could include modelling the variance for the additive models by using REML, where the binomial weights proposed in Chapter 2 are incorporated.

6.2 Motivation

Despite the theoretical advantage of fitting a regression model to the mortality surface, high order interactions that are required to capture the non-linearities are difficult to interpret, prevent sensible predictions, and may mask legitimate cohort effects. The male fitted model (4.4.2m) produces a close fit to the observed data, but fails to capture the diagonal pattern in the derivatives (see Figure 5.11) which is suggestive of a strong cohort effect.

The models examined in this chapter do not enable standard errors to be generated in an easy way as they are for the regression model. In addition, the statistical assumptions underlying the regression model are not incorporated during the fitting procedure. However, the approach of this chapter enables close examination of the effects of age, period *and* cohort on mortality. Separation and quantification of a cohort effect not only enables decomposition of historical patterns, but can raise questions of an epidemiological nature regarding the long-term health impacts of periods of debilitation, and can lead to projections that incorporate cohort as well as age and year as predictors.

The models generated in this chapter are based on techniques proposed by Wilmoth (1990) and Lee and Carter (1992). Due to the identification problem associated with age-period-cohort models, a model cannot be estimated that takes age, period and cohort effects into account simultaneously. Wilmoth (1990) approaches the problem by fitting a model in age and year, including interactions between age and year, and then decomposing the residuals of the fit to search for cohort effects.

6.3 Fitting the Additive Models

For this chapter we use a different notation: the subscript x,t refers to age x and year t, rather than age and cohort as it had done in previous chapters.

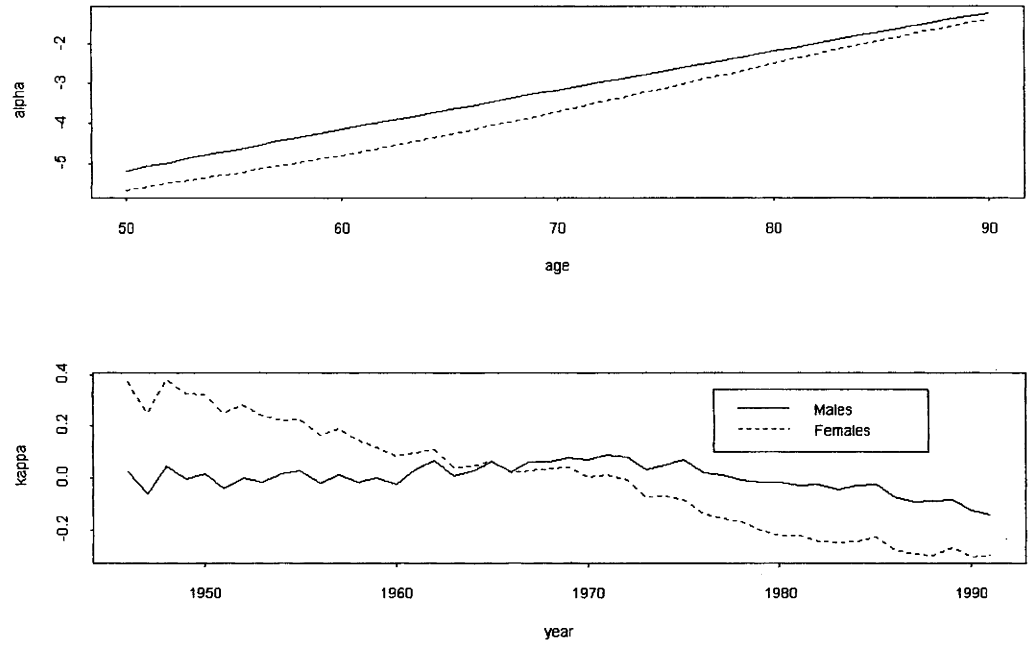
The first step in this procedure is to generate an additive model as below:

$$f_{x,t} = \alpha_x + \kappa_t + \varepsilon_{x,t} \tag{6.3.1}$$

Wilmoth uses $f_{x,t} = \log\left(\frac{q_{x,t}}{1-\frac{1}{2}q_{x,t}}\right)$, where, although it is not explicitly stated in his paper, it is believed that $q_{x,t}$ is the rate of mortality estimated along periods. For consistency with the analysis performed in Chapters 1-5 of this thesis, we use $f_{x,t} = \log\left(\frac{q_{x,t}}{1-q_{x,t}}\right)$, where $q_{x,t}$ is estimated along cohorts as previously. The use of cohort rates of death is more appropriate than period rates when examining patterns in mortality, particularly with regard to uncovering cohort effects. This is argued in Section 6.7. A unique solution can not be found to equation (6.3.1) unless constraints are placed on the model. The method outlined in the appendix of Wilmoth (1990) is followed, whereby α_x is determined as the average of the log(odds) over the period 1946 to 1991, and κ_t is then estimated $\frac{1}{I}\sum_{x=50}^{90}(f_{x,t} - \alpha_x)$, where I is 41, being the number of ages 50 to 90.

The results for males and females are presented in Figure 6.1. It shows that female mortality is lower than male mortality, and that the rate of decline post-1945 has been much more dramatic for females than males. The cardiovascular hump for males is apparent in the plot of kappa.

Figure 6.1 Dutch Males and Female. Alpha and Kappa from additive fit.



6.3.1 Accounting for Variation in Age over Time

An alternative approach to that of Wilmoth is to assume a basic level of interaction to account for the fact that the change in age over time varies from age to age. The basic model proposed by Lee and Carter (1992) introduces a new variable 'beta' that takes into account the variation at different ages over time:

$$f_{x,t} = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t} \qquad (6.3.2)$$

The three effects were estimated using the procedure of Lee and Carter, which, like that of Wilmoth, imposes constraints in order to produce unique estimates; α_x and κ_t were estimated as in Section 6.3. β_x was estimated by regressing κ_t on $f_{x,t} - \alpha_x$ for each t over all ages x.

The results for β_x for males and females are presented in Figure 6.2.

Figure 6.2 Dutch Males and Female. Beta from Lee-Carter fit.

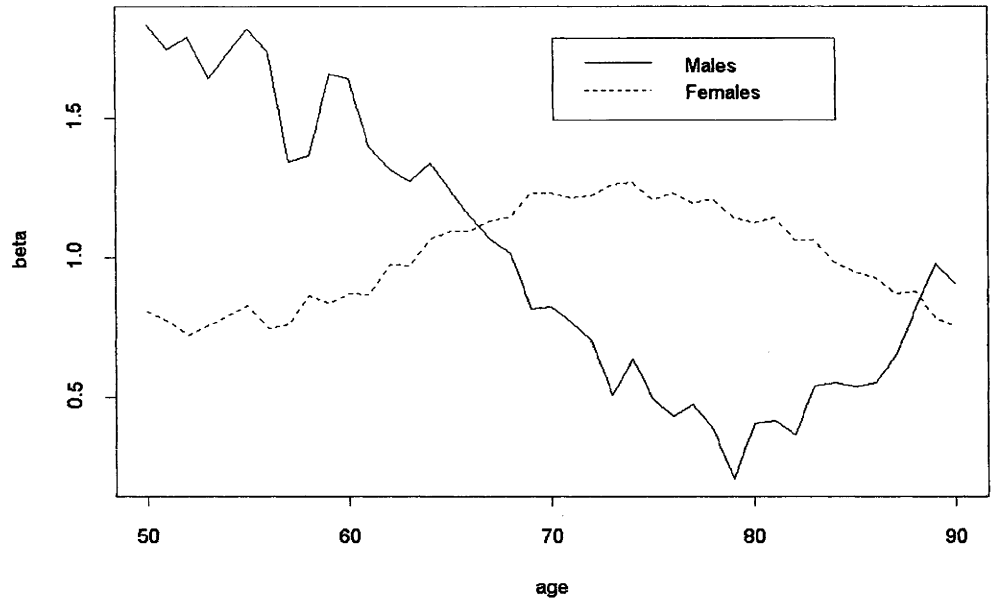


Figure 6.2 shows that the most dramatic change in mortality for males has occurred at age 50, whereas for females, the largest changes have occurred around age 75.

The β_x improves the fit for a number of ages in particular years, however, both fits do not adequately capture the trends in mortality over time. An inadequacy of model (6.3.2) is that it assumes a linear interaction between age and year, whereas the interaction between age and year is highly non-linear (as supported by the significance of the interaction terms between year and the polynomial in age in the regression models of Chapter 4).

The $\beta_x \kappa_t$ represents an interaction between age and time, and as such, may reflect a cohort effect to some extent. Although there is no clear way to ascertain this, some insight may be achieved by proceeding with a decomposition and examination of the residuals following the fits to date.

6.4 Incorporating Interactions and Determining a Cohort Effect

Wilmoth searches for a cohort effect by decomposing the residuals from the additive fit, and searches for a pattern in the residuals that appears to relate to cohort. This is achieved by carrying out singular value decomposition on the residuals, and examining the pattern in terms associated with the significant singular values.

When applying the technique to French male mortality data for ages 0 to 89, Wilmoth (1990) finds that the first two singular values are the most significant, but that when these values are presented as a contour plot they display an irregular pattern of interactions between age and year, rather than some definite cohort effect. To accommodate this pattern, the additive fit is supplemented by a multiplicative component of the first two singular values that accounts for the main interactions between age and year. The residuals from the new fit contained a pattern that appeared to correlate strongly with cohort. Based on the graphical evidence, Wilmoth derives a new term θ_c, that is calculated as the mean of the residuals along each cohort c. The final model is given below:

$$f_{x,t} = \alpha_x + \kappa_t + \sum_{m=1}^{p} \phi_m \gamma_{xm} \delta_{tm} + \theta_c + \varepsilon_{x,t} \qquad (6.4.1)$$

Rather than finding a pattern of non-linearity like that for the French male data by Wilmoth, there appears to be a definite correlation with cohort in the Dutch male residuals from the simple additive model given by equation (6.3.1). In contrast, on inspection there is no definite cohort effect apparent in the female residuals. The residuals for the Dutch male and female additive fits are given in Figure 6.3.

For the sake of comparison, the residuals from the Lee-Carter model (6.3.2) for males and females were decomposed. The male residuals display a similar pattern for both models, whereas there appears to be more evidence of a cohort effect for females though the relationship is still questionable.

After decomposing the male residuals, the first two singular values showed the same diagonal effect for both equations (6.3.1) and (6.3.2). The diagonal pattern observed is consistent with the pattern in the derivatives seen in Figure 5.13 from the previous chapter.

Both Figure 6.3 and Figure 6.4 point to males being subject to a cohort effect. Given the pattern that has emerged from the residuals, rather than fitting a multiplicative term $\sum_{m=1}^{p} \phi_m \gamma_{xm} \delta_{tm}$ as done by Wilmoth (1990), we proceed by estimating the term θ_c for each cohort c directly from the residuals of the additive fit. Given the lack of a definite pattern in the female residuals, θ_c is only estimated from the residuals for both fits for males. When θ_c is below zero this implies that the additive fit has overestimated mortality. Conversely, when θ_c is above zero, the additive fit has underestimated mortality.

The features that are immediately noticeable from Figure 6.4 are the strong periodic relationship between θ_c and cohort and the similarities in θ_c for the two models. Discussion of this pattern is aided through examination of the fitted models after allowing for θ_c (see Section 6.5).

Incorporating an interaction term for females improves the additive fit, but there is insufficient evidence to suggest that a cohort effect θ_c should be included in the model. As a result, the remainder of this chapter involves exploration of Dutch male mortality exclusively.

Figure 6.3 **Dutch Males and Female. Residuals from the additive fit.**

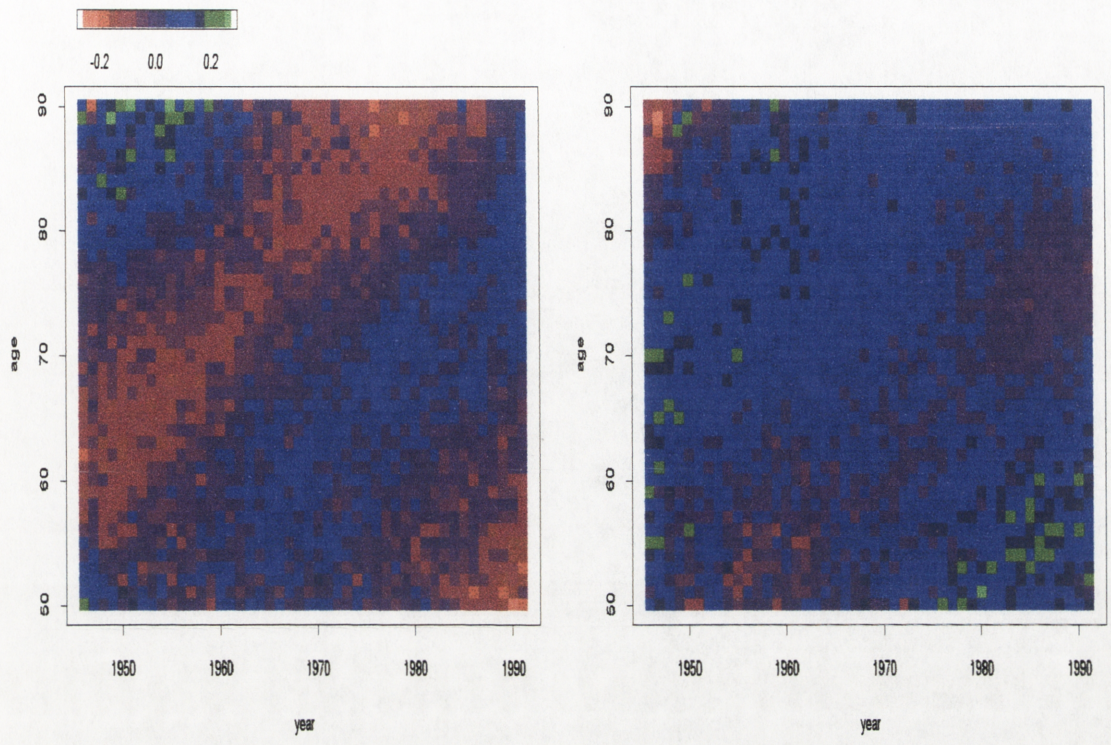
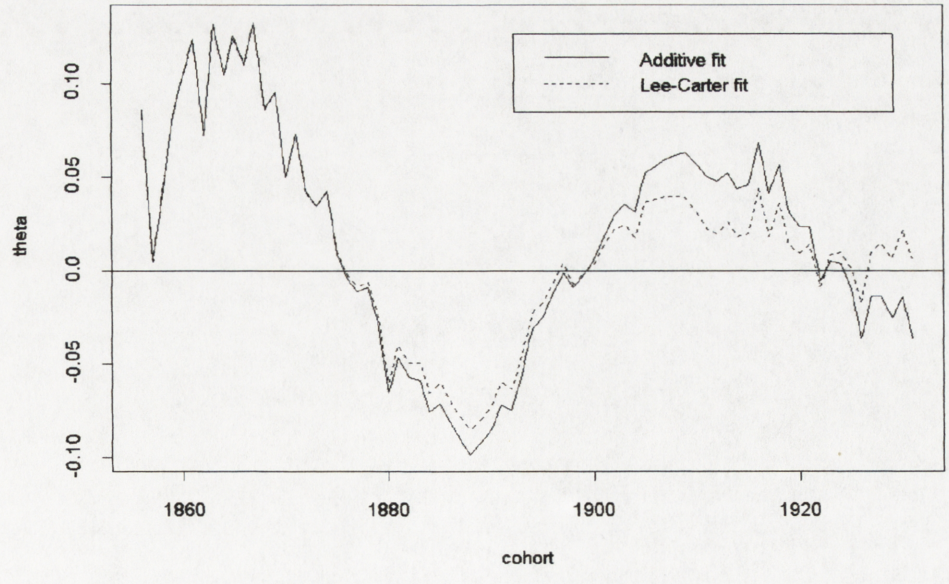


Figure 6.4 **Dutch Males. θ_c = mean of residuals along cohorts after additive fit (6.3.1) and after Lee-Carter fit (6.3.2).**



6.5 Assessing the Fits

Given that no interaction terms have been included, the additive model is determined entirely by the formula:

$$f_{x,t} = \alpha_x + \kappa_t + \theta_c + \varepsilon_{x,t} \tag{6.5.1}$$

whereas the modified Lee-Carter model is:

$$f_{x,t} = \alpha_x + \beta_x \kappa_t + \theta_c + \varepsilon_{x,t} \tag{6.5.2}$$

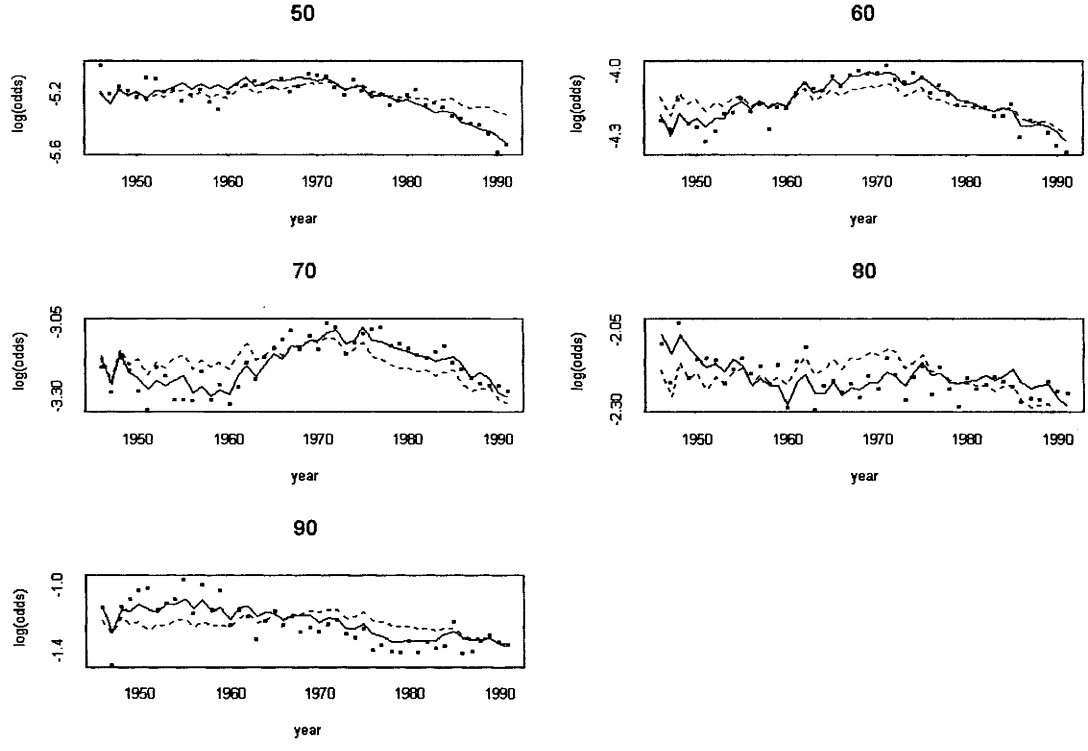
The quality of the fitted models can be determined by examining how the estimated log(odds) follow the observed log(odds) Figure 6.5 shows that inclusion of θ_c has a dramatic effect in improving the additive fit. θ_c has the effect of raising the mortality for cohorts between 1860 and 1875 and between 1900 and 1920 and lowering it for cohorts outside these ranges.

The vast improvements of fit for ages 50 to 90 after incorporating θ_c suggests that the trend over time since World War II can be explained, to a large extent, by differences in mortality due to birth cohort. Figure 6.5 shows that the suspected cardiovascular hump is greatly exaggerated across many ages due to the impact of cohort variation. This raises an interesting question, namely: to what extent are unfavourable conditions in adolescence responsible for the cardiovascular epidemic. Is it due to the impact of psychological stress on the immune system, or physical stress caused by exposure to disease? These questions are better answered by virologists, epidemiologist and researchers in the medical profession, than being taken up here.

The closeness of fit following the incorporation of θ_c suggests that the non-linearities in the trend in mortality over time at ages 50 to 90 can be explained by cohort variation. However, as illustrated by the alternative model based on the Lee-Carter parameterisation,

if a simple interaction is included between age and year, a new theta can be produced that can also be justified.

Figure 6.5 **Dutch Males. Log(odds), additive fits with and without θ_c. (Bold= Additive fit with θ_c (equation 6.5.1); Dotted= Additive fit without θ_c (equation 6.3.1))**



Both the additive model and modified Lee-Carter model with θ_c fit the data closely. The two different, yet both plausible fits, serve as a reminder that any model is simply an approximation of reality. Many different models can adequately capture the patterns in a system and yet not represent the workings of the system. Although both the additive model with cohort effect, and the Lee-Carter model with cohort effect provide adequate fits (for the most part) to the male data, determining whether the theta provided by either model is correct is not possible.

The quality of the fits can also be assessed by examining the residuals $f_{x,t} - (\alpha_x + \kappa_t + \theta_c)$ and $f_{x,t} - (\alpha_x + \beta_x \kappa_t + \theta_c)$ for bias.

Despite the closeness of the fitted model, there are a number of regions of the mortality surface where the fit is poor. The fixed effects models can be improved by including interaction terms to accommodate the deficiencies in the additive fit not captured by the cohort component. The purpose of the fixed effects models in this chapter, however, has been predominantly to explore the impact of cohorts on the observed log(odds), rather than an attempt to produce as close a fit as possible.

The results for the Dutch males support the conclusions that were reached through inspection of the derivatives in Chapter 3 and Chapter 5, and the bias in the residuals by cohort in Chapter 4; that survivors born in 1880-1895 and adjacent cohorts are relatively more robust than males born near 1905-1915. A possible cohort effect, though less apparent, is found when a similar analysis was performed on mortality data for females aged 50 to 90. The cohort effect noted for males is consistent with the findings of Caselli (1990) who noted a cohort effect for French and Italian males serving in World War I and also found evidence of hardship on adolescents born between 1900 and 1905, but found no such effect for females. Since Holland was neutral during WWI, it is likely that any cohort effect is due to the impact of the Spanish Influenza of 1918 rather than from war-time service.

The effect that an extraordinary event has on cohort mortality has been studied in detail by Vaupel et al. (1988). They simulate a number of situations where a debilitating event has varying impact on a cohort's mortality and composition depending on the heterogeneity of the cohort prior to and following the event. Vaupel et al (1988) show that when a heterogeneous cohort experiences an adverse event that affects the cohort in a non-uniform way, but rather has a debilitating effect on the more frail individuals in the cohort, the survivors will subsequently experience lower mortality rates than otherwise due to the selection that occurred. Poppel et al (1994:7) show that for Dutch males, there is evidence of reduced mortality for the 1850 cohort following the 1918 influenza epidemic. It is plausible that the pattern in Dutch mortality for the 1880-1895 cohorts can be explained in this context, since the weakest individuals aged between approximately 25 and 40 would have died during the Influenza epidemic, thereby reducing the average frailty of the surviving population. A question can be raised as a consequence of this proposition: will

the mortality of a cohort stay at a relatively low level or will it rise as if selection had not occurred ? In other words, does a catastrophe affect the survivors so as to eventually increase their mortality ? (see Henry and Ulijaszek (1996) for a discussion of the impact of early environment on long term mortality). It is plausible that this is the case for the later cohorts (1905-1915), who were aged between 3 and 13 during the Influenza epidemic. The extra mortality for these cohorts may be evidence that they have suffered lasting debilitation, rather than the selection effect of the earlier cohorts. We could speculate that age is responsible for the differences in this instance; where a virus outbreak such as the Spanish Influenza has a lasting influence on the health of young children, but less so on young adults who survive the virus. A similar effect may emerge in the mortality data for cohorts exposed to poor nutrition due to the Hunger-Winter of 1944-1945.

It is important to stress that these conclusions are very speculative, and are based on the subjective decisions that year and age are additive effects, that θ_c represents a cohort effect, and that the general shape of θ_c is due to mortality variation brought on by the Spanish Influenza.

Although it is not undertaken here, the variance of the models could be extracted for use in determining confidence and prediction intervals. Following the findings of the previous chapters, there is an argument for decomposing the variance of the error into two components: one based on the theoretical binomial variance, and another based on a random effect. In essence we could replace the fixed effects models (6.5.1) and (6.5.2) with a mixed model with a 3-way classification. REML (restricted maximum likelihood) could then be used to find the variance structure (see, for example, Searle et al. (1992) for a discussion of REML as applied to mixed models).

6.6 Reducing the Dimensionality

A criticism of the fitted model $f_{x,t} = \alpha_x + \kappa_t + \theta_c + \varepsilon_{x,t}$ is the large number of degrees of freedom taken up by the additive components. The dimensionality of the model can be reduced by modeling each of the effects as a function of their respective variables. For example, the additive model can be re-written,

$$f_{x,t} = g(x) + h(t) + l(c) + \varepsilon_{xt}$$

where the functions $g(.)$, $h(.)$ and $l(.)$ are polynomial or non-linear functions.

As discussed in Chapters 3 and 4, age can be modeled by a high order polynomial, or a non-linear function such as the Perks model. The Perks has the advantage of having a theoretical foundation, though it is more appropriate along cohorts rather than periods.

To capture the pattern in κ_t a 4th degree polynomial can be shown to provide a close fit.

Modeling θ_c is more difficult due to the large number of turning points.

If we take the approach of using polynomials exclusively, models 6.3.1 and 6.5.1 can be written, respectively,

$$f_{x,t} = g(x) + h(t) + \varepsilon_{x,t} = \sum_{i=0}^{3} a_i x^i + \sum_{i=1}^{4} a_{i+3} t^i + \varepsilon_{x,t} \qquad (6.6.1)$$

$$f_{x,t} = g(x) + h(t) + l(c) + \varepsilon_{x,t} = \sum_{i=0}^{3} a_i x^i + \sum_{i=1}^{4} a_{i+3} t^i + \sum_{i=1}^{9} a_{i+7} c^i + \varepsilon_{x,t} \qquad (6.6.2)$$

Alternatively, the polynomial in age could be replaced by a Perks model,

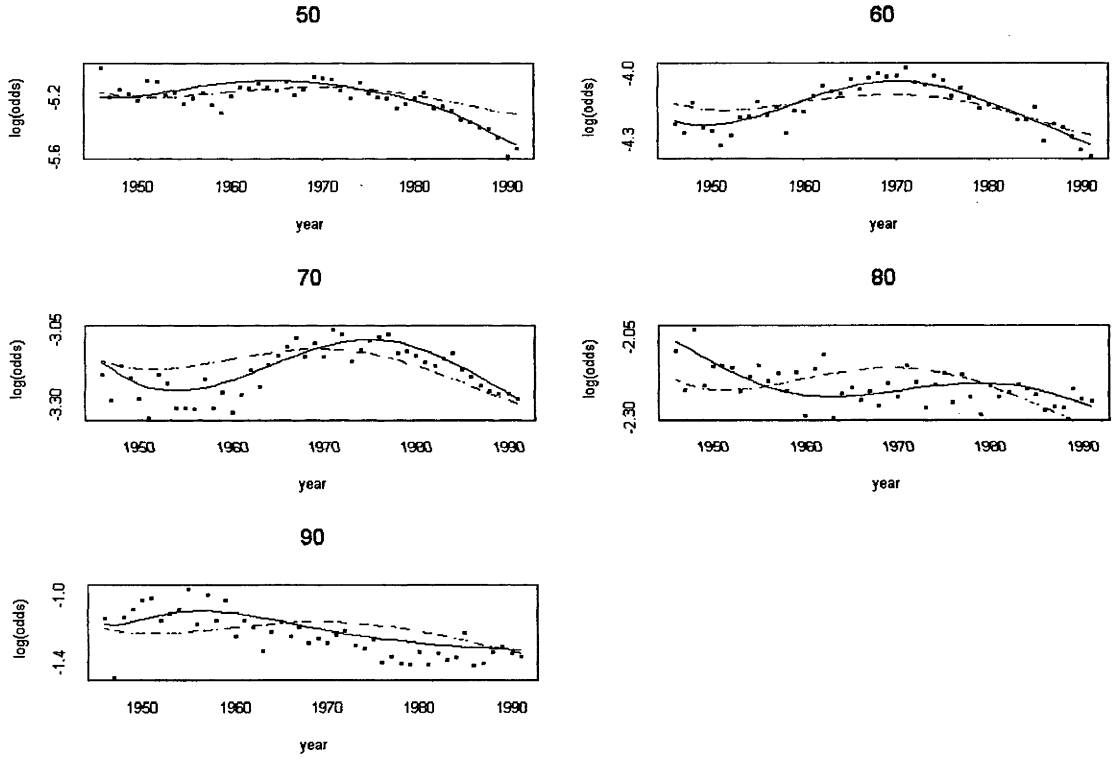
$$\frac{A + B\exp(ux)}{1 + C\exp(ux)}$$

Alternative parameterisations for $h(.)$ and $l(.)$ could involve combinations of logistic and/ or sinusoidal components.

Ultimately, the purpose of reduction of dimensionality could be for comparison with the mortality schedules of other countries, for smoothing in the construction of cohort-life-tables, or for descriptive purposes.

Comparison of the fitted models using these method with those of Chapter 4 shows the superiority of the regression model (4.4.2m) over the additive model for certain sections, and the closer fit of the additive model for others (see Figure 6.7). In particular, the regression model outperforms the fixed effects models for ages 50, 60 and 90, whereas the fixed effects model seems to capture the log(odds) for 70 and 80 year olds more accurately.

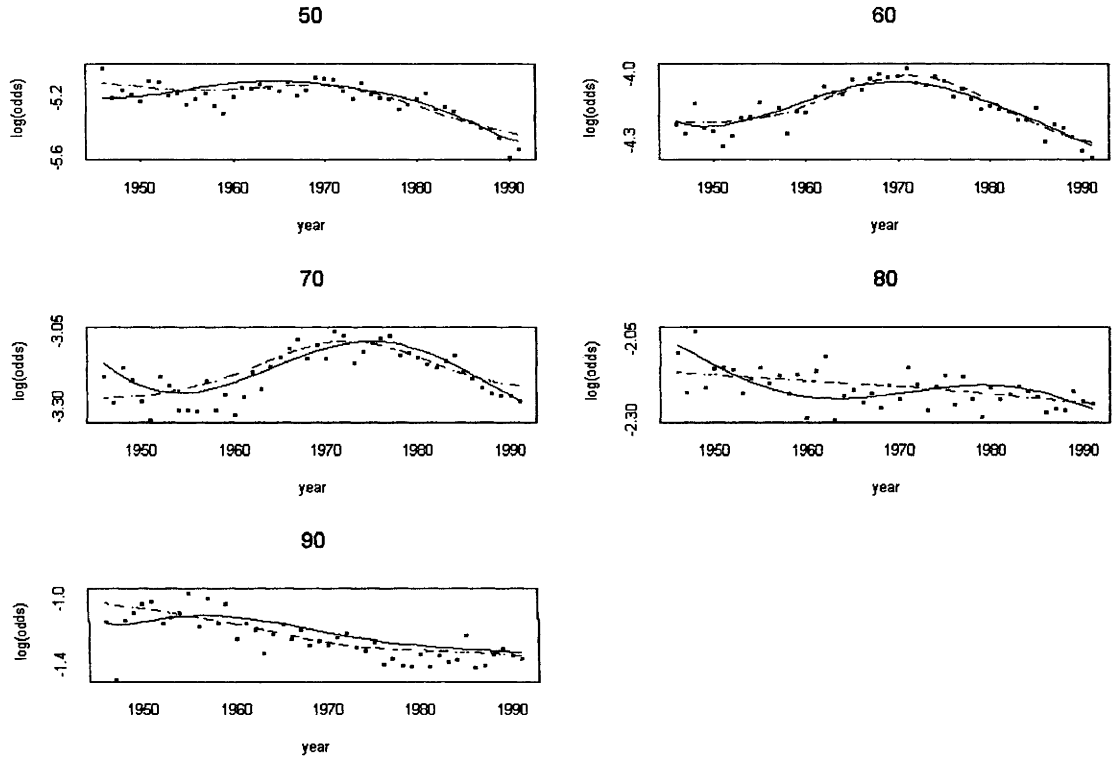
Figure 6.6 **Dutch Males. Log(odds), additive fits with and without θ_c based on smoothed effects model. (Bold= Smoothed additive fit with θ_c (equation 6.6.2); Dotted= Smoothed Additive fit without θ_c (equation 6.6.1))**



The superior fit of the regression model over some of the age range is due to the inclusion of interactions between age and the year and the normal density term. A regression model that includes numerous higher order interactions between age and year performs better than both the models in Figure 6.7, particularly for the oldest ages. As suggested by the additive model, many of the interactions in the male regression models of Chapter 4 would seem to

pertain to cohort effects. Others, however, relate to non-linearities between age and year that have not been explained by inclusion of θ_c.

Figure 6.7 **Dutch Males. Log(odds), additive with θ_c based on smoothed effects and regression model (4.4.2m). (Bold= Smoothed additive fit with θ_c (equation 6.6.2) ; Dotted= 4.4.2m)**



6.7 Modelling using Period Rates

As stated in Section 6.4, the magnitude of the cohort effects found in the analysis above are much stronger than those uncovered by Wilmoth in his examination of French male data. It is hypothesised that the difference between the magnitude of the cohort effects is to a large degree due to the use of cohort rather than period-derived mortality rates in construction of $f_{x,t}$.

The analyses above use rates derived along cohorts, whereas the majority of demographic analysis performed uses rates derived along periods.

A simple illustration of the potential differences between period and cohort rates is provided in Table 6.1 and Table 6.2. Table 6.1 contains the number of survivors at each year (t) and age (x) for a hypothetical population.

Table 6.1 **Number of survivors for a hypothetical population**

	t	t+1	t+2	t+3
x+3			932	913
x+2		960	941	
x+1	990	970		
x	1000			

Along the cohort t-x, the calculated period and cohort mortality rates based on survivor numbers from Table 6.1 are :

Table 6.2 **Approximate mortality rates by period and cohort for cohort t-x**

	period rates	cohort rates
$q_{x,t}$	1%	3%
$q_{x+1,t+1}$	1%	3%
$q_{x+2,t+2}$	1%	3%

The period rates of Table 6.2 are estimated from survivor numbers within an individual year, whereas the cohort rates are estimated from survivor numbers along the cohort. For example, for period rates, $q_{x,t}=\frac{\ell_{x,t}-\ell_{x+1,t}}{\ell_{x,t}}$, whereas for cohort rates $q_{x,t}=\frac{\ell_{x,t}-\ell_{x+1,t+1}}{\ell_{x,t}}$.

Although there is a large increase in mortality along the cohort, the period rates underestimate mortality significantly at each age in this example.

To test this hypothesis, male period mortality rates could be extracted from the doubly-classified Dutch data, where, $f_{x,t} = \log\left(\frac{q_{x,t}}{1-q_{x,t}}\right)$ and $q_{x,t}$ are the initial rates of mortality ***estimated along periods***. The generation of an additive model following the techniques discussed above could be undertaken and, in particular, the residuals could be analysed for possible cohort effects. In the interest of brevity, this was not undertaken here, but is worthy of future work.

6.8 Conclusions

Through the fitting and exploration of the models in this chapter a number of interesting conclusions have emerged:

- there exists a strong cohort effect in Dutch male mortality that explains many of the non-linearities observed in the trend of mortality with time between 1946 and 1991. This is not the case for females, however, who require a complex interaction term independent of cohort to capture the interaction between age and year over time;

- the suspected cohort effects can be explained in the context of debilitation and selection, and changed public health circumstances. A catastrophe can cause a dramatic instantaneous increase in mortality. In the realistic instance where the cohort is heterogeneous, those individuals who survived the event are generally more robust and have a lower mortality. In the long-term, however, the mortality of the survivors may increase as a result of health consequences of the event.

- a conclusion from this analysis is that survivors born between 1885-1895 are relatively more robust than males born in the subsequent cohorts. The Spanish Influenza may have had a long-term debilitating impact on the future mortality of adolescents exposed during 1918, thereby increasing the frailty of these cohorts relative to adjacent cohorts, whereas the same outbreak may have led to a decrease in average frailty for the 1885-1895 cohorts due to removal of the frail members;

- the dimensionality of the fixed effects models can be reduced by using polynomial or non-linear functions to capture the age, year and/or cohort components;

- the fitted models compare closely to the fitted regression models of Chapter 4. Many of the interactions in the male regression model (4.4.2m) can be interpreted as relating to cohort patterns in light of the findings of this chapter, though there are other non-linearities captured by the regression model that have not been linked to cohorts.

The next chapter takes the regression models of Chapter 4, and the models of this chapter, and modifies them so as to enable feasible projections of mortality. Period and cohort life expectancies are generated, as are cohort specific conditional probabilities of survival based on projections from the regression models. In addition, the implications of uncertainty in mortality projections on the future population structure are considered.

7 Projections and Conclusions.

7.1 Introduction

Mortality is forecast using the fitted regression models of Chapter 4. It is shown that simple extrapolations of the regression surfaces from 1946 to 1991 produce questionable forecasts for some ages, and alternative projections are produced based on differing lengths of historical time. Following the method of Heathcote and Higgins (2001b), it is shown how qualitative opinions can be incorporated into projections by altering the parameters of the regression models. Period and cohort life expectancies are generated, as are cohort-specific conditional probabilities of survival from the projected regression models. An important conclusion is that a series of alternative, yet plausible, assumptions lead to dramatically different mortality forecasts. The modified Lee-Carter model introduced in Chapter 6 is projected to illustrate how cohort effects can alter forecasts.

Population projections are undertaken (under the assumption of zero migration) based on a series of plausible mortality projections from the regression models. It is shown that differing projections have a dramatic impact on future elderly population numbers, which has ramifications for policy planning of health care, social security and retirement policy. The results add weight to the need for developing stochastic models of mortality in order to quantify the uncertainty in future projections. Finally, a summary of the main conclusions of the thesis is presented.

7.2 Projecting mortality

7.2.1 Extrapolations from the Regression Models.

As stated in earlier chapters, an advantage of the parsimonious models of Chapter 4 is that they involve a linear term in year. In the simplest instance, estimation of future mortality can be found by extrapolating from the fitted models (see Figure 7.1).

Manton (1993) discusses the restrictions of forecasting by extrapolation. Extrapolation does not account for future interventions such as sudden improvement in medical technology, or the possibility of reduced risk from genetic manipulation.

Prediction intervals for forecasts can be calculated in a number of ways. One way is to use the variance of the residuals and the parameter estimates that are extracted from the MLE fit. This is the approach taken for the projections of the regression model given in this section (see Figure 7.1). Forecasted intervals based on this method would be accurate if we could predict future interventions, be they social, environmental or medical, or if we could predict divergence from the historical rates of change. Since this is not plausible, including prediction intervals in this way may be misleading if those interpreting the forecasts are not aware of the methodological constraints. Another method of generating prediction intervals would be to estimate the variance for a model developed without indicators or non-linear terms, since long-term prediction of extraordinary events and non-linearities is not possible. For example, it could be argued that to determine more realistic intervals, the normal density ordinate that is used to model the mortality hump in post-WWII Dutch males should be omitted. The variance of the model would then be significantly greater than before[6], yet it would provide us with a less misleading measurement of uncertainty.

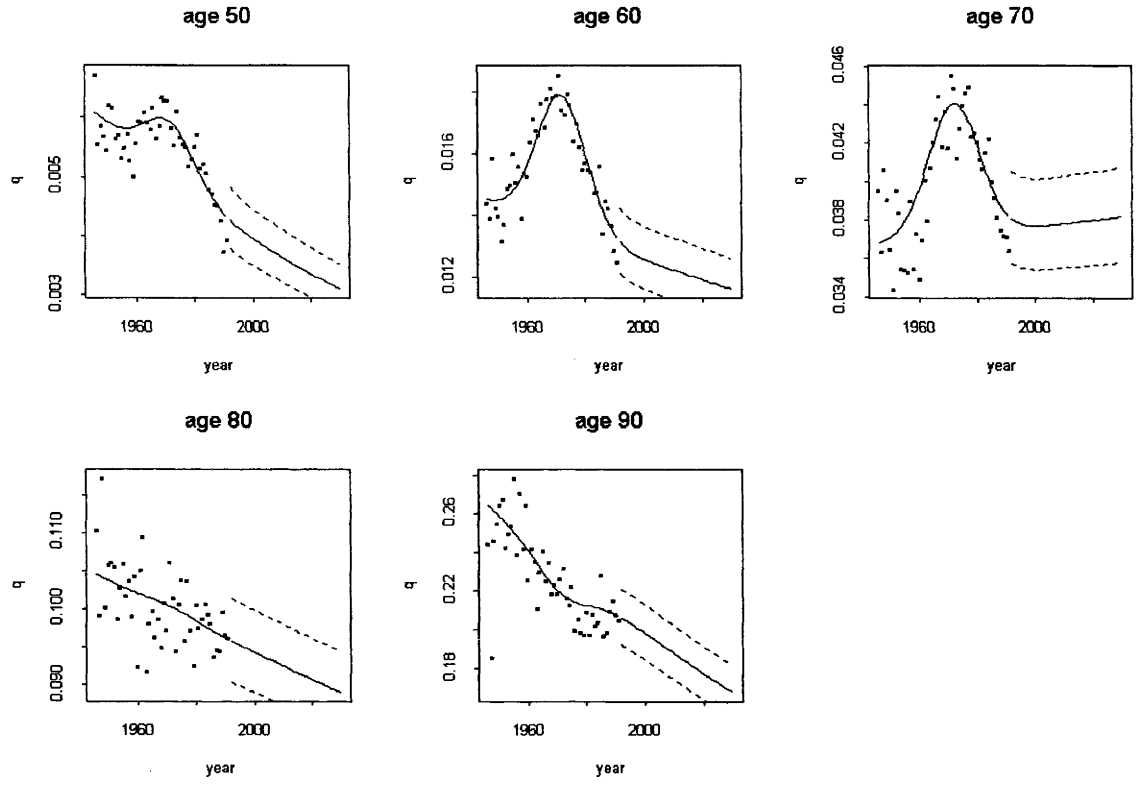
A feature of the regression models is that the boundaries of the fitted surface are a poorer fit to the observed data than the parts of the surface close to the mean of the predictors. Since it is reasonable to suppose that the most recent observed mortality rates are the most accurate indicators of future mortality, this raises an immediate criticism of projecting by extrapolating from fitted regression surfaces.

To illustrate the limitations of extrapolative methods and the prediction intervals from Figure 7.1, similar models to those of Chapter 4 could be fit to a reduced surface, say from 1960 to 1990, and projections could then proceed by extrapolation from the new models.

---

[6] Lee and Carter (1992) include a dummy for the flu epidemic of 1918 in their time-series model of U.S. mortality and note that including the epidemic in error estimation of mortality index would increase the predicted interval in the year 2065 by 57%.

Figure 7.1 **Dutch males. Observed, fitted and extrapolated q_x to 2030, with 95% prediction intervals.**



The results from this approach are presented below with comparisons of simple extrapolations of the fitted models. It is clear from Figure 7.2 that the variability in historic mortality trends leads to extrapolative forecasts that are highly uncertain. The range of variability of forecasts for Dutch females (see Figure 7.3) is lower than that for males due to less variability in the slope of female mortality decline since WWII.

As consequence of the empirical data and the male mortality model (4.4.2m), mortality for male 70 year olds rises in the long-term when the model is extrapolated. In addition, long-term projected male and female mortality rates (see Figure 7.4) and life-expectancies (see Section 7.3) are implausible when compared in light of historical patterns and future expectations. In particular, as shown in Figure 7.4, male and female rates for most ages are projected to diverge according to extrapolations of the regression models, whereas recent empirical data may suggest a narrowing of the mortality differential.

Figure 7.2 **Dutch males. Observed, fitted and extrapolated q_x to 2030, with alternative extrapolations.**

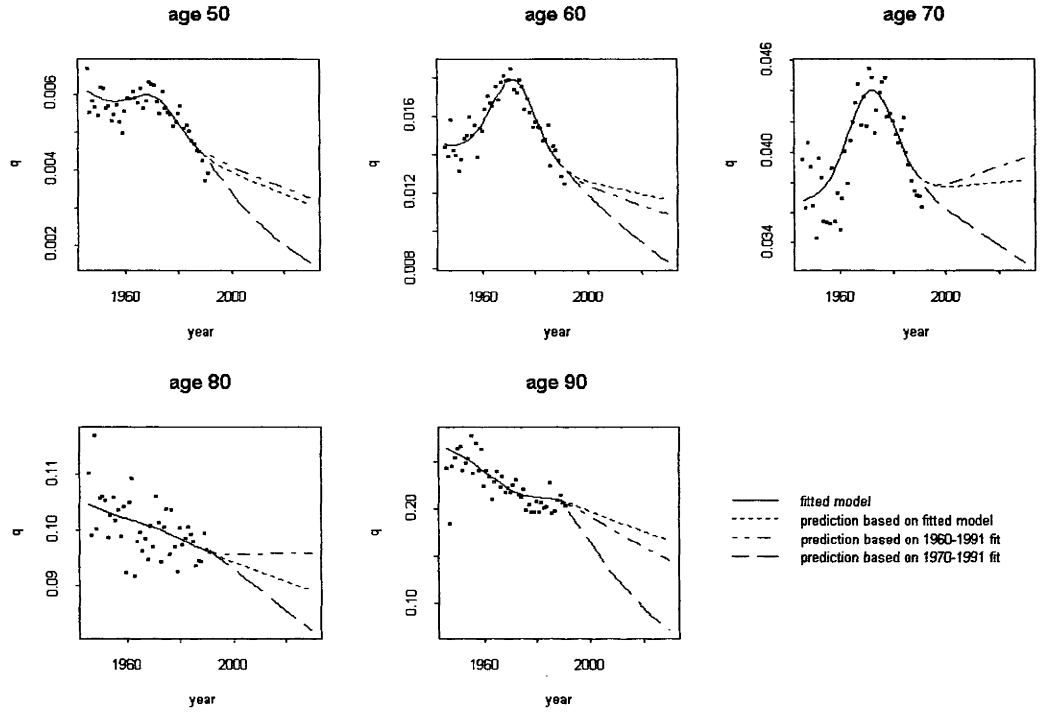


Figure 7.3 **Dutch females. Observed, fitted and extrapolated q_x to 2030, with alternative extrapolations.**

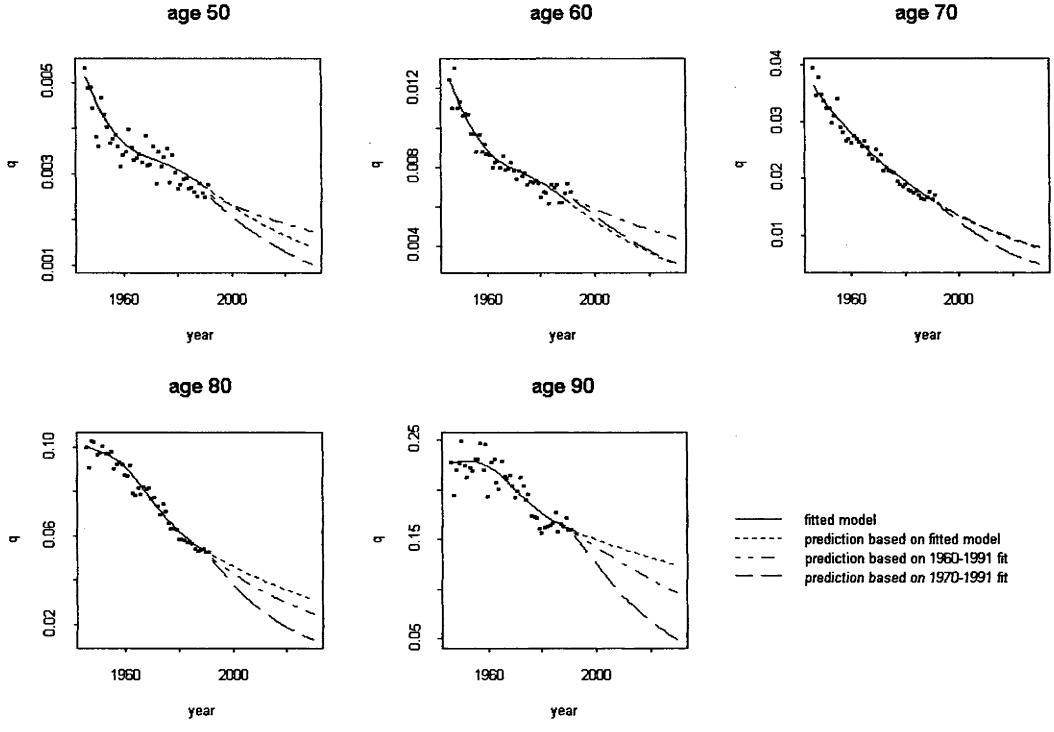
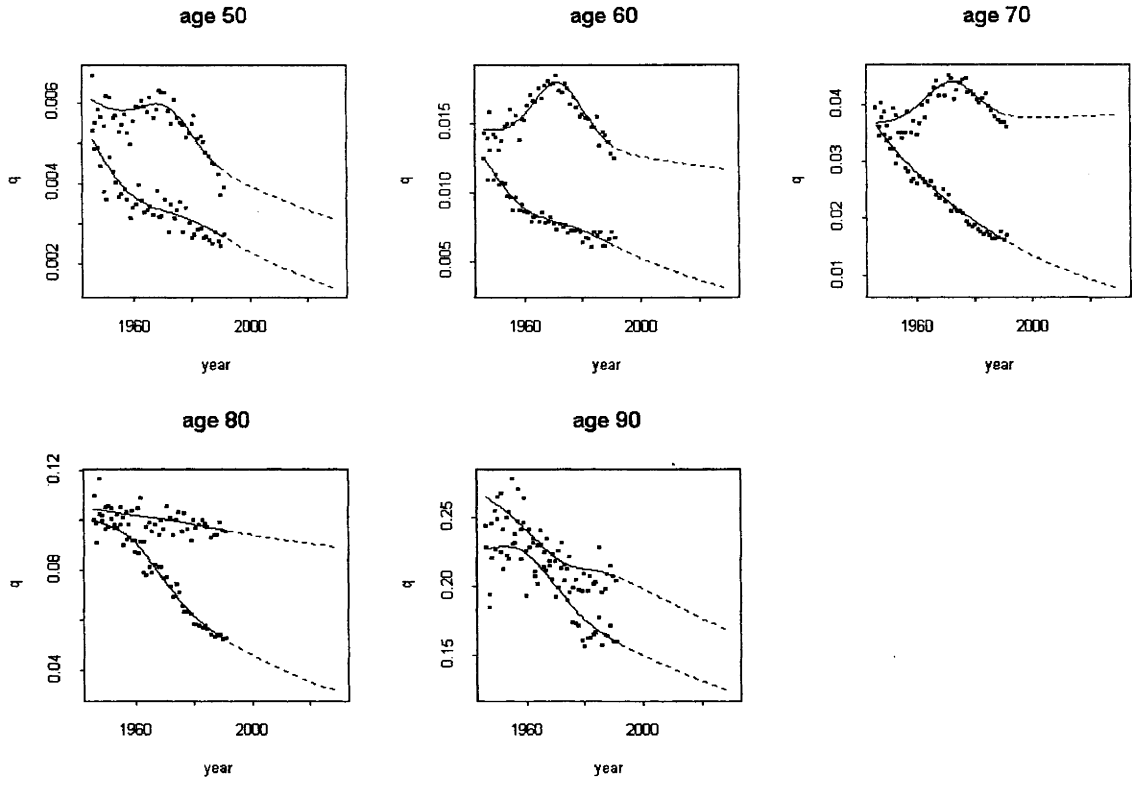


Figure 7.4 **Dutch male and female observed and projected mortality based on extrapolations from (4.4.2m) and (4.4.2f).**



From these observations, it is clear that a purely extrapolative method is fraught with danger. Further support for this conclusion is provided in Section 7.2.2 where age-specific rates of mortality for a range of historical periods are extracted and compared.

Despite the subjectivity inherent in the process, when faced with unknown covariates and a dynamic non-stationary process, plausible forecasts require human judgement. Van Hoorn and De Beer (2001) argue that for long-term forecasts, qualitative arguments about life expectancy levels based on expert opinion should carry more weight than historical extrapolations. They use a mixture of trend extrapolation of age-specific mortality rates and expert judgement, whereby assumptions are made regarding mortality in 2050 from time series analysis, an examination of the literature and expert opinion.

Projections incorporating a numerical 'target' level of mortality for a specified date in the future can be obtained by generating a regression surface using constrained weighted least squares. An alternative means of incorporating expert opinion based on regression surfaces

can be achieved by altering the parameters of the fitted model. This was the approach taken by Heathcote and Higgins (2001b), where an assessment of the future rate of decline in mortality for the two sexes was made in the context of their historical pre and post-WWII mortality trends and the coronary epidemic of the late 20th century. In particular, it is plausible that the pre-WWII parallelism (see Figure 3.13) of male and female mortality was destroyed by the presumed coronary hump, and that the long-term divergent mortality suggested by extrapolation of the fitted surfaces for some ages is implausible. A return to parallelism is supported by Figure 7.4, which shows that for some ages, in particular for ages 50 and 90, the projected rate of decline in mortality for both sexes is almost identical. For the other ages, it is plausible that the divergent projections are the result of distortions caused by the mortality hump. There is an argument, therefore, for a possible realignment of decline in male and female mortality, or even a narrowing of the differential. Van Hoorn and De Beer (2001) argue for a reduced long term sex differential in the Netherlands based on observations and predictions of male and female smoking habits (2001:220).

The approach taken by Heathcote and Higgins (2001b) was to generate a predictive model identical to the descriptive models, with the constraint that the coefficient of year is the average of those for the two sexes. Support for this approach also comes from empirical female data. For females the decrease in mortality has recently slowed for some ages. Since data for the late 1980s and early 1990s is at the edge of the regression surface, the trend has not been adequately captured by the regression model (4.4.2f) (see Figure 4.8 and the residuals by year in Figure 4.15). If this recent trend is a real effect, then there is an argument to modify the female fit for projections in addition to the males.

Heathcote and Higgins (2001b) modified the coefficient of year, but left the year-age interaction unchanged. The regression models (4.4.2m) and (4.4.2f) are more complex in that they involve high-order interactions between age and year, and extrapolations of the unmodified fitted surfaces are more plausible than those in Heathcote and Higgins (2001b). Therefore, rather than taking the average of the coefficient in year, the male coefficient was only slightly reduced, and the female coefficient slightly increased. The intercepts were also changed to ensure continuity with the fitted surfaces. In this example, the coefficient of year was modified by taking:

$$\beta_{\text{mod}} = \left(\frac{6}{7}\right)\beta_t^{males} + \left(\frac{1}{7}\right)\beta_t^{females} \tag{7.2.1}$$

for males, and,

$$\beta_{\text{mod}} = \left(\frac{6}{7}\right)\beta_t^{females} + \left(\frac{1}{7}\right)\beta_t^{males} \tag{7.2.2}$$

for females. The ratios chosen were purely subjective, and were selected in preference to other ratios from visual inspection of the resulting extrapolations. Although not pursued here, a formal approach in selecting a parameter modification could be based on a Bayesian argument whereby a particular prior distribution is assumed for the slope of future mortality decline.

Results are presented in Table 7.1 and projections in terms of q_x are given in Figure 7.5. Because of the higher-order interactions between age and year, modifying the regression model in this way results in changes in the slope of mortality that vary by age. In particular, from Figure 7.5 it can be seen that the modifications have a larger impact on age 90 than on other ages.

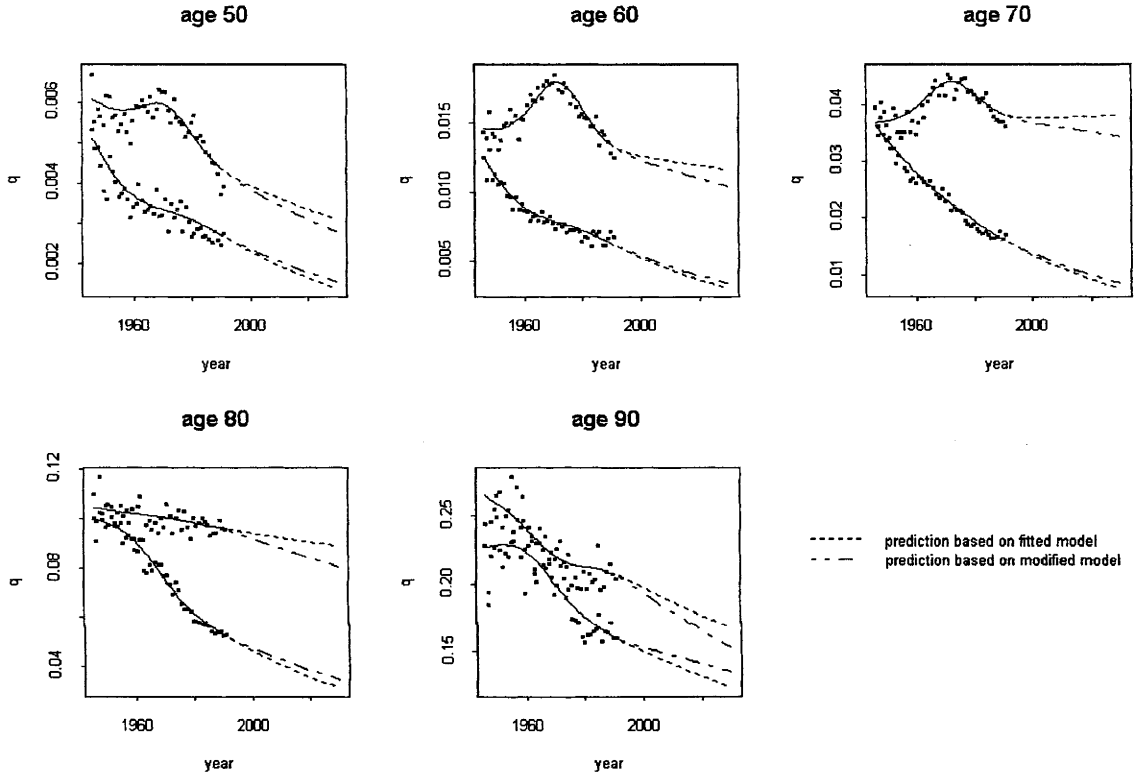
Table 7.1 **Modified parameters for models (4.4.2m) and (4.4.2f).**

Modified Parameters	Value
Intercept (males)	-669.7
Intercept (females)	-1412
coef for t (males)	0.334
coef for t (females)	-0.729

It is worth reiterating that the modifications to the coefficients, and in particular, the magnitude of the modifications are based on achieving changes to the projected mortality slopes that appear plausible and consistent with approximate parallelism between the sexes. Alternative modifications could include varying the interaction terms between age and year. However, complications may arise when trying to achieve continuity.

Extrapolations from both the original models from Chapter 4 and the modified models are used in construction of unconditional probabilities of survival and period and cohort life expectancies in Section 7.3.

Figure 7.5 **Dutch male and female observed and projected mortality to 2030 based on extrapolations from (4.4.2m) and (4.4.2f) and extrapolations from models with modified parameters.**



7.2.2 Projections from Age-Specific Rates of Improvement

As stated in Chapter 1, a standard method of projecting mortality is to determine age-specific rates of mortality change from historical trends and extrapolate these into the future. As with all methods reliant on extrapolation, the forecasts will depend on the length of the historical period from which the rates of change have been deduced.

Annual rates of change in q_x were generated from the fitted male and female regression models (4.4.1m, 4.4.2m, 4.4.1f and 4.4.2f) for five different periods: the five sets were

based on the change in mortality between 1991 and 1890, 1946, 1960, 1970 and 1980. Since the regression model performs most poorly at the edges of the surface, and particularly for 1991, the rates of improvement estimated from the regression models are not consistent with estimates that would arise from, say, a spline-smoothed surface. The main purpose of this exercise, however, is to provide insight into the inherent uncertainty in extrapolating mortality from historic data, rather than to produce accurate forecasts.

Table 7.2 and Table 7.3 contain the improvement factors for a selection of ages. As an illustration of the meaning of the factors, the first cell in Table 7.2 means that there is, on average, a 1.26% decrease in mortality per annum for 50 year old men over the 101 years between 1890 and 1991.

Table 7.2 **Dutch males. Percentage mortality improvement factors for a selection of ages and years.**

Years / Age	1890-1991	1946-1991	1960-1991	1970-1991	1980-1991
50	-1.26	-0.76	-0.97	-1.53	-1.71
60	-0.77	-0.20	-0.54	-1.41	-1.69
70	-0.51	+0.08	-0.12	-0.66	-0.83
80	-0.40	-0.19	-0.20	-0.22	-0.23
90	-0.37	-0.54	-0.48	-0.30	-0.24

Table 7.3 **Dutch females. Percentage mortality improvement factors for a selection of ages and years.**

Years / Age	1890-1991	1946-1991	1960-1991	1970-1991	1980-1991
50	-1.43	-1.43	-1.03	-1.06	-1.32
60	-1.30	-1.56	-1.17	-1.19	-1.45
70	-1.25	-1.82	-1.81	-1.81	-1.83
80	-0.94	-1.45	-1.79	-1.78	-1.57
90	-0.52	-0.78	-1.08	-1.08	-0.89

It is apparent that the average annual rates of change of mortality vary considerably depending on the period used for derivation. Mortality projections based on these rates of change closely resemble those of Figure 7.2 and Figure 7.3, and so have not been included here.

7.2.3 Projections from the Modified Lee-Carter Model.

Rather than projecting from the fixed effects models of Chapter 6 for comparison with the regression models, this section is included to illustrate the impact that cohort effects may have on mortality projections.

The steep declines in mortality for males aged 90 in the late 1970s is related to the fall in the mortality hump, and possibly due to the relatively favourable mortality experience of the 1885-1890 cohorts as discussed in Chapter 6. The subsequent flattening and even slight rise of mortality in the late 1980s may reflect the poorer mortality of the 1900-1905 cohorts that is suggested from the analysis undertaken in Chapter 6. In light of the cohort variation suggested from the previous chapter, it is conceivable that, for the 1900-1905 cohorts, the mortality in the late 1990s and the first decade of the 21st century will continue to reflect poorer than average mortality.

Projections of Dutch male mortality were generated using the modified Lee-Carter model (6.5.2) from Chapter 6, where a cohort effect denoted by θ_c is incorporated into the model (following Wilmoth (1990)) after decomposing the residuals from the standard Lee-Carter fit. In order to achieve continuity between the last data and the forecasts, α_x is taken to be the log of the odds of death for the year 1991.

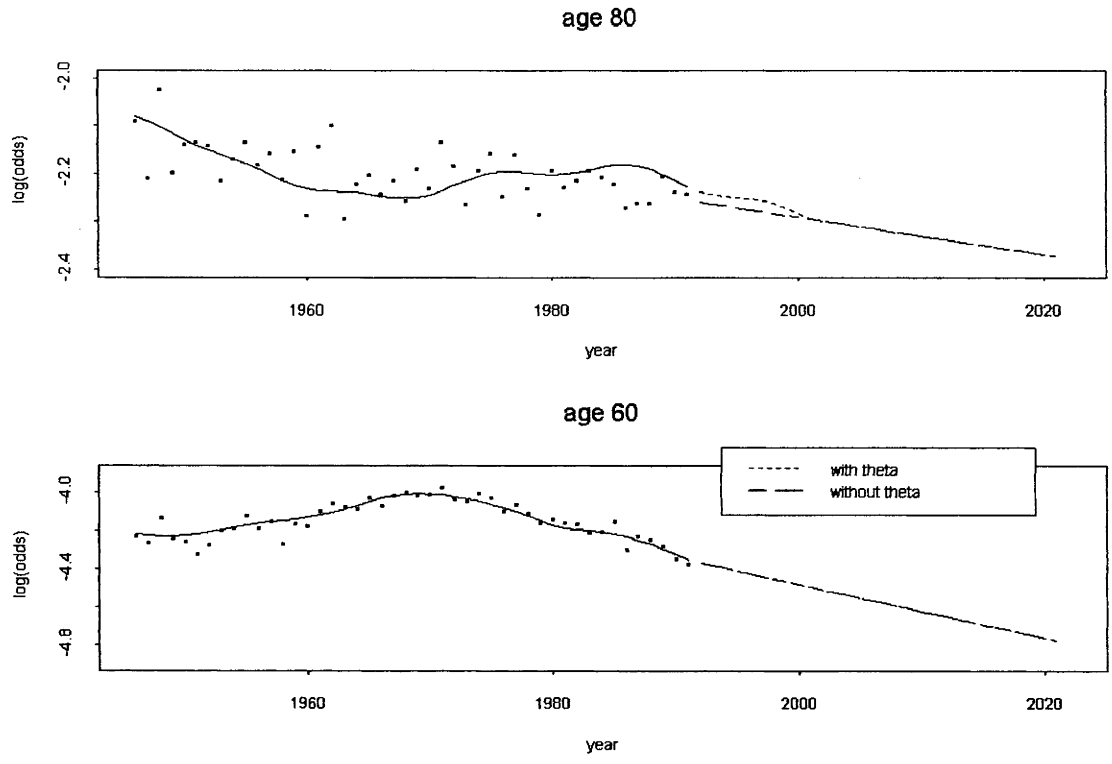
Projections of the mortality surface involve making judgements regarding the shape of the parameters in the future. If we ignore expert judgement in this example, a continuation of the pattern with respect to age is assumed (ie. α_x and β_x are left unchanged). Future levels of κ_t have the largest impact on mortality projections. Following the approach of Lee and Carter, time-series techniques were employed on κ_t to determine an appropriate fit, and an

ARIMA(0,1,0) model (autoregressive integrated moving average) was found most appropriate for the range 1946 to 1991 as well as most shorter ranges such as 1970 to 1991. As with all models that rely on historical trends, judgement is required to determine the range of the historic κ_t to use in projections. Since the time-series models chosen in this instance are linear, the projections are the same as those generated from simple age-specific cross-sectional projections, and are very similar to those of the regression models. In particular, as a consequence of extrapolation, the variability in the time-series projections is comparable with the variability in projections based on the regression approach when different historical periods are used in formulating the fit.

The implication of including θ_c in projections is illustrated in Figure 7.6 where the observed, fitted and projected log(odds) for males are given for ages 60 and 80. The projections are based on a κ_t derived from historical data from 1970 to 1991. For age 80, inclusion of θ_c increases the log(odds) between 1991 and 2000, corresponding to the 1910 to 1920 birth cohorts. As discussed in Chapter 6, the increased log(odds) for these cohorts may be the result of long-term debilitation brought on by the Spanish Influenza epidemic of 1918. A similar bump is apparent between 1990 and 2010 in the projections for age 90. For cohorts following the 1920 birth cohort, θ_c is assumed to be zero and, therefore, does not influence projections for ages 70 and below.

It is plausible that the regression models developed in Chapter 4 could be modified in a similar way to the models of Chapter 6 in order to isolate cohort effects. Any potential cohort effects could then be added to predictions as was done in the modified Lee-Carter model above. It is, however, questionable whether inclusion of a cohort effect adds useful information to projections; the shift in predicted curve due to the cohort effect is likely to be overwhelmed by uncertainty associated with determination of the legitimacy of the cohort effect and the variance of the fixed effects.

Figure 7.6 **Dutch males. Observed, fitted and projected log(odds) based on the modified Lee-Carter model**



7.3 Expectation of Life

A common misconception is that period life tables provide an estimate of expected future lifetime. Period life tables are calculated with cross-sectional data; they only take into account the mortality pattern for the year in question, and do not consider any possible changes in future mortality. Cohort life expectancies are calculated from mortality levels along each incomplete cohort, and so may incorporate the possibility of future changes. Forming cohort life expectancies, therefore, involves projecting mortality.

Male and female period life-expectancies are generated based on models (4.4.1m), (4.4.2m), (4.4.1f) and (4.4.2f) for ages 50, 60, 70, 80 and 90, for years ranging from 1890 to 2030, and additional life expectancies are generated from the modified regression models based on equations (7.2.1) and (7.2.2). Section 2.7.3 of Chapter 2 contains the theory underlying the life expectancy calculations including standard errors.

Period life expectancies are generated using mortality rates up to age 95 for most of the surface. Since the fitted models only cover ages 50 to 90, they are extrapolated from in order to generate rates for ages 91 to 95. Therefore, the precision of life expectancies for the very oldest ages is questionable and dependent on the accuracy of our model for ages above 90. For completion, period life expectancies based on observed rates are included in the figures that follow.

Table 7.4 **Male PERIOD life expectancies (standard errors in brackets). (1) is projection based on fitted model (4.4.2m). (2) is projection based on modified model.**

Year / Age	1890	1910	1930	1950	1970	1990	2010 (1)	2010 (2)	2030 (1)	2030 (2)
50	20.41 (0.12)	22.18 (0.10)	23.86 (0.08)	25.76 (0.08)	24.85 (0.06)	26.39 (0.05)	*26.92 (0.05)*	*27.38 (0.05)*	*27.27 (0.05)*	*28.24 (0.05)*
60	13.84 (0.12)	14.94 (0.09)	16.07 (0.08)	17.72 (0.07)	16.98 (0.06)	18.02 (0.05)	*18.36 (0.05)*	*18.78 (0.05)*	*18.58 (0.05)*	*19.47 (0.05)*
70	8.44 (0.11)	9.01 (0.09)	9.68 (0.08)	10.86 (0.08)	10.74 (0.05)	11.21 (0.05)	*11.47 (0.05)*	*11.81 (0.05)*	*11.70 (0.05)*	*12.43 (0.05)*
80	4.57 (0.10)	4.78 (0.08)	5.21 (0.08)	5.74 (0.08)	6.20 (0.05)	6.42 (0.04)	*6.77 (0.04)*	*7.00 (0.04)*	*7.17 (0.05)*	*7.68 (0.05)*
90	NA NA	NA NA	2.65 (0.09)	2.73 (0.08)	3.25 (0.05)	3.44 (0.04)	*3.79 (0.05)*	*3.92 (0.05)*	*4.15 (0.06)*	*4.44 (0.06)*

Note: Life expectancies for age 90 prior to 1930 were not calculated due to lack of empirical data.

Table 7.5 **Female PERIOD life expectancies (standard errors in brackets). (1) is projection based on fitted model (4.4.2f). (2) is projection based on modified model.**

Year / Age	1890	1910	1930	1950	1970	1990	2010 (1)	2010 (2)	2030 (1)	2030 (2)
50	22.03 (0.17)	23.19 (0.10)	24.40 (0.09)	27.03 (0.08)	29.67 (0.06)	32.32 (0.06)	*34.67 (0.06)*	*34.26 (0.06)*	*36.74 (0.06)*	*35.94 (0.06)*
60	14.86 (0.16)	15.68 (0.10)	16.58 (0.08)	18.55 (0.08)	20.90 (0.06)	23.44 (0.06)	*25.53 (0.05)*	*24.16 (0.05)*	*27.39 (0.06)*	*26.65 (0.06)*
70	8.96 (0.16)	9.44 (0.09)	10.05 (0.09)	11.28 (0.08)	13.06 (0.05)	15.27 (0.05)	*16.92 (0.05)*	*16.59 (0.05)*	*18.41 (0.06)*	*17.76 (0.06)*
80	4.81 (0.14)	5.00 (0.08)	5.48 (0.09)	6.08 (0.09)	7.06 (0.06)	8.50 (0.05)	*9.50 (0.05)*	*9.25 (0.05)*	*10.46 (0.06)*	*9.93 (0.06)*
90	NA NA	NA NA	2.78 (0.09)	3.02 (0.09)	3.58 (0.06)	4.17 (0.04)	*4.70 (0.06)*	*4.55 (0.06)*	*5.17 (0.08)*	*4.86 (0.08)*

Male expectation of life increases only slightly for all ages between 1990 and 2030 (by approximately half of one year) according to extrapolations from model (4.4.2m). In contrast, the dramatic drop in log(odds) following 1945 for females results in a steady increase (by between one and four years depending on age) in expectation of life between 1946 and 2030 when model (4.4.2f) is extrapolated. Under the modified models, male life expectancies increase more rapidly, and females more slowly than under the extrapolated fitted surfaces.

Figure 7.7 **Dutch males. Complete PERIOD expectation of life. Observed, fitted and projected from models (4.4.1m) and (4.4.2m), and projected from modified model to 2030.**

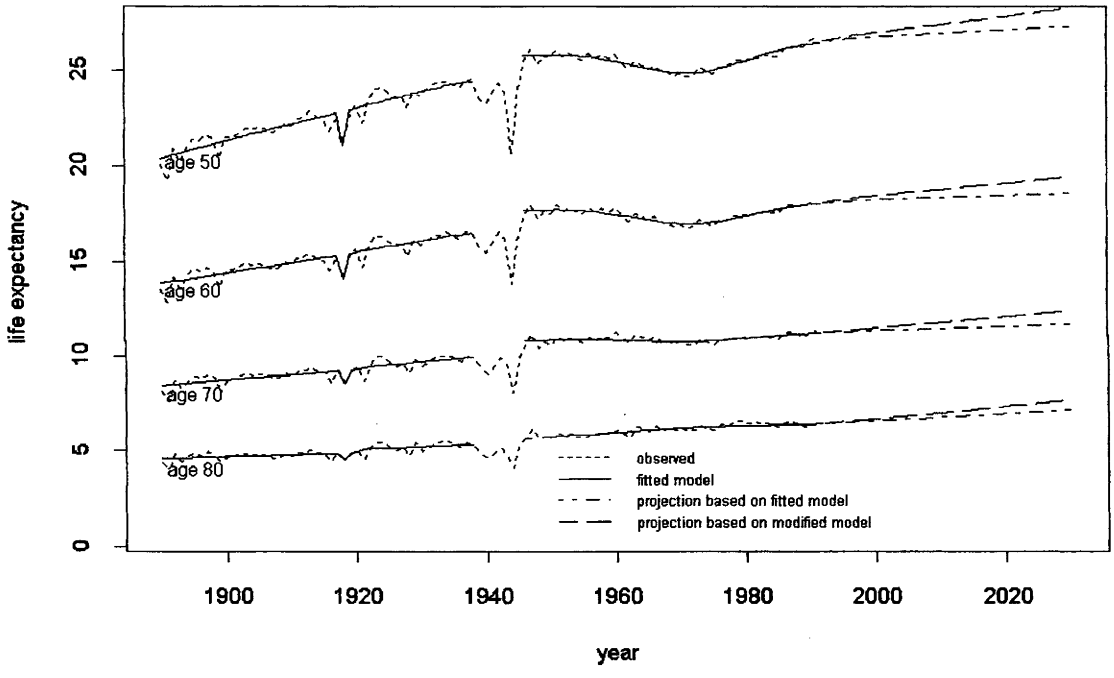


Figure 7.8 **Dutch females. Complete PERIOD expectation of life. Observed, fitted and projected from models (4.4.1f) and (4.4.2f), and projected from modified model.**

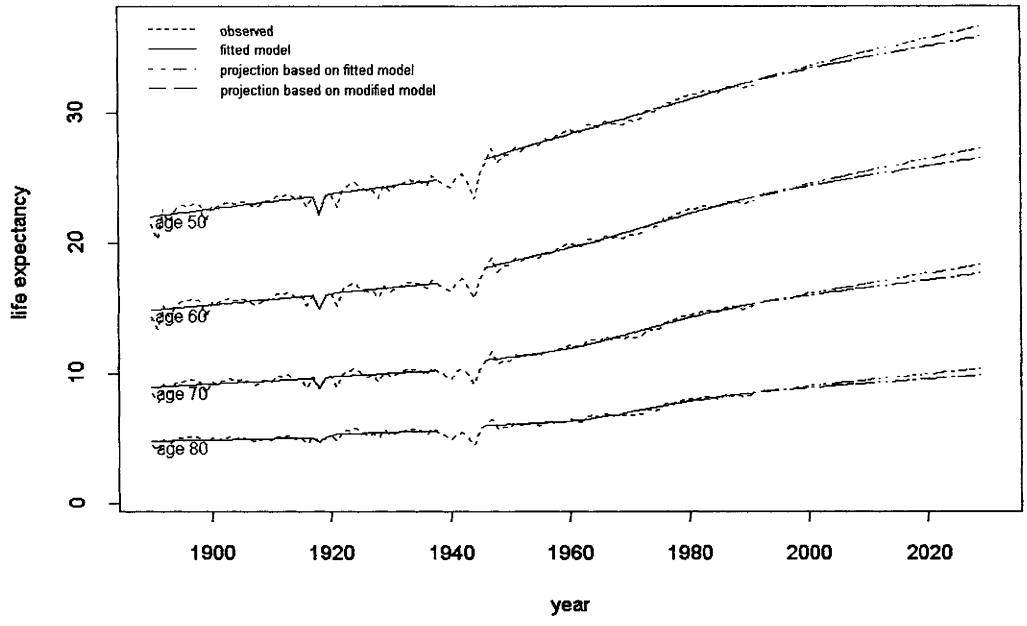
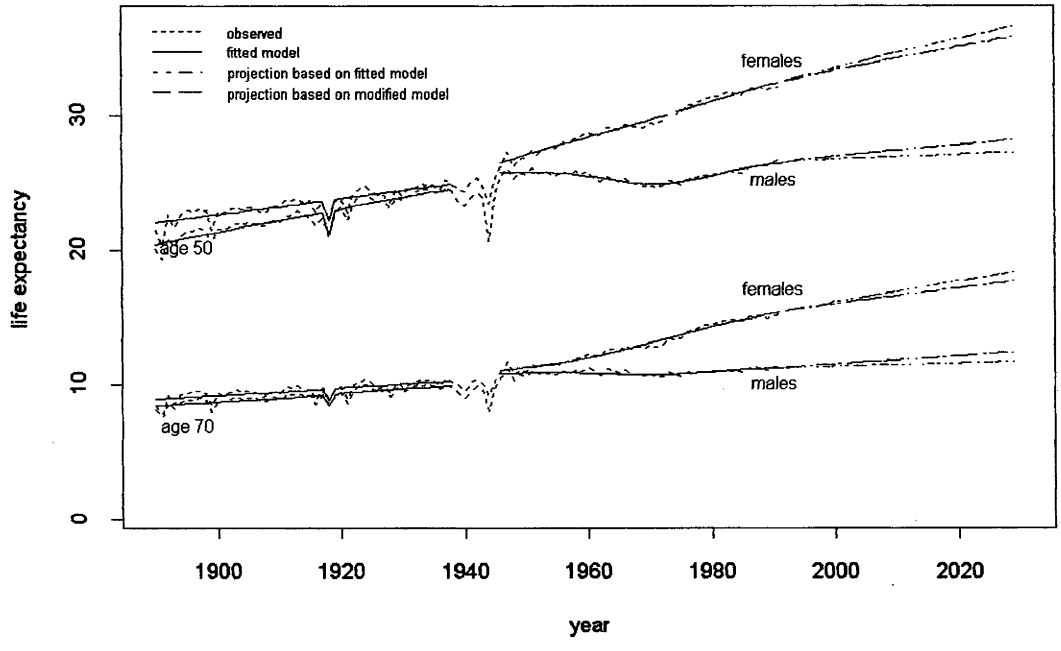


Figure 7.9 **Dutch males and females. Complete PERIOD expectation of life. Observed, fitted and projected from fitted models, and projected from modified models.**



Projections based on cohort life expectancies are calculated and the results are presented in Table 7.6, Table 7.7, Figure 7.10 and Figure 7.11. The male cohort life-expectancies are similar to the corresponding period life-expectancies since male mortality extrapolated from the fitted model (4.4.2m) is fairly unchanging. The female cohort life-expectancies are much higher than the corresponding period life-expectancies due to extrapolation of the steep decline in female mortality captured by the fitted model (4.4.2f).

Table 7.6 **Dutch males. COHORT life expectancies (standard errors in brackets). (1) is projection based on fitted model (4.4.2m). (2) is projection based on modified model.**

Year / Age	1890	1910	1930	1950	1970	1990 (1)	1990 (2)	2010 (1)	2010 (2)	2030 (1)	2030 (2)
50	21.44 (0.11)	23.08 (0.09)	25.13 (0.07)	25.33 (0.06)	25.82 (0.05)	*26.89 (0.05)*	*27.39 (0.05)*	*27.29 (0.06)*	*28.34 (0.06)*	*27.67 (0.06)*	*29.29 (0.06)*
60		15.36 (0.09)	16.51 (0.06)	17.70 (0.06)	17.31 (0.05)	*18.34 (0.05)*	*18.62 (0.05)*	*18.61 (0.05)*	*19.37 (0.05)*	*18.85 (0.06)*	*20.13 (0.06)*
70			9.67 (0.06)	11.03 (0.07)	10.85 (0.05)	*11.40 (0.05)*	*11.53 (0.05)*	*11.66 (0.05)*	*12.17 (0.05)*	*11.91 (0.05)*	*12.83 (0.05)*
80				5.87 (0.07)	6.28 (0.05)	*6.52 (0.04)*	*6.56 (0.04)*	*6.89 (0.05)*	*7.20 (0.05)*	*7.30 (0.05)*	*7.90 (0.05)*
90					3.28 (0.05)	*3.50 (0.05)*	*3.51 (0.05)*	*3.83 (0.05)*	*3.98 (0.05)*	*4.20 (0.06)*	*4.50 (0.06)*

Table 7.7 **Dutch females. COHORT life expectancies (standard errors in brackets). (1) is projection based on fitted model (4.4.2f). (2) is projection based on modified model.**

Year / Age	1890	1910	1930	1950	1970	1990 (1)	1990 (2)	2010 (1)	2010 (2)	2030 (1)	2030 (2)
50	22.85 (0.11)	23.94 (0.09)	26.14 (0.07)	30.18 (0.06)	32.68 (0.05)	*35.02 (0.05)*	*34.46 (0.05)*	*37.22 (0.06)*	*36.23 (0.06)*	*39.15 (0.06)*	*37.79 (0.06)*
60		16.08 (0.09)	17.10 (0.06)	20.20 (0.06)	22.93 (0.05)	*25.10 (0.05)*	*24.74 (0.05)*	*27.15 (0.05)*	*26.37 (0.05)*	*28.97 (0.06)*	*27.83 (0.06)*
70			10.07 (0.06)	11.77 (0.07)	14.17 (0.05)	*16.08 (0.05)*	*15.90 (0.05)*	*17.73 (0.05)*	*17.17 (0.05)*	*19.25 (0.05)*	*18.35 (0.05)*
80				6.17 (0.07)	7.46 (0.05)	*8.80 (0.04)*	*8.73 (0.04)*	*9.79 (0.05)*	*9.44 (0.05)*	*10.77 (0.05)*	*10.13 (0.05)*
90					3.70 (0.05)	*4.31 (0.05)*	*4.29 (0.05)*	*4.77 (0.05)*	*4.60 (0.05)*	*5.25 (0.06)*	*4.91 (0.06)*

The cohort life expectancies for some ages depend on projections into the distant future and should be treated with caution. For example, cohort life expectancies for 50 year olds in 2030 require mortality projections to the year 2075.

Figure 7.10 **Dutch males. Complete COHORT expectation of life. Fitted and projected from model (4.4.2m), and projected from modified model to 2030.**

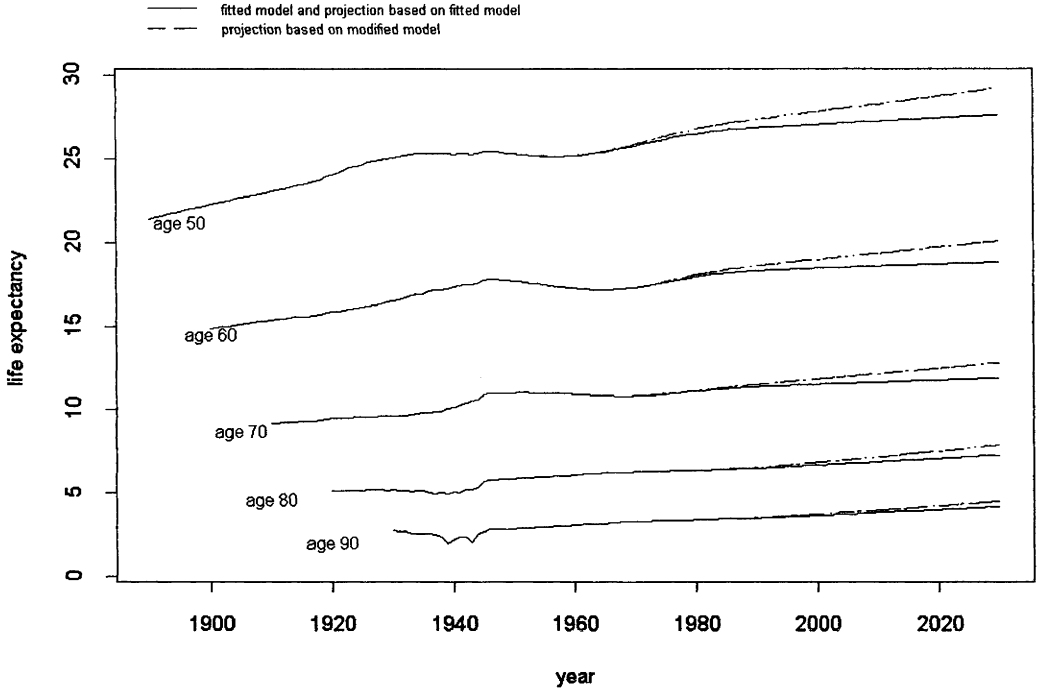
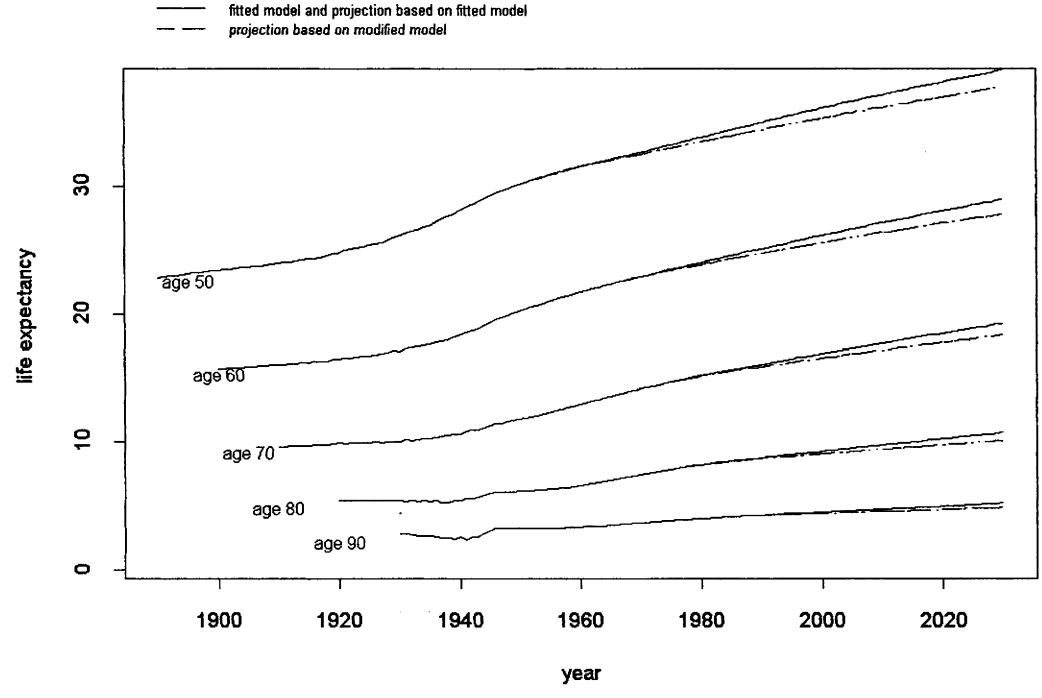
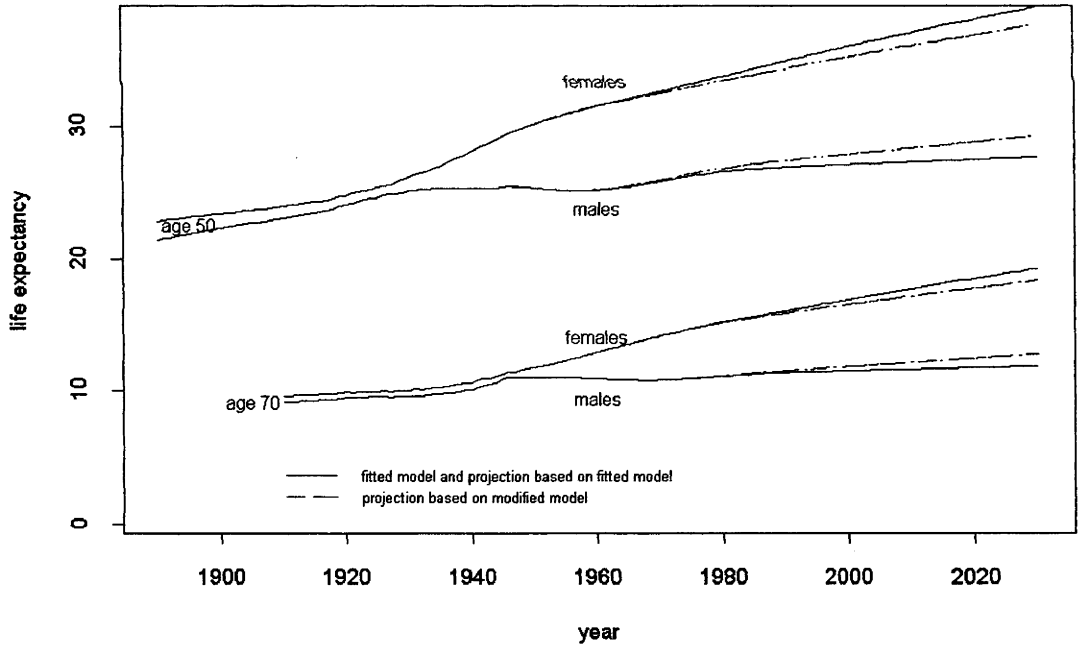


Figure 7.11 **Dutch females. Complete COHORT expectation of life. Fitted and projected from model (4.4.2f), and projected from modified model to 2030.**



As discussed in Chapters 1 and 6, there are two schools of thought on the long-term effects of catastrophic events on a cohort. One theory is that the event will not only result in the immediate death of those experiencing it, it will also adversely affect the cohort's future mortality. A conflicting argument is that of selection, whereby the frail members of a cohort die, leaving behind a relatively healthy, more robust population with a lower average mortality rate and an increased expectation of life. Selection mortality may play a significant role amongst Dutch men and women, if our cohort expectation of life plots have been interpreted correctly. Figure 7.10 and Figure 7.11 show a jump in life-expectancy coinciding with the Hunger Winter of 1944/45. This may be explained in terms of selection in that survivors of the Hunger Winter would be the more robust of the cohort, and so would have a higher chance of survival and thus a higher expectation of life. A selection effect is supported further by the low level of mortality in 1947 identified in Chapter 3.

Figure 7.12 **Dutch males and females. Complete COHORT expectations of life. Fitted and projected from fitted models, and projected from modified models to 2030.**



As seen in Figure 7.9 and Figure 7.12, for period and cohort life expectancies the sex differential changes substantially when projections are based on the modified models.

The life expectancy trends have been presented here primarily to illustrate the capabilities of the regression model in forecasting. For analysis and decomposition of Dutch life expectancy and mortality patterns see Nusselder and Mackenbach (1997, 2000).

7.4 Conditional Probabilities of Survival

Cohort-specific conditional probabilities of death are presented for males and females based on extrapolations of the fitted regression models and the modified models. The theory behind estimation of the probabilities and their standard errors is given in Section 2.7.2. Table 7.8 and Table 7.9 contain the probabilities of surviving to ages 60,70,80 and 90, given survival to age 50 for a range of years.

The probabilities are calculated along cohorts rather than periods. For example, referring to Table 7.8, the probability that a 50 year old man alive in 1890 would live to at least the age of 90 in 1930 is 5%. In contrast, assuming that mortality rates will change in the future according to extrapolations from model (4.4.2m), a 50 year old alive in 2030 would have a 17.9% chance of being alive in 2070. For the oldest ages in particular, the female probabilities of survival are extremely high, reflecting their favourable mortality and the projected improvements from the regression models. The implication to population projections is briefly discussed in Section 7.5. The same warning that followed the cohort life expectancies in Section 7.3 applies; namely, that certain conditional probabilities of survival rely on very long term projections and are less reliable than other estimates.

Table 7.8 **Dutch males. Conditional COHORT-specific probabilities of survival from age 50 to age y (standard errors in brackets). (1) is projection based on fitted model (4.4.2m). (2) is projection based on modified model. (values in %).**

Year y	1890	1910	1930	1950	1970	1990 (1)	 (2)	2010 (1)	 (2)	2030 (1)	 (2)
60	85.2 (0.3)	89.0 (0.3)	91.8 (0.2)	92.9 (0.1)	92.4 (0.1)	*94.6* *(0.1)*	*94.6* *(0.1)*	*95.2* *(0.1)*	*95.5* *(0.1)*	*95.8* *(0.1)*	*96.2* *(0.1)*
70	61.8 (0.5)	68.2 (0.4)	72.9 (0.2)	74.8 (0.3)	75.3 (0.2)	*79.2* *(0.2)*	*79.7* *(0.2)*	*80.1* *(0.2)*	*81.5* *(0.2)*	*80.9* *(0.2)*	*83.1* *(0.2)*
80	29.7 (0.6)	35.8 (0.5)	43.5 (0.4)	42.2 (0.3)	44.9 (0.3)	*47.5* *(0.3)*	*49.3* *(0.3)*	*48.1* *(0.3)*	*51.7* *(0.3)*	*48.5* *(0.3)*	*54.0* *(0.3)*
90	5.0 (0.2)	6.1 (0.2)	10.9 (0.3)	11.4 (0.2)	13.1 (0.2)	*15.1* (0.2)	*17.1* (0.2)	*16.5* (0.2)	*20.3* (0.2)	*17.9* (0.3)	*23.6* *(0.3)*

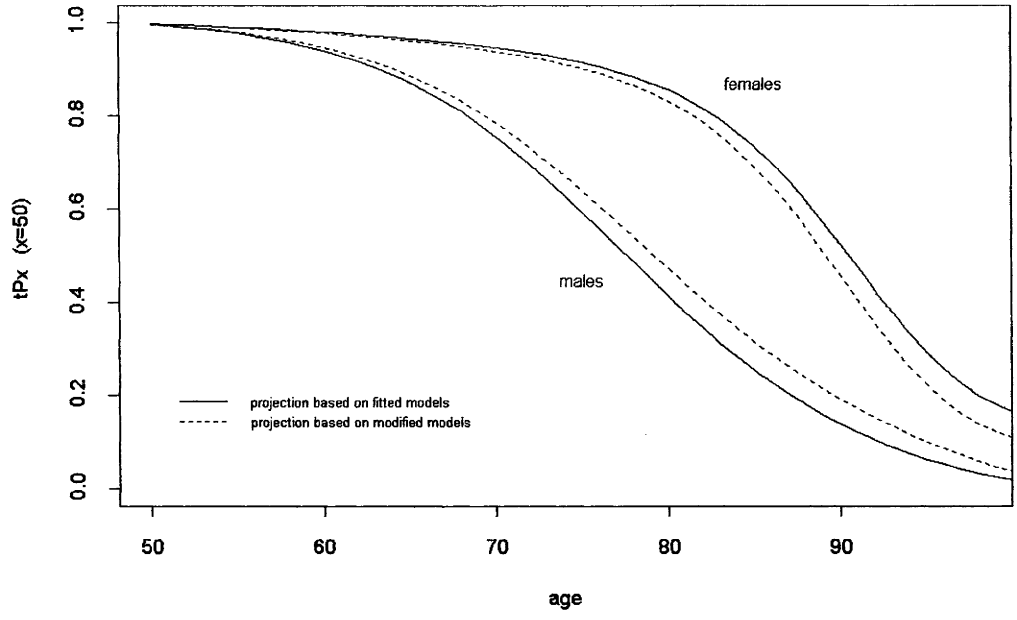
Table 7.9 **Dutch females. Conditional COHORT-specific probabilities of survival from age 50 to age y (standard errors in brackets). (1) is projection based on fitted model (4.4.2f). (2) is projection based on modified model. (values in %).**

Year y	1890	1910	1930	1950	1970	1990 (1)	 (2)	2010 (1)	 (2)	2030 (1)	 (2)
60	88.4 (0.4)	90.2 (0.2)	91.9 (0.2)	95.2 (0.1)	96.2 (0.1)	*97.0* *(0.1)*	*96.9* *(0.1)*	*97.8* *(0.1)*	*97.7* *(0.1)*	*98.4* *(0.1)*	*98.2* *(0.1)*
70	67.2 (0.5)	71.0 (0.4)	75.5 (0.2)	84.7 (0.2)	87.9 (0.2)	*91.0* *(0.1)*	*90.8* *(0.1)*	*93.6* *(0.1)*	*93.1* *(0.1)*	*95.5* *(0.1)*	*94.9* *(0.1)*
80	34.3 (0.6)	38.7 (0.6)	47.6 (0.4)	61.8 (0.3)	70.7 (0.3)	*78.0* *(0.2)*	*77.1* *(0.2)*	*84.0* *(0.2)*	*82.4* *(0.2)*	*88.4* *(0.1)*	*86.6* *(0.1)*
90	6.5 (0.3)	7.7 (0.3)	13.8 (0.3)	26.1 (0.3)	35.9 (0.4)	*45.1* (0.4)	*42.3* (0.5)	*53.9* (0.5)	*49.4* (0.5)	*61.8* (0.5)	*55.9* *(0.5)*

Figure 7.13 is a survival curve for the 1980 cohort (corresponding to age 50 in 2030). It shows the projected conditional probabilities of survival for males and females based on extrapolations from the fitted regression models and the modified models. Although the curves for the different models are similar, the areas under these curves can be calculated to form life expectancies, so differences accumulate in life expectancy. Probabilities for ages above 90 were formulated from extrapolations of the regression models, and are highly uncertain.

The debate on whether or not a biological limit exists to human life expectancy is well-known (see for example, Manton (1993); Warner (1993)) and is discussed in Chapter 1. Rectangularization of survival curves is often used as evidence that a limit is being approached. Rectangularization of survival curves is associated with decreasing variability in the distribution of ages at death. For interest, Wilmoth and Horiuchi (1999) present ten measures of rectangularity of the survival curve.

Figure 7.13 **Conditional COHORT-specific probabilities of survival from age 50 for 1980 cohort. Projections from fitted models (4.4.2f) and (4.4.2m), and projections from modified models.**

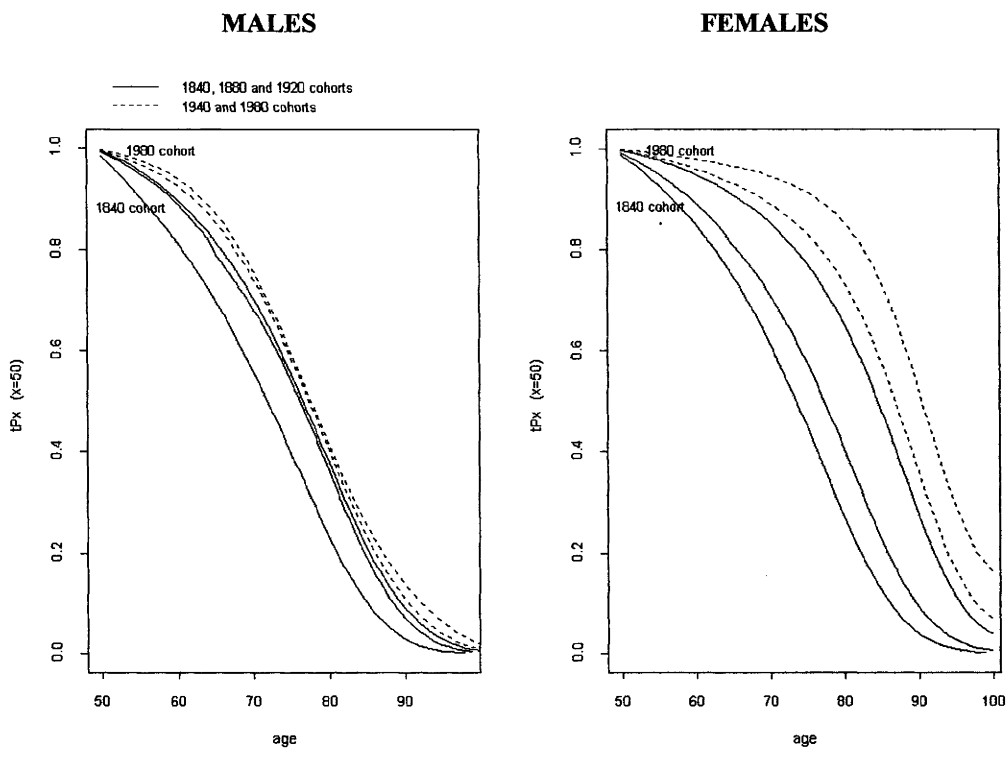


Nusselder and Mackenbach (1997) noticed rectangularization of period-specific survival curves in the Netherlands during the 1980s as opposed to the 1970s. They analysed mortality patterns by cause of death and concluded that the rectangularization was due to decreases in heart disease, cerebrovascular diseases, and lung cancer in men, coupled with an increase in mortality from certain diseases at old ages (such as emphysema). This is supported by our analysis of mortality patterns in Chapter 3 where it was noticed that there had been little change in mortality at the upper ages, but dramatic improvements at middle ages during the 1980s.

Figure 7.14 presents the conditional cohort-specific probabilities of survival from age 50 for a number of cohorts. According to the projections based on the fitted regression models, the probabilities at the upper ages continue to rise, suggesting no evidence of rectangularization. However, as stated earlier, there is considerable uncertainty surrounding estimates above age 90. A proper assessment of evidence for rectangularization requires

accurate data and models for the oldest ages and so is not pursued here.

Figure 7.14 **Conditional COHORT-specific probabilities of survival from age 50. Projected from models (4.4.1f), (4.4.2f), (4.4.1m) and (4.4.2m). The curves are descending with cohort: eg. 1980 cohort is highest; 1840 cohort is lowest.**

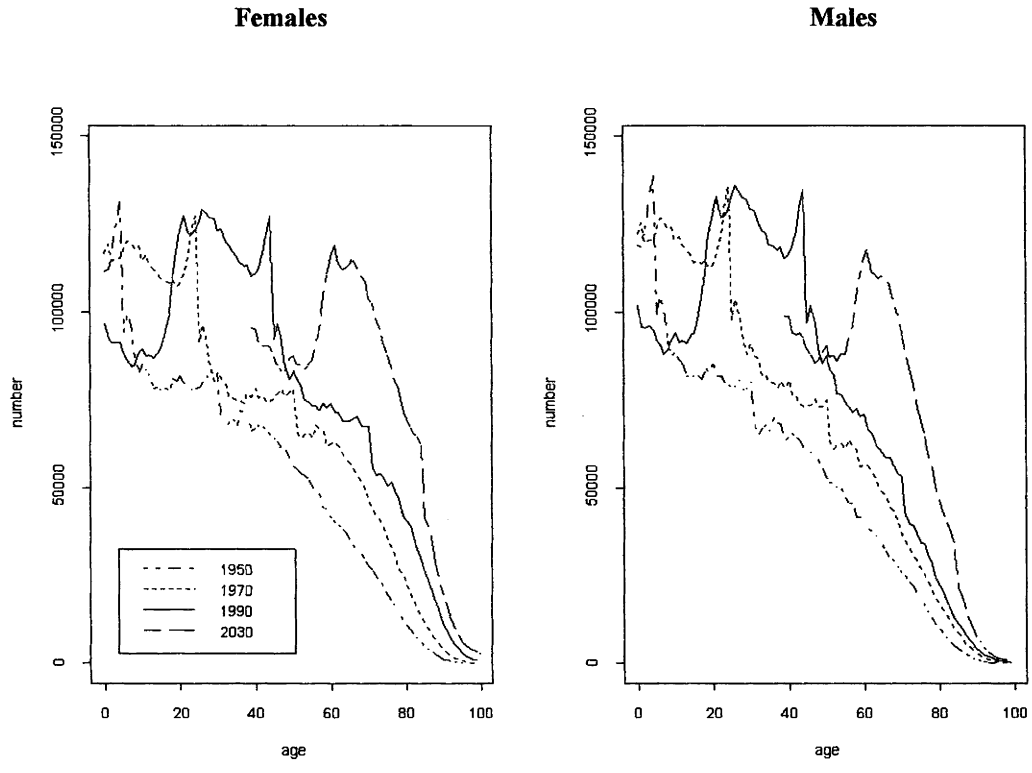


7.5 Impact of Uncertainty on Population Projections

An important conclusion of the above analysis is that a great deal of subjectivity, and hence uncertainty, exists in mortality projections. This feeds into uncertain population projections. The ageing of the population is a phenomenon that has been identified throughout the developed world that will persist well into the middle of the next century with no prospect of reversing or halting in the meantime. Due to increased birth rates following World War II much of the developed world is experiencing the ageing of a significant cohort of individuals. The 'baby-boomers' as they are known, are approaching retirement, and how this will affect Government social security funding, health care and investment markets is an issue of great importance. The size of the Dutch population for different ages and years is plotted in Figure 7.15 for males and females. The noticeable bumps and curves in these

charts point to a large difference in population numbers pre and post age 45 in 1990, corresponding with births after World War II.

Figure 7.15 **Dutch Males and Females. Population by age at 1950,1970, 1990 and 2030 (projected from empirical rates at 1991 assuming no change in mortality)**



If we project the population forward in time, and ignore migration, the curve will shift to the right as shown in Figure 7.15 for year 2030. The estimate for 2030 was constructed by projecting survivors assuming that the empirical mortality rates for 1991 will remain constant in the future. It is immediately obvious, that even if mortality rates cease improving, which is unlikely given historic trends, the number of elderly in the population will increase substantially.

Of particular interest is the variation in future population projections due to different mortality assumptions. We project Dutch population numbers for 50 to 90 year olds from 1991 to 2030 under three scenarios for future mortality rates: constant at 1991 levels (as above); based on extrapolations of the fitted models (4.4.2m) and (4.4.2f); and, based on extrapolations of the modified models introduced in Section 7.2.1. Results are presented in

Table 7.10 and Table 7.11. These forecasts are included for illustration of the effect of uncertainty in mortality, and are not considered to be reliable estimates of the Dutch population. Migration would need to be accounted for in an accurate population projection.

The analysis shows that a relatively small change in forecast assumptions can lead to significant differences in population composition at the oldest ages. The longer the projection period, the larger the influence of alternative projections.

Table 7.10 **Dutch Females. Population forecasts for specific ages under 3 scenarios. Percentage increase in 1991 figures is given in brackets underneath population number.**

	Age 60	Age 70	Age 80	Age 90
1991 population	72,630	66,540	40,470	11,820
2030 (no improvement)	115,780 (59.4%)	103,910 (56.2%)	69,780 (72.4%)	19,480 (64.8%)
2030 (fitted model)	117,940 (62.4%)	111,440 (67.5%)	84,430 (108.6%)	30,150 (155.1%)
2030 (modified model)	117,680 (62.0%)	110,680 (66.3%)	82,610 (104.1%)	27,680 (134.2%)

Table 7.11 **Dutch Males. Population forecasts for specific ages under 3 scenarios. Percentage increase in 1991 figures is given in brackets underneath population number.**

	Age 60	Age 70	Age 80	Age 90
1991 population	69,390	52,180	22,030	4,160
2030 (no improvement)	115,010 (65.7%)	93,820 (79.8%)	46,320 (110.3%)	7,460 (79.3%)
2030 (fitted model)	116,200 (67.5%)	94,230 (80.6%)	46,140 (109.4%)	9,460 (127.4%)
2030 (modified model)	116,850 (68.4%)	96,520 (85.0%)	49,520 (124.8%)	11,100 (166.8%)

7.6 A Measurement of 'Old Age' ?

Following World War II, the Netherlands introduced a universal social security system that included an act enabling the payment of benefits to men who had reached age 65 (General Old Age Benefit Act 1957 or AOW) (van Oorschot,1998). As mortality rates have decreased, the average Dutch lifespan has increased, yet, although particulars of the AOW have changed substantially (for example, women are also entitled to benefits, and means testing has been introduced), the entitlement age of 65 that was valid in 1957 remains the same today.

A paper recently published titled "How Old is Old ? Revising the Definition Based on Life Table Criteria", examined methods of comparing statistics between life tables with respect to a definition of "old". In particular, they posed the question 'If 65 was considered appropriate (as the point of entry into "old age") four decades ago, what is the corresponding age today ?' (Denton and Spencer,1996b).

Higgins and Thorburn (1999) utilised the methodology developed by Denton and Spencer to pose a similar question for Australia. We follow this tradition, and apply the methodology to the Netherlands to answer the question: 'If 65 was considered as the point of entry into old age in 1957 (corresponding to social security eligibility), what is the corresponding age today ?'. In particular, we calculate a male and female equivalent age based on the fitted regression models for 1991.

Denton and Spencer conduct nineteen tests based on criteria ranging from common statistics such as the mean, median and mode, to modern mathematical tests involving 'fuzzy' distributions. They found that, despite the differences in the tests conducted, the results achieved were quite similar.

Following Higgins and Thorburn (1999), for the Dutch data we have conducted 6 tests: mean years of life remaining; mean as a proportion of years already lived; median years of life remaining; and k-year survival rate comparisons for k=5,10 and 20. The results are presented in Table 7.12.

Table 7.12 Results of tests to determine equivalent age in 1991 given age 65 in 1957.

	Male equivalent age in 1991 given age 65 in 1957	Female equivalent age in 1991 given age 65 in 1957
Mean years of life remaining	65.60	70.12
Mean as a proportion of years already lived	65.44	68.93
Median years of life remaining	65.21	70.51
5 year survival rate	65.28	70.34
10 year survival rate	65.15	70.35
15 year survival rate	65.27	70.23
Average	**65.3**	**70.1**

If eligibility for age related benefits in the Netherlands was to change solely based on a reflection of life expectancy increases, then the male age should remain constant at 65 years, whereas the female age should increase to 70. In practice, other considerations would need to be addressed, including quality of health and evidence for compression of morbidity, changes in the level of age benefits provided, and entitlements to supplementary benefits. Regardless, the analysis provides an interesting summary of the changes in Dutch mortality since the late 1950s. In particular, the relatively small increase in equivalent age for males relative to females is reflects the almost steady mortality for males and the steep decline for females since 1957.

7.7 Conclusions and Future Work

The regression models proposed in Chapter 4 and explored in Chapter 5 have the advantage of being founded on plausible statistical assumptions involving the distribution of survivors within and between cohorts.

Under the classical assumptions of independence and homogeneity, the variance of the errors from a smooth regression model can be determined. Divergence from the expected variance allows us to determine the magnitude of the extra-variability, though the source cannot be identified in detail due to the absence of disaggregated data. It is speculated, however, that the failure of the assumption of homogeneity within cohorts is responsible for the extra variance, and the effect on aggregate mortality of a mixture of distributions at each age within cohorts is discussed in the context of selection and debilitation. Traditional approaches correct for extra-binomial variance by inclusion of a fixed inflation factor. Since there is evidence that the extra variance is not constant over the mortality surface, the preferred approach is to model the extra variance using a suitably flexible parametric surface.

An advantage of asymptotically normal errors is that maximum likelihood methods can be used to produce consistent estimators for the parameters of both the mean model and the model for the extra variance component. The information matrix from the MLE fit allows us to easily extract standard errors which enable estimation of confidence and prediction intervals for individual parameters, the fitted surface, or predictions.

A particularly interesting feature that is identified in Chapter 5 for Dutch females is an increase in the age at which the rate of change in mortality reaches a maximum between the 1950s and the present.

The fitted models developed in Chapter 4 are parsimonious in order to provide sensible bases for projections. A consequence of a parsimonious approach is that complex trends within the data are not modelled. Inclusion of high order interactions improves the quality of the fits, though at the expense of interpretation and predictability. In particular, perfect collinearity between age, period and cohort means that it is impossible to uniquely determine the separate effects. A potential cohort effect is identified in Chapter 4 and Chapter 5 by exploring the residuals and derivatives of the male model (4.4.2m). Incorporating cohort effects within a mortality model is covered in Chapter 6, where the technique of Wilmoth (1990) is applied to Dutch males, and it is shown how a Lee-Carter model (1992) can be modified to incorporate allowance for cohort effects.

Projections from the regression models are undertaken in Chapter 6. It is shown that the choice of historic period used in formulation of the models greatly affects the direction of forecasts. Projections from extrapolation of the regression models of Chapter 4 are questionable for some ages, and alternative extrapolations are generated by modifying the parameters of the fitted models to incorporate human judgement. Life expectancies and conditional probabilities of survival are generated to illustrate the different scenarios. Finally it is shown that uncertainty in mortality has a large influence on population projections for the oldest sections of the community.

It would be possible in theory to extend the regression model developed in this thesis to the entire age surface. A parameterisation such as that proposed by Heligman and Pollard could be modified to allow variation with year. Explorations of extra-binomial variance over the entire age surface may provide some insight into the nature and causes of the extra variability. An analysis of the entire age-pattern of mortality utilising the techniques described in Chapter 6 may provide useful information that could influence forecasts. In particular, it would be informative to assess whether the Hunger Winter of WWII has had a long-term debilitating consequence on any cohorts, or, indeed whether selection effects have produced a more robust population of survivors.

Large sample approximation for the distribution of the errors enables the regression technique to be applied in a wide variety of areas. An extension of the model to disability states was carried out by Davis et al. (2001), where health expectancies were estimated from cross-sectional data by weighted least squares using log(odds). The main issue restricting progress with the development of mortality models is the lack of detailed data. If sufficient longitudinal data were available, regression surfaces could be applied to individual causes of death or individual states of health, where cohorts consist of multiple hazard functions, each representing a different state. With the appropriate disaggregated data, developing regression methods like those advocated here would lead to an improved understanding of sources of variation, and ultimately a deeper understanding of the processes of ageing and dying.

APPENDICES

Appendix A

The following code was used to generate the negative log-likelihood for the post-WWII male model (4.4.2m). Similar code was used for the female data and the pre-WWII male data. The log-likelihood was given in Chapter 4 and is reproduced here:

$$\sum_{x=50}^{90}\sum_{t=J1}^{J2}\log f\left(t,x;\underline{\hat{\beta}},\underline{\hat{\alpha}}\right)=-\frac{n}{2}\log 2\pi-\frac{1}{2}\sum_{x=50}^{90}\sum_{t=J1}^{J2}\log\left(\sigma^2(t,x)+g\left(t,x;\underline{\hat{\alpha}}\right)\right)-\frac{1}{2}\sum_{x=50}^{90}\sum_{t=J1}^{J2}\left(\frac{\left(Y_{x,t-x}-\delta(t,x;\underline{\hat{\beta}})\right.}{\sigma^2(t,x)+g\left(t,x;\underline{\hat{\alpha}}\right.}\right.$$

Since the first term in the log-likelihood is independent of the parameters to be estimated we can remove it from the equation. Maximising the log-likelihood is equivalent to minimising the negative log-likelihood:

$$-\sum_{x=50}^{90}\sum_{t=J1}^{J2}\log f\left(t,x;\underline{\hat{\beta}},\underline{\hat{\alpha}}\right)\approx\frac{1}{2}\sum_{x=50}^{90}\sum_{t=J1}^{J2}\log\left(\sigma^2(t,x)+g\left(t,x;\underline{\hat{\alpha}}\right)\right)+\frac{1}{2}\sum_{x=50}^{90}\sum_{t=J1}^{J2}\left(\frac{\left(Y_{x,t-x}-\delta(t,x;\underline{\hat{\beta}})\right)^2}{\sigma^2(t,x)+g\left(t,x;\underline{\hat{\alpha}}\right)}\right)$$

The data objects used in the log-likelihood code are defined below.

a1-a13 = the parameters to be estimated corresponding to the mean (log(odds)). The subscript c refers to centred variables.

$$\delta(t,x;\underline{b})=b_1+b_2x_c+b_3x_c^2+b_4x_c^3+b_5x_c^4+$$
$$+b_6t_c+b_7x_ct_c+b_8t_cx_c^2+b_9t_cx_c^4+$$
$$+dnorm(t,1972,9)\left(b_{10}+b_{11}x_c+b_{12}x_c^2+b_{13}x_c^3\right)$$

b1-b5 = the parameters to be estimated corresponding to the extra variance,

$$g\left(t,x;\underline{a}\right)=\exp\left(a_1+a_2x_c+a_3t_c+a_4t_c^2+a_5t_cx_c\right)$$

Maledata = data set containing the observed data for males for 1945-1991, ages 50-90.

Weights = the binomial weights used in the model (equivalent to $1/\sigma^2(t,x)$)

logitq = the observed log(odds)

agec = the centred variable for age. This was equal to the age minus 70 (being the mean age).

Yearc = the centred variable for year. This is equal to year minus 1968.5 (being the mean year for the post-WWII data)

dnorm = the normal density ordinate used to model the cardiovascular epidemic.

The S-plus code for the function includes the first derivatives of the non-linear model and the hessian, which is the sum of the second derivatives array over the observations.

```
Lprobg_function(b1,b2,b3,b4,b5,b6,b7,b8,b9,b10,b11,b12,b13,a1,a2,a3,a4,a5)
{
.expr1 <-(1/weights+exp(a1+a2*agec+a3*yearc+a4*yearc^2+a5*yearc*agec))
.expr2 <- (logitq-
(b1+b2*agec+b3*agec^2+b4*agec^3+b5*agec^4+b6*yearc+b7*agec*yearc+b8*yearc*agec^2+b9*
yearc*agec^4+dnorm(year,1972,9)*(b10+b11*agec+b12*agec^2+b13*agec^3)))
.expr3 <- 0.5/.expr1
.expr4 <- -0.5*(.expr2^2)/(.expr1^2)
.expr5 <- -.expr2/.expr1
.expr6 <- exp(a1+a2*agec+a3*yearc+a4*yearc^2+a5*yearc*agec)
.expr7 <- 0.5*(1/weights)*.expr6/(.expr1^2)+ .expr4*.expr6*(1-2*.expr6/.expr1)
.expr8 <- (.expr2/(.expr1^2))*.expr6
.value <- 0.5*log(.expr1)+0.5*((.expr2)^2)/.expr1
.grad <- array(0, c(length(.value), 18),
list(NULL,c("b1","b2","b3","b4","b5","b6","b7","b8","b9","b10","b11","b12","b13","a1","a2","a3
","a4","a5")))
.grad[, "b1"] <- .expr5;                    .grad[, "b2"] <- .expr5*agec;
```

```
.grad[, "b3"] <- .expr5*agec^2;                    .grad[, "b4"] <- .expr5*agec^3;

.grad[, "b5"] <- .expr5*agec^4;                    .grad[, "b6"] <- .expr5*yearc
.grad[, "b7"] <- .expr5*yearc*agec;                .grad[, "b8"] <- .expr5*yearc*agec^2
.grad[, "b9"] <- .expr5*yearc*agec^4;              .grad[, "b10"] <- .expr5*dnorm(year,1972,9)
.grad[, "b11"] <- .expr5* dnorm(year,1972,9)*agec;
.grad[, "b12"] <- .expr5* dnorm(year,1972,9)*agec^2;
.grad[, "b13"] <- .expr5* dnorm(year,1972,9)*agec^3;
.grad[, "a1"] <- (.expr3 + .expr4)*.expr6;
.grad[, "a2"] <- (.expr3 + .expr4)*.expr6*agec
.grad[, "a3"] <- (.expr3 + .expr4)*.expr6*yearc
.grad[, "a4"] <- (.expr3 + .expr4)*.expr6*yearc^2
.grad[, "a5"] <- (.expr3 + .expr4)*.expr6*yearc*agec
attr(.value, "gradient") <- .grad


.hess<-array(0,c(length(.value),18,18))
dimnames(.hess)<-list(NULL,
c("b1","b2","b3","b4","b5","b6","b7","b8","b9","b10","b11","b12","b13","a1","a2","a3","a4","a5
"),
c("b1","b2","b3","b4","b5","b6","b7","b8","b9","b10","b11","b12","b13","a1","a2","a3","a4","a5
"))


.hess[, "b1","b1"]<-(1/.expr1);                    .hess[, "b1","b2"]<-(1/.expr1)*(agec);
.hess[, "b1","b3"]<-(1/.expr1)*(agec^2);           .hess[, "b1","b4"]<-(1/.expr1)*(agec^3)
.hess[, "b1","b5"]<-(1/.expr1)*(agec^4);           .hess[, "b1","b6"]<-(1/.expr1)*(yearc)
.hess[, "b1","b7"]<-(1/.expr1)*(yearc*agec);       .hess[, "b1","b8"]<-(1/.expr1)*(yearc*agec^2)
.hess[, "b1","b9"]<-(1/.expr1)*(yearc*agec^4);     .hess[, "b1","b10"]<-
(1/.expr1)*dnorm(year,1972,9)
.hess[, "b1","b11"]<-(1/.expr1)*(dnorm(year,1972,9)*agec)
.hess[, "b1","b12"]<-(1/.expr1)*(dnorm(year,1972,9)*agec^2)
.hess[, "b1","b13"]<-(1/.expr1)*(dnorm(year,1972,9)*agec^3)
.hess[, "b1","a1"]<- .expr8
.hess[, "b1","a2"]<- .expr8*(agec)
.hess[, "b1","a3"]<- .expr8*(yearc)
.hess[, "b1","a4"]<- .expr8*(yearc^2)
.hess[, "b1","a5"]<- .expr8*(yearc*agec)


for(I in 1:18){
.hess[, "b2",I]<- .hess[, "b1",I]*agec
.hess[, "b3",I]<- .hess[, "b1",I]*agec^2
```

```
.hess[, "b4",I]<- .hess[, "b1",I]*agec^3
.hess[, "b5",I]<- .hess[, "b1",I]*agec^4
.hess[, "b6",I]<- .hess[, "b1",I]*yearc
.hess[, "b7",I]<- .hess[, "b1",I]*yearc*agec
.hess[, "b8",I]<- .hess[, "b1",I]*yearc*agec^2
.hess[, "b9",I]<- .hess[, "b1",I]*yearc*agec^4
.hess[, "b10",I]<- .hess[, "b1",I]*dnorm(year,1972,9)
.hess[, "b11",I]<- .hess[, "b1",I]*dnorm(year,1972,9)*agec
.hess[, "b12",I]<- .hess[, "b1",I]*dnorm(year,1972,9)*agec^2
.hess[, "b13",I]<- .hess[, "b1",I]*dnorm(year,1972,9)*agec^3}

for(I in 1:13){
.hess[,"a1",I]<- .expr8*.hess[,"a1",I]*.expr1
.hess[,"a2",I]<- .expr8*.hess[,"a1",I]*.expr1*agec
.hess[,"a3",I]<- .expr8*.hess[,"a1",I]*.expr1*yearc
.hess[,"a4",I]<- .expr8*.hess[,"a1",I]*.expr1*yearc^2
.hess[,"a5",I]<- .expr8*.hess[,"a1",I]*.expr1*yearc*agec}

.hess[, "a1","a1"]<-.expr7;
.hess[, "a2","a1"]<-.expr7*agec;
.hess[, "a3","a1"]<-.expr7*yearc;
.hess[, "a4","a1"]<-.expr7*yearc^2;
.hess[, "a5","a1"]<-.expr7*yearc*agec;

for(I in 14:18){
.hess[, I,"a2"]<- .hess[, I,"a1"]*agec;
.hess[, I,"a3"]<- .hess[, I,"a1"]*yearc;
.hess[, I,"a4"]<- .hess[, I,"a1"]*yearc^2;
.hess[, I,"a5"]<- .hess[, I,"a1"]*yearc*agec}

attr(.value, "hessian") <- .hess
.value}
```

The parameters are estimated by minimising the negative log-likelihood function through the S-plus function 'ms' which minimizes the sum of nonlinear functions over parameters in a data frame. If we define an object 'fM' to store the results of the minimisation, then:

```
fM_ms(~lprobg(b1,b2,b3,b4,b5,b6,b7,b8,b9,b10,b11,b12,b13,a1,a2,a3,a4,a5),Maledata,trace=T)
```

Following minimisation of the negative log-likelihood, the fitted mean (fitM) can be found, in this instance, by:

```
fitM_ (fM$parameter[1] + fM$parameter[2] *agec+ fM$parameter[3]*agec^2+
    fM$parameter[4]*agec^3+ fM$parameter[5]*agec^4+ fM$parameter[6]*yearc+
    fM$parameter[7]*agec*yearc+fM$parameter[8]*yearc*agec^2+
    fM$parameter[9]*yearc*agec^4+dnorm(year,1972,9)*(fM$parameter[10] +
    fM$parameter[11]*agec+ fM$parameter[12] *agec^2+ fM$parameter[13] *agec^3))
```

The fitted extra variance component $\gamma^2(t,x) = g(t,x;\hat{\underline{\alpha}})$ (gamM) can be found, in this instance, by:

```
gamM_exp(fM$parameter[14]+fM$parameter[15]*agec+fM$parameter[16]*yearc+
    fM$parameter[17]*yearc^2+ fM$parameter[18]*yearc*agec)
```

Despite its incorrect label, the code below gives the *inverse* of the information matrix (Venables and Ripley, 1994:244). This is used in determining the variance of the individual parameter estimates and the variance of the fitted model (see Chapter 2 for the theory underlying the derivation of the variance).

```
summary(fitM)$Information
```

BIBLIOGRAPHY

Alho,J.M.,1990, Stochastic methods in population forecasting. *International Journal of Forecasting*, 6:521-530.

Alho,J.M.,B.D.Spencer, 1985, Uncertain Population Forecasts. *Journal of the American Statistical Association*, 80(390):306-314.

Alho,J.M.,B.D.Spencer, 1990, Effects of Targets and Aggregation on the Propagation of Error in Mortality Forecasts. *Mathematical Population Studies*, 2(3):209-227.

Alho,J.M.,1992, Estimating the Strength of Expert Judgement: The Case of US Mortality Forecasts. *Journal of Forecasting*, 11:157-167.

Bell,W.R., B.Monsell, 1991, Using Principal Components in Time Series Modeling and Forecasting of Age-Specific Mortality Rates. *1991 Proceedings of the Social Statistics Section*, American Statistical Association, Alexandria, VA, 154:159.

Bennett, N.G., S.J.Olshansky, 1996, Forecasting US Age Structure and the Future of Social Security: The Impact of Adjustments to Official Mortality Schedules. *Population and Development Review*, 22(4):703-727.

Bulatao, R.A., P.W.Stephens, 1992, *Global Estimates and Projections of Mortality by Cause*, Policy Research Working Papers, Population, Health and Nutrition, The World Bank.

Burtless,G., 1998, *Increasing the Eligibility Age for Social Security Pensions*, Testimony for the Senate Special Committee on Aging, July 15,1998.

Carey, J.R., 1997, What Demographers can learn from fruit fly actuarial models and biology. *Demography*, 34(1): 17-30.

Carnes, B.A., S.J. Olshansky, D.Grahn, 1996, Continuing the Search for a Law of Mortality. *Population and Development Review*, 22(2): 231-264.

Caselli,G., 1990, The Influence of Cohort Effects on Differentials and Trends in Mortality. *Measurement and analysis of mortality*, J.Vallin, D'Souza and Palloni (eds.)

Chambers J.M., D.M.Bates, 1993, Nonlinear Models. *Statistical Models in S*, J.M.Chambers, T.J.Hastie (eds.), ch10, Chapman and Hall, London.

Chambers J.M., 1993, Linear Models. *Statistical Models in S*, J.M.Chambers, T.J.Hastie (eds.), ch4, Chapman and Hall, London.

Chiang, C.L.,1968, *Introduction to Stochastic Processes in Biostatistics*, Wiley, New York.

Chiang, C.L.,1984,*The Life Table and its Applications*, Robert E. Krieger Publishing Company, Malabar, Florida.

Coale,A.J., 1996, Age Patterns and Time Sequence of Mortality in National Populations with the Highest Expectation of Life at Birth. *Population and Development Review*, 22(1):127-135.

Davies, R.B., 1987, The limitations of cross sectional analysis. *Longitudinal Data Analysis*, Surrey Conferences on Sociological Theory and Methods 4, R.Crouchley (ed.), Gower Publishing Company, Avebury.

Davis, B.A, C.R.Heathcote, T.J.O'Neill, 2001, Estimating Cohort Health Expectancies from Cross-Sectional Surveys of Disability. *Statistics in Medicine*, 20.

Denton,F.T., B.G.Spencer, 1996a, *Population Aging and the Maintenance of Social Support Systems*, Independence and Economic Security of the Older Population Research Paper No.9, McMaster University, Canada.

- 1996b, *How Old is Old ? Revising the Definition Based on Life Table Criteria*, IESOP Research Paper No.2, McMaster University, Canada.

Gage, T.B., 1994, Population Variation in Cause of Death: Level, Gender, and Period Effects. *Demography*, 30(2):271-296

Gompertz, B., 1825, On the nature of the function expressive of the law of human mortality and on a new mode of determining life contingencies, *Philosophical Transactions of the Royal Society of London*,115:513-85.

Hagenaars, J.A, 1990, *Categorical Longitudinal data*, Sage Publications,. Newbury Park,California.

Hartmann,M.,1987, Past and Recent Attempts to Model Mortality at All Ages. *Journal of Official Statistics*, Statistics Sweden, 3(1):19-36.

Heathcote,C.R. and I.M.,McDermid, 1994, Projections of cohort life expectancy based on weighted least squares methods. *Advances in health expectancies*, C.Mathers, J.McCallum.,J.M.Robine (eds.), Canberra: Australian Institute of Health and Welfare.

Heathcote, C.R., T.Higgins, 2001a, A Regression Model of Mortality with Application to the Netherlands. *Forecasting Mortality in Developed Countries*, E.Tabeau, A.van den Berg Jeths, C.Heathcote (eds.), Kluwer Academic Publishers, the Netherlands: 59-82.

Heathcote, C.R., T.Higgins, 2001b, Forecasting Mortality from Regression Models: the Case of the Netherlands. *Forecasting Mortality in Developed Countries*, E.Tabeau, A.van den Berg Jeths, C.Heathcote (eds.), Kluwer Academic Publishers, the Netherlands: 83-103.

Heligman,L., J.H.Pollard, 1980,The Age Pattern of Mortality, *Journal Institute of Actuaries*, vol.107, 49-75.

Henry, C.J.K, and S.J.Ulijaszek, 1996, *Long-term consequences of early environment*, Society for the study of human biology symposium 37, Cambridge University Press, Cambridge.

Higgins, T., 1998, *Australian Mortality: Improvement and Uncertainty in an Ageing Population*, Sixth Annual Colloquium of Superannuation Researchers, Melbourne.

Higgins,T., C.Thorburn, 1999, *Life in Abundance: The Australian Life Tables 1995-1997 and Some Observations on Australian Mortality*, Seventh Annual Colloquium of Superannuation Researchers, Melbourne.

Horiuchi,S., A.J.Coale, 1990, Age Pattern of Mortality for Older Women: An Analysis using the Age-Specific Rate of Mortality Change with Age, in *Mathematical Population Studies*, 2(4): 245-267.

Horiuchi,S., J.R.Wilmoth, 1998, Deceleration in the Age Pattern of Mortality at Older Ages. *Demography*, 35(4):391-412.

Jacobsen,G.,1998, More Cancer Victims defying Death Sentence, in *Sydney Morning Herald*, May 11, 1998:5.

Jain,S.K.,1994, *Trends in Mortality*, Australian Bureau of Statistics, ABS Catalogue No.3313.0, NCEPH and ABS,Canberra

Jeanneret,O.,1983 Sex Differentials in Mortality and Health Care Delivery: A tentative exploration of some relationships. *Sex Differentials in Mortality*, A.D.Lopez, L.T.Ruzicka(eds.), Australian National University Press, Canberra,409:425.

Kendall, M., A.Stuart, 1977, *The Advanced Theory of Statistics. Volume 1. Distribution Theory*, Charles Griffin and Company Ltd, London.

Keyfitz,N., 1982, Choice of Function for Mortality Analysis: Effective Forecasting Depends on a Minimum Parameter Representation. *Theoretical Population Biology*, 21:329-352.

Land,K.C.,1986, Methods for National Population Forecasts: A Review. *Journal of the American Statistical Association*, 1986, 81(396):888:901.

Lee,R.D.,L.W.Carter, 1992, Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87(419):659-675.

Lee,R.D,1992, Stochastic Demographic Forecasting. *International Journal of Forecasting*, 8:315-327

Lee,R., S.,Tuljapurkar, 1998, *Death and Taxes: How Longer Life Will Affect Social Security*, Demography of Aging Working Papers, No.4.,University of California Berkeley.

Lilienfeld, A.M.,1976, *Foundations of Epidemiology*, Oxford University Press, New York.

Lopez,A.D., 1983,The Sex Mortality Differential in Developed Countries. *Sex Differentials in Mortality*, A.D.Lopez, L.T.Ruzicka(eds.), Australian National University Press, Canberra, 53-120.

Manton,K.G., 1993,Biomedical Research and Changing Concepts of Disease and Aging: Implications for Long-Term Health Forecasts for Elderly Populations. *Forecasting the Health of Elderly Populations*, K.G.Manton, B.H.Singer, R.M.Suzman (eds.), Springer-Verlag, New York, 319-365.

Manton,K.G., B.H.Singer, R.M.Suzman, 1993, The Scientific and Policy Needs for Improved Health Forecasting Models for Elderly Populations. *Forecasting the Health of Elderly Populations*, K.G.Manton, B.H.Singer, R.M.Suzman (eds.), Springer-Verlag, New York.

Manton,K.G.,E.Stallard,J.W.Vaupel, 1986, Alternative Models for the heterogeneity of Mortality Risks Among the Aged. *Journal of the American Statistical Association*, 81(395): 635-644.

Mathers,C.D., 1990, *Disability-free and handicap-free life expectancy in Australia*, Health Differentials Series No.1, Australian Institute of Health and Welfare, Canberra.

McCullagh,P., J.A.Nelder, 1989, *Generalized Linear Models*, Chapman & Hall, London.

Moore,D.F.,A.Tsiatis, 1991, Robust estimation of the Variance in Moment Methods for Extra-Binomial and Extra-Poisson Variation. *Biometrics*, 47, 383-401.

McNown, R., A.Rogers, 1989, Forecasting Mortality: A Parameterised Time Series Approach. *Demography*, 26(4):645-660.

McNown, R.,1992, Comments on Modeling and Forecasting U.S. Mortality, by Lee,R.D.,L.W.Carter. *Journal of the American Statistical Association*, 87(419):659-675.

Nusselder,W.J., J.P.Mackenbach, 1997, Rectangularization of the Survival Curve in the Netherlands: An Analysis of Underlying Causes of Death. *Journal of Gerontology: Series B; Social Sciences*, 52B(3):145-154.

Nusselder,W.J., J.P.Mackenbach, 2000, Lack of Improvement of Life Expectancy at Advanced Ages in the Netherlands, International Journal of Epidemiology, 29:140-148.

Olshansky, S.J., B.A.Carnes, 1997, Ever Since Gompertz. *Demography*, 34(1):1-15.

Olshansky,J.,1988,On forecasting mortality. *Milbank Quarterly*, 66,482-530.

Pollard,A.H.,1970, Random Mortality Fluctuations and the Binomial Hypothesis. *Journal of the Institute of Actuaries*, 96, 251-264.

Pollard, A.H., F.Yusuf, G.N.Pollard, 1974, *Demographic Techniques* (3rd ed.), A.S.Wilson, Australia.

Pollard, J.H., 1987, Projection of Age-Specific Mortality Rates, *in Population Bulletin of the United Nations*, 21/22: 55-69.

Poppel, F.van, E.Tabeau, F.Willekens, 1994, *Male Excess Mortality in the Netherlands since 1850:A Life Table Approach*, Paper presented at workshop: Life tables in Europe: Data, methods, and models. Belgium 1994, Netherlands Interdisciplinary Demographic Institute, The Hague, the Netherlands.

Rao, C.R., 1973, *Linear Statistical Inference and its Applications*, John Wiley and Sons, Toronto.

Rice,D.P.,1983, Sex Differences in Mortality and Morbidity: Some Aspects of the Economic Burden. *Sex Differentials in Mortality*, A.D.Lopez, L.T.Ruzicka (eds.), Australian National University Press, Canberra,335:369

Ruzicka,L.T, Alan D.Lopez,1983, Sex Differentials in Mortality:Conclusions and Prospects. *Sex Differentials in Mortality*, A.D.Lopez, L.T.Ruzicka (eds.),Australian National University Press, Canberra, 477:486.

Ryder,N.B.,1980,*The Cohort Approach*, Arno Press, New York (dissertation for PhD).

Ryan, T.P., 1997, *Modern Regression Methods*, John Wiley and Sons Inc., New York.

Salhi,M.,G.Caselli,J.,Duchene,V.Egidi,A.Santini,E.,Thiltges,G.,Wunsch, 1995, Assessing Mortality Differentials Using Life-Histories: A Method and Applications, *Adult mortality in developed countries: From Description to explanation*, A.Lopez, G.Caselli, T.Valkonen (eds.),Oxford.

Schervish,M.J.,1995, *Theory of Statistics*, Springer-Verlag, New York.

Searle, S.R, G.Cassella, C.E.McCulloch, 1992, *Variance Components*, John Wiley & Sons, New York.

Smith, D.W.E, 1993, *Human Longevity*, Oxford University Press, New York.

Stefanyuk,A., 1994, The Problem of Nonparametric Estimation of Mortality Risk Function. *Modeling Techniques for Uncertain Systems*, A.B.Kuzhanski, V.M.Veliov (eds), International Institute for Applied Systems Analysis:53-68.

Tabeau, E., F.van Poppel, F.Willekens, 1994, *Mortality in the Netherlands: The Data Base, Report no. 36*, Netherlands Interdisciplinary Demographic Institute, The Hague, the Netherlands.

Tabeau, E., 2001, A Review of Demographic Forecasting Models for Mortality. *Forecasting Mortality in Developed Countries*, E.Tabeau, A.van den Berg Jeths, C.Heathcote (eds.), Kluwer Academic Publishers, the Netherlands: 1-32.

Tabeau, E., P.Ekamper,C.Huisman,A.Bosch, 2001a, Predicting Mortality from Period, Cohort or Cause-Specific Trends: a Study of Four European Countries. *Forecasting Mortality in Developed Countries*, E.Tabeau, A.van den Berg Jeths, C.Heathcote (eds.), Kluwer Academic Publishers, the Netherlands: 159-187.

Tabeau, E., A.Van Den Berg Jeths, C.Heathcote, 2001b, Towards an Integration of the Statistical, Demographic and Epidemiological Perspectives. *Forecasting Mortality, in Forecasting Mortality in Developed Countries*, E.Tabeau, A.van den Berg Jeths, C.Heathcote (eds.), Kluwer Academic Publishers, the Netherlands: 281-303.

Thiltges E., J. Duchene, G. Wunsch, 1995, Causal Theories and Models in the Study of Mortality. *Adult Mortality in Developed Countries: From Description to Explanation*, Lopez, Caselli, Valknonen (eds.), Clarendon Press, Oxford: 21-36.

Thorburn, C., 1999, Births, Deaths, Passports and Pensions: Australia's Demographics. *Australian Actuarial Journal*, Institute of Actuaries of Australia, 5(4):531-624

Tickle, L., 1997, *The Sex Differential in Mortality Rates and Underlying Factors*, Actuarial Studies and Demography Research Paper Series 4, March 1997, 6(97), Macquarie University.

Tuljapurkar, S., C.Boe, 1998, Mortality Change and Forecasting: How Much and How Little Do We Know. *North American Actuarial Journal*, 2(4).

Tuljapurkar, S., R.Lee, 2000, Demographic Uncertainty and the OASDI Fund. *Intergenerational Economic Relations and Demographic Change*, A.Mason and G.Tapinos (eds.),Oxford University Press.

Venables,W.N., B.D.Ripley, 1994, *Modern Applied Statistics with S-Plus*, Springer-Verlag, New York.

Van Oorschot, W., 1998, *The reconstruction of the Dutch Social Security System1980-2000: retrenchment and modernization*, Centre for Comparative Welfare Studies Working Papers 1998, 3, Aalborg University.

Van Den Berg Jeths, A., R.Hoogenvenn, G.De Hollander, E.Tabeau, 2001, A Review of Epidemiological Approaches to Forecasting Mortality and Morbidity. *Forecasting Mortality in Developed Countries*, E.Tabeau, A.van den Berg Jeths, C.Heathcote (eds.), Kluwer Academic Publishers, the Netherlands: 33-56.

Van Genugten, M., R.Hoogenvenn, G.De Hollander, 2001, Incorporating Risk Factor Epidemiology in Mortality Projections. *Forecasting Mortality in Developed Countries*, E.Tabeau, A.van den Berg Jeths, C.Heathcote (eds.), Kluwer Academic Publishers, the Netherlands: 189-204.

Van Hoorn, W., J.De Beer, 2001, Projecting Mortality in Population Forecasts in the Netherlands. *Forecasting Mortality in Developed Countries*, E.Tabeau, A.van den Berg Jeths, C.Heathcote (eds.), Kluwer Academic Publishers, the Netherlands: 205-226.

Vaupel,J.W, A.Yashin, K.Manton, 1988, Debilitation's Aftermath: Stochastic Process Models of Mortality. *Mathematical Population Studies*, 1(1):21-48, Gordon and Breach Science Publishers.

Vaupel,J.W, A.Yashin, 1985, Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics. *The American Statistician*, 39(3): 176-184.

Vaupel, J.W., K.Manton, E.Stallard, 1979, The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, 16(3): 439-454.

Vaupel, J.W., V. Kannisto, J. Lauritsen, A.R.Thatcher, 1994, Reductions in Mortality at Advanced Ages: Several Decades of Evidence from 27 Countries. *Population and Development Review*, 20(4):793-809

Vaupel,J.W., 1986, How Change in Age-specific Mortality Affects Life Expectancy. *Population Studies*, 40:147-157.

Venables, W.N., B.D.Ripley, 1994, *Modern Applied Statistics with S-Plus*, Springer-Verlag, New York.

Warner,H.B., 1993,Molecular Biological Approaches to Understanding Aging and Senescence. *Forecasting the Health of Elderly Populations*, K.G.Manton,B.H.Singer, R.M.Suzman (eds.), Springer-Verlag, New York,307:318.

Weisberg, S., 1985, *Applied Linear Regression*, John Wiley and Sons, New York.

Williams, D.A.,1982, Extra-binomial Variation in Logistic Linear Models. *Applied Statistics*, 31(2):144-148.

Willekens, F., 2001, Gompertz in Context: the Gompertz and Related Distributions. *Forecasting Mortality in Developed Countries*, E.Tabeau, A.van den Berg Jeths, C.Heathcote (eds.), Kluwer Academic Publishers, the Netherlands: 105-126.

Wilmoth,J.R., 1990, Variation in vital rates by age, period and cohort. *Sociological Methodology*, 20, Clifford C Clogg (ed.), Basel Blackwell Ltd, Oxford.

Wilmoth,J.R., S.Horiuchi, 1999, Rectangularization Revisited: Variability of Age at Death Within Human Populations, in Demography, 36(4):475-495.

World Health Organisation, 1997, *World Health Report 1997*, http://www.who.ch/whr/1997/.

Yashin, A.I., K.G.Manton, J.W.Vaupel, 1985, Mortality and Aging in a Heterogeneous Population: A Stochastic Process Model with Observed and Unobserved Variables. *Theoretical Population Biology*, 27: 154-175.

Yashin, A.I., 2001, Mortality Models Incorporating Theoretical Concepts of Ageing. *Forecasting Mortality in Developed Countries*, E.Tabeau, A.van den Berg Jeths, C.Heathcote (eds.), Kluwer Academic Publishers, the Netherlands: 261-280.