ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

METHODS OF EXPANDING ABRIDGED LIFE TABLES: EVALUATION AND COMPARISONS

By

Vangelis Panousis

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
2001

ISBN: 960-7929-90-x

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

Methods of Expanding Abridged Life Tables: Evaluation and Comparisons

By

Vangelis Panousis

A THESIS

Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
May 2001

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

Μέθοδοι Εξάπλωσης Συνεπτυγμένων Πινάκων Επιβίωσης: Εφαρμογές και Αξιολόγηση των Μεθόδων

Βαγγέλης Πανούσης

ΔΙΑΤΡΙΒΗ
Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Μάιος 2001

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

A Thesis submitted in partial fulfillment of

the requirements for the degree of

Master of Science

METHODS OF EXPANDING ABRIDGED LIFE TABLES:
EVALUATION AND COMPARISONS

Vangelis Panousis

Supervisor:
A. Kostaki
Lecturer

External examiner:
Professor John Hsieh,
Department of Statistics & Department of Public Health Sciences, University of Toronto, CANADA

Approved by the Graduate Committee
Athens, November 2003

Michael Zazanis, Associate Professor
Director of the Graduate Program

DEDICATION

To my parents and my brother

ACKNOWLEDGEMENTS

First of all, I want to thank my supervisor, Lecturer of Statistics, Ms Anastasia Kostaki, for her valuable assistance, comments and great patience. Also I want to thank my friends for supporting me all these years in several ways. Finally, I thank my parents and brother for their financial and psychological support.

VITA

I was born in the 6th of September 1975 in Spata of Attica. There I went to school and in June 1993 I graduated. In October 1993 I began my studies in Statistics at the Athens University of Economics and Business. In October 1997 I joined the Master's program of Statistics of the same department during I attended classes in the Katholieke Universiteit Leuven in Belgium. My MSc thesis was on the field of Demography and Statistics, which is presented, in the following pages.

ABSTRACT

Vangelis Panousis

Methods of Expanding Abridged Life Tables: Evaluation and Comparisons

May 2001

A common problem faced by demographers is the estimation of the age-specific mortality pattern when data are given in age groups. Data can be provided in such a form usually because of systematic fluctuations caused by age heaping. This is a phenomenon usual to death registrations related to age misstatements, i.e. preferences of ages ending in multiples of zero and five. In this study we review, evaluate and compare the several methods for expanding an abridged life table to a complete one. In order to provide accurate evaluations and comparisons of the various methods, we apply the different techniques to several empirical data sets from different populations and different time periods.

ΠΕΡΙΛΗΨΗ

Βαγγέλης Πανούσης

Μέθοδοι Εξάπλωσης Συνεπτυγμένων Πινάκων Επιβίωσης: Εφαρμογές και Αξιολόγηση των Μεθόδων

Μάϊος 2001

Σύνηθες πρόβλημα για τους δημογράφους είναι η εκτίμηση της σειράς των κατά ηλικία πιθανοτήτων θνησιμότητας ενός πληθυσμού όταν τα δεδομένα που έχουν στη διάθεση τους αφορούν ομάδες ηλικιών. Σκοπός είναι η εκτίμηση ενός αναλυτικού πίνακα επιβίωσης του υπό-ανάλυση πληθυσμού από τον αντίστοιχο συνεπτυγμένο πίνακα. Τα δεδομένα ενός πίνακα επιβίωσης δημοσιεύονται σε συνεπτυγμένη μορφή, διότι τα αντίστοιχα κατά ηλικία δεδομένα είναι επιφορτισμένα από συστηματικά σφάλματα. Τα σφάλματα τα οποία παρουσιάζονται στη μέτρηση των εμπειρικών δεδομένων, οφείλονται συνήθως στη συστηματική δήλωση κατά τη ληξιαρχική καταγραφή ενός θανάτου, ηλικιών που καταλήγουν σε ψηφία πολλαπλάσια του 5. Σαν αποτέλεσμα τέτοιων εσφαλμένων καταγραφών, η γνωστή καμπύλη της θνησιμότητας του πληθυσμού παρουσιάζει συστηματικά υψηλότερα επίπεδα θνησιμότητας σε αυτές της ηλικίες. Το φαινόμενο αυτό είναι γνωστό στη βιβλιογραφία σαν “age heaping”. Σ’αυτή την εργασία γίνεται μια προσπάθεια να παρουσιαστούν, να αξιολογηθούν και να συγκριθούν οι μέθοδοι που έχουν έως τώρα προταθεί στη βιβλιογραφία σαν λύσεις στην αντιμετώπιση του παραπάνω προβλήματος. Οι διάφορες μέθοδοι αξιολογούνται και συγκρίνονται με την εφαρμογή τους σε εμπειρικά δεδομένα θνησιμότητας διαφόρων πληθυσμών.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Mortality and fertility form the two general causes of the physical notion of a human population. Both causes are object of great attention and analysis by demographers. A demographer constructs a life table in order to describe the survivorship of a population as subjected to the risk of death. This table describes the effect of mortality. A mortality "law" is useful for describing mortality of a population and also it provides the basis for useful projections.

The problem of estimating the age specific mortality pattern, when the data are provided in age intervals has been extensively discussed in demographic, biostatistical, as well as in actuarial literature. The main reasons for providing data in an abridged form are related to the phenomenon of *"age heaping"*, caused by the age misstatements in data registration and also the unstable mortality probability estimates provided by insufficiently small samples. The most typical case of age-misstatements is that of the preference of ages ending in multiples of five. Such misstatements cause the appearance of age heaps. An example of graphically presenting such age preference (*age heaps*), is the one that follows. This is the example of the *observed* age-specific mortality of Greece for 1960. It is obvious (from Figure 1) that bars referring to ages ending in multiples of five, or at even numbers, tend to be larger than it is expected to be observed.

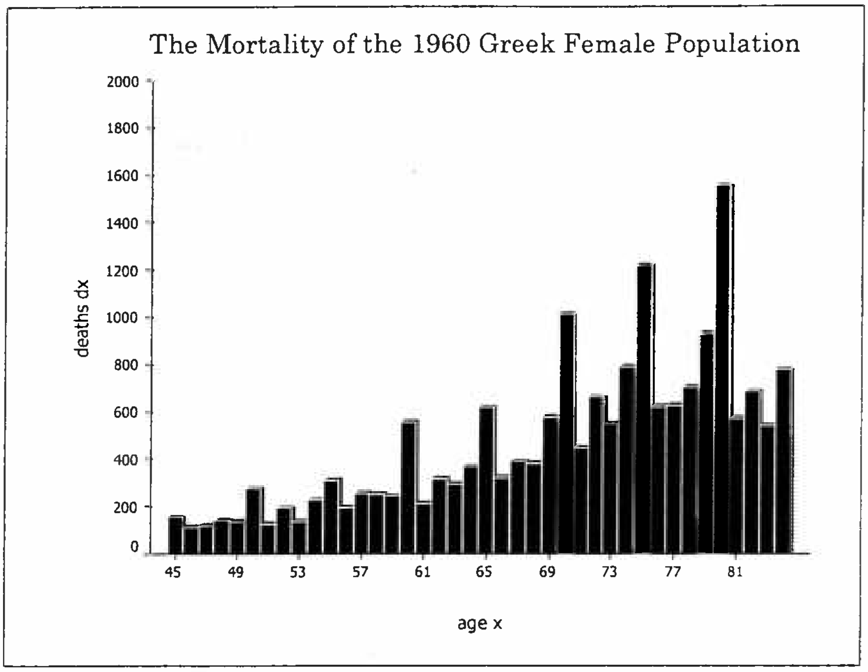

Figure 1: Observed death distribution by age x (from ages $x = 45$ to 84) for the Greek Female Population of 1960.

The observed mortality pattern surely contains great error which is enforced by the insufficiently small samples used in practice. A typical way for adjusting this errors is by grouping the data usually in quinquennial age groups, resulting to an *"abridged life table"*. An abridged life table does not provide information on the age specific mortality. We may estimate the last by expanding the table.

In this study we gather and analyze the several methods that have been appeared so far in the literature as tools of expanding an abridged life table. In Chapter 2 we refer to useful concepts and formulae of the theory of life tables. Sources and types of errors that appear in mortality data are referred in Chapter 3. The presentation of expanding methods starts at Chapter 4. We first make a brief presentation of the several methods starting by presenting an old method proposed by Reed, which was and is still used for the construction of the Greek Official life tables by the National Statistical Service of Greece, as it is described in detail by V.Valaoras (1984). A relatively new one which requires five-year age groups, presented by J.Pollard (1989) follows. We call as a *"parametric method"*, the application of parametric laws of mortality as expanding tools (see e.g., Kostaki, 1991, 1997). A non - parametric method, see Kostaki (2000), in the sense that it does not require the use of a parametric model, is also considered, which relates the target abridged life table with an existed complete one. The application of some interpolation formula to the survivors function l_x of the abridged life table is also presented. A conventional interpolation technique is the application of a six-point Lagrangean interpolation. A set of some six- point interpolation formulae is also presented, see Beers (1944). Spline interpolation (see e.g. Wegman and Wright (1983) or Hsieh (1991), e.t.c.) is a case of an osculatory interpolation technique which has lately received great attention. Because of the great literature referring to spline functions, we deal with them in a separate chapter. In Chapter 5 we start by providing some theoretical background on splines and present their involvement in interpolation of demographic data.

The efficiency of the several methods on empirical life tables of several countries (e.g. Sweden, Norway, Finland, Italy, New Zealand, e.t.c.) is evaluated in Chapter 6. A method comparison and some concluding remarks of this study are added there too.

Chapter 2

THEORY OF LIFE TABLES

A life table is a statistical model, designed essentially to measure mortality. In practise it is employed by a variety of specialists in a variety of ways. Life Tables are used by actuaries, vital statisticians and medical researchers to determine life insurance premiums, pension values, gains in life expectancy of a people and decreased probabilities of dying from improved medicines and surgical techniques. In its simplest form, the entire table is generated from age-specific mortality rates (m_x) , resulting to a set of useful functions, which in general determine mortality, survivorship and life expectation. Life tables are in a sense, one form of combining mortality rates of a population at different ages into a single statistical model.

2.1 Categories of Life Tables

Life Tables are distinguished in two general categories.

A. Cohort Life Tables.

B. Period Life Tables.

In general the existence of a number of l_0 live births is assumed . This is of a *"closed"* nature (a closed population).

A demographer will observe how this cohort of $l_0$persons diminutes as it grows older.

The last table presents the age pattern of mortality of this birth cohort.

The other case, which is more realistic from the point of the ability to construct it, is a *Period Life Table*. Since the first case may take 100 to 110 years or more to be constructed, we do an approximation by case B.

This is a more practical solution, since it can give as a table for every year. Life Tables are also distinguished in two categories:

A. Full (Complete) Life Tables

They present the age - specific mortality experience by single age x.

B. Abridged Life Tables

They present mortality for groups of ages. The usual representation is for age 0 separate and for the groups 1-4, 5-9, 10-14, ..., 85+.

2.2 The Construction of an Abridged and a Full Life Table

We start from l_0=100.000, or 10.000, which is the *hypothetical* cohort. It is referred, as the *root* (radix) of the table.

The basic life table function is l_x, $0 \leq x \leq w$ (age where the cohort extincts), which describes the survivors of death exposure at the exact age x. Usually from the death registrations, we get the central death rate ${}_nm_x$ for any value x in $[x, x+n)$. From that we can approximate q_x, the probability of dying, which is the only information we need to know in order to construct our life table.

The *Abridged life table* formulae are introduced here.It is simple to derive from the following, the expressions corresponding to a full life table. It only requires to equate n with *one (=1)*.

${}_nq_x$: this is the probability of someone of age x to die before reaching age $x+n$, i.e. to die in the age interval $[x, x+n)$

${}_np_x$: this is the probability of someone of age x to survive through the interval $[x, x+$

n).

With,

$$ {}_n p_x = 1 - {}_n q_x $$

l_x: this is the number of people surviving at exact age x.

With,

$$ S(x) = \frac{l_x}{l_0} $$

denoting the survival probability for the age interval $(0, x]$.

${}_n d_x$: this is the number of deaths for the age interval $[x, x+n)$.

With,

$$ {}_n d_x = l_x - l_{x+n} $$

or,

$$ {}_n d_x = l_x \, {}_n q_x $$

${}_n L_x$: this the total number of years of life, that the l_x-people of the population experience in the age interval $[x, x+n)$.

Each person that survives through the age interval $[x, x+n)$, contributes n-years of life, and each one that dies (since a uniform distribution of deaths is assumed) contributes approximately $n/2$ of years of life.

Then,

$$ {}_n L_x = n l_{x+1} + \frac{1}{2} d_x $$

and since,

$$ {}_n d_x = l_x - l_{x+1} $$

then,

$$ {}_n L_x = \frac{1}{2}(l_x + l_{x+1}) $$

In the continuous case, we will have,

$$_nL_x = \int_x^{x+n} l(t)dt = \int_0^n l(x+t)dt \simeq n(\frac{l_x + l_{x+1}}{l_x})$$

T_x: this is the total number of years, that the l_x-people of the population are about to live in the interval $[x, w)$, where w is an age difficult to reach ($w-1$, the greater age).

With,

$$T_x = \sum_{i \geq x} L_i$$

also,

$$T_{x+n} = T_x - \ _nL_x$$

and,

$$_nL_x = T_x - T_{x+n}$$

And in the continuous case,

$$T_x = \int_0^{w-x} l(x+t)dt$$

$\overset{0}{e}_x$: is called expectation of life or life expectancy at age x. It is the expected remaining life of persons of age x.

With,

$$\overset{0}{e}_x = \frac{T_x}{l_x}$$

Then the expected age at death of a person of age x is equal to, $x + \overset{0}{e}_x$

And in the continuous case,

$$\overset{0}{e}_x = \frac{1}{l_x} \int_0^{w-x} l(x+t)dt$$

2.3 Approximations of Life Table functions.

We present here some useful relations between mortality measures ${}_nq_x$ and, ${}_nm_x$. It is usual in practise to be unable to obtain ${}_nq_x$ directly by the empirical data, but one can approximate it by ${}_nm_x$.

2.3.1 Approximations of ${}_nq_x$, ${}_nm_x$

Assume that ${}_na_x$ is the expected number of years that a random person of age x will be surviving in the age interval $(0 \prec_n a_x \prec x)$. Then,

$$ {}_nf_x = \frac{{}_na_x}{n}, \; 0 \leq_n f_x \leq 1 $$

is the expected percent of years lived by such person in the interval *[x,x+n)*. So,

1. n, the number of years that each person which survives through the age interval *[x,x+n)*, contributes to ${}_nL_x$. The number of l_{x+n} people will have total contribution of a number of nl_{x+n} years of life.

2. ${}_na_x$, the number of years that each person which dies in the age interval *[x,x+n)*, contributes to ${}_nL_x$. Then their total contribution will be a number of $({}_na_x).({}_nd_x)$ years of life.

and,

$$
\begin{aligned}
{}_nL_x &= nl_{x+n} +_n a_{xn} d_{xn} L_x = \\
&= n(l_x -_n d_x) + n_n f_{xn} d_x \Rightarrow \\
&\Rightarrow \; {}_nL_x = n[l_x -_n d_x(1 -_n f_x)]
\end{aligned}
$$

From the previous, the mortality ratio ${}_nm_x$, will take the following form,

$$ {}_nm_x = \frac{{}_nd_x}{{}_nL_x} \Rightarrow \; {}_nm_x = \frac{{}_nd_x}{n[l_x -_n d_x(1 -_n f_x)]} $$

But, since ${}_nd_x = l_x\, {}_nq_x$,

$$ {}_nm_x = \frac{l_x\ {}_nq_x}{nl_x[1 - {}_nq_x(1 - {}_nf_x)]} \quad {}_nm_x = \frac{{}_nd_x}{n[1 - {}_nq_x(1 -{}_n f_x)]} $$

Assuming now uniform distribution of deaths in each age interval we will have, ${}_na_x = \frac{n}{2}$ and, ${}_nf_x = \frac{1}{2}$ so,

$$ {}_nm_x = \frac{{}_nq_x}{n[1 -{}_n q_x \frac{1}{2}]} $$

or,

$$ {}_nq_x = \frac{{}_nm_x}{n[1 +{}_n m_x \frac{n}{2}]} $$

The age - specific ${}_nm_x$ and ${}_nq_x$ values can be derived from the above by setting *n=1*. Such approximations are very useful in the application of certain expanding methods (*see Pollard, 1989, Reed (see Valaoras, 1984, King, 1914*).

2.4 Force of Mortality μ_x (or Hazard Function).

Another very useful mortality function, which we do not take usually in a life table, is the force of mortality μ_x. We refer to a theoretical measure, denoted by μ_x, which is the limit of the age - specific mortality rate ${}_{\Delta x}m_x$, when $\Delta x \to 0$. It describes the intensity of dying for a random person of age x when age is *assumed to be continuous*. This is called *Force of Mortality or Intensity of Mortality or Hazard Function*. It will be seen later, that it is basic to the illustration of some certain expanding methods, (*see Pollard, 1989*). In general we have for μ_x,

$$ \mu_x = \lim_{\Delta x \to 0} {}_{\Delta x}m_x = \lim_{\Delta x \to 0} \frac{{}_{\Delta x}d_x}{{}_{\Delta x}L_x} = \lim_{\Delta x \to 0} \frac{l(x) - l(x + \Delta x)}{\int\limits_x^{x+\Delta x} l(t)dt} \Rightarrow $$

$$ \Rightarrow \quad \mu_x = \lim_{\Delta x \to 0} \frac{l(x) - l(x + \Delta x)}{l(x)(x + \Delta x - x)} = \lim_{\Delta x \to 0} \frac{l(x) - l(x + \Delta x)}{l(x)(\Delta x)} $$

from a function's derivative definition,

$$\mu_x = -\frac{dl(x)}{dx}\frac{1}{l(x)} = -\frac{d\ln l(x)}{dx} \Rightarrow -\int_0^x \mu(t)dt = \ln l(x) + k \Rightarrow \exp(-\int_0^x \mu(t)dt) = l(x)e^k$$

setting $\frac{1}{e^k} = c$,

$$l(x) = c\exp(-\int_0^x \mu(t)dt)$$

and for $x = 0$,

$$l(0) = c, l(x) = l(0)\exp(-\int_0^x \mu(t)dt)$$

Now, since ${}_np_x = \frac{l(x+n)}{l(x)}$, we have from the previous,

$${}_np_x = \exp(-\int_x^{x+n} \mu(t)dt) \Rightarrow 1 -_n p_x =_n q_x = 1 - \exp(-\int_x^{x+n} \mu(t)dt)$$

Assuming stable intense of mortality for the age interval [x,x+1), we get,

$$\int_x^{x+1} \mu(t)dt = \mu_x$$

and for $n = 1$,

$$p_x = \exp(-\mu_x) \Rightarrow \mu_x = -\ln p_x$$

This last expression is used in order to produce μ_x-values for each age x, to be included in a life table, see *Pollard (1989)*.

2.5 Standard(or Model) Life Table

Usually we use as a tool to the estimation procedure applied a table of reference (*see Chaper 4, section 4.6. The Non - Parametric Method*). This is also a life table which is not required to satisfy some certain conditions or to have some particular characteristics. If the last suffices then we will cal the table, a standard life table. A standard table will demonstrate a mortality pattern similar to that of the population we analyze.

Chapter 3

ERRORS IN CENSUSES AND VITAL STATISTICS

Despite the care taken to ensure quality of the demographic data collected by censuses or by vital registration systems, the final tabulation of the observed data will sometimes give obvious indication of errors.

The different types of errors, that appear in observed demographic data can be distinguished in three categories:

A. Sampling Errors

B. Errors of Coverage

C. Errors of Content

Demographic data usually arise from a census or a vital registration system, so we focus on cases of error B. and C.

We study then two categories of error:

A. Errors of Coverage

These are concerned with losses or double registrations when we talk about vital statistics registrations, e.g. number of deaths. Usually in death registrations, losses of registration dominate the existed error, e.g. young borns which die, and no registration appears on their death or birth, so losses occur on both death and birth registrations.

and,

B. Errors of Content

These concern the refuse of someone to answer or a wrong answer to be given to a registration certificate or a questionnaire(in a census).

The most typical type of error that arise, is the *misstatements of age.*

3.1 Misstatement of age

Misstatements of age observed in census, or in vital registrations can be classified as follows:

1. an overstatement of ages of very young children, to which reference has already been made.
2. a heaping at the ages of majority, that is at 21 years, particularly among males, and at 18 years for females.
3. a heaping at the ages ending in multiples of five, and at the even ages.
4. an overstatement of ages before majority, and an understatement of ages after majority.
5. an overstatement of ages by persons just below 65 years.
6. an overstatement of ages toward the extreme life.
7. ages not stated or reported as unknown.

3.1.1 Heaping at the ages ending in multiples of five, and at the even ages

We adopt again the graphical representation of a Greek mortality experience that was presented previously in the Introduction. It is observed that heaps appear at ages ending in multiples of five and at even ages. These characteristics of the graph are undoubtedly the result of misstatement of age, for their regularity, they could hardly have been produced by corresponding periodic fluctuations in births in past years, or by systematic waves of immigration.

3.2 Measurement of the tendency to round - off age returns

Roberto Bachi (1951,1954) suggests three measures for which he considers the *exact*, *approximate* and *"blended"* solutions. These and various others can furnish measures of the amount of misstatements of age at three levels,

1. at each age (*e.g.* preference for age 50 or dislike for age 51),
2. at all ages with the same unit digit (*e.g.* preference for all ages ending in 0),
3. at all ages together (showing the proportions of persons in the population with any inaccurate digit).

Bachi (1954) proposes measurement methods for the cases *2,3*.

3.2.1 Measures

Let $x = 10v + u$ to be any value of age x, where v equals the tens digit, and u the units digit. So, *e.g.* if $u = 5$ then x may take values of ages ending to five. Additionally r_{10v+u} equals the number of people *returning* certain age x, and $\sum_v r_{10v+u}$, the total number of people returning ages with unit digit u. In the same sense t_{10v+u} , and $\sum_v t_{10v+u}$, define the *true* number of people with certain age x and ages ending to u.

1. Relative excess or deficiency of number of persons returned at age x over the true number of persons aged x.

$$\frac{r_x - t_x}{t_x}$$

2. Relative excess or deficiency of number of persons returned at all ages with the same unit digit, over the true number in those ages.

$$\frac{\sum_v r_{10v+u} - \sum_v t_{10v+u}}{\sum_v t_{10v+u}}$$

3. Synthetic measure of total consequences of preferences or dislikes for certain unit digits.

$$\frac{\sum_{+}(\sum_v r_{10v+u} - \sum_v t_{10v+u})}{N}$$

r_i and t_i where $i = x$, or $10v+u$, are the returned and true person counts respectively.

Because the different scope of this study, the reader interested in details may consult *Bachi (1951,1954)*.

3.3 Adjustments of Errors (Abridging the data)

Spiegelman (1976), Keyfitz (1985), Benjamin and Pollard (1980) make reference to some adjustments of the observed errors. Common to all is the suggestion of grouping the data. Since there is no way of ascertaining the true ages of those who have contributed towards the heaping at selected digits, the adjustment for this type of errors can hardly be unique. The general approach to adjusting has been to form groups of the numbers at successive individual ages. These grouped data (**abridged data**), rather than the individual age data, are then used for the analysis. The usual grouping is the one of five year age groups, so chosen that the grouped data would be free in some extent of error. Several tests have been devised to ascertain which age grouping minimizes greater the observed error. These do not point to any particular age grouping. It is anyhow followed, to present first age (age at birth) alone, with the age interval 1-4, and five year age groups (*e.g. 5-9, 10-14 e.t.c.*) to follow, until the extreme ages like 75 or 85, for which an open age interval is preferred because of the greater error observed there. This is the grouping adopted in data collection by several services e.g., the *World Health Organization*, the *National Statistical Service of Greece*. The demographer interested in the age-specific data solves this problem by applying an expanding method to the abridged data. He estimates, thus the unobserved age -specific experience by reference to an abridged one. The various methods that so far the bibliography suggests are described in the following chapters.

Chapter 4

EXPANDING ABRIDGED LIFE TABLES

4.1 Methods - A Brief Overview.

Several methods have been suggested in the literature for the estimation of the age specific mortality pattern from grouped data. A suggested solution is the application of some interpolation formula or a graduation procedure to the observed data. Since data contain great *"systematic"* fluctuations except of *"random"* ones the above solution is not preferred. Demographers as described in the previous chapter abridge the data, in order to eliminate systematic errors and produce the unobserved but real grouped data. From the last they try to estimate the age - specific data.

We consider here methods that have appeared so far in the literature as tools of expanding an abridged life table to a complete one. Valaoras (1984) presents in detail an old method which was initially presented by Reed (any reference , of previous publication of this method is not given in Valaoras, 1984) . A relatively new one which requires five-year age groups is presented by J.Pollard (1989). A method in which a parametric model of mortality is utilized (is presented by Kostaki ,1991 and Helingman&Pollard, 1980) it is called here as the parametric method. In our applications the eight-parameter Heligman-

Pollard formula and a nine - parameter version of it (see Kostaki, 1992) are adopted. A short reference on a Bayesian version of the HP-8 formula (see, Dellaportas, Smith and Stavropoulos (1997)) it is also done. A new method presented by Kostaki (2000), which is a non parametric one, in the sense that it does not require the use of a parametric model, is also considered. It is a method, which relates the target abridged life table with a standard complete one. The rest of the methods presented, are the application of some interpolation formula to the survivors function l_x of the abridged life table. A conventional interpolation technique used is the application of a six-point Lagrangean interpolation. A set of some other six- point interpolation formulas is also presented by Beers (1944). King (1914) presented a method of separate interpolation on ${}_nd_x$ and ${}_nE_x$ values of the abridged life table. Spline interpolation (see e.g. Wegman and Wright (1983) or Hsieh (1991)) is a case of an osculatory interpolation technique which has lately received great attention. Because of the great literature referring to spline functions, we deal with them in a separate chapter.

4.2 The Parametric Method-Parametric Models of Mortality as Expanding Tools.

4.2.1 Parametric Models of Mortality

Mathematical descriptions of schedules of demographic rates (e.g. mortality rates), called parametric models (of mortality-for mortality schedules-), offer usually an efficient means of condensing the amount of information to be specified as a set of assumptions, with the last imposed by a set of parameters and functions. The information required for a parametric model adopted to represent a mortality schedule is included in a *"Life Table"*.

Parameterised model mortality schedules describe the remarkably persistent regularities in age pattern that are exchibited by a set of empirical age - specific mortality rates. A mortality schedule will normally exchibit a great shift of mortality following birth age, a sudden drop (ages 10 to 15), followed by another sudden shift, known as *"accident hump"* (for ages 20 to 40). Finally after recalling to a minimum level, it will rise at an increasing pace until the last years of life.

The search for a "mortality law" has occupied the attention of statisticians and demographers for over a century. The attempt of representing mortality via a parametric model starts by Gompertz (see, Pollard (1991)). This is the well known Gompertz law of mortality initially proposed for modelling mortality at the elderly, but also adopted for the earlier adult ages (see, Pollard (1989)). The earliest attempt to represent mortality at all ages is that of Thiele (1872), (see, Kostaki (1999b)) who combined three functions each one representing a different part of the mortality schedule.

In the same sense Heligman&Pollard (1980) set out analogously a function of mortality, as represented by the odds of mortality $\frac{q_x}{p_x}$ at age x as an eight parameter formula of age x. This followed to be the classical eight - parameter HP model (HP8) for representing the age pattern of mortality. The problem that demographers usually phase is that of reproducing the age - specific pattern from incomplete or grouped data. Kostaki (1991) solves this problem by proposing the use of an adequate parametric model.

The rationale of the method is simple. If there is a model that adequately represents mortality of a population, then this model can also estimate in an adequate way the *"complete"* mortality experience, when an abridged one is only available.

There is no particular choice of a parametric model. Its adequacy of representing mortality is only required. Kostaki (1991) presents how such model choice can form an efficient solution for estimating age - specific mortality schedules. She introduces the method by adopting the application of the eight - parameter HP model (HP8), but another adequate model is also permissible. Kostaki (1997) later developed the nine - parameter version of the HP model (HP9) as a tool of the method.

The Expanding Procedure

We have our model for the mortality experience

$$\frac{q_x}{p_x} = F(x; \Theta)$$

where $F(x; \Theta)$ the right hand side of the equation with Θ being the vector of the parameters of the model, $\Theta = (A, B, ..., H)$.

From the model we get for the one-year odds of dying, and as concluded for the one year probabilities of dying,

$$q_x = \frac{F(x; \Theta)}{1 + F(x; \Theta)} = G(x; \Theta)$$

and the relation

$$_nq_x = 1 - \prod_{i=0}^{n-1}(1 - q_{x+i})$$

implies the following model for the death probabilities in the abridged life table

$$_nq_x = 1 - \prod_{i=0}^{n-1}(1 - G(x + i; \Theta)) =_n G(x; \Theta)$$

where we consider $_nG(x; \Theta)$ as an explicit but complicated function of Θ, x, n.

Then given the abridged (grouped) mortality experience one starts by estimating the parameters contained in vector Θ in least squares sense by minimizing,

$$\sum_x (\frac{\widehat{q_x}}{q_x} - 1)^2$$

Where the summation is over all relevant values of x. Get the estimates of the parameters as they come from this minimization procedure and insert them to the mortality formula. The estimated model may produce now a complete life table.

4.2.2 The Eight - parameter H&P Formula as an Expanding Tool (HP8)

A recent attempt to represent mortality over the course of the entire life span with a single analytical expression, has been made by Heligman&Pollard (1980). They propose a non - linear model of eight - parameters relating the odds of dying with age x. The rationale behind the model is simple. Basicely the model is distinguished in three parts representing, three classes of death causes. So, three causes can be seen, namely those affecting childhood, early and middle adult life, and the old age.

The formula utilizes a model that presents the odds of mortality $\frac{q_x}{p_x}$ as a parametric function of age x according to the formula,

$$\frac{q_x}{p_x} = A^{(x+B)^C} + D\exp(-E(\ln(\frac{x}{F}))^2) + GH^x$$

The idea is that we can decompose the odds that an individual of age x dies before he or she reaches age $x+1$ into three parts: a child mortality curve, an accident hump in early adult life, and an adult mortality curve.

q_x , is the modelled probability of an individual of age x to die before reaching age x+1. The modelled quantity, as referred previously, are the odds of dying than not dying before age x+1 , when the individual of study is of age x.

The three causes of death are described by the three terms that appear into the model respectively.

1. first a rapidly decreasing exponential, reflects the fall in mortality at the infant and early childhood ages. This term has three parts, A, which is nearly equal to q_1, measures the level of mortality, B, is ana age displacement to account for infant mortality, and C, measures the rate of mortality decline in childhood

2. a function similar to the log-normal density, reflects the middle life mortality pattern. It reflects the accident mortality for males and the accident plus maternal

mortality for the females. The accident appears generally between ages 10 and 40 and has three parameters, F indicating location, E representing spread, D the severity, and,

3. an exponential term as that of Gompertz seems adequate to model mortality at the later ages i.e. > 40 years of age.

The parameter G represents the base level of senescent mortality while H reflects the rate of increase of that mortality.

The right hand side of the mathematical formula H&P (1980) suggest is interpreted as $A^{B^C} + G$ for $x = 0$. H&P (1980) gave the above interpretation of the parameters which are all of positive value and are estimated by least squares. We seek for those parameter values minimizing a sum of squares, usually the following which has been applied to applications of the model referred later,

$$\sum_x \left(\frac{\hat{q}_x}{q_x} - 1\right)^2$$

where $\hat{q}_x$ the estimated probability and q_x the empirical observed probability value.

The quantity to minimize has been initially proposed by H&P in 1980. Anyhow attempting to fit many parameters simultaneously by a nonlinear algorithm, as it is demonstrated by practise can lead to inefficient model solutions (fits). HP8 has provided satisfactory representations of a variety of mortality patterns. Recent experience with fitting this model, like Congdon (1993), Forfar and Smith (1987), and Rogers (1986), suggest that the model is in certain cases overparameterised. Such cases usually suggest that the mortality pattern does not exhibit an accident hump, or a severe accident hump. In such a case with eight parameters to estimate, very similar distributions can be obtained with different combinations of parameter values (parameters that refer to the second part of the model are not required).

Kostaki (1991) proposes this model as a good solution to an expanding problem, since it provides, what is called an adequate representation of the unknown age - specific mortality pattern. United Nations (1988a, 1988b), independently developed this model application, within the MORTPAK and MORTPAK - LITE software for mortality analysis.

Several applications of this formula on a wide variety of mortality experiences e.g. Australia (Heligman and Pollard, 1980), and also Sweden by the same authors, have shown that the model provides a quite satisfactory representation of the age pattern of mortality. Kostaki (1985) also applied this formula to Swedish and Greek national mortality schedules concluding also to a successful model performance.

4.2.3 The Nine - parameter H&P Formula as an Expanding tool (HP9)

In general, HP8 is an adequate solution to the expanding of an abridged life table. Anyway, it fails to reproduce correctly the accident hump, since it estimates its beginning at a later age. That is more obvious in the cases of data where a severe accident hump exists (see e.g Swedish data).

In such cases, a nine - parameter version of the H&P formula which is introduced by Kostaki (1991) will perform better, since it improves the estimates provided by HP8 at that part of the curve. The HP9 model makes the model more flexible in the accident hump by modifying the middle model term. The suggested formula is:

$$\frac{q_x}{p_x} = A^{(x+B)^C} + D\exp(-E_i(\ln(\frac{x}{F}))^2) + GH^x$$

$i = 1, 2$, with $i = 1$ for $x \leq F$ and $i = 2$ for $x > F$.

The new parameters are the E_1, E_2 terms related to the spread of the accident hump at the left and right respectively.

Model is estimated by nonlinear least squares and the fit is again connected to the problem of overparameterisation...

Kostaki (1992) presents the HP9 as a tool for expanding an abridged life table and also as a solution to incomplete data problems. It is simple to define the method since it is an application of the originally proposed parametric also by Kostaki (1991), where now as the model $F(x;\Theta)$ we use HP9 (vector Θ is now of length nine).

4.2.4 Bayesian HP8 formula

Dellaportas, et. al. (1997) adopt a bayesian inference approach to the eight - parameter H&P (HP8) fit which according to the authors, it has several advantages compared to the classical solution. Firstly, it resolves the problem of overparameterisation (see Chapter 6), very usual to a least squares fitting of a model, by the use of an informative prior distribution. Secondly, the non normality of the likelihood surface in the parameterization usually adopted means that the least squares estimates are inadequate. Thirdly, it is applied as an expanding tool by routinely applying a simulation - based bayesian computation methodology. Here we carry out a small reference from a theoretical only view point, since this thesis extends only to solutions of classical statistics.

The Expanding Procedure

From a bayesian perspective, the abridged life table problem can be seen as an incomplete data problem, or as adopted by the authors as a constrained parameter problem. A general approach using an MCMC strategy is used, see Gelfang et al. (1992). As for applying graduation procedures to statistical data (e.g., demographic data) using a bayesian estimation strategy, see Carlin (1992).

In general, we assume that the eight - parameter H&P (HP8) model describes the true underlying the data age - specific pattern (the one year $q_x's$). We also consider as $\Theta = (A, B, C, D, E, F, G, H)$ the parameter vector, and by d_x the age-specific death counts which is binomially distributed with par's E_x and q_x. The only known non-random quantities are the grouped population counts (the exposed to the risk of death for each group) E_x, and the population with age 0 at last birthday and the only observed data the grouped death counts, ${}_nd_x$. Quantities as Θ, $d_x^{\prime}s$ (,or $q_x^{\prime}s$) are considered as unknowns.

Let ${}_nE_x$, ${}_nd_x$, E_x, d_x, be the vectors of the previous. E_x, d_x, do not contain the values for age $x = 0$. So, the full model for the data and the unknowns given the known quantities is of the form,

$$\begin{aligned} p({}_nd_x, d_x, d_0, E_x, \Theta / {}_nE_x, E_0) &= p({}_nd_x / d_x, E_x, \Theta,_n E_x, E_0, d_0) p(d_x / E_x, \Theta,_n E_x, E_0, d_0) \\ &\quad p(d_0 / E_x, \Theta,_n E_x, E_0) p(E_x, \Theta / {}_nE_x, E_0) \end{aligned}$$

We note that ${}_nd_x$ depends only on d_x, since

$$ {}_nd_x = \sum_{i=x}^{x+n} d_i $$

so for the first conditional density we get a product of indicator functions, ranging over $x = 1, 5, 10, \ldots, w$ (where w is the age where a generation extincts, as already noted in previous chapters).

$$ p({}_nd_x / d_x, E_x, \Theta,_n E_x, E_0, d_0) = \prod p({}_nd_x / d_x, \ldots, d_{x+k}) $$

When E_x, Θ are known the next conditional density is derived where d_x depends only on the previous, with each components being independent binomial distributions.

$$ p(d_x / E_x, \Theta,_n E_x, E_0, d_0) = p(d_x / E_x, \Theta) = \prod_{x=1}^{n} p(d_x / E_x, q_x), $$

where q_x the HP8 one year estimated probabilities.

Similarly,

$$ p(d_0 / E_x, \Theta,_n E_x, E_0) = p(d_0 / E_0, q_0) $$

q_0 the first year HP8 estimate.

And finally, we consider *a-priori* that Θ is independent of anything else, and that E_x

given ${}_nE_x$ is independent of Θ.

$$p(E_x, \Theta / {}_nE_x, E_0) = p(\Theta) p(E_x / {}_nE_x, E_0)$$

$p(\Theta)$ is the prior distribution for ϑ, the HP8 model parameters.

$p(E_x / {}_nE_x, E_0)$ is the prior for the exposed to the risk of death population counts.

For a greater detail someone may consult Dellaportas, et. al. (1997).

4.3 An Additional Adjustment applied to the results of the Parametric Method

A good expanding method must imply that if we abridge the estimated one - year probabilities, they will be equal to the original abridged. Since the estimates that the H&P model provides, or in general the estimates of a parametric model do not have that property, we perform an adjustment.

Theoretically, the expanded one-year probabilities of dying constructed by our technique or any expanding technique, it is expected to have corresponding n - year probabilities which have values close to the original abridged life table values. In practice, usually this is not observed. We may change the one - year probabilities by values that satisfy the desired property. Kostaki (1991) proposed the following, simple solution.

$$\hat{q}'_{x+i} = 1 - (1 - \hat{q}_x)^K$$

where K,

$$K = \frac{\ln(1 - {}_n q_x)}{\sum_{i=0}^{n-1} \ln(1 - \hat{q}_{x+i})}$$

and $\bar{q}_x$ the new series of probabilities.

We conclude to that using the following sense.

Since the n-year probabilities ${}_n\widehat{q}_x$, which correspond to the Heligman & Pollard formula one-year probabilities ${}_1\widehat{q}_x$do not satisfy the following equation,

$$ {}_n\widehat{q}_x = 1 - \prod_{i=0}^{n-1}(1 - \widehat{q}_{x+i}) \approx_n q_x $$

then we produce some $\widehat{q}_x'$values with corresponding n-year probabilities the adjusted Heligman & Pollard probabilities such that they satisfy the previous equation.

So we have for ${}_n\widehat{q}_x'$,

$$ {}_n\widehat{q}_x' = 1 - \prod_{i=0}^{n-1}(1 - \widehat{q}_x') =_n q_x $$

It is easy to explain the rationale behind this particular choice of adjustment. It amounts to assuming that the force of mortality $\mu'(x)$ underlying the original abridged life table is, in each n-year interval $[x, x+n)$, a constant multiple, say $K\mu(.)$, of one say $\mu(.)$, of those infinitely many forces of mortality which produce the complete life table obtained by the applied expanding process.

The additional adjustment presented by Kostaki (1991), was proposed as a tool that can be applied to the results of her technique. However it can be applied to the results of expanding technique (see e.g. Appendix C where we apply it on the results of Reed's technique).

4.4 The Pollard's Model

The usual representation of abridged data is the one with groups of age $x = 0, 1-4, 5-9, ..., 75+$,or 85+. Pollard (1989) proposes a *method* for deriving a full life table from mortality data given for each single year until the age of 5 and then for five - year age groups. The usual form of grouping is close to the one the method requires.

To describe the method we also assume that deaths during a given calendar year are reported according to age x of last birthday at time of death. Because of non-uniform exposure of deaths over the 5-year age group, ${}_5\widehat{m}_x$ as produced by the mean population over an age interval will provide a biased estimate of the mortality rate. To overcome this problem two assumptions are adopted :

Exponential variation of the force of mortality μ_x(at age **x**) within the age interval and that the population remains stable within this age interval.

Pollard's method, adopts the use of the classical Gompertz's law of mortality (Benjamin and Pollard (1980)), for degrouping mortality in the whole life span.

Gompertz's law assumes the above described exponential variation of μ_x *(at x) within* an age interval, i.e.

$$\mu_x = Bc^x$$

This was usually applied only to increasing mortality at adult ages, now it is revised as a tool for describing the decreasing mortality of younger age groups too. In the last case we assume that $c < 1$.

4.4.1 The Procedure

Having values of deaths and central populations for the ages 1, 2, 3 and 4, allow the immediate estimation of $\mu(1.5), \mu(2.5), \mu(3.5), \mu(4.5)$. Values of $\mu(x)$ at the pivotal ages 7.5, 12.5, ..., 82.5 are provided by the ratios , $\overset{o}{e}_5$, ${}_5\overset{o}{m}_{10}, \ldots, {}_5\overset{o}{m}_{80}$, without adjustment at the younger ages, and after adjustment according to formula,

$$\mu(x+2.5) = \frac{{}_5\,\overset{o}{m}_x}{\left[1 - \frac{25}{12}({}_5\overset{o}{m}_x + r)\ln c + \frac{25}{24}(\ln c)^2\right]}$$

for ages as above, say, 30. In order to have values of $m(x)$ at intervening ages we may apply *interpolation* (described later). If no smoothing or graduation is required then, linear interpolation on $\ln \mu_x$ is recommended. Also for ages $x < 7.5$ we may apply an *extrapolation*. Then **interpolation** and **extrapolation** appear to be the main tools of this method. Since μ_x implies of a theoretical and not an empirical valued measure, any value given refers to this function's approximation.

To describe the construction of the *adjustment formula*, we outline the following.

We start from the assumption of exponential variation in the μ_x.

- Exponential variation of μ_x

The survival probability from age x to $x + t$ can take the form,

$$_tp_x = \exp\left\{-\int_0^t \mu_{x+u} du\right\}$$

and after some calculations we deduce to,

$$_tp_x = \exp\left[\frac{t(\mu_x - \mu_{x+t})}{\ln(\mu_{x+t}/\mu_x)}\right]$$

Which is an exact formula when μ_x varies exponentially.

Concentrating now on the application on the adult ages, we get by $c^t = e^{\ln c}$, and $\ln c$ approximated by 0.09, that,

$$_tp_x \doteq \exp\left\{-\mu_x\left[t + \frac{1}{2}t^2 \ln c\right]\right\}$$

The last assumes *quinquennial age groups.*

Now under the stability assumption over the age range $(x, x+5)$ with intrinsic growth rate r, the number of persons in the age interval may be written,

$$P(x, x+5) = K \int_{-2.5}^{2.5} e_x^{-rt} p_{x+2.5} dt$$

and under μ_x exponentially varying we may write,

$$P(x, x+5) \doteqdot 5K\left[1 - \frac{25}{24}\mu_{x+2.5}\ln c + \frac{25}{12}(\mu_{x+2.5} + r)^2\right] \tag{4.1}$$

and,

$$\begin{aligned} D(x, x+5) &= K\int_{-2.5}^{2.5} e_t^{-rt} p_{x+2.5}\mu_{x+2.5+t}dt \doteqdot \\ & 5K\mu_{x+2.5}\left[1 - \frac{25}{24}\mu_{x+2.5}\ln c + \frac{25}{12}(\mu_{x+2.5} + r - \ln c)^2\right] \end{aligned} \tag{4.2}$$

Writing ${}_5\overset{o}{m}_x$ for $D(x, x+5)/P(x, x+5)$, dividing the two above formulae and replacing $\mu_{x+2.5}$ in the correction term by ${}_5\overset{o}{m}_x$, we conclude to the adjustment formula.

4.5 The Reed's Model

A simple tool for expanding was also introduced by *Reed* (see Valaoras, 1984). This method's main tool is the five - year mortality rate.

$${}_5m_x = \frac{{}_5d_x}{{}_5M_x}$$

where ${}_5d_x$ represents the number of deaths in the five year age interval $[x, x+5)$ and ${}_5M_x$ the mean population at the same age interval.

This method was adopted in the construction of some Greek Life Tables by Valaoras (1984) who presented an analytical description of this method and it's performance on Greek data.

4.5.1 The Procedure

We have the five - year mortality rates, ${}_5m_x$. Then the one - year probabilities of dying q_7, q_{12}, q_{17} , ... are calculated, using the following approximate formula,

$$q_{x+2} = \frac{2{}_5m_x}{2 +_5 m_x}$$

In order to estimate the full series of the one year probabilities the following formula is then adopted,

$$\frac{q_x}{K^x} = a + bx + cx^2 + dx^3, \;\; x \geq 5$$

Using the already estimated one year probabilities ($q_7, q_{12}, q_{17}, \ldots$), the complete q_x-series is estimated by least squares.

In order to estimate the complete probability series the following procedure is used.

The formula is fitted twice. First, in order to produce the complete probability series for the age range $5 \leq x \leq 20$ it is fitted to $q_7, q_{12}, q_{17}, q_{22}$ values with $K = 0.989943$. Then for $x \geq 25$, the formula is fitted to using the $q_{22}, q_{27}, q_{32}, \ldots$ values with $K = 1.0251234$.

Finally, for the ages 21 to 24, a linear combination of the two fitted equations is used in order to estimate the one-year probabilities:

$$\begin{aligned} q_{21} &= 0.8q_{21}^{'} + 0.2q_{21}^{''} \\ q_{22} &= 0.6q_{22}^{'} + 0.4q_{22}^{''} \\ q_{23} &= 0.4q_{23}^{'} + 0.6q_{23}^{''} \\ q_{24} &= 0.2q_{24}^{'} + 0.8q_{24}^{''} \end{aligned}$$

Where the first and second terms of the above equations are the fitted probability values for $K = 0.989943$ and $K = 1.0251234$ respectively. The two of K values were proposed by Reed.

4.5.2 Properties and Problems

This technique is not applicable to the early childhood ages ($x < 5$). Adequate as well as it does not produce successful results for the early adult ages (see Kostaki, 1992). However, it is effective for the adult ages.

4.6 The Non Parametric Method

Kostaki (1998) describes a non-parametric method in the sense it does not require the adaptation of a parametric model, adequate for describing the mortality pattern, like in Kostaki (1997) and (1991).

It is considered as a **relational technique** since it *relates* an abridged life table with another complete one used as a reference table. The *reference* table is not necessary to be a standard one (*see Chapter 2*). The performance of the method will always differ as the reference table changes in period or in the population it describes. Anyhow the differences, as application demonstrate, will be of slight extent (see Appendix C, Tables **C1a, C1b**).

4.6.1 The Procedure

We have the ${}_nq_x$ probabilities, elements of our abridged life table and the $q_x^{(R)}$of the reference (e.g. a standard table) life table, which is of complete form.

Under the assumption that the force of mortality μ_x, underlying the abridged life table is, in each n-year age interval $[x, x+n)$, a constant multiple of the one underlying the reference (complete) life table in the same interval $\mu_x^{(R)}$, i.e.,

$$\mu(x) = {}_n K_x \mu^{(R)}(x)$$

where,

$$K = \frac{\ln(1 -_n q_x)}{\sum_{i=0}^{n-1} \ln(1 - q_{x+i}^{(R)})}$$

Then the one year probabilities of dying $q_{x+i}, i = 0, 1, ..., n-1$ are in each n-year interval equal to

$$1 - (1 - q_{x+i}^{(R)})^{{}_nK_x}$$

Therefore from the ${}_nq_x$ and the ${}_1q_x^{(R)}$we calculate ${}_nK_x$ and then estimate the one - year probabilities, q_x underlying the target abridged life table.

4.7 Polynomial or Piecewise Polynomial Interpolation

4.7.1 Interpolation Techniques

Since mortality data grouped in single years (or even narrower) intervals usually are not available we study techniques of expanding the interval data (abridged data) to single year values, based on the application of *mathematical interpolation.*

A usual method is to fit to the interval data a single polynomial. We interpolate values of a function $f(x)$, where $x = age$, which we have tabulated and produce function's values at intervening ages.

Interpolation

The notion of interpolation is to estimate a non-tabulated value of a function from tabulated values. This method is useful when the corresponding function, here usually a survivors function l_x, is not mathematically specified (usual case in life tables, e.g. we have certain values on certain ages only).

4.7.2 Lagrangean interpolation of six terms (points).

The power of an interpolation formula is denoted by the number of terms that is consisted of. Six term formulae are most famous in the literature, that means formulae that comprise six successive data points or else six successive tabulated values of a function.

In general Lagrangean interpolation formula assumes that the fitted polynomial, the function of interest say $u(x)$ which is of degree k, is a linear combination of $k+1$ tabulated values of this function. When only two tabulated values are interpolated by another then $k+1 = 2$ and $k = 1$, the fitted polynomial is of first degree and the interpolation is called *linear*. With the same rationale we have cubic interpolation when the values to be interpolated are four and the polynomial of third degree. So,

$$u(x) = \sum_{i=1}^{k} \frac{\prod_{j\neq i}(x - x_j)}{\prod_{j\neq i}(x_i - x_j)} \cdot u(x_i)$$

Then if we have $x_1 = 1$, $x_2 = 3$, the formula takes the above expression with coefficients, -0.5 and 0.5 respectively for u_{x_1} and u_{x_3} when the interpolated value is u_{x_2}. These are the coefficients for the linear interpolation of u_{x_1} and u_{x_3} and the following is the above derived linear interpolation formula :

$$u(x) = \frac{(x - x_2)}{(x_1 - x_2)}u(x_1) + \frac{(x - x_1)}{(x_2 - x_1)}u(x_2)$$

It is interesting to observe that coefficients add up to one. That happens for every value that $k+1$ and then k take. It is justified by the fact that if all tabulated values have to be equal to a constant then the non tabulated ones have to be equal to the same constant also.

A conventional interpolating technique that is usually applied, is the six point Lagrangean interpolation method. It is very applicable, fast, and very simple to handle, since it does not require great computational skills by the researcher.

The six term Lagrangean interpolation formula applied on the existing values of the survivors function l_x expresses each non tabulated value as a linear combination of six particular polynomials in x, each of degree five. As Johnson & Johnson (1990) describe,

$$l(x) = \sum_{i=1}^{6} \frac{\prod_{j \neq i}(x - x_j)}{\prod_{j \neq i}(x_i - x_j)} \cdot l(x_i)$$

where $x_1, x_2, ..., x_6$ are the tabular ages nearest to x.

When the x_i's are equally spaced this formula can be expressed in simpler forms.

Elandt and Norman Johnson (1990), and Abramowitz, and Stegun (1972), tabulated the above equation's coefficients, which we also provide in the **Appendix C** (see **Table C4**) of this study. Lagrangean interpolation is a very famous case of interpolation, adopted by several authors, like *e.g.,* Namboodiri, Suchindran (1987), when conversation comes to expanding an abridged life table.

4.7.3 Other Six-Term Formulae for (Actuarial) interpolation.

The literature on interpolation techniques contains too many formulae from which few are used to real data applications. The most applied, is the one previously described, Lagrangean formula. Formulae that comprise six terms of the interpolated function as said before, are more famous.

Henry Beers (1944) gathers some of these and compares their performance. It is reasonable that all comprise cases of piecewise polynomial interpolation. The formulae briefly stated are:

1. The elementary fifth difference formula.

2. The curve-of-sines osculatory formula.

3. Sprague's fifth-difference osculatory formula.

4. Shovelton's tangential formula.

5. The minimized-fifth-difference formula.

6. Henderson's famous near-osculatory formula.

7. Jenkin's formula.

As $f(s)$ we present the function to be interpolated, where our usual choice is the survivors function (l_{x+s} or $l(x+s)$).

In general **six - point interpolation** formulae (*most of them here*) have *three common characteristics.*

They are *symmetric* in the sense that they give the same results whether they are applied forward or backward, i.e. $f(s)$ or $l(x+s)$, is the same function of $f(-2)$, $f(-1)$, $f(0)$, $f(1)$, $f(2)$ and $f(3)$ as $f(-s)$ is of $f(2)$, $f(1)$, $f(0)$, $f(-1)$, $f(-2)$ and $f(-3)$ with s positive and less but not equal to *1* , and with $f(s)$ denoting the value u_{x+s} (or, $l(x+s)$) of the function we interpolate. The rest $f(.)$ values take similar expressions. They are *correct up to four differences,* $\Delta^4 f$, i.e. when fifth differences are zero, they give the same results as the classical elementary fourth difference formula applied to the same given values. They are *correct to fifth differences on the average,* $\Delta^5 f$, i.e. the sum of the interpolated values in each age interval is equal to the sum computed from the same given values by the classical elementary fifth difference formula (*presented later*). This characteristic is usually a sufficient guaranty of the reliability of the results of interpolation performed by one of these formulae. The reader interested to mathematical details on those characteristics may follow the already referred work by, Beers (1944).

Formulae Mathematical Expressions

1. **The elementary fifth differences formula** is considered as the best formula when fifth differences are constant.

Formula:

$$l_{x+s} = f(s) = \frac{(s+2)(s+1)s(s-1)(s-2)}{120}$$

1. The **curve of sines osculatory formula** was used in the course of the construction from population statistics of some of the early English life tables. It is called "osculatory" because first and second derivatives of consecutive curves are equal at points of junction. The term osculatory will be considered in detail. later, when the category of "osculatory " interpolation techniques will be described.

Formula:

$$l_{x+s} = f(s) = \frac{(s+1)s(s-1)(s-2)(1-\cos\pi s)}{48}$$

1. **Sprague's fifth difference osculatory** formula has the same first and second derivatives at each point of junction as the fourth degree curve through that point and the two given values on each side of it.

Formula:

$$l_{x+s} = f(s) = s^3 \frac{(1-2)(7-5s)}{24}$$

1. **Shovelton's tangential formula** is characterized like that, as tangential, because it makes the first derivatives (but not the second) of consecutive curves equal at points of junction.

Formula:

$$l_{x+s} = f(s) = s^2 \frac{(1-s)(5-s)}{48}$$

1. **The minimized fifth difference** formula was derived for the special purpose of minimizing the sum of squares of the fifth differences of the interpolated values. It has no algebraic expression for l_{x+s}, or $f(s)$.

2. **Henderson's** (famous, as given by the author) **near-osculatory formula** (presented in TASA,Transactions of the American Society of Actuaries,Vol. IX,219-20) is expressed as

Formula:

$$\begin{aligned} l_{x+s} &= f(s) = l_x + s\Delta l_x + \frac{s(s-1)}{2}(\Delta^2 l_{x-1} - \frac{1}{6}\Delta^4 l_{x-2}) \\ &+ \frac{(s-1)s(s+1)}{6}(\Delta^3 u_{x-1} - \frac{1}{6}\Delta^5 l_{x-2}) \end{aligned}$$

1. Jenkin's **formula** (presented in RAIA, the Records of the American Institute of Actuaries,Vol. XV, 89) is expressed as

Formula:

$$\begin{aligned} l_{x+s} &= f(s) = l_x + s\Delta l_x + \frac{s(s-1)}{2}\Delta^2 l_{x-1} \\ &+ \frac{(s+1)s(s-1)}{6}\Delta^3 l_{x-1} \\ &- \frac{s^3(s-1)}{12}\Delta^4 l_{x-1} \\ &+ \frac{s(s-1)^3}{12}\Delta^4 l_{x-2} \end{aligned}$$

Measures of comparison and formulae evaluation.

The author compares the smoothness of formulae 1 to 5 by calculating their fifth differences, $\Delta^5 f$. In general differences of a certain degree may be adopted as a comparison tool of two, or more interpolation formulae (*for formulae 1 to 7*). As Beers (1944) suggested, for routine actuarial interpolation it is highly desirable, or even necessary to have a formula that can be applied with some assurance without preliminary analysis of the particular characteristics of the differences of the given values of the function to be interpolated. Now, if fifth differences are negligible in size, there is no problem since formulae with that characteristic will provide almost the same results. Problems arise when the fifth differences of the given values are too large to neglect but too irregular to be plausible. Another criterion of the smoothness of an interpolation formula is required. Beers (1944), suggested the smallness of the sum of squares of the fifth differences of the interpolated values.

$$\sum \left(\Delta^5 l_x\right)^2,$$

where l_x the interpolated function.

Some of the formulae presented by Beers (1944) are also osculatory. That means, that they osculate at points of junction x. This is the case we deal within the next sections.

4.8 Osculatory Interpolation

The usual case of interpolation is the one called, *ordinary piecewise interpolation*. As it is described we fit polynomials to pieces of the data which we connect, one by one, at points called joins, and produce a continuous function. The problem is that derivatives at the joints are discontinuous. Osculatory interpolation techniques overcome this problem by ensuring that important derivatives (usually the first two) will be continuous to the whole range of values. In that case polynomials join smoothly or "kiss", hence the name "osculatory". Osculatory interpolation techniques introduce here, a method by *King* and the one of the application of *Spline Models.*

4.8.1 King's Method

This is the method of osculatory interpolation that was proposed by George King (1914), mainly in order to reduce the effects of age misstatements in mortality data. It was applied so to graduate mildly the data and produce some of the published English Life Tables. The method itself is less applicable and largely for historical interest. It requires the assumption of a small effect of age misstatement since it has a small graduating power. The original method is applied separately to the exposed to risk population ${}_nE_x$ and death counts ${}_nd_x$ although it can be applied directly to the mortality probabilities

${}_nq_x$, but only with the assumption of mild amount of error to suffice. If ${}_nE_x$, ${}_nd_x$ do not obtain a typical pattern, then they do not require just a mild graduation. In such case King's technique will apply better to the ${}_nq_x$ - values of the abridged life table. The probability of dying usually demonstrates a typical behavior.

The Method.

Originally the method alternates in the next five steps:

1. The Exposed to risk are grouped into quinquennial age groups.
2. A pivotal exposed to risk value is calculated for the central age of each group, using King's **pivotal value formula**.
3. Graduated Exposed to risk values at the remaining ages are found by osculatory interpolation, using **King's osculatory** formula.
4. Graduated deaths are obtained by applying steps 1 to 3 to the observed deaths.
5. Graduated mortality rates are found by division.

Next the method's two basic tools are presented via mathematical expressions.

A. King's pivotal value formula.

We consider a third degree polynomial u_x and define,

$$w_{-1} = [n]\, u_{-n};\, w_0 = [n]\, u_0;\, w_1 = [n]\, u_n$$

u_x is the value of our function for age x (e.g., ${}_nE_x, {}_nd_x$ or ${}_nq_x$).

$[n]$ is a summation operator and it means the *n-term* simple moving average applied on the series of values of u_x.

In general the method uses the notion of simple moving averages applied to a series of function values (here, ${}_nE_x, {}_nd_x$ or ${}_nq_x$)

The usual case is n=*5*, which implies the use of quinquennial age groups. So,

$$\begin{aligned} w_{-1} &= u_{-7} + u_{-6} + u_{-5} + u_{-4} + u_{-3} \\ w_0 &= u_{-2} + u_{-1} + u_0 + u_1 + u_2 \\ w_1 &= u_3 + u_4 + u_5 + u_6 + u_7 \end{aligned}$$

A formula for u_0 in terms of w_{-1}, w_0, w_1 is then considered,

$$w_0 = [n]\, u_0 = nu_0 + \frac{n(n^2-1)}{24}\Delta^2 u_{-1} \tag{4.3}$$

$$w_{-1} + w_0 + w_1 = [3n]\, u_0 = 3nu_0 + \frac{3n(9n^2-1)}{24}\Delta^2 u_{-1} \tag{4.4}$$

Where (4.3) is an alternative expression for a simple moving average using the Δ^i - operator (difference operator). The last denotes the *i-th differences* of a series of function values.

By substracting three times equation (4.3) from equation (4.4) we come up with the following,

$$\Delta^2 w_{-1} = n^3 \Delta^2 u_{-1}$$

Solve the above with respect to $\Delta^2 u_{-1}$ and substitute the result to equation (4.5). We finally solve with respect to u_0 and conclude to King's pivotal value formula,

$$\mathbf{u} = \frac{1}{n}\left\{ w_0 - \frac{(n^2-1)}{24n^2}\Delta^2 w_{-1} \right\} \tag{4.5}$$

B. King's osculatory interpolation formula.

In order to describe the formulae construction, assume as Benjamin & Pollard(1980) do in their work, that we have four successive values of a function $f(s)$ or u_x (the function we want to interpolate, e.g., E_x, d_x, q_x). The four points may be denoted as *A,B,C,D* with that their physical order.

The idea is to fit a quadratic through points A, B, C ,

$$u_x = (1+\Delta)^{x+1}u_{-1} = u_{-1} + (x+1)\Delta^2 u_{-1}$$

with slope at position B equal to,

$$\left[\frac{d}{d_x}u_x\right]_{x=0} = \Delta u_{-1} + \frac{1}{2}\Delta^2 u_{-1}$$

Plus a quadratic through B, C, D ,

$$u_x = (1+\Delta)^x u_0 = u_0 + x\Delta u_0 + \frac{1}{2}x(x+1)\Delta^2 u_0$$

with slope at point C equal to,

$$\left[\frac{d}{d_x}u_x\right]_{x=1} = \Delta u_0 + \frac{1}{2}\Delta^2 u_0 = \Delta u_{-1} + \frac{3}{2}\Delta^2 u_{-1} + \frac{1}{2}\Delta^3 u_{-1}$$

And a cubic through B, C:

$$ax^3 + bx^2 + cx + d$$

with gradient,

$$3ax^2 + 2bx + c$$

The pivotal values or the joins of the three polynomials are points B,C. As it is reasonable, it is unable to estimate a function of third degree using two values. We do that by imposing constraints on the slope values of the fitted polynomials.

By equating ordinates and gradients at the pivotal points B and C, we deduce that:

$$\begin{aligned} d &= u_0\,, && \text{(ordinate at B)} \\ c &= \Delta u_{-1} + \frac{1}{2}\Delta^2 u_{-1}, && \text{(gradient at B)} \\ u_1 &= a+b+c+d\,, && \text{(ordinate at C)} \end{aligned}$$

$$3a + 2b + c \quad = \quad \Delta u_{-1} + \frac{3}{2}\Delta^2 u_{-1} + \frac{1}{2}\Delta^3 u_{-1}\ , \qquad \text{(gradient at C)}$$

and by solving with respect to *a,b,c,d* we deduce the interpolating formula of King,

$$\mathbf{u}_x = \mathbf{u}_0 + \mathbf{x}\Delta\mathbf{u}_{-1} + \frac{x+x^2}{2}\Delta^2\mathbf{u}_{-1} + \frac{x^2-x^3}{2}\Delta^3\mathbf{u}_{-1} \tag{4.6}$$

Benjamin & Pollard(1980) describe all the above theoretical considerations via a numerical example.

4.9 Spline Functions

Piecewise polynomials that osculate at specific values known as *knots*, produce a polynomial function called *spline*. They present interesting properties when applied as a smoothing tool but also as an interpolation tool, which is the case analyzed here.

Chapter 5

EXPANDING AN ABRIDGED LIFE TABLE BY A SPLINE FUNCTION

There is some enormous literature on splines, most of it concerning their numerical analytic properties, see DeBoor (1966), Greville (1969),Wahba (1975), Fröberg (1969), Tyrtyshnikov (1997), rather than statistical properties, see e.g., Green, Silverman (1995), McNeil, Trussel and, Turner (1977), or Härdle (1992). Special interest has been shown lately to such osculatory polynomials on the course of demography. So we deal, here, with the concept of spline interpolation. Spline functions are introduced along with their properties and their use on interpolating demographic data, in specific mortality data, is studied in detail.

5.1 Theoretical Background

A spline, as Wahba (1975), and McCutcheon (1981) describe, symbolized by S, is a function derived after joining a sequence of polynomial arcs. Then it is a **piecewise polynomial function**, for which as later we explain the maximum possible number of derivatives exists. Suppose that $a = x_0 < x_1 < ... < x_{n+1} = b$, then a spline S of degree **k** defined on the interval $[a, b]$ with internal "*knots*" $x_1, x_2, ..., x_n$ is a function such that, if $0 \leq i \leq n$ and $x_i \leq x \leq x_{i+1}$, then $S(x) = p_i(x)$ where $p_i(x)$ is a polynomial in **x** of degree not greater than **k**. The polynomial arcs meet at pivotal values called "*knots*". The number of knots chosen must be less or equal the data points used. In interpolation problems the knot number equals the data.

Spline functions demonstrate substantial advantages against the traditional interpolation formulae, such as greater smoothness, and continuity of the greatest possible number of derivatives. The choice of the possible number of spline curves through the data points is infinite. An arbitrary choice of a spline function may produce unsatisfactory results.

Although splines are a fairly simple mathematical concept, their mathematical theory is relatively new. The idea behind the mathematical spline can be traced back to an old technique used by draftsmen. For many years long thin strips of flexible material have been used by draftsmen in the same manner as French curves to fair in a smooth curve between some specified points. These strips, or splines are anchored in place by attaching weights called "ducks" at points along the spline. The mathematical spline function is similar to the draftsman's spline in that its graph resembles the curve drawn by a mechanical curve.

Little work was done with splines until the 1960's. However, the literature has since proliferated rapidly and splines currently receive a great deal of attention in approximation theory (see, Barndorff-Ole and Cox, 1995). In fact these functions are considered as adequate to represent what is called "*structural change*". The use of splines in representing structural change is clearly motivated in a much different way that are the traditional uses of splines in approximation theory.

We called spline functions as piecewise polynomials. An example is the linear spline, which is a continuous piecewise linear polynomial function. The idea behind that is that there exists a linear relation, which has been subjected to possible structural change at the knots. Then knots indicate the points where our function changes it's typical behavior. We may generalize this when the underlying relation is not a linear one e.g. a cubic deriving a cubic spline (a spline of a third degree) which is presented later.

5.2 The Roughness Penalty Approach

Splines are related to a roughness penalty approach which Green and Silverman (1995) consider when the attempt is to fit a curve S to a set of data.

The roughness penalty approach to curve estimation is stated as follows:

Given any twice - differentiable function S defined on $[a, b]$, and a smoothing parameter a or $\lambda > 0$, define the modified sum of squares,

$$L(S) = \sum_{i=1}^{n} [Y_i - S(t_i)]^2 + a \int_a^b [S''(x)]^2 dx$$

The penalized least squares estimator $\widehat{S}$ is defined to be the minimizer of the functional $L(S)$ over the class of all twice - differentiable functions S. This minimizer $L(S)$ is the case of a cubic spline. This case we will discuss about later.

The addition of a roughness penalty term $\left[a \int S''^2\right]$ in the above sum of squares ensures that the cost $L(S)$ of a particular curve is determined not only by its goodness-of-fit to the data as quantified by the residual sum of squares $\sum_{i=1}^{n}[Y_i - S(t_i)]^2$ but also by its roughness $\int S''^2$. The term a is what we call a smoothing parameter. That quantifies the fitted curves roughness. It represents the rate of exchange between residual error and local variation and gives the amount in terms of summed square residual error that corresponds to one unit of integrated squared second derivative. If a is large then the main component in $L(S)$ will be the roughness penalty and hence the minimizer $\widehat{S}$ will display very little curvature (a typical pattern). In the limiting case of a tending to infinity the term $\int S''^2$ will be forced to zero and the fitted curve will approach a linear regression fit. Now if a is relatively small then the residual sum of squares will dominate the minimizing sum of squares and the curve will track the data closely even if it is at the expense of being rather variable. In the other limiting case of a tending to zero, $\widehat{S}$ will approach a smooth interpolating curve. Since the local variation does not exist, when modelling two successive data points by a deterministic function. The question of how to choose the smoothing parameter value that agrees with a certain data set, is a very important task, which can be addressed, for example, by a *cross - validation (CV)* approach, or a *Generalized Cross - Validation (GCV)*, (see Craven and Wahba, 1979).

5.3 Cubic Splines

A widely used case of spline functions is that of cubic splines. We will describe what a cubic spline is and then explain how it arises in interpolation.

Suppose we are given $x_1, x_2, ..., x_n$ ...on some interval $[a, b]$, satisfying $a < x_1 < ... < x_n < b$. A function S on $[a, b]$ is a *cubic spline*, if two conditions are satisfied.

Firstly, on each of the following intervals $(a, x_1), (x_1, x_2), ..., (x_n, b)$, g is a cubic polynomial and secondly the polynomials fit together at the points $x_i, i = 1, ...n$ smoothly. Then in a way that the function itself and its first and second derivative are continuous at each x_i, and then on $[a, b]$. It is easily stated from the previous that a cubic spline is the minimizer $L(S)$, to the minimizing problem stated in the previous section.

5.3.1 Equivalent representations of a (cubic) Spline

There are many essentially equivalent ways of specifying a cubic spline, or in general a spline.

1. Green and Silverman (1995) consider the case of giving the four polynomial coefficients for each polynomial fit.

2. McCutcheon (1981) and Hsieh (1991), consider the *Hermite* representation of a cubic spline. See Fröberg (1969), or Tyrtyshnikov (1997), for details on concepts like, Hermite polynomials, Hermite interpolation, e.t.c.

3. Another equivalent way is to represent a cubic spline as a linear combination of a set of basis functions. This is the most famous. It is called a *B-spline* representation and it is the one authors, like McCutcheon (1981), DeBoor (1966), and Greville (1969), adopt in their works.

4. An alternative and very handling case is that of viewing a cubic spline fit as a linear regression fitting procedure. See Benjamin and Pollard (1980), Speirs (1986), McNeil, Trussel, and Turner (1977). It is simple to see that an interpolating spline is similar to fitting a deterministic regression model to each data piece defined by two successive knots.

5. Green and Silverman (1995) consider also the case of presenting a cubic spline by giving it's values $S(x)$, and second derivatives $S''(x)$ at x_i, $i = 1, ...n$, where x_i the knot values. Suppose that S is the cubic spline fit with successive knots $x_1 \prec ... \prec x_n$. Define,

$$S_i = S(x_i),\ S'' = S''(x_i),\ for\ i = 1, ...n$$

Additionally, Rensaw (1995), explains how a cubic spline, is related to Generalized Linear Models. Two cases of the cubic spline are considered in this study.

5.3.2 The Natural (cubic) Spline

We introduce here the term of the *natural spline.*

From a problem that Shoenberg first considered, find a function S in the Sobolev space W_m of functions with *(m-1) - continuous* derivatives and the m - *th* derivative square integrable, so to minimize,

$$\int_a^b \left(S^{(m)}(x)\right)^2 dx$$

subject to $S(x_i) = S_i, i = 1, 2, ..., n$. He showed that if $n \geq m$, this minimizer is unique (see, uniqueness theorem) and also what we call the *natural polynomial spline.* That is one of the desirable properties splines have, and means that from the class of functions defined on the interval $[a, b]$ and for which we have values at $x_1, ..., x_n$ with $a = x_0 < x_1 < ... < x_{x+1} = b$, there is a unique function that minimizes the above integral and is smooth. The terminology "natural" comes from the fact that if the previous minimization problem is replaced by:

$$\int_{x_1}^{x_n} \left(S^{(m)}(x)\right)^2 dx,$$

then the solutions to the two problems will coincide in $[x_1, x_n]$, with the solution to the latter satisfying the so called Neumann or "natural " boundary conditions,

$$S^{(j)}(x_i) = S^{(j)}(x_n) = 0, j = m,\ m + 1, ..., 2m - 1.$$

It is easy to derive the cubic natural interpolant by setting $m = 3$.

5.3.3 The Complete (cubic) Spline

Hsieh (1991) in his work considers another case, that of a complete (cubic) spline. According to Hsieh, it appears to be a more accurate solution when applied to a Life Table construction problem. A natural spline, or in general a spline processes optimum approximation - minimum norm (maximum smoothness) and best approximation (high order of accuracy). A complete solution adds fast convergence to the above properties.

A complete cubic is a cubic spline with *two certain end conditions.* It has the same definition of the natural one but with no natural conditions applied to the ends of the data. Consult Hsieh (1991), for further details on the end conditions.

5.4 Interpolating Spline

Our main emphasis here is on interpolation problems. The attempt to expand an abridged life table introduces such a problem. The subject of interpolation is perhaps more familiar to numerical analysts than to statisticians. Works such as Pollard (1989), McNeil, Trussel, and Turner (1977), Hsieh (1991), e.t.c. present the adaptation of interpolation in general, but especially spline interpolation for demographic data.

The *problem of interpolation* is easily stated as follows:

Suppose we are given values, $z_1, z_2, ..., z_n$, at the points, $x_1, x_2, ..., x_n$. We wish to find a smooth curve S such that interpolates the points (x_i, z_i), that is to say $S(x_i) = z_i$ for all $i = 1, ...n$. obviously there are many ways of constructing a sensible interpolating function S. For example Pollard (1988) in his proposal of expanding method, uses a linear interpolation solution on the $\ln \mu_x$ values. That is the simplest and most widely used solution, the one of connecting all points by straight lines. But it does not yield a smooth curve since the fitted function has discontinuous derivatives at each data point. A suitably chosen smooth interpolant will do a much better job on approximating the true underlying curve than will do the piecewise linear interpolant. Such a good solution is introduced by a spline interpolant.

From Green's and Silverman's (1995) definition of roughness penalty:

Let $S[a, b]$ be the space of all functions S on $[a, b]$ that have two continuous derivatives, and call a function smooth if is in that space. If we wanted the "smoothest possible" curve that interpolated the given points, then a natural choice would be to use as our interpolant the curve that had the minimum value of $\int S''^2$ among all smooth curves that interpolate the data.

It turns out that among all curves S in $S[a, b]$ interpolating the points (x_i, z_i), the one minimizing $\int S''^2$ is a natural cubic spline with knots x_i. Furthermore provided $n \geq 2$ there is exactly one such natural spline interpolant. Thus the problem of finding the interpolant with minimal $\int S''^2$ is precisely that of finding the *unique* natural cubic spline that has knots at the points x_i and values $S(x_i) = z_i$ for all i. The uniqueness of the natural spline solution is known in the spline bibliography by the uniqueness theorem. Works such as, Greville (1969), Wahba (1975), or Green, Silverman (1995), introduce the following theorem. The reader interested to the proof may consult the previous.

Theorem 1 *(Uniqueness Theorem). Suppose that $n \geq 2$ and that $x_1 < ... < x_n$. Given any values $z_1, z_2, ..., z_n$, there is a unique natural cubic spline S with knots at the points x_i satisfying,*

$$S(x_i) = z_i, \text{ for } i = 1, ...n$$

5.5 Interpolating Spline in Demographic and Actuarial Applications.

A small review is conducted here on the applications of spline functions that arise in demographic and actuarial literature.

The work of Wahba forms the core of much of the work on splines (especially for smoothing purposes) in Statistics in general, (see for example Wahba, 1975a, b). McNeil, Trussel and Turner (1977), present the first analytic application of spline interpolation on demographic data. They carry out the effort of making known and understood the application of spline functions in demographic data problems. A common problem poses the need to represent a a smooth curve an age - specific fertility schedule for which average fertility by standard five year age groups are only available to the demographer. Equivalently important is the problem of interpolating mortality data, in which we focus in this thesis. McNeil, et. al. (1977) present and adopt the famous cubic spline solution on a demographic interpolation problem concerning fertility for Italian women of 1955. Such an application has to deal with rates that begin by age $x = 15$, and not by ages $x = 0, 1, 2, 3, 4, ...,$ *e.t.c* which is the case when we have to interpolate the l_x - values of a life table.

Benjamin and Pollard (1980) work is the first which presents in a greater detail spline smoothing and interpolation in mortality analysis. By adopting the natural cubic spline they produce the graduated mortality rates for a certain life table example. McCutcheon (1981), deals with certain aspects of splines at a greater length. Applications are concerned with the construction of some U.K. National Life Tables. Details of that may also be found in the work of McCutcheon and Eilbeck (1977). Anyhow as far concerning those interested to study the subject of spline functions further, they may consult Greville (1969), DeBoor (1966), Härdle (1992), or Green, and Silverman (1995). An interesting survey article on Splines in Statistics is the one of Wegman, and Wright (1983), which attempts to synthesize a variety of works on Splines in Statistics. That extends from the theory of interpolating and smoothing splines to a discussion of the role of splines on Time Series Analysis. Anyhow this is also general and does not concern a certain application.

Later developments of the cubic spline mostly deal again with fertility data. Nanjo (1986) describes the use of rational spline functions, which include cubic spline as a special case in obtaining data for single years of age for births given by five year age groups of mothers. Barkalov (1988) also describes the use of rational splines for demographic data. Bergström, and Lamm (1989) apply cubic spline interpolation in order to recover event histories. The method is used to adjust U.S. age at marriage schedules explaining a substantial part of the discrepancy in the 1960 and 1970 censuses. The "1980-82 New Zealand life tables" may be considered as famous to life table bibliography. These are derived by R. Speirs, which applied in 1986 the cubic spline by means of graduation. This thesis includes the application of the cubic spline interpolant (natural and a complete one) for the derivation of the same life table example.

It is of great interest to focus on the works of Hsieh (1991a,b) presenting applications on Canadian Data. The author develops a set of new life table functions which also includes in the formulation of the "complete" cubic spline interpolation method as applied to the expansion of abridged life tables. According to the applications included in this thesis there is no obvious evidence of a greater accuracy, according to Hsieh (1991 a,b), of the complete spline as compared to the natural one. The differences are insignificant small, as indicated by the results (see Appendix A and B).

Applications concerning splines in demography and also life insurance continue to appear later in the literature, but with all of them to concern the smoothing spline case. Biller, and Farhmeir (1997) connect spline type smoothing with GLM's from a full Bayesian approach. Based on the roughness penalty approach of Green, and Silverman (1995), they present among other the application of smoothing splines on a data example of the last authors. This deals with the crude death rates of a population of retired American white females. Renshaw (1995) also considers splines for actuarial use. As he notes, many of the graduation methods that are based on parameterized formulae as practised by actuaries in the construction of life tables are specific instances of the application of the class of GLM's. Applications with the graduation of the force of mortality μ_x , and probability of dying q_x are introduced here. Finally the more recent application of spline functions in the course of demography is presented by Haberman & Rensaw (1999). They present a simple graphical tool for the comparison of two mortality experiences. They start by graphically displaying the difference of the log crude morality rates plotted against age. This concludes to the difference of the forces of mortality of the two mortality experiences involved. Fitting cubic smoothing splines to such plots is particularly effective for targeting the underlying age specific pattern.

Chapter 6

METHOD EVALUATION AND COMPARISONS

6.1 Evaluations

In this chapter, we provide some applications of the various expanding techniques presented in the two previous chapters.

This is done in order to evaluate and compare the performance of these methods. We use empirical data of different countries and different time periods for both sexes. We have the empirical q_x - values which we initially abridge in wider age intervals for the ages 0, 1-4, 5-9, e.t.c. Then we apply each one of the expansion methods considered to these abridged data sets. At a final step, we compare the resulting sets with the corresponding complete empirical ones.

For the purpose of the evaluation and the comparison of the performance of the different techniques used, we provide graphical representations of the results and we calculate the values of two different criteria. These are the sums of squares of the respective absolute and relative deviations between the resulting and the empirical q_x - values.

$$\sum_x (\hat{q}_x - q_x)^2 \tag{6.1}$$

$$\sum_x \left(\frac{\hat{q}_x}{q_x} - 1\right)^2 \tag{6.2}$$

The sum of relative differences (6.2) is the one Heligman & Pollard (1980) propose as the quantity to minimize when trying to estimate **HP8**. The same loss function is also used by Kostaki (1991 and 1992a), Hartmann (1987), Forfar & Smith (1987) and other authors. Tables 1 and 2 provide the values of (6.1) and (6.2) of our application. Kostaki (1991) makes also reference to another evaluation criterion, where the residuals are weighted by some transformation of the E_x(the exposed to the risk of death population values). This is,

$$\sum_x \frac{E_x}{q_x(1-q_x)}(\hat{q}_x - q_x)^2 \tag{6.3}$$

where the quantities $\frac{E_x}{q_x(1-q_x)}$ are equal to the reciprocals of the binomial distribution variances of q_x.

The last sum requires the knowledge of the E_x - values in the actual population and it is more complicated than (6.1), and (6.2). Kostaki (1991, 1992b, 2000) uses both (6.1), and (6.2) because they complete each other as criteria since their values are indicative for the performance of a technique at different parts of the total age interval. The first one takes equally into consideration the residuals for the whole range of ages and since q_x - values for the later adult ages are on a much higher level than those for the younger ages, the values of this sum are more indicative for the later part of the age interval. The second sum gives greater consideration to the residuals of the childhood and the early adult ages, where the weights are heavier as they correspond to smaller q_x - values. The third corresponds to an alternative weighting of the residual values. It takes into consideration mostly those residuals, which correspond to q_x - values with the smallest observed fluctuations.

The literature does not suggest a theoretical based criterion as a tool for evaluating the performance of an expanding technique. The various authors only illustrate a graphical representation of their results, see e.g., Heligman & Pollard (1980), Pollard (1989), Forfar & Smith (1987). In the literature of the lagrangean and spline interpolation, no reference is either done for some evaluation criterion. Some authors e.g. McCutcheon & Eilbeck, (1977), Congdon (1993), use a sum similar to (6.3) as a loss function, which is approximated by a chi -square (X_n^2) distribution. Spline interpolation is a relatively recently used technique in demography but there does not exist any criterion suggestion for evaluating its performance.

In the face of the parametric technique of Kostaki (1991) we used the Heligman & Pollard model of eight (HP8) and nine parameters (HP9) as it was proposed by Kostaki (1991, and 1992b respectively).

Since the Heligman&Pollard model is a non-linear model of age x a non-linear least squares procedure was adopted in order to estimate the parameters. This is accomplished using an iterative estimation algorithm. **E04FDF** routine of the **NAG** library is an easy-to-use algorithm for finding the unconstrained minimum of the supplied sum of squares, which is defined here by formula **(6.2)**. The applied Heligman&Pollard models require constraints for the parameters. Since this algorithm searches for the unconstrained solution, it is expected to occur in some cases one or more negative parameter values. All parameters are constrained to take non negative values. Neither of our applied data sets demonstrated such problems. Additionally to the use of the minimization routine we supplied the subroutine LSFUN1 in order to evaluate the loss function at each age x.

In order to provide adequate starting values for the parameters the UNABR procedure of the MORTPAK package is used. In some few cases (e.g. Finland's data) MORTPAK does not suggest adequate initial values. In such cases the algorithm fails to converge and stops.

In order to obtain estimates for the standard errors of the parameters of HP8 and HP9, we used **E04YCF** routine also supplied by the **NAG** library. The E04YCF routine provides estimates of the elements of the variance-covariance matrix of the estimated parameters. The estimates are derived from the Jacobian of the loss function's value at the given solution. In all cases that we got estimates for the parameters, we did not have problem to obtain an estimate for their standard errors. A floating point occurred by the algorithm in some cases but we got adequate results for the parameters and for their standard errors. Several applications of the HP8, like Rogers (1986), Forfar and Smith (1987), Congdon (1993), or Karlis & Kostaki (2000) suggest that the model is overparameterized. An overparameterized model means that one, or several parameters are not required, or else, that several parameters are statistically insignificant. Experience by demographers so far, demonstrates that mortality caused by accidents is in general less severe to females and nonexistent to old female and some old male data sets. We comment that it becomes increasingly intense to recent data sets, also for the females. The last denotes the importance of the HP9 model for graduating recent mortality experiences. As Forfar and Smith (1987) comment, in cases where the "accident hump" is not severe, the model will try to find an accident hump at the later ages. That comes to justify the cases of females of Italy or Norway where a great value for the parameters *F* and, or *E* were obtained. A great value especially for the *F* parameter means that the model estimated a shift of the curve at a later age. Such a great value cannot be interpreted, since *F* is equivalent to the age where the accident hump of the mortality curve is reached. Such estimated values will have also a great standard error, suggesting that they are insignificant. In this case, the larger the initial value of the parameter *E* or *F*, the larger the estimated. It is reasonable to avoid fitting HP9 in such cases. On the other hand, it might be reasonable in such a case to try to reduce the HP8 model by subtracting its middle term. Problems raised when we tried to estimate HP9 for several data sets. There were cases where HP8 was estimated correctly, but the algorithm was unable to provide results for the HP9. Several runs of the program for HP9 on the Finland female data set didn't conclude to any results. In this case, a modest accident hump exists.

In general, HP8 is an adequate solution to the expanding of an abridged life table. In cases with an intense accident hump, the HP9 model will exhibit a better performance than the eight - parameter one. We outline values for the minimized loss function (formula 6.2) by the least squares algorithm which also is used as a criterion of the efficiency of the expanding technique. We also outline values for formula (6.1). As the values of the criteria indicate, (see Tables 1 and 2 at the end of this section), in all the cases where we had results for both models, the HP9 performance was superior to HP8. The additional adjustment will improve the probability estimates provided by the two models. The adjusted now results of HP9 will still be better from the corresponding of

HP8. That is indicated by the superiority of the corresponding criterion values (see again Tables 1, and 2).

Kostaki (1991) in order to improve the estimations obtained by a parametric model (e.g., HP8, HP9) performs this adjustment. It is interesting to observe that this is an adjustment to be performed on the age-specific $\hat{q}_x$'s which arise from an expanding procedure. For the additional adjustment, a simple logistic package like EXCEL97, or a simple statistical package like MINITAB are sufficed.

The non - parametric technique of Kostaki is a relational one since it relates the abridged data set with an age-specific one of another life table, which is used as a reference. It was also obtained in EXCEL97, and in MINITAB for the same reasons as before. Kostaki's applications (see, Kostaki, 2000b) as well as our applications showed that it is not required for the reference table to exhibit a pattern similar to the one which underlines the abridged life table used. In other words, we are not obligated to search for a standard life table. It is also not required for the reference table to represent a smooth pattern, or to refer to the same gender. Tables C1a and C1b of the Appendix support the above comments. These presents the technique's performance on the Italian life tables of 1990-91 for several choices of a reference table, as these were judged by the values of the two criteria used in this application. Values of Tables 1 and 2 can be adopted as tools for the comparison of the non – parametric technique and the adjustment performed on the results of the parametric method. Judging by the previous comments, the non-parametric technique can be considered as a fine solution to an abridged data problem.

As an alternative to reproducing the whole age range by a single parametric model, or in general a single polynomial function, we estimated regions of the data as defined by two successive data values, by applying piecewise polynomials. This is the case of performing cubic spline interpolation on the abridged mortality data. The cubic spline was applied twice, once for the case of the natural, and once for the case of a complete cubic spline interpolant. Interpolation in general requires values of a function, which is not mathematically specified. The survivor's function l_x indicates such a function. The interpolating polynomial spline applied on the l_x - values of the abridged life table was obtained in S-plus (see Chambers and Hastie, 1992), by using the {**spline (parameter1, parameter2,...)**} interpolation tool offered by the program. This tool has the following complete form, {**spline (x, y, n = , periodic = F, boundary = 0, or 1, xmin = min(x), xmax = max(x))**}. The "*boundary*" parameter introduced here, sets out the choice for a natural, or a complete spline. A **boundary** parameter equal to zero **(0)** will conclude to the **natural cubic spline interpolant**, and so the **boundary** equal to one **(1)** will conclude to *a case* of a **complete cubic spline interpolant**. The later represents a complete cubic spline case, in the sense that it poses certain end conditions to the interpolating function. It is not the case proposed by Hsieh (1990, 1991). Anyhow, it sets clearly out the superiority of the complete spline, when compared to the natural, as Hsieh (1990, 1991) comments. This comes also out, from the following tables (Tables 1, and 2) where the two splines are compared via values of the sum of squares of the relative deviations (see formulae 6.1 and 6.2). It is obvious that the complete spline is in almost all cases better than the natural one. The values of the sums of squared residuals (6.1. and 6.2) are usually higher for the natural spline (see Tables 1 and 2). It is interesting to evaluate how the two cases of spline interpolation behave at the accident hump, especially when they are compared to the HP8, HP9 parametric models. We notice that for data that were smooth from the beginning, the performance of splines was comparable to the one of parametric models, and though better at the accident hump part. Graphs presented in the Appendix A (see Figures A1.1, 1.2, 7.1, 7.2, 8.1, and 8.2), or Appendix B, (see Figure B2.2) can assist such comparisons. The superiority of the complete spline *vs* the natural is also verified there. We should note that there were cases of data sets, e.g. Italy's, or New Zealand's males, (see Table 1), where the performance of the complete spline was not superior to the natural (according to the values of formula 6.1, see Tables 1, and 2). That might suggest that the defined end conditions may be not the appropriate ones in order to improve the natural end conditions. The last comes out by the fact that a great value for formula 6.1. possibly indicates some systematic fluctuation of the $\hat{q}_x$'s at the later adult ages.

A conventional technique of applying polynomial interpolation on the l_x - values of the abridged life table, is the Lagrangean technique of six terms. It uses six successive values of a function, which it interpolates, by a single value. Because of the great number of values used for one value interpolation, the method does not achieve to reproduce the whole life table. The extent of the estimated age interval depends on the age-range of the observed data. In the cases where we used data for the ages, $x \leq 85$ the Lagrangean interpolation produced estimates only for $x \leq 70$. The last becomes $x \leq 59$, for observed data for the ages $x \leq 75$ (e.g. the data of Finland and Norway). The method is simple to handle and it produces results very fast. It requires the use of certain coefficient values, known as the Lagrangean coefficients, which are tabulated by Ambramowitz, and Stegun (1972), and Elandt - Johnson, and Johnson (1980). These are also presented in Table C4 of the Appendix C. The method can be performed easily by a simple computer package like EXCEL97. Lagrangean interpolation is proposed to estimate mortality only for the age interval of $x \leq 70$. Our calculations showed that the method underestimates mortality at the first ages ($x = 1$), and overestimates until age, $x = 5$. A great and sudden drop of the estimated curve at about age $x = 10$ is evident to the performance of this method in all our applications. It appears also that the method estimates a greater in length of time accident hump. Usually, it will overestimate the mortality curve before the observed hump but it will follow it correctly later. That sudden shift of the estimated mortality curve is more severe and abnormal, when the accident hump is a modest one, e.g. for the data of Norway and the females data of Italy. This technique behaves in a satisfactory way for the old aged, but we remind that it doesn't extent to ages above 70. An important point here is that this method provides estimates, which we can abridge again and get the initially observed ${}_{n}q_{x}$'s. So the Lagrangean interpolation estimated $\hat{q}_x$'s, have the desired property for an expanding technique. We concluded to that, when we tried to perform the additional adjustment of Kostaki (1991) on the Lagrangean estimated probabilities.

Getting back to the spline interpolation case, we may do a comparison with the lagrangean technique, since both are cases of a piecewise polynomial interpolation.

Splines estimate mortality for the whole age span and provide smooth results, because of the conditions applied at the adjacent points of the piecewise polynomials. But in general, both methods for certain ages demonstrate a common behavior. This common behavior is judged by the underestimation of mortality at age $x = 1$ and the overestimation for some ages before the accident hump that both interpolation techniques demonstrate. A "sudden" drop at age $x = 10$ is observed, but that is obviously smaller for the spline case. Tables 1 and 2 can assist us on doing a comparison of the lagrangean interpolation with the spline interpolation judging by the values of the two criteria 6.1. and 6.2. In all cases the superiority of the spline interpolation to the one of Lagrange is obvious. Additionally, one can consult figures presented in Appendices A and B. Figures B1.1. and B2.1. in the Appendix B compare the Lagrangean interpolation with the famous natural cubic spline. The problems of the lagrangean technique at the childhood and the early adult ages can readily be observed. Other methods are also considered in this application. Reed's technique for example is a simple old tool. The only reference that exists is by Valaoras (1980) who described and applied the method for the construction of the official life tables of Greece. The technique was simply performed in EXCEL97. It is comprised of some regression applications of the $\frac{q_x}{K^x}$ values on age x, where K takes certain values given by Reed. It was easy to obtain the regression equations in EXCEL97. Judging the methods performance, we would say that it faces problems at estimating correctly the accident hump. In almost all the applied datasets, problems were observed at the ages $x = 21$, to 24. These were more evident when the accident hump was not severe, which is of a common case for the female data. It seems that the method tries to model a hump at a later age interval, so an overestimation of the mortality curve at the sequel is observed (see performance at the age interval $x = 25$ to 42). Where the accident hump exists, the performance of the method at the ages $x = 21$ to 24 is better, but an underestimation occurs later for the ages $x = 25$, to 42. An underestimation is again observed for the old aged, $x = 60$, to 65 (see figures of Part 9 of Appendix A). Studying the values of Tables 1 and 2 at the end of this section, one can observe the great improvement that Kostaki's expanding technique can offer to every expanding technique's estimates. We performed that additional adjustment of Kostaki (1991) on Reed's estimated probabilities and the improvement was obvious.

In general, the method estimates mortality from the ages greater than five. It performs badly until age $x = 10$. The same performance is observed for senescent mortality. Pollard's expanding technique is not possible to be applied since it requires the existence of the ${}_n\overline{P}_x$, central population counts for each five year age group. We tried anyway to apply the method by using the exposed to the risk of death populations, E_x instead of ${}_n\overline{P}_x$, in order to calculate the *r*-rate introduced by the method. Every data set we used does not have the form required. The expanding method requires the age-specific data until age the $x = 4$. We suggest that the method can perform better on data where the accident hump does not exist if a linear interpolation on the $\ln \mu_x$ - values is only applied. We used again S-plus for this experimental application, and the {**approx(parameter1, parameter2, ...)**} tool offered for linear interpolation by the program (see Chambers, and Hastie, 1992). The complete form of the S-plus command is {**approx(x, y, xout, method="linear", n=50, rule=1, f=0)**}, where the parameter "*method*" defines whether we use the {linear}, or another interpolation scheme.

We conclude with a comment on those methods that do not participate in this application. These concern old expanding tools that do not appear anymore in practice. The several six – point interpolation techniques that were presented by Beers (1944), it seems that they stop at that year. No reference on one or more of these is done later. The same stands also for the osculatory interpolation technique of King, which seems more useful as a graduation technique, than as an expanding one.

Table[1] 1: Values of the sum of squares of the *absolut* deviations between the resulting and the empirical q_x -values (see, criterion ***6.1.***), multiplied by 10^6 and calculated for the common age interval of all tested methods, **age** ***x=[5,70]***.

	Expanding Method									
	HP8	**HP8 Adjusted**	**HP9**	**HP9 Adjusted**	**Lagrange**	**Non-Parametric[2]**	**Natural Spline**	**Complete spline**	**Reed**	**Reed Adjusted**
Life Table										
Italy females 1990-91	13.250	1.488	-	-	0.710	2.040	0.756	0.600	4.900	2.363
Italy males 1990-91	25.740	6.750	25.680	6.700	1.840	8.630	1.250	1.300	103.300	7.680
Sweden females 1991-95	11.160	0.820	-	-	0.360	1.010	0.410	0.390	5.620	1.067
Sweden males 1991-95	4.090	0.890	1.740	0.800	0.810	3.340	0.700	0.690	41.340	3.411
New Zealand females 1975-77	21.940	2.263	21.920	2.240	0.420	1.940	0.300	0.230	14.540	4.853
New Zealand males 1975-77	10.790	1.405	10.720	1.400	2.280	6.670	1.640	1.780	342.32	23.648
New Zealand females 1980-82	6.750	0.994	6.724	0.990	0.380	1.800	0.280	0.250	30.32	2.192
New Zealand males 1980-82	22.640	2.720	22.610	2.700	1.810	5.400	1.340	1.470	28.205	15.380
Norway[3] females 1951-55	2.000	0.242	-	-	0.200	0.358	0.198	0.177	1.170	0.200
Norway males 1951-55	3.000	0.257	-	-	0.340	0.970	0.203	0.190	7.300	0.330

[1] The dashes in each table refer to cases where HP9 was not estimated, since the accident hump of these data was dissappeared.
[2] The table of the opposite gender is used as the table of reference in each case for the non - parametric method
[3] Criterion values refer to ages $x \leq 75$, except for the case of the lagrangean interpolation, where values refer only to ages $x \leq 59$.

Table 2: Values of the sum of squares of the *relative* deviations between the resulting and the empirical q_x -values (see, criterion *6.2.*), calculated for the common age interval of all tested methods, **age** ***x=[5,70]***.

	EXPANDING METHOD									
	HP8	HP8 Adjusted	HP9	HP9 Adjusted	Lagrange	Non-Parametric	Natural Spline	Complete Spline	Reed	Reed Adjusted
LIFE TABLE										
Italy females 1990-91	0.310	0.057	-	-	0.573	0.295	0.149	0.095	3.856	0.339
Italy males 1990-91	0.196	0.050	0.189	0.049	0.815	0.339	0.179	0.148	2.262	1.149
Sweden females 1991-95	1.006	0.907	-	-	1.085	1.205	0.862	0.853	4.560	1.139
Sweden males 1991-95	0.938	0.707	0.728	0.530	2.015	1.246	1.158	1.091	4.707	2.904
New Zealand females 1975-77	0.231	0.055	0.230	0.055	0.468	0.421	0.378	0.244	0.913	0.315
New Zealand males 1975-77	0.171	0.066	0.136	0.065	1.582	0.584	0.743	0.681	4.157	2.297
New Zealand females 1980-82	0.240	0.102	0.240	0.101	0.705	0.300	0.479	0.306	1.841	0.762
New Zealand males 1980-82	0.254	0.063	0.175	0.060	1.691	0.328	0.737	0.633	4.653	2.408
Norway females 1951-55	0.163	0.074	-	-	0.410	0.334	0.126	0.081	0.139	0.074
Norway males 1951-55	0.147	0.056	-	-	0.464	0.399	0.106	0.087	0.261	0.179

6.2 Conclusions

In this study, we reviewed the methods one can adopt when is faced with the problem of expanding an abridged life table to a complete one.

Summarizing, we distinguished these methods in three categories:

1) Methods which require the use of a parametric model. The application of the Heligman & Pollard models of eight and nine parameters was reviewed here. Anyhow as already commented the choice of the parametric model is open.
2) Piecewise polynomial interpolation techniques. Here we reviewed the famous case of spline interpolation. Two cases of splines interpolants were applied. The Lagrangean interpolation is also a piecewise interpolation technique but simpler in comparison to spline interpolation.
3) Finally other methods which cannot be clustered in a certain category. Here we reviewed methods like the non – parametric of Kostaki (2000) which is the newest, or Reed's technique, which it is considered as the oldest one.

Parametric methods and Splines require advanced software in order to be obtained. The rest can be handled more easily. A set of basic properties in order to select the best method solution should be:

1) The ability of the technique to estimate the whole mortality pattern. The several parameterized approaches (e.g. HP8, HP9), polynomial interpolation techniques (e.g. splines) and the non-parametric relational technique are proposed for reproducing the whole mortality pattern. The famous Lagrangean interpolation technique although still simple to use, does not provide estimates of the complete series of the mortality probabilities. It requires the abridged life table to extend to more ages. As for the age interval $x \leq 85$, used in almost all cases here, it will give us estimates of the complete mortality until the age $x = 70$. Pollard's technique also estimates a part of the table when data are not provided in a certain form. It achieves to extend to age $x = 77$, but because of a certain grouping for the first ages that it requires, it fail to estimate mortality for the ages below $x = 8$. Reed (see Valaoras, 1984) developed

a method that begins by estimating mortality at age $x = 5$ extending to the whole age range

2) The ability of a method to obtain smooth results defines our second important property. A parametric model surely will provide smooth results. A spline interpolant will also provide a smooth mortality curve, since that property distinguishes a spline from other piecewise interpolation techniques. The rest of the methods included in this study will provide a series of probabilities that are not graduated.
3) An important property for an expanding technique is to get again the initial abridged life table after abridging the estimates. That can be provided by the results of the Lagrangean interpolation and the non – parametric technique of Kostaki (2000). By applying this additional adjustment to anyone of the expanding techniques, the resulting probabilities will obtain this property.
4) A method to be simple to handle or really complex. A complex method will require a great deal of computational task in order to be applied. Usually it will require advanced software e.g. a parametric model needs the application of a computer algorithm. On the other hand a method can be considered as complex also if it requires a certain form of the original data set. Pollard's technique application for example it is based on many assumptions, e.g., approximation of the μ_x - values, uniform exposure of deaths within an age interval and certain age grouping for the first ages. It also requires the existence of the ${}_n\overline{P}_x$, central population counts in each five year age group. Splines on the other hand will require advanced software in order to construct the piecewise interpolants.

Summarizing we should mention the following:

Parametric model solutions offer graduation to the data by obtaining a smooth central curve, which extends to the whole age interval. Additionally, parametric models provide a more flexible solution since they facilitate projections of the mortality pattern. The last can assist the analysis and the extrapolation of demographic trends and the setting of assumptions about such trends. It can also facilitate comparisons over space (e.g.

spatiotemporal modeling). Splines also obtain a smooth curve interpolating the grouped data. A finding of our applications is that the performance is better than the HP8's at the accident hump (see Figures B1.2. and B2.2. at Appendix B). These are cases where data were already graduated. Concerning the accident hump mortality, we must note again that it becomes greater and more intense, as recent mortality experiences suggest, especially in female population experiences. Someone may compare the old data of Norway 1951-55, with Sweden's 1991-95, or Italy's 1990-91 data used in this study. The need for a model that studies that part well is pronounced. The nine-parameter Heligman&Pollard model, or a cubic spline interpolant suggest some cases that can deal well with this problem.

A systematic fluctuation of the spline interpolant is tracked at the ages about $x = 10$ to 15. This is a sudden drop of the mortality curve that the interpolation technique of Lagrange also presents in a greater extent. Lagrangean interpolation always presents that systematic fluctuation at the ages before the accident hump. A sudden drop at the birth ages mortality is added to the previous observation, for both interpolation techniques.

The additional adjustment will improve the estimates of every applied expanding technique. As it appears by the several figures at Parts 2 or 4 and 5 of the Appendix A, a usual observation for the adjusted results of the applied parametric models, or the results of the non - parametric relational technique is that they overestimate mortality at the very elderly. The improvement offered by the additional adjustment can be easily observed by applying this to the inaccurate results of Reed's technique. It is important to note here that Greece still uses Reed's expanding technique as a tool to construct its complete life tables.

In general non - parametric techniques as Splines, or Kostaki's (1998) relational one lack at the interpretability part that can be obtained using the set of parameter values in order to provide spacio - temporal comparisons. It is worth again to note that splines can produce accurate and smooth results and reproduce the whole mortality curve. However, this technique requires advanced software in order to be applied to real data. Nevertheless, it avoids problems that are related to an algorithm, used to estimate the parameters comprised in a parametric model. Therefore, it is suggested that the actuary, or biostatistician, who needs only to provide analytical and accurate mortality estimations can rely on estimations obtained by such non - parametric techniques.

APPENDIX A:

EVALUATION FIGURES

PART 1

HP8 MODEL

HP8 MODEL

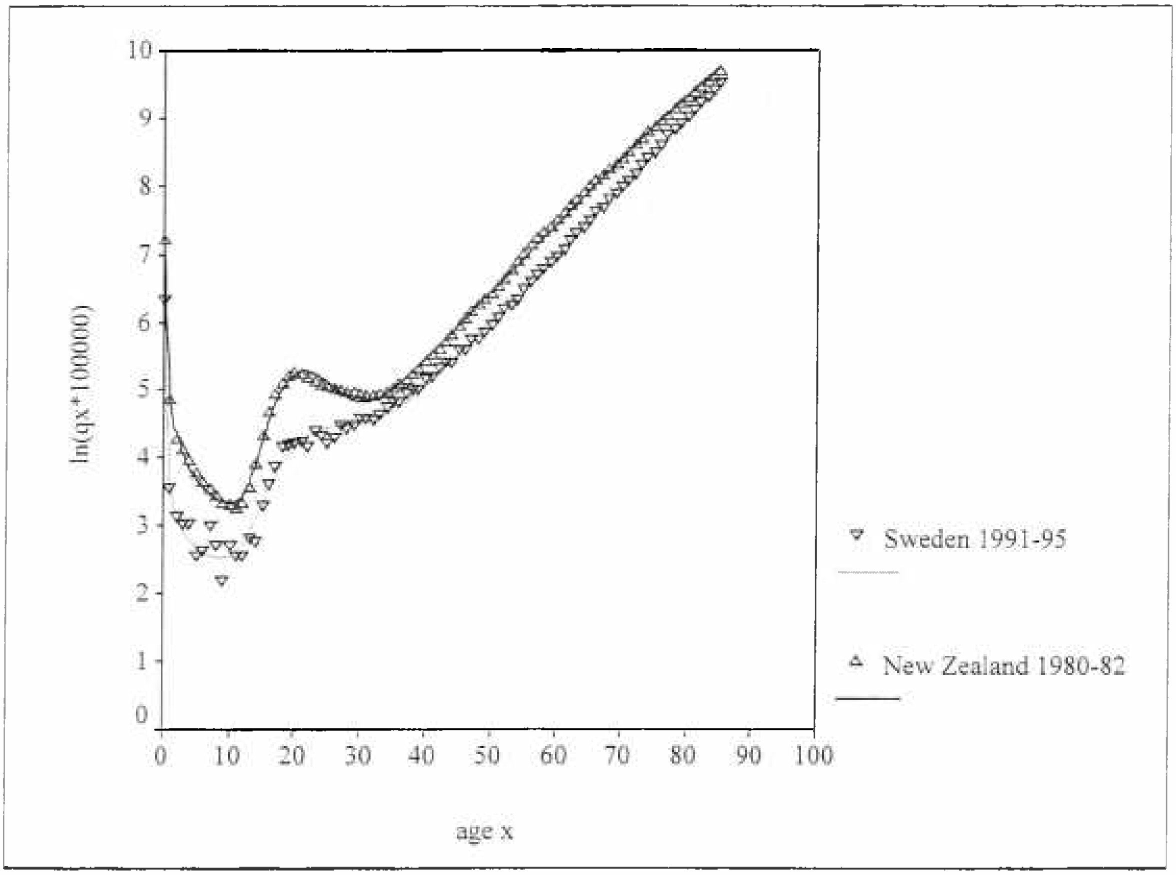

Figure A1.1: Empirical q_x - values (***points***) and estimations by HP8 formula (***solid lines***) for New-Zealand's 1980-82 and Sweden's 1991-95 females life table.

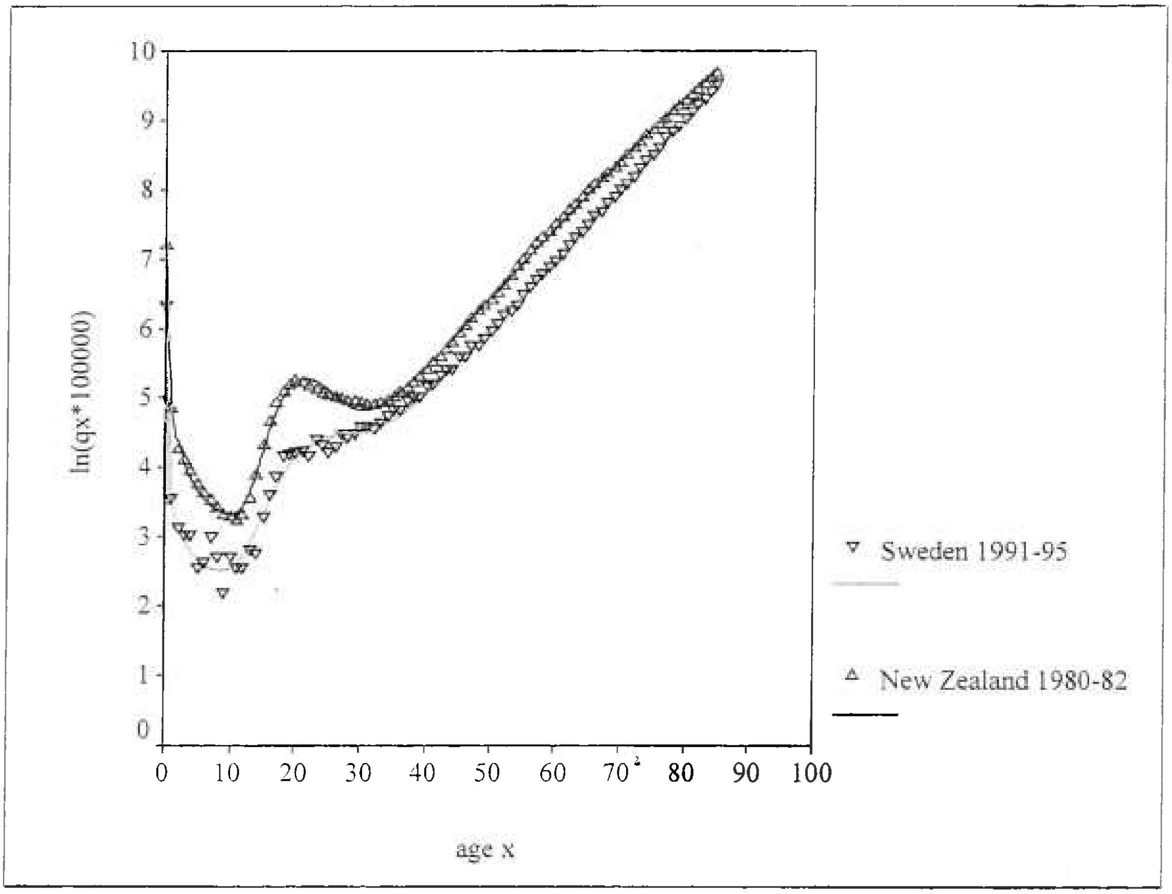

Figure A1.2: Empirical q_x - values (***points***) and estimations by HP8 formula (***solid lines***) for New-Zealand's 1980-82 and Sweden's 1991-95 males life table.

HP8 MODEL

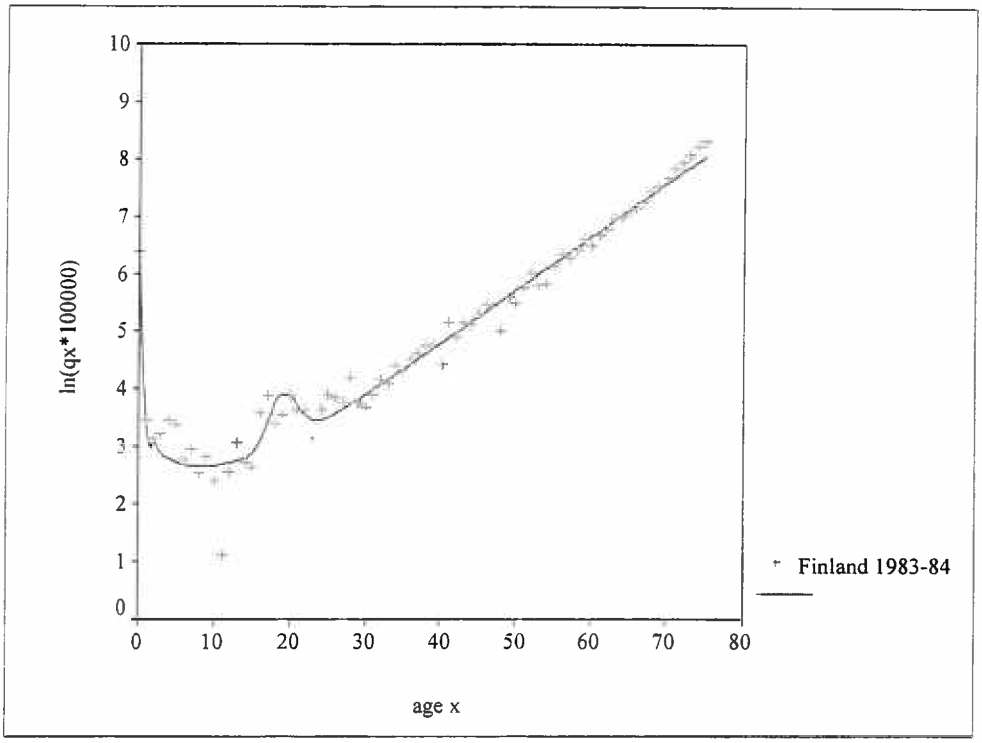

Figure A1.3: Empirical q_x - values (***points***) and estimations by HP8 formula (***solid line***) for Finland's 1983-84 females life table.

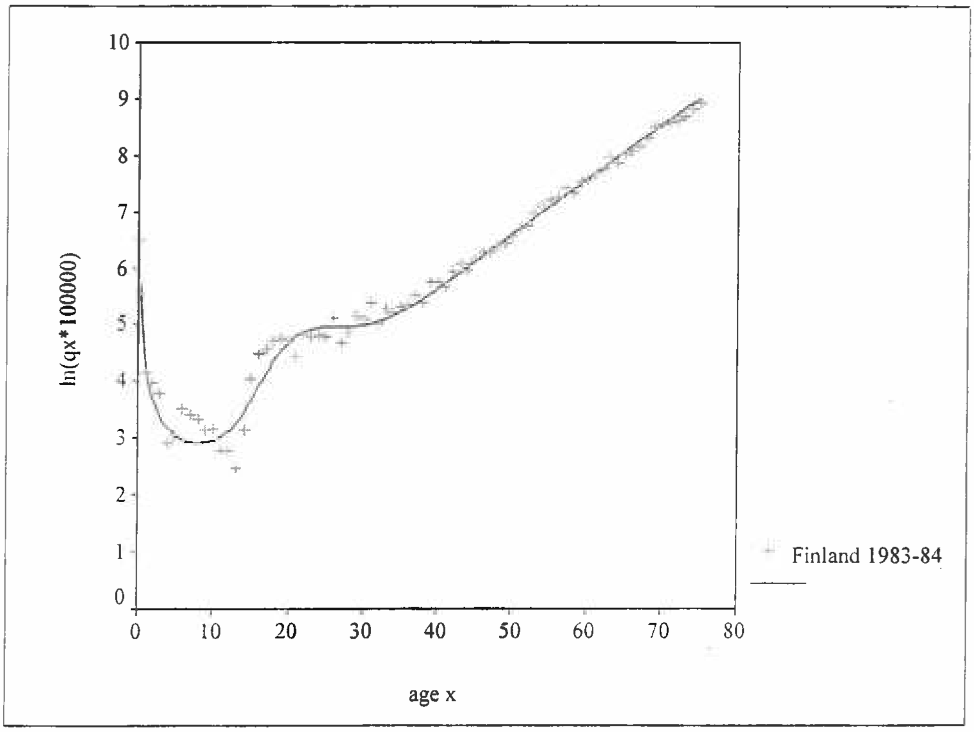

Figure A1.4: Empirical q_x - values (***points***) and estimations by HP8 formula (***solid line***) for Finland's 1983-84 males life table.

HP8 MODEL

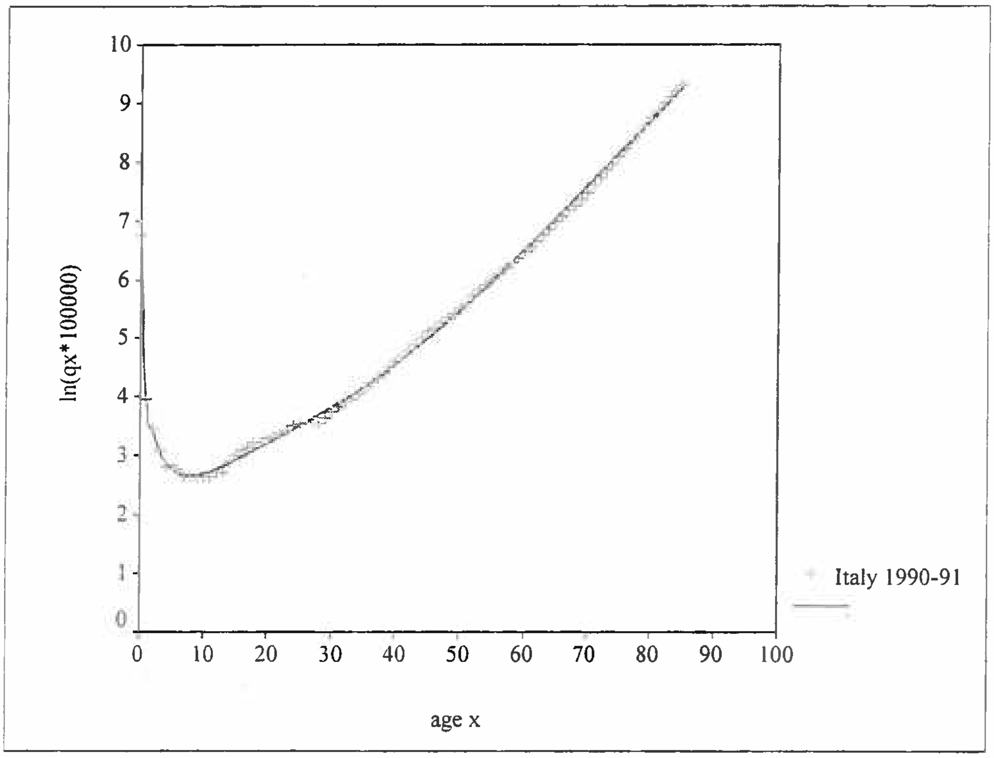

Figure A1.5: Empirical q_x - values (***points***) and estimations by HP8 formula (***solid line***) for Italy's 1990-91 females life table.

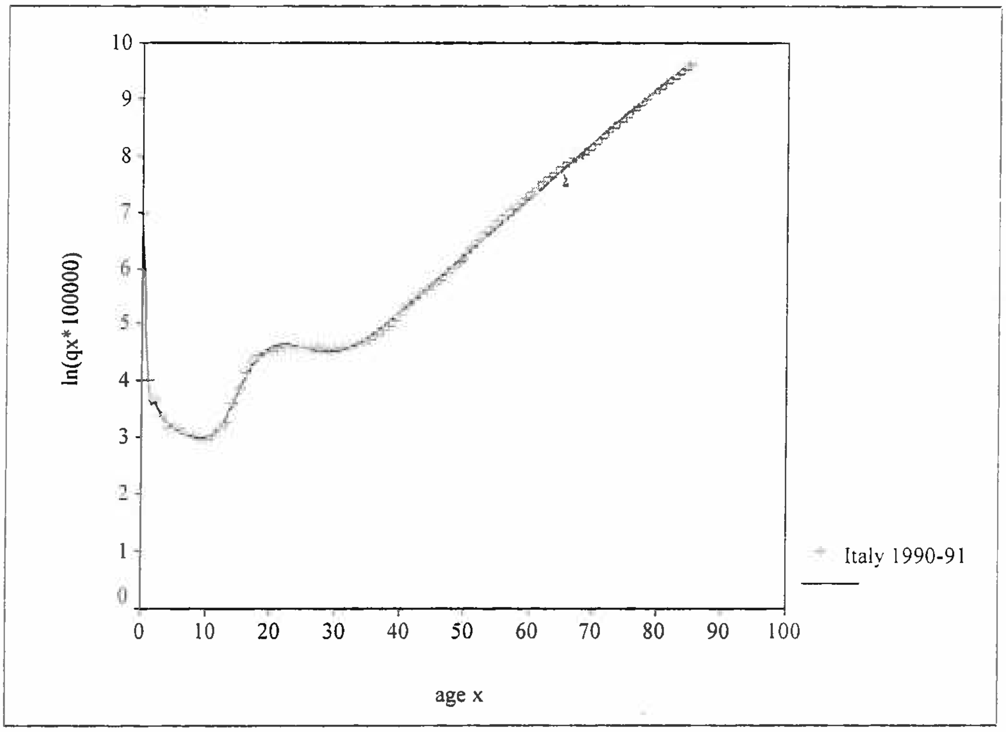

Figure A1.6: Empirical q_x - values (***points***) and estimations by HP8 formula (***solid line***) for Italy's 1990-91 males life table.

PART 2

HP8 ADJUSTED MODEL

HP8 ADJUSTED MODEL

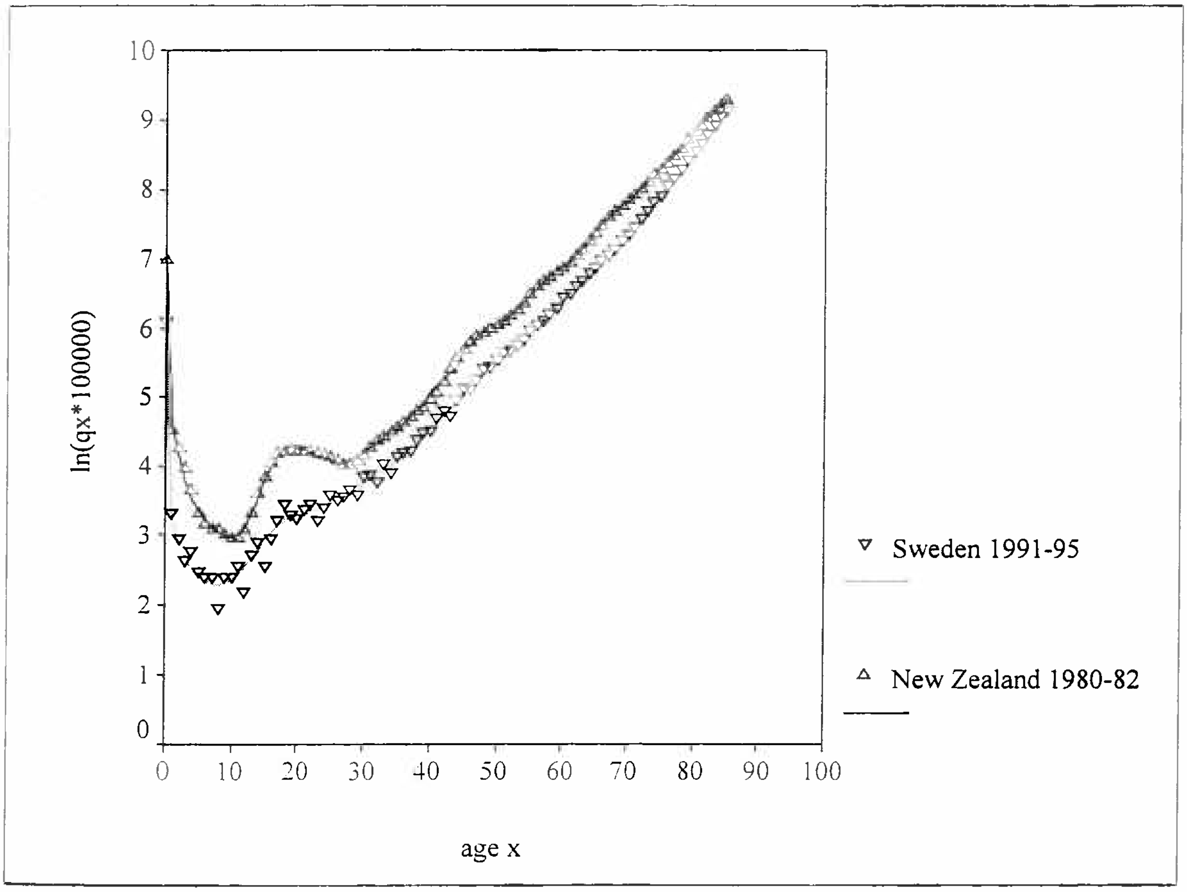

Figure A2.1: Empirical q_x - values (***points***) and estimations by HP8 adjusted formula (***solid lines***) for New-Zealand's 1980-82 and Sweden's 1991-95 females life table.

HP8 ADJUSTED MODEL

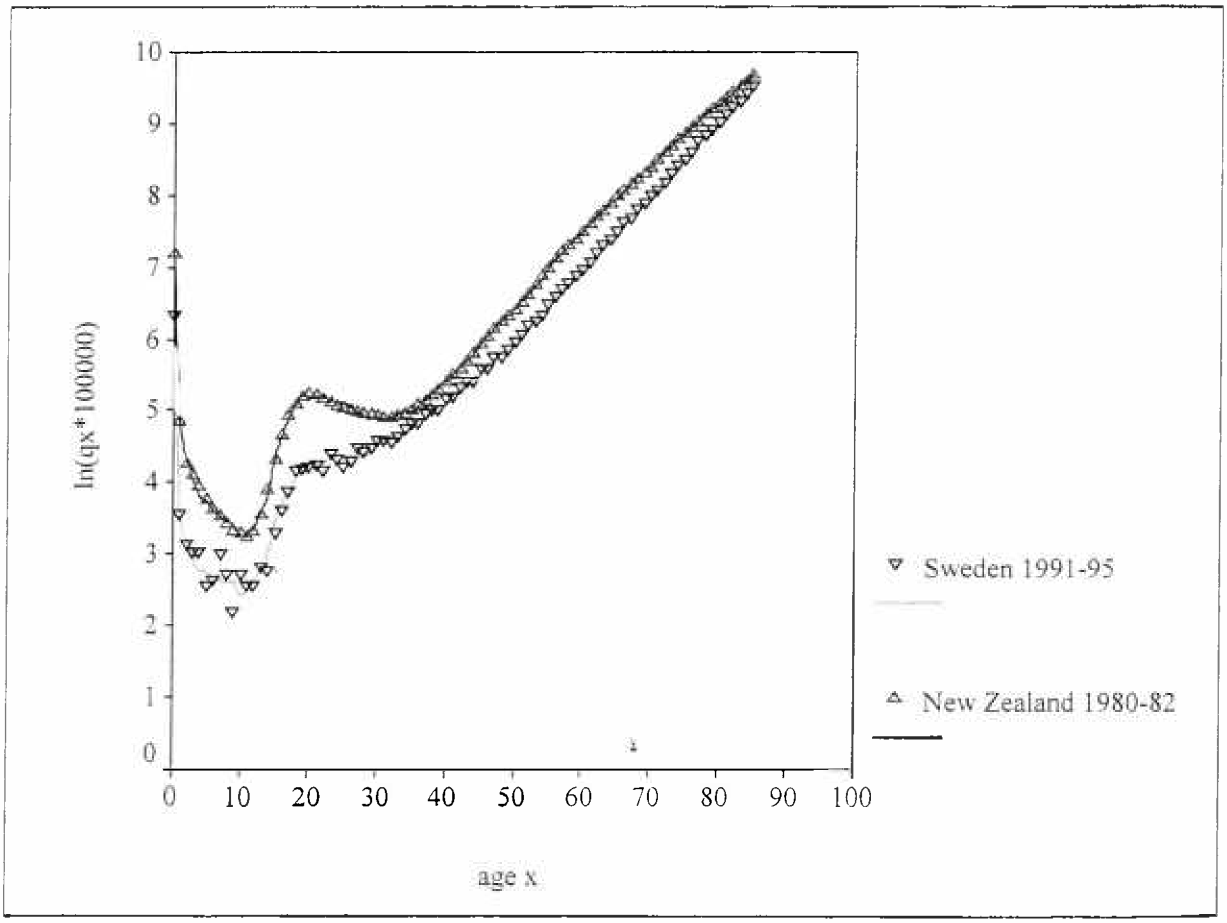

Figure A2.2: Empirical q_x - values (***points***) and estimations by HP8 adjusted formula (***solid lines***) for New-Zealand's 1980-82 and Sweden's 1991-95 males life table.

HP8 ADJUSTED MODEL

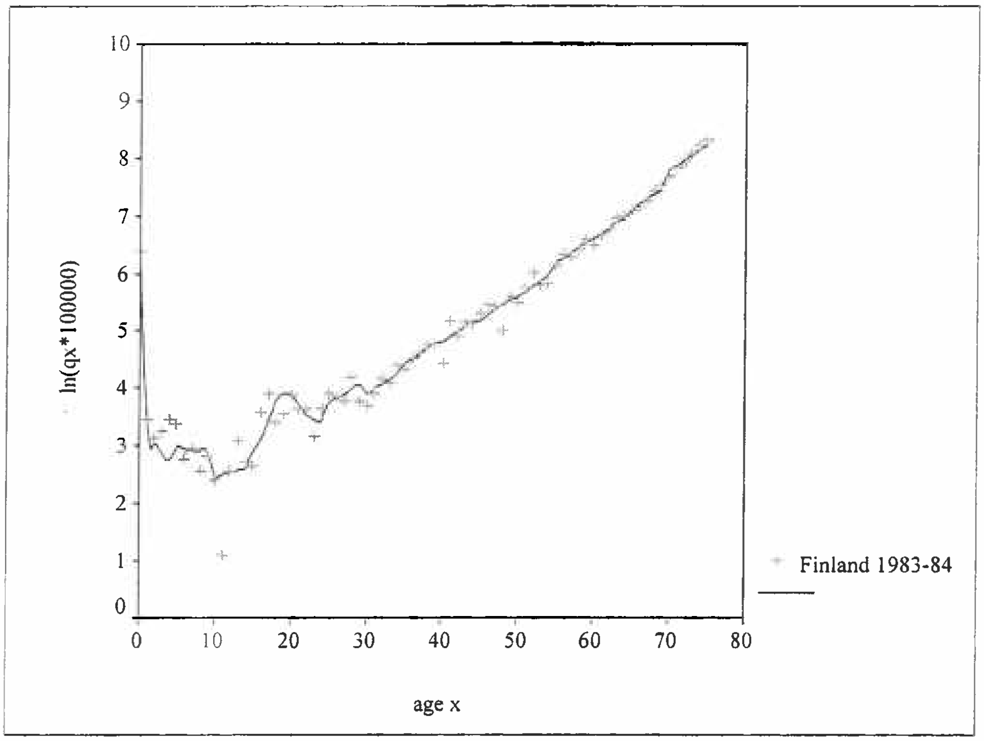

Figure A2.3: Empirical q_x - values (***points***) and estimations by HP8 adjusted formula (***solid line***) for Finland's 1983-84 females life table.

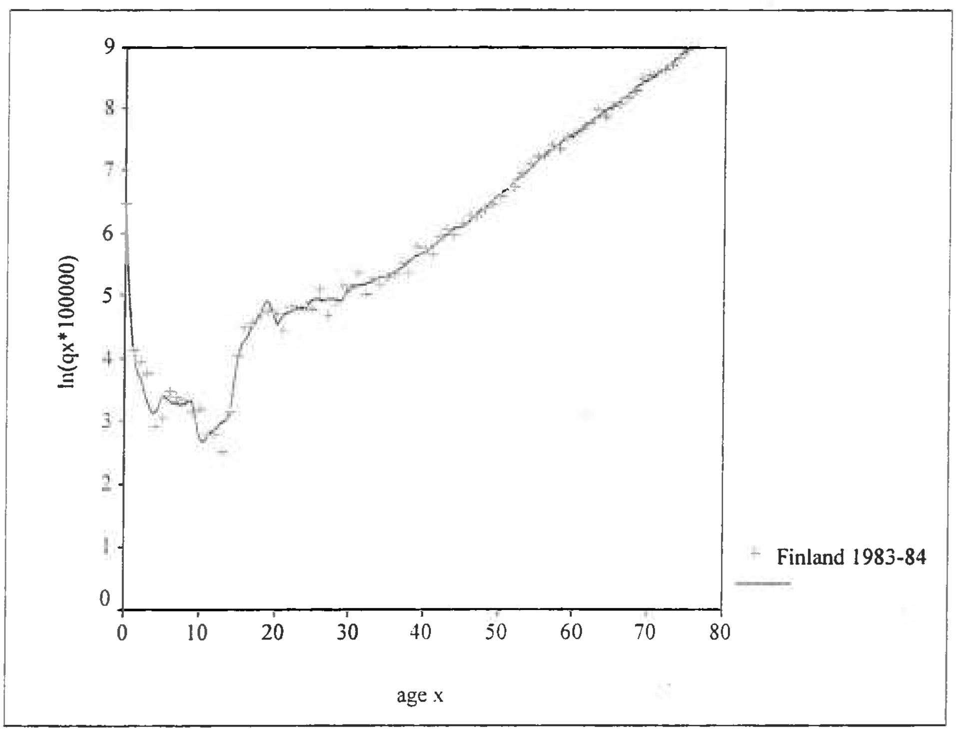

Figure A2.4: Empirical q_x - values (***points***) and estimations by HP8 adjusted formula (***solid line***) for Finland's 1983-84 males life table.

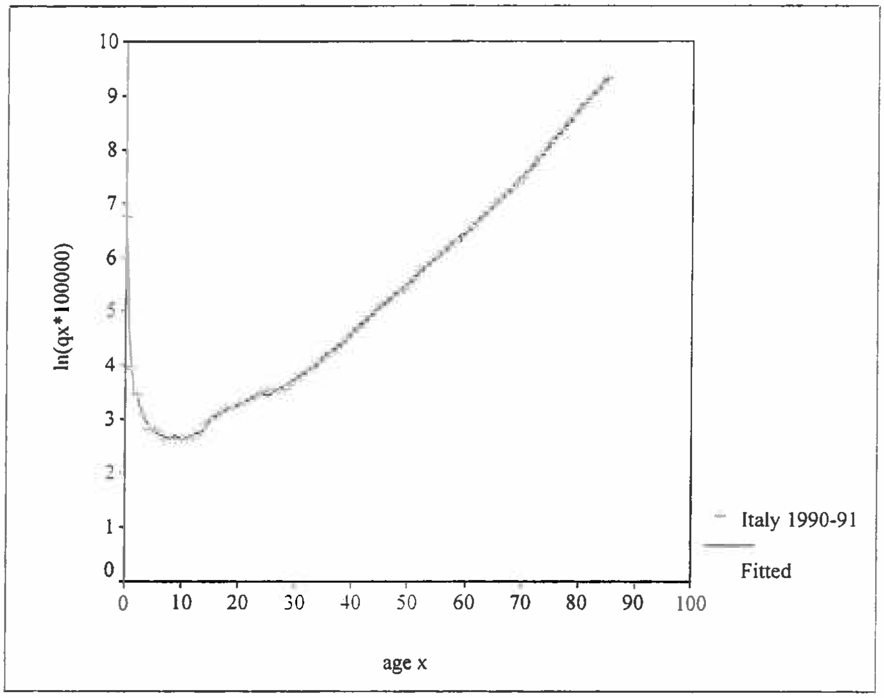

Figure A2.5: Empirical q_x - values (***points***) and estimations by HP8 adjusted formula (***solid line***) for Italy's 1990-91 females life table.

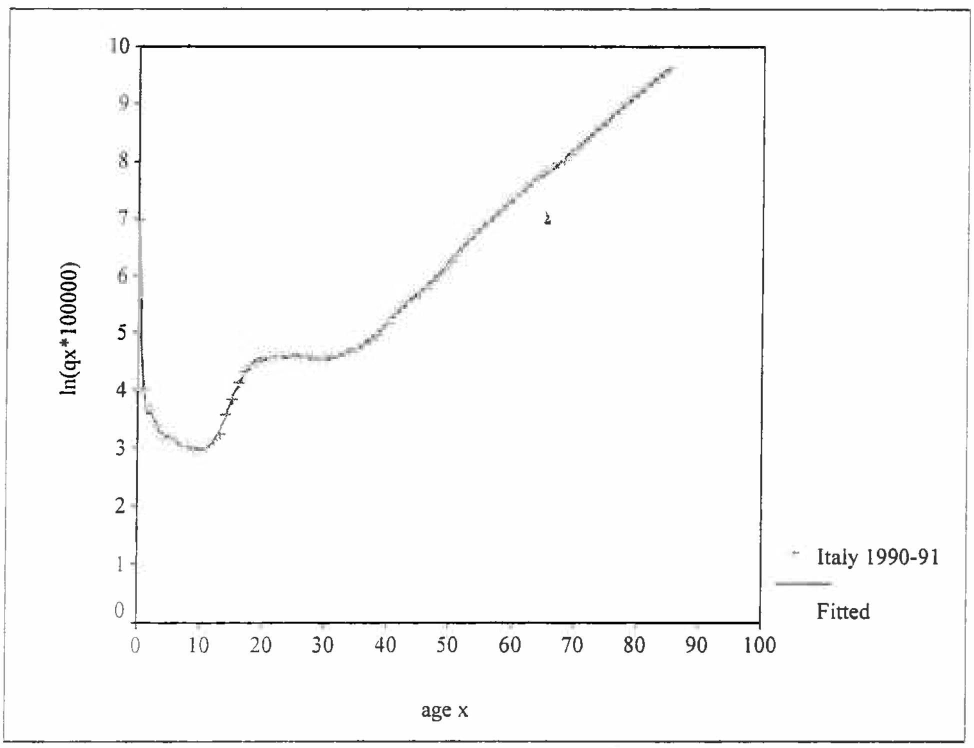

Figure A2.6: Empirical q_x - values (***points***) and estimations by HP8 adjusted formula (***solid line***) for Italy's 1990-91 males life table.

PART 3

HP9 MODEL

HP9 MODEL

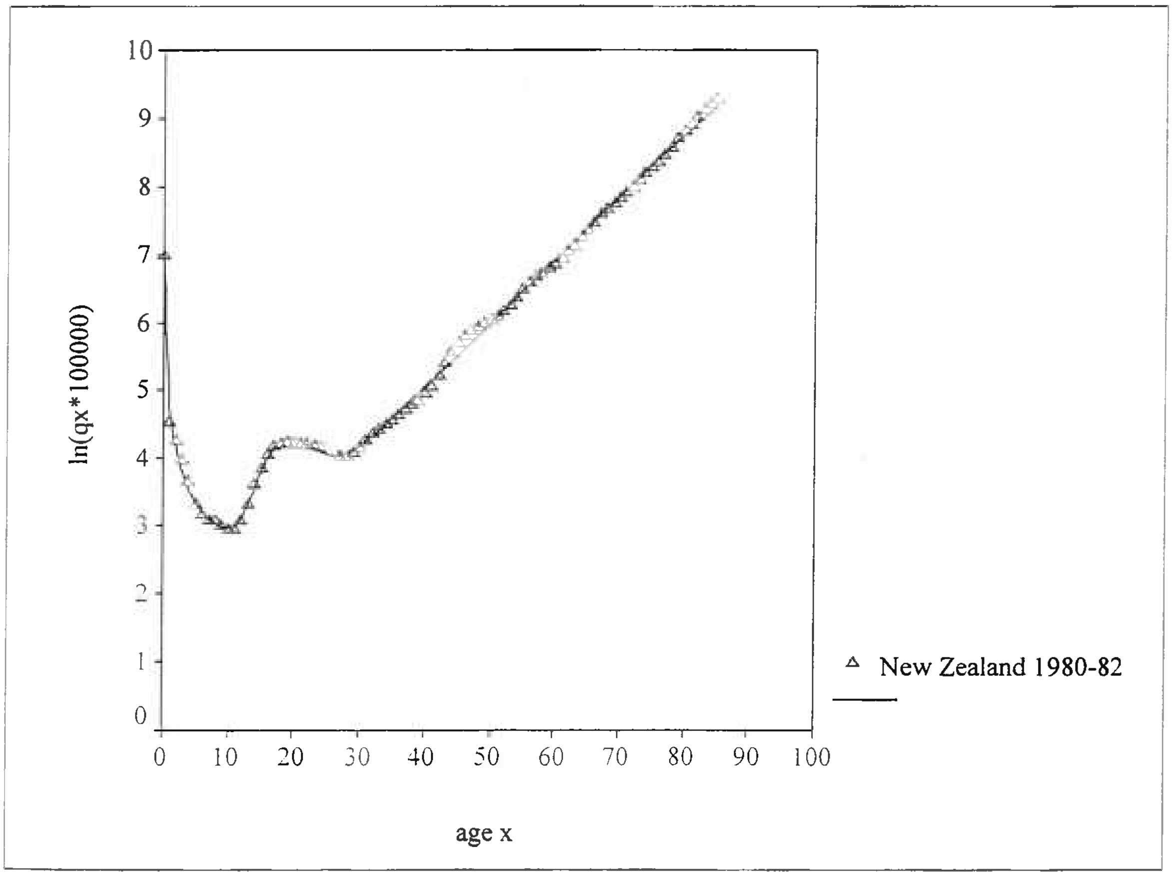

Figure A3.1: Empirical q_x - values (***points***) and estimations by HP9 formula (***solid line***s) for New-Zealand's 1980-82 females life table.

HP9 MODEL

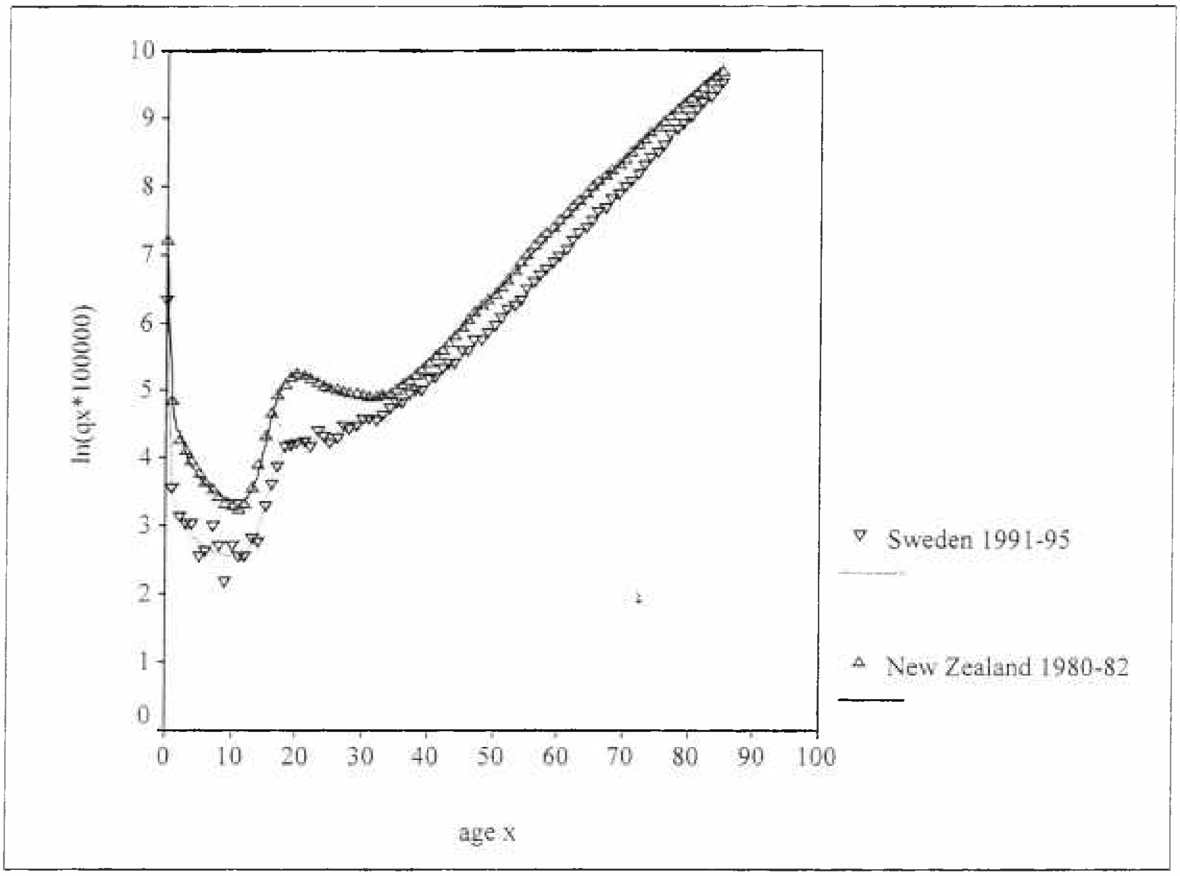

Figure A3.2: Empirical q_x - values (***points***) and estimations by HP9 formula (***solid line***s) for New-Zealand's 1980-82 and Sweden's 1991-95 males life table.

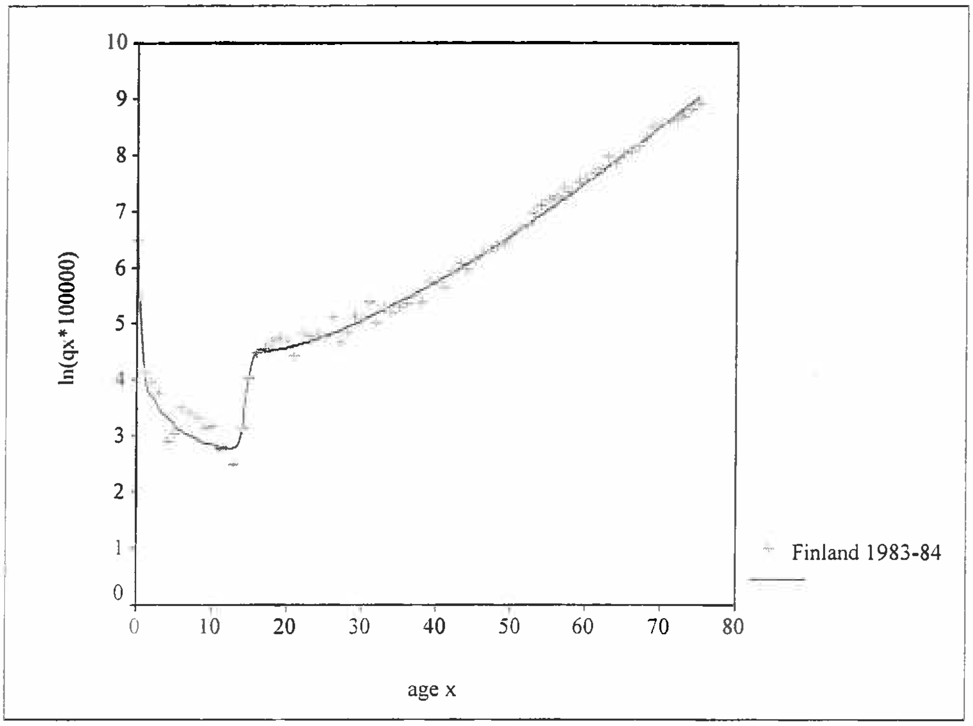

Figure A3.3: Empirical q_x - values (***points***) and estimations by HP9 formula (***solid line***) for Finland's 1983-84 males life table.

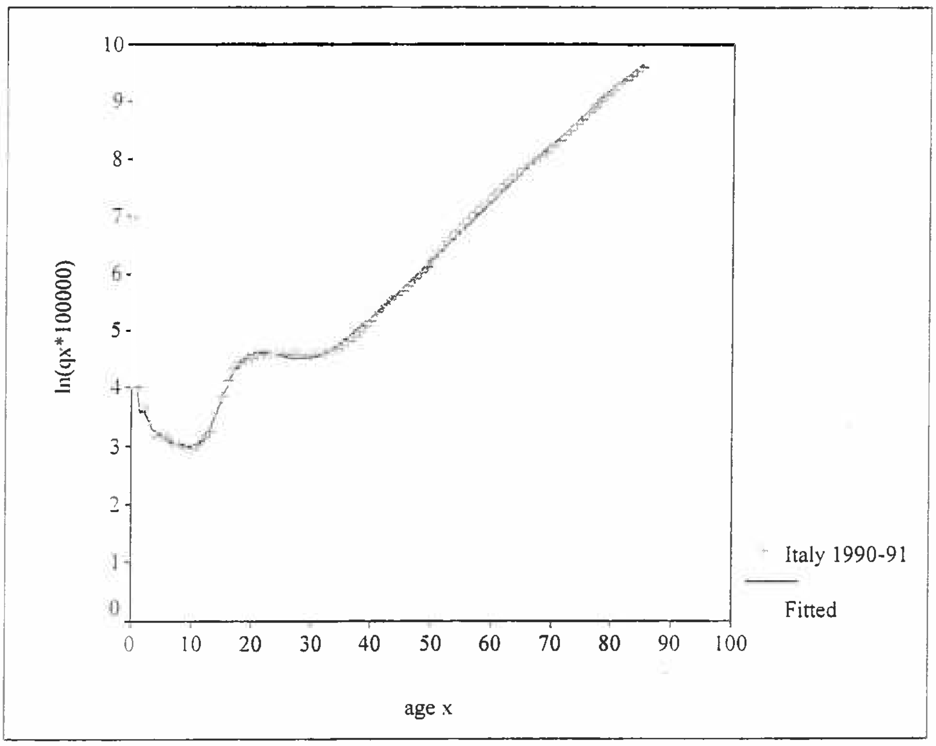

Figure A3.4: Empirical q_x - values (***points***) and estimations by HP9 formula (***solid line***) for Italy's 1990-91 males life table.

PART 4
HP9 ADJUSTED MODEL

HP9 ADJUSTED MODEL

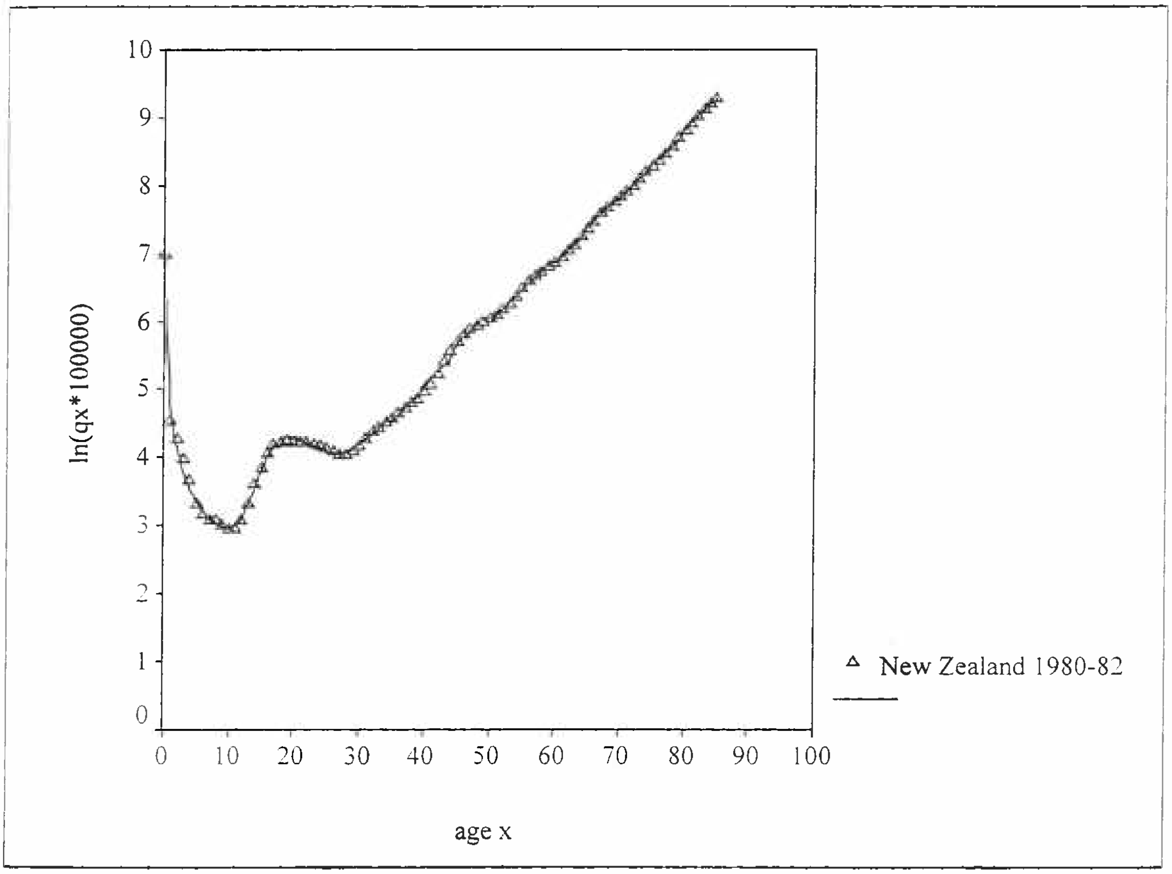

Figure A4.1: Empirical q_x - values (***points***) and estimations by HP9 adjusted formula (***solid line***s) for New-Zealand's 1980-82 females life table.

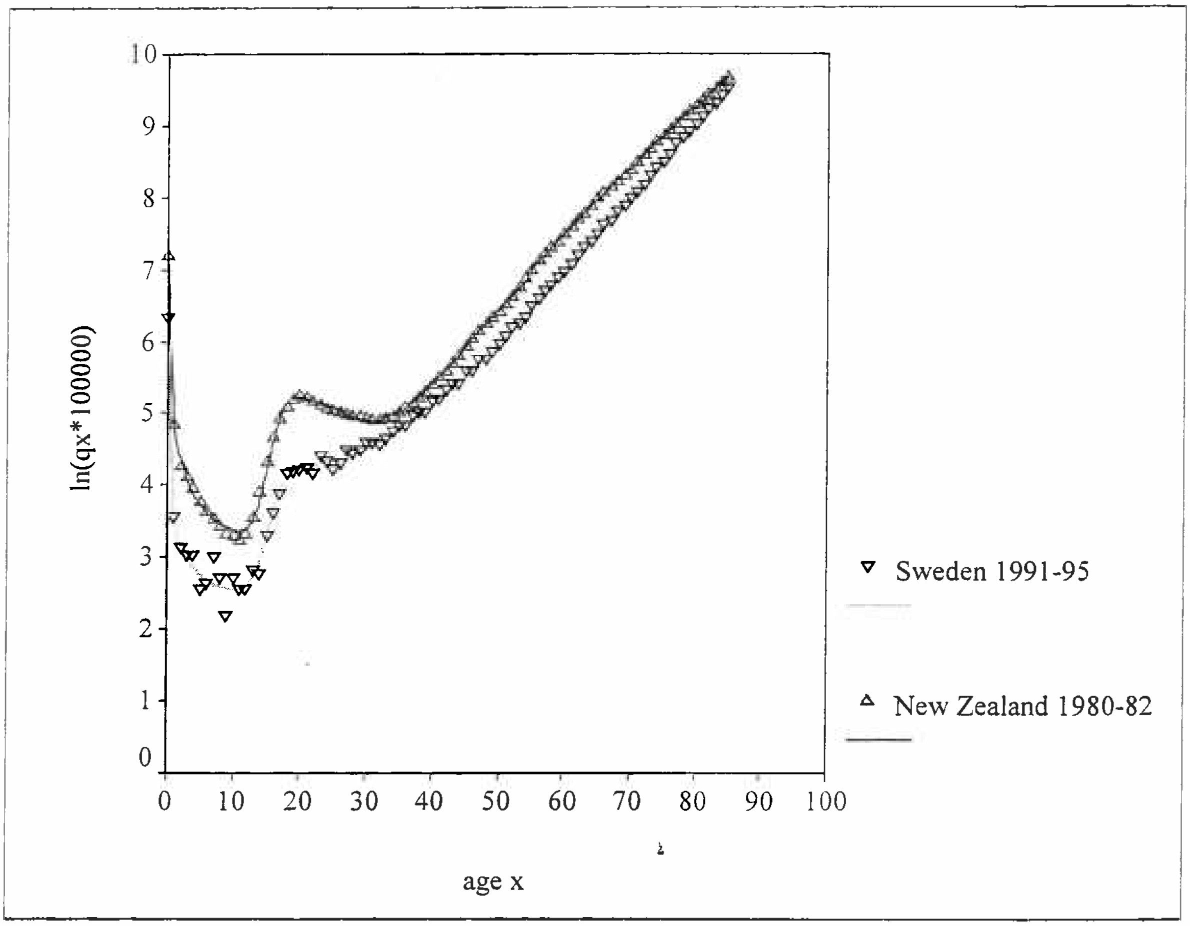

Figure A4.2: Empirical q_x - values (***points***) and estimations by HP9 adjusted formula (*solid lines*) for New-Zealand's 1980-82 and Sweden's 1991-95 males life table.

HP9 ADJUSTED MODEL

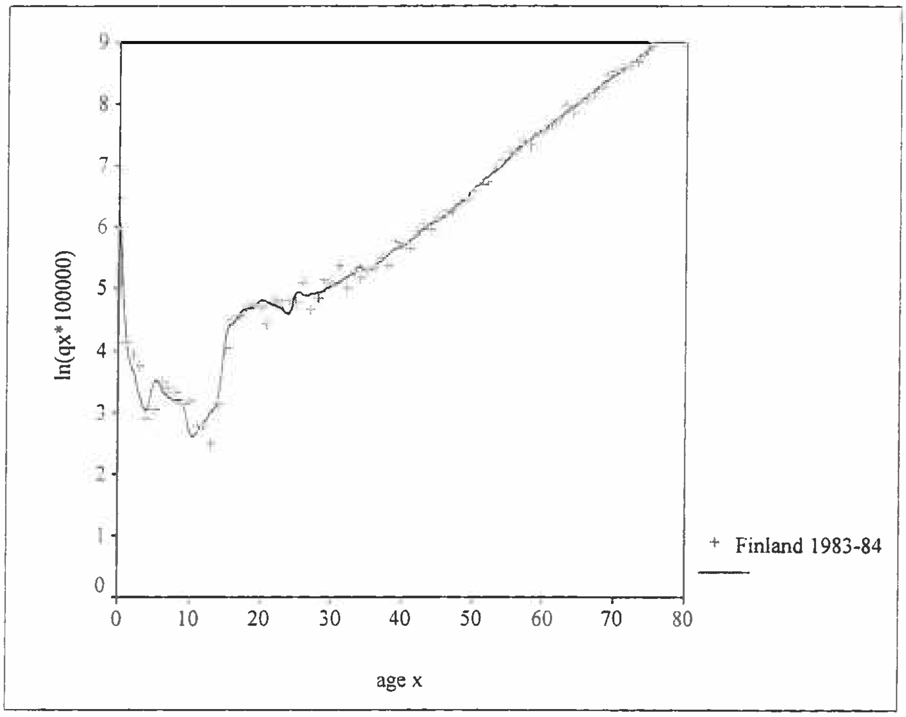

Figure A4.3: Empirical q_x - values (***points***) and estimations by HP9 adjusted formula (***solid line***) for Finland's 1983-84 males life table.

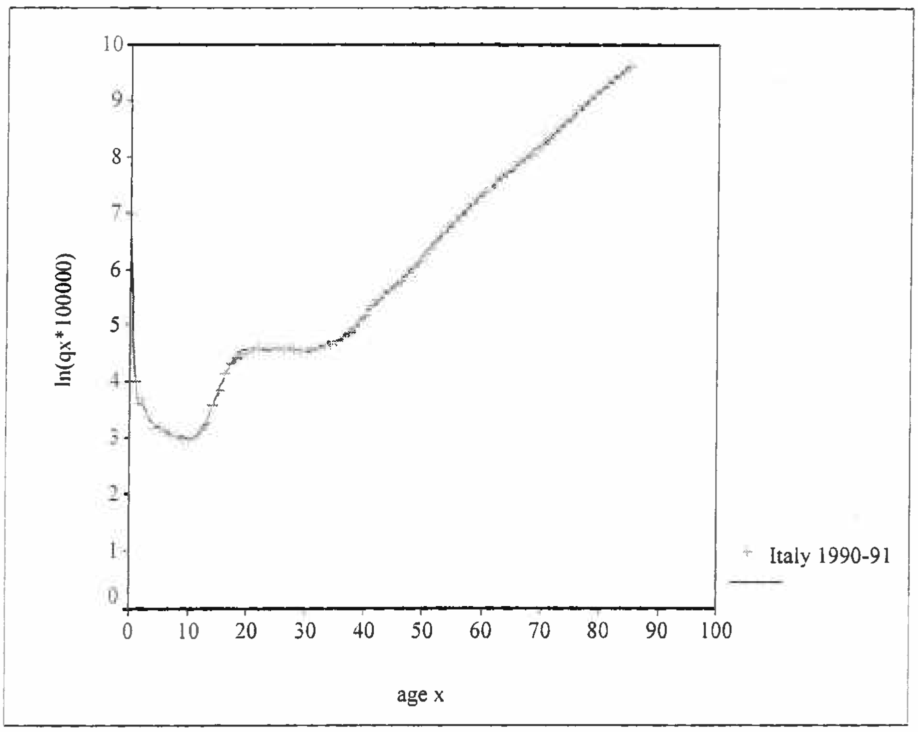

Figure A4.4: Empirical q_x - values (***points***) and estimations by HP9 adjusted formula (***solid line***) for Italy's 1990-91 males life table.

PART 5

THE NON - PARAMETRIC RELATIONAL TECHNIQUE

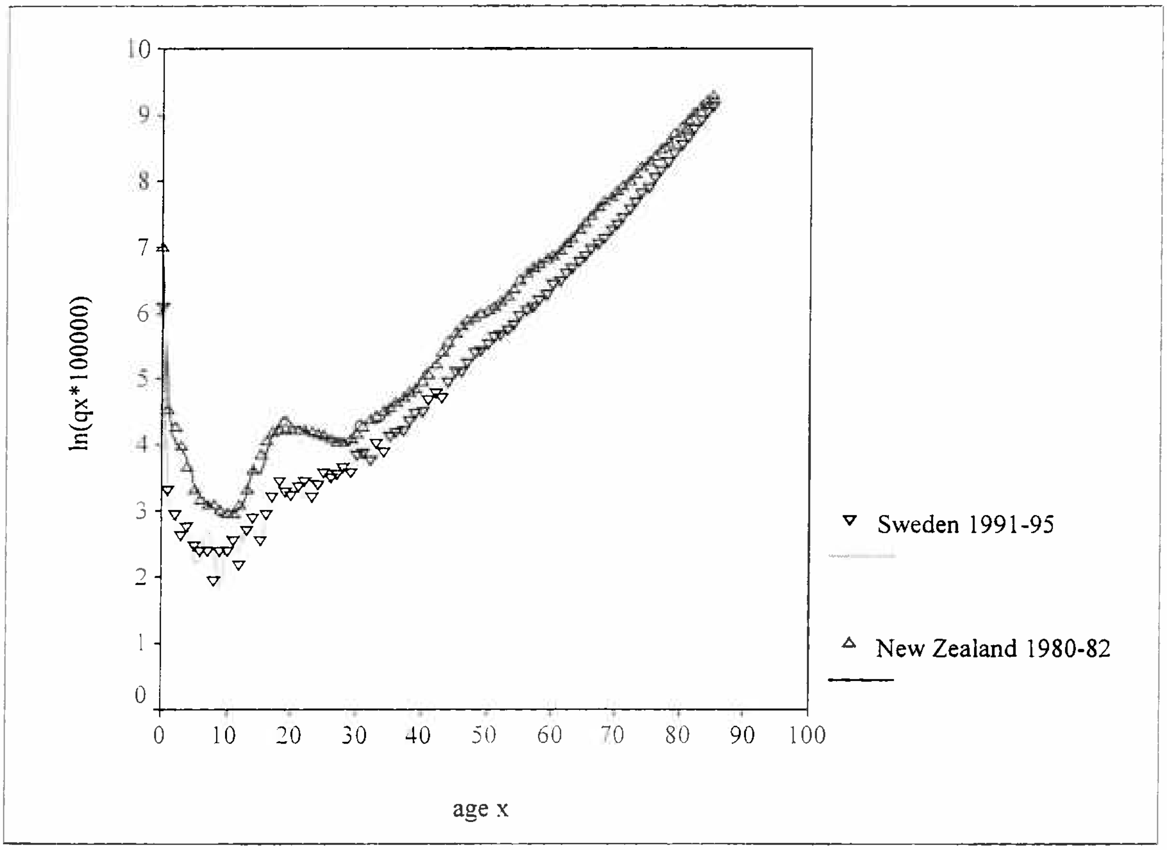

Figure A5.1: Empirical q_x - values (***points***) and estimations by Kostaki's non parametric technique (***solid lines***) for New-Zealand's 1980-82 and Sweden's 1991-95 females life table.

Figure A5.2: Empirical q_x - values (***points***) and estimations by Kostaki's non parametric technique (***solid lines***) for New-Zealand's 1980-82 and Sweden's 1991-95 males life table.

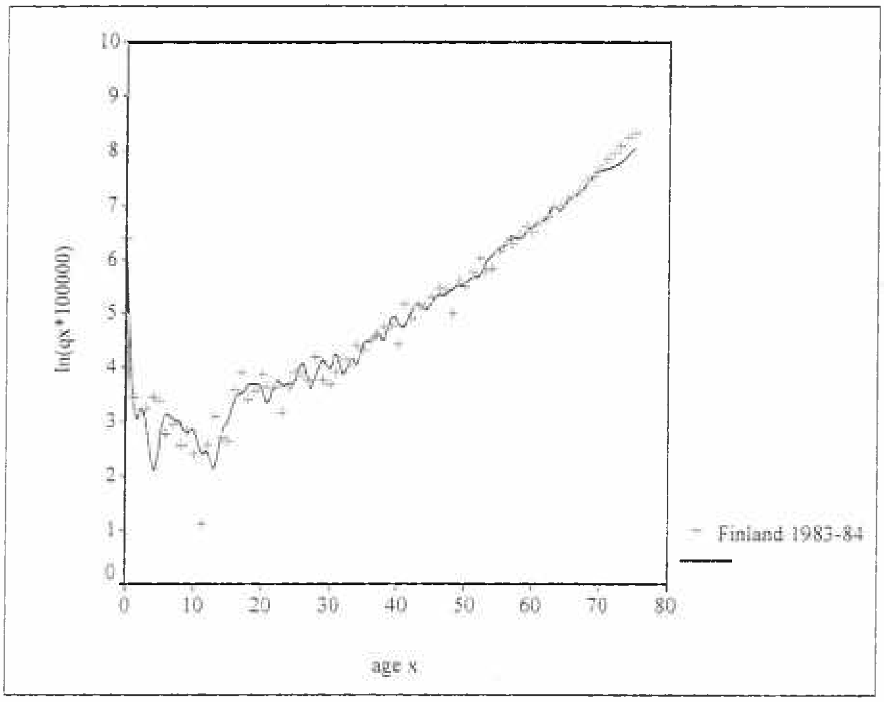

Figure A5.3: Empirical q_x - values (***points***) and estimations by Kostaki's non parametric technique (***solid line***) for Finland's 1983-84 females life table.

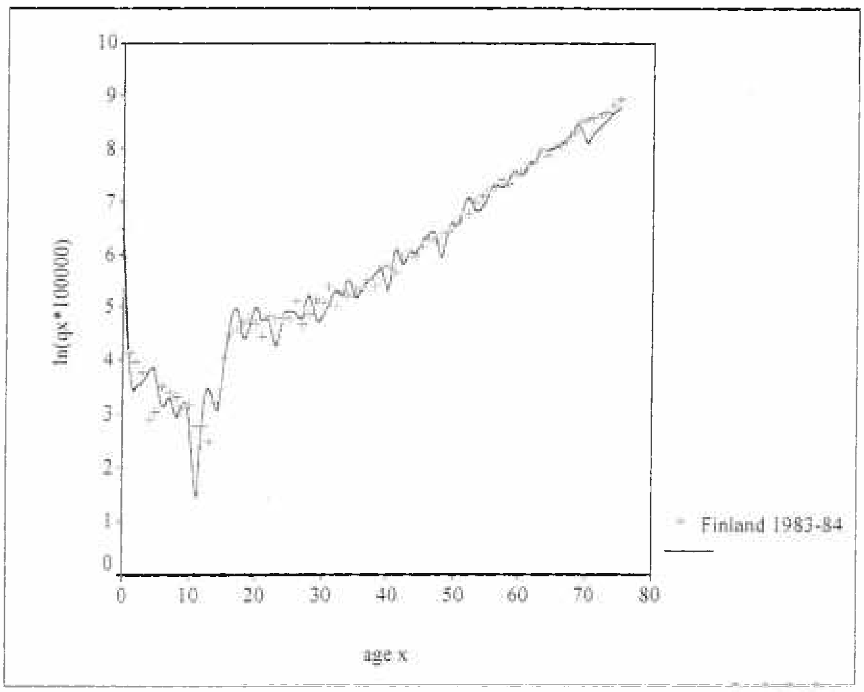

Figure A5.4: Empirical q_x - values (***points***) and estimations by Kostaki's non parametric technique (***solid line***) for Finland's 1983-84 males life table.

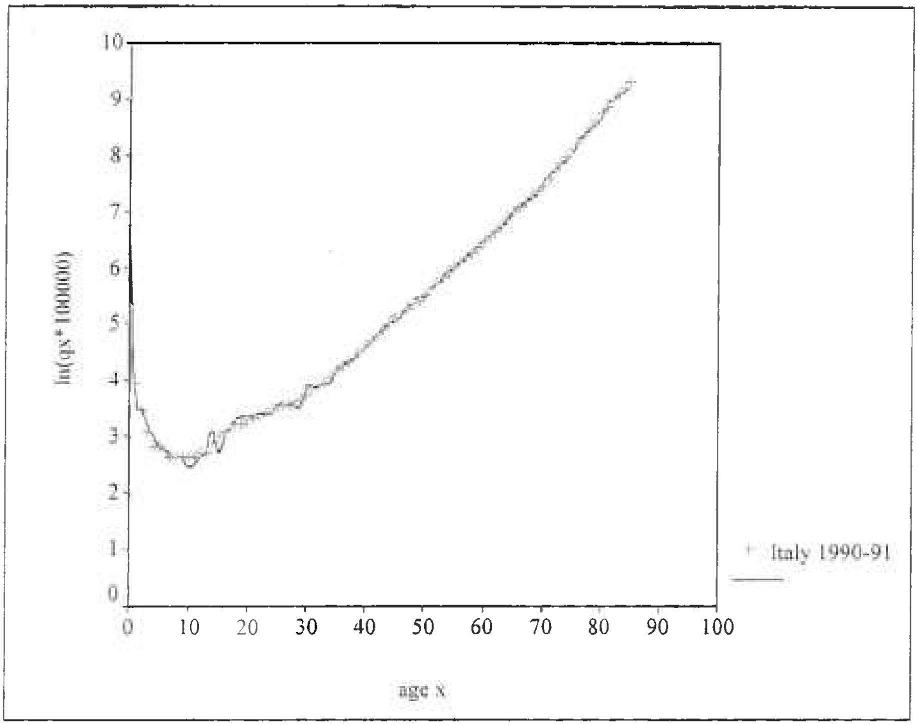

Figure A5.5: Empirical q_x - values (***points***) and estimations by Kostaki's non parametric technique (***solid line***) for Italy's 1990-91 females life table.

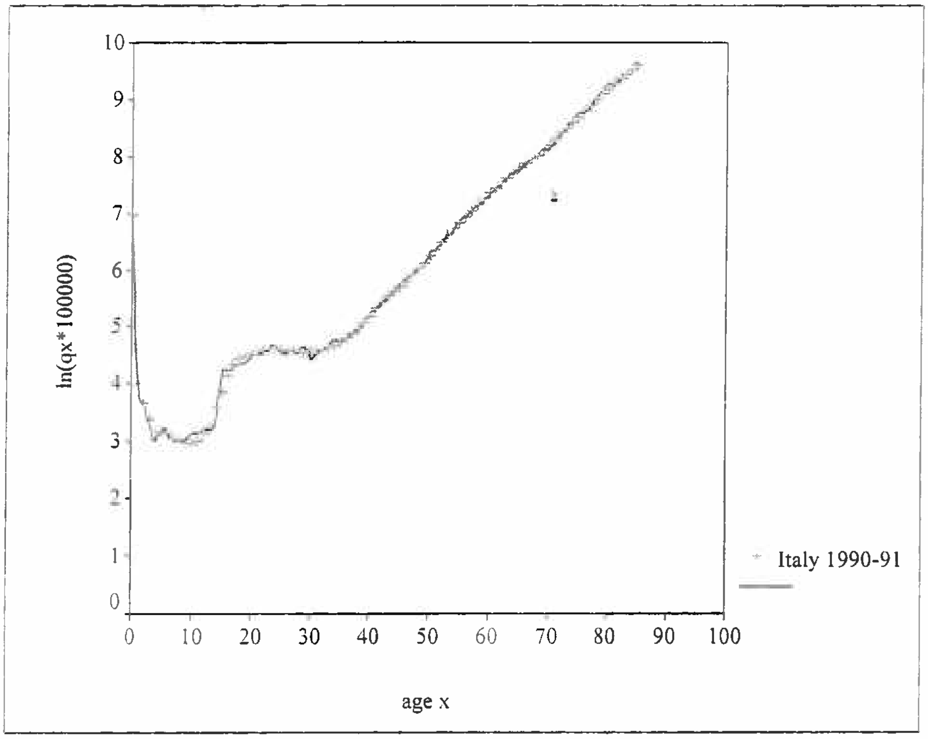

Figure A5.6: Empirical q_x- values (***points***) and estimations by Kostaki's non parametric technique (***solid line***) for Italy's 1990-91 males life table.

PART 6

SIX - POINT LAGRANGEAN INTERPOLATION

Figure A6.1: Empirical q_x- values (***points***) and estimations by six-point lagrangean interpolation (***solid lines***) for New-Zealand's 1980-82 and Sweden's 1991-95 females life table.

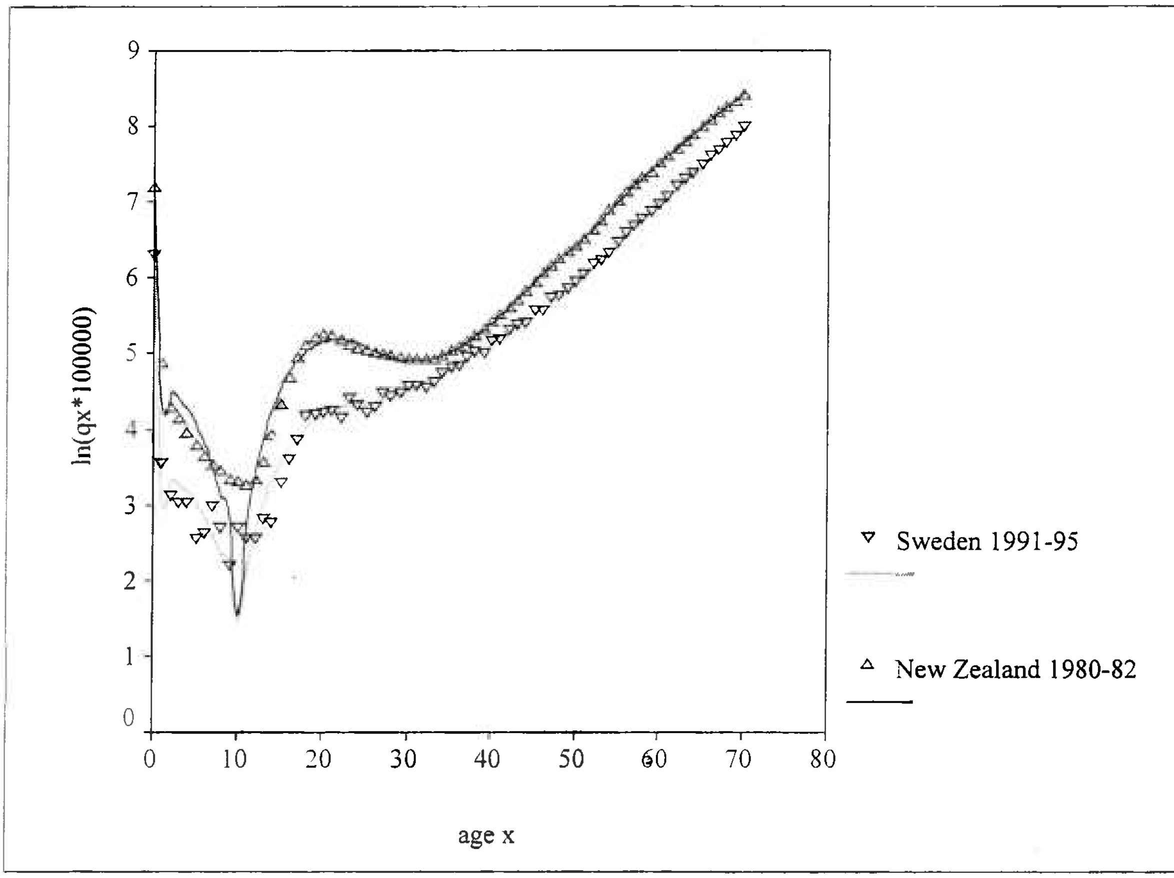

Figure A6.2: Empirical q_x- values (***points***) and estimations by six-point lagrangean interpolation (***solid lines***) for New-Zealand's 1980-82 and Sweden's 1991-95 males life table.

SIX - POINT LAGRANGEAN INTERPOLATION

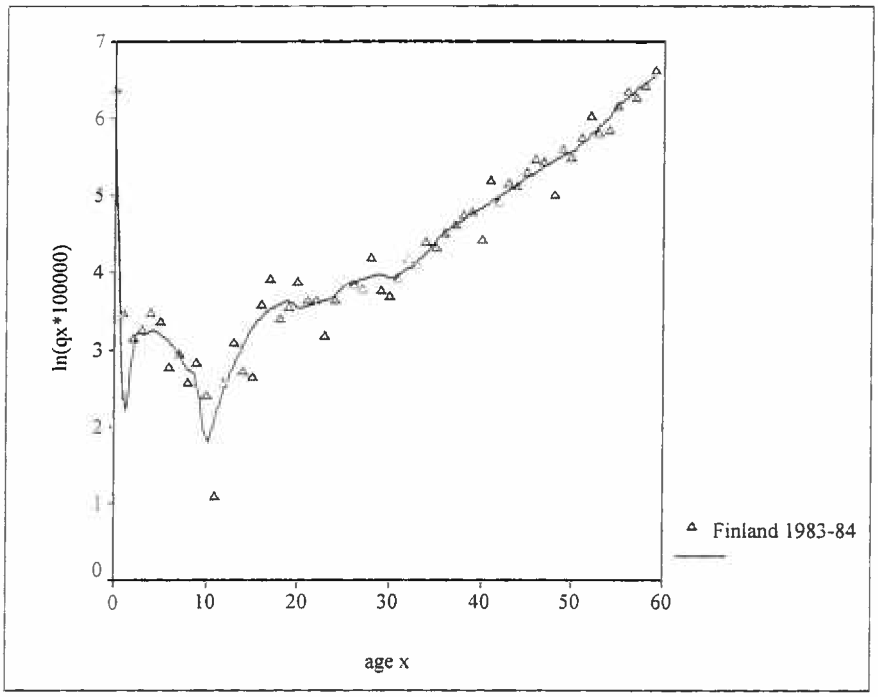

Figure A6.3: Empirical q_x- values (***points***) and estimations by six-point lagrangean interpolation (***solid line***) for Finland's 1983-84 females life table.

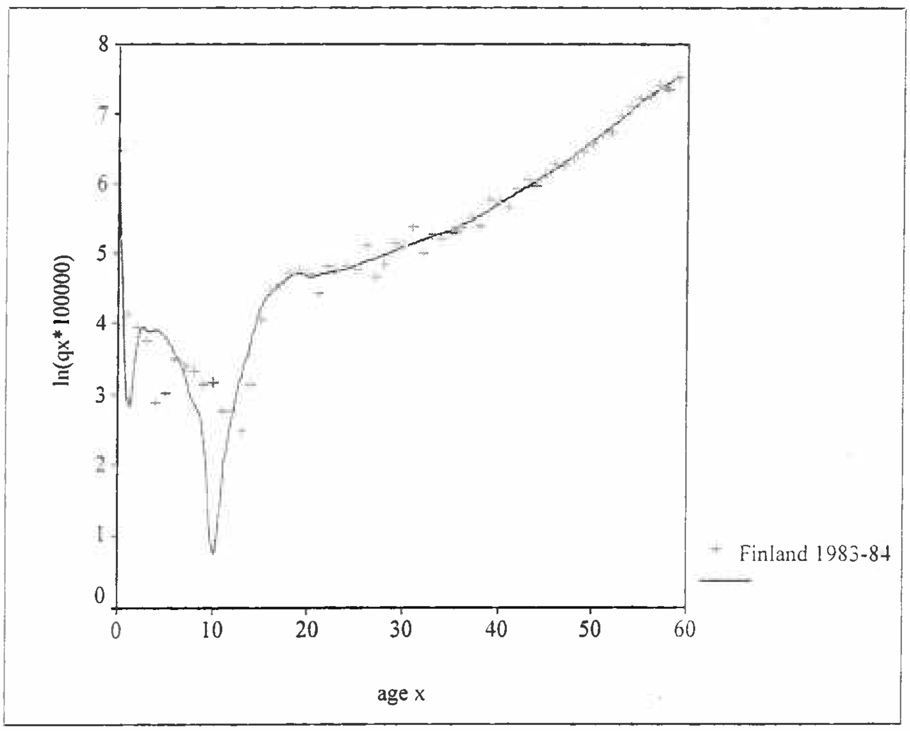

Figure A6.4: Empirical q_x- values (***points***) and estimations by six-point lagrangean interpolation (***solid line***) for Finland's 1983-84 males life table.

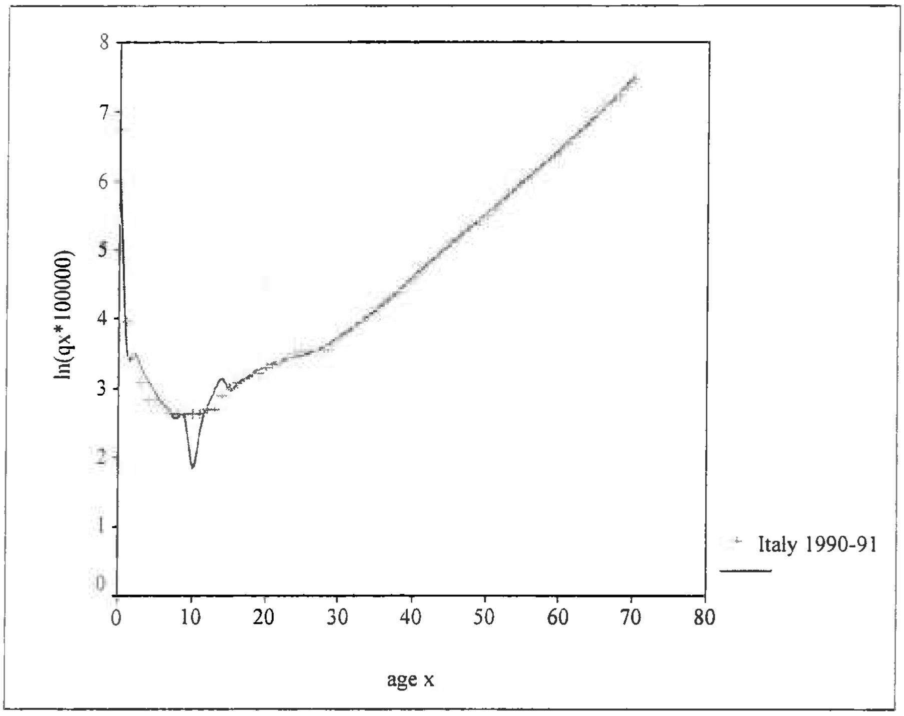

Figure A6.5: Empirical q_x - values (***points***) and estimations by six-point lagrangean interpolation (***solid line***) for Italy's 1990-91 females life table.

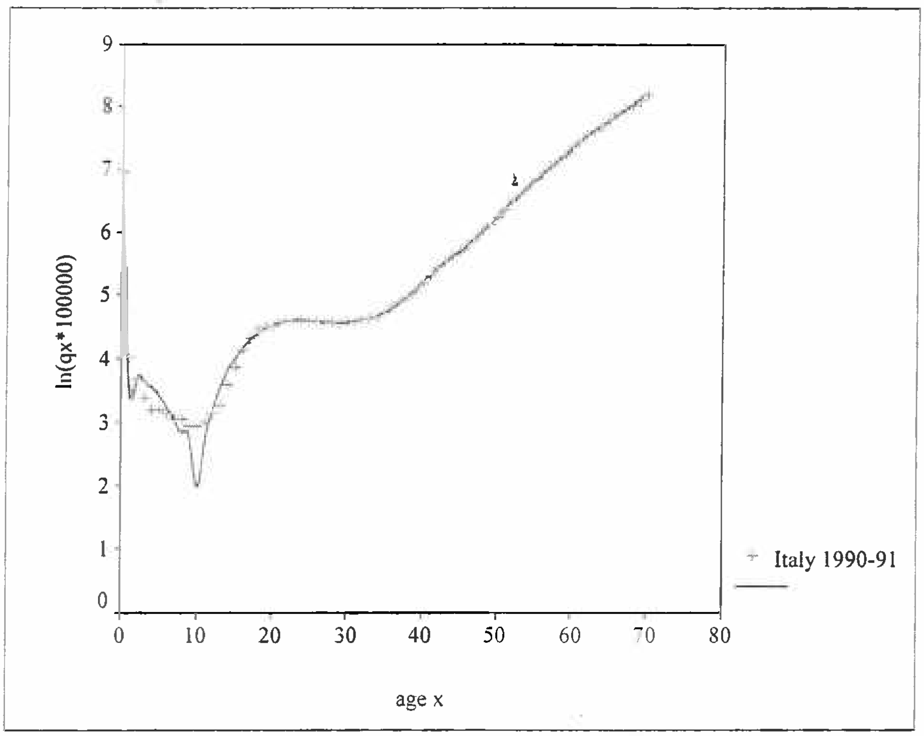

Figure A6.6: Empirical q_x - values (***points***) and estimations by six-point lagrangean interpolation (***solid line***) for Italy's 1990-91 males life table.

PART 7

NATURAL CUBIC SPLINE INTERPOLATION

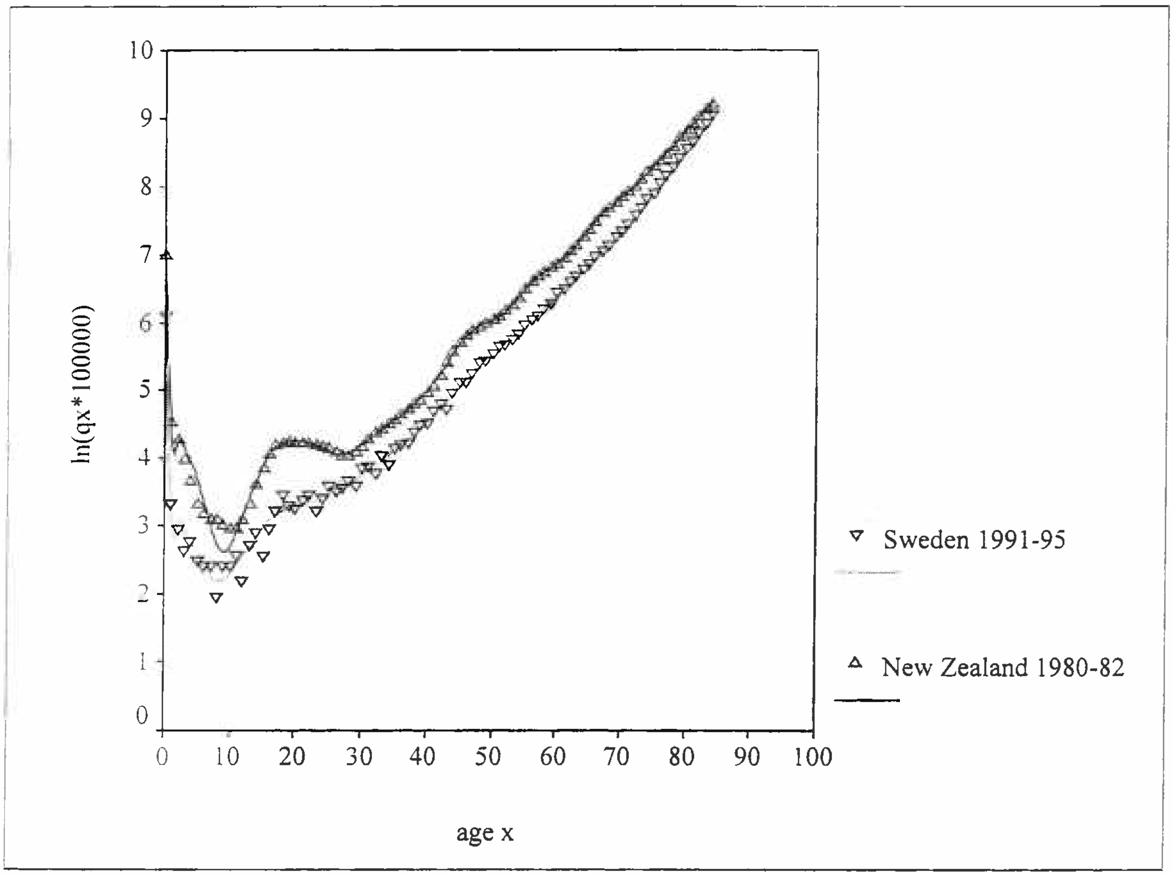

Figure A7.1: Empirical q_x - values (***points***) and estimations by natural cubic spline interpolation (***solid line***s) for New-Zealand's 1980-82 and Sweden's 1991-95 females life table.

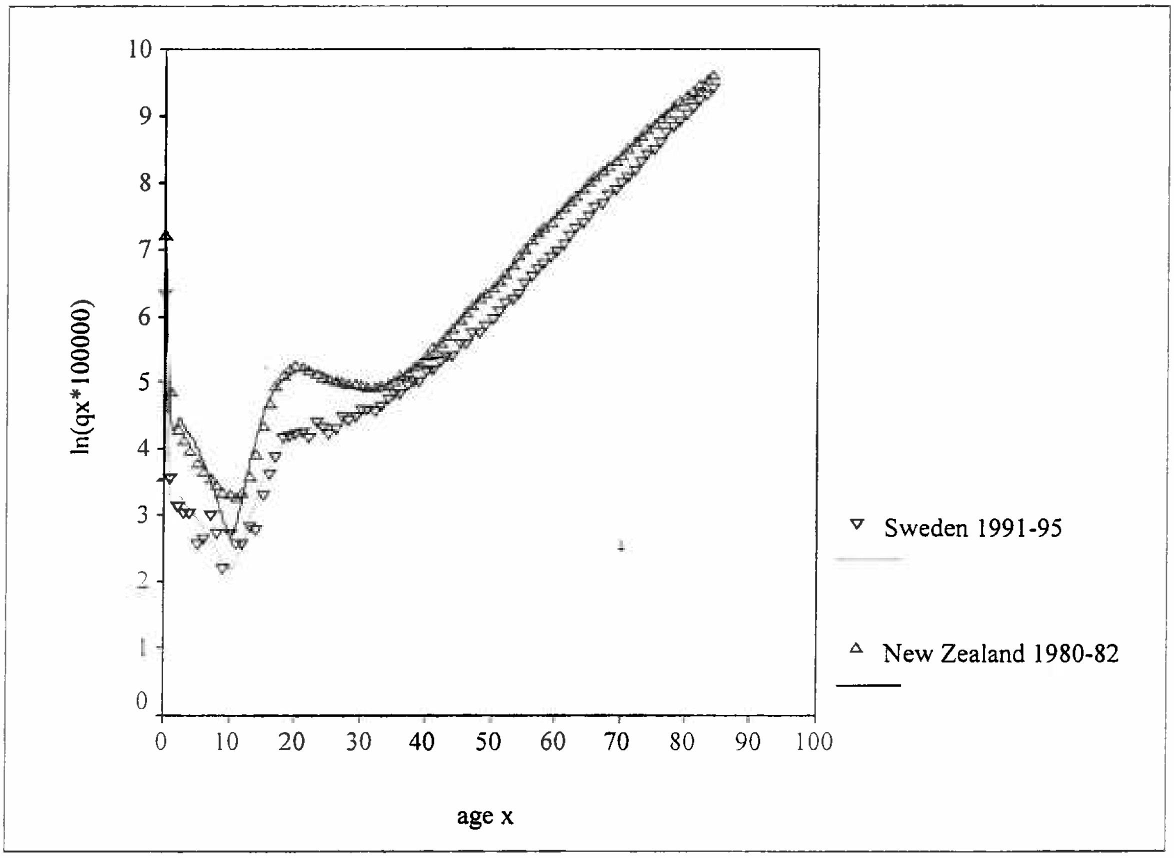

Figure A7.2: Empirical q_x - values (***points***) and estimations by natural cubic spline interpolation (***solid lines***) for New-Zealand's 1980-82 and Sweden's 1991-95 males life table.

NATURAL CUBIC SPLINE INTERPOLATION

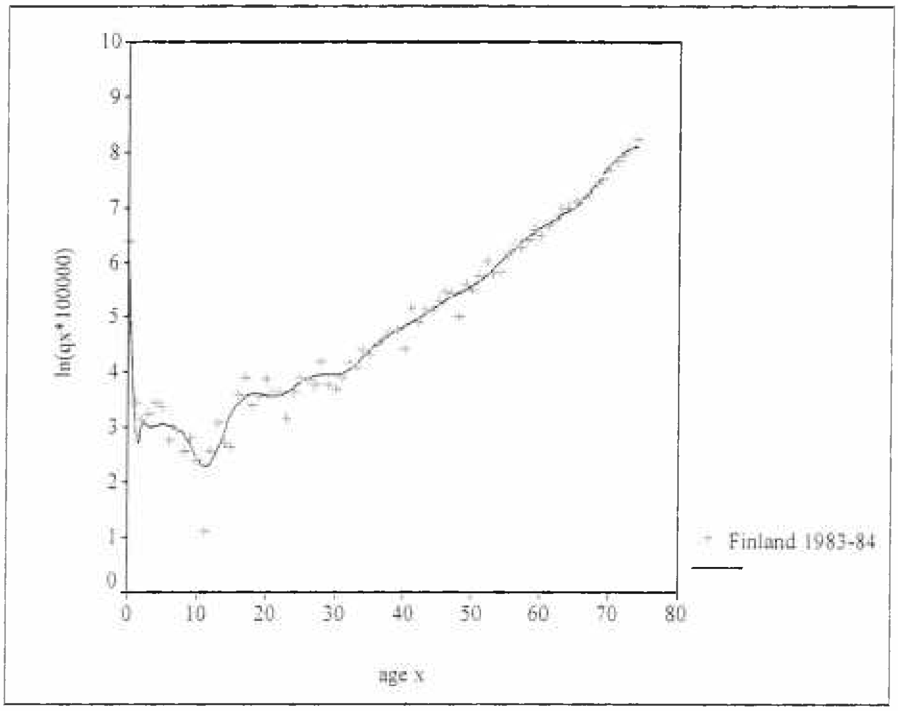

Figure A7.3: Empirical q_x - values (***points***) and estimations by natural cubic spline interpolation (***solid line***) for Finland's 1983-84 females life table.

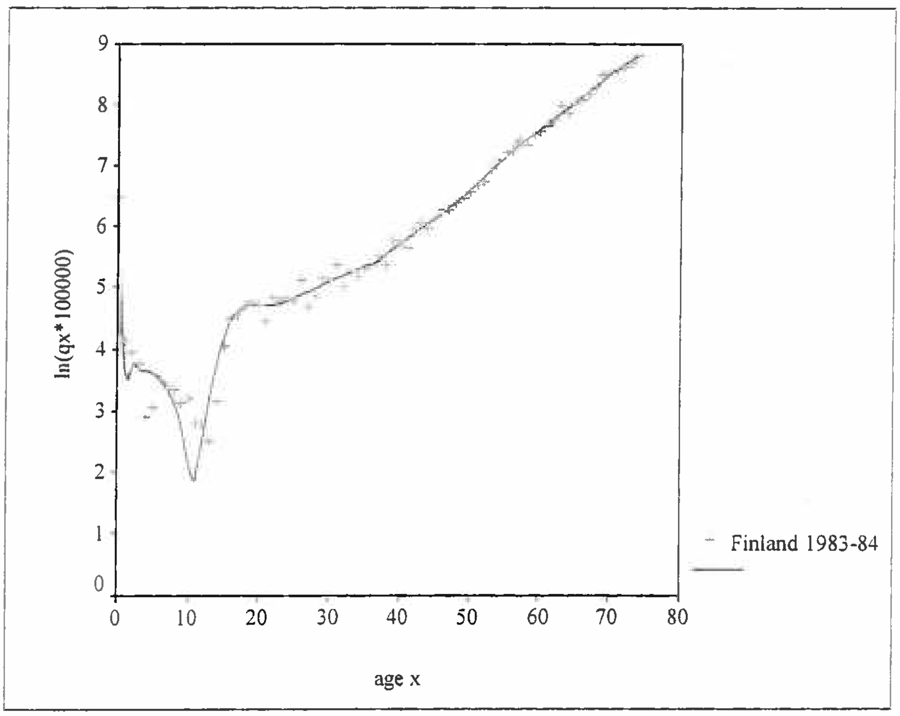

Figure A7.4: Empirical q_x - values (***points***) and estimations by natural cubic spline interpolation (***solid line***) for Finland's 1983-84 males life table.

NATURAL CUBIC SPLINE INTERPOLATION

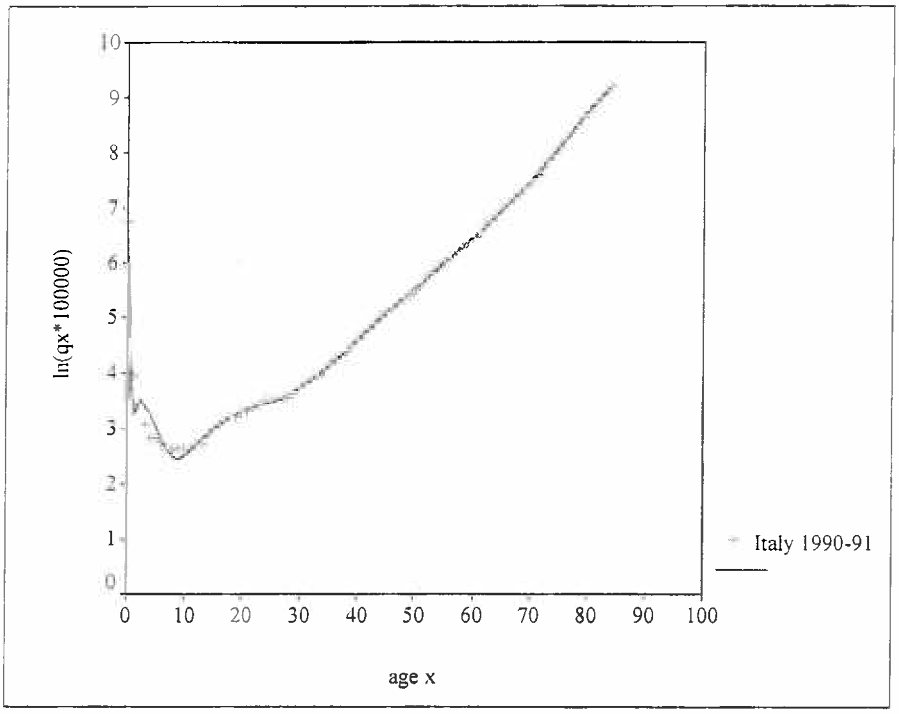

Figure A7.5: Empirical q_x - values (***points***) and estimations by natural cubic spline interpolation (***solid line***) for Italy's 1990-91 females life table.

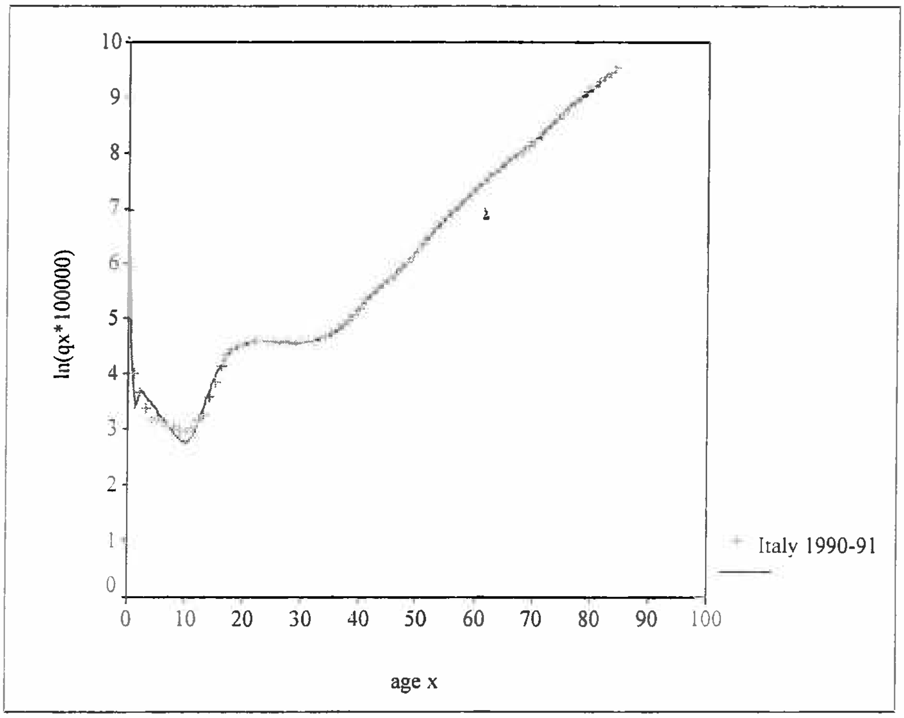

Figure A7.6: Empirical q_x - values (***points***) and estimations by natural cubic spline interpolation (***solid line***) for Italy's 1990-91 males life table.

PART 8

COMPLETE CUBIC SPLINE INTERPOLATION

COMPLETE CUBIC SPLINE INTERPOLATION

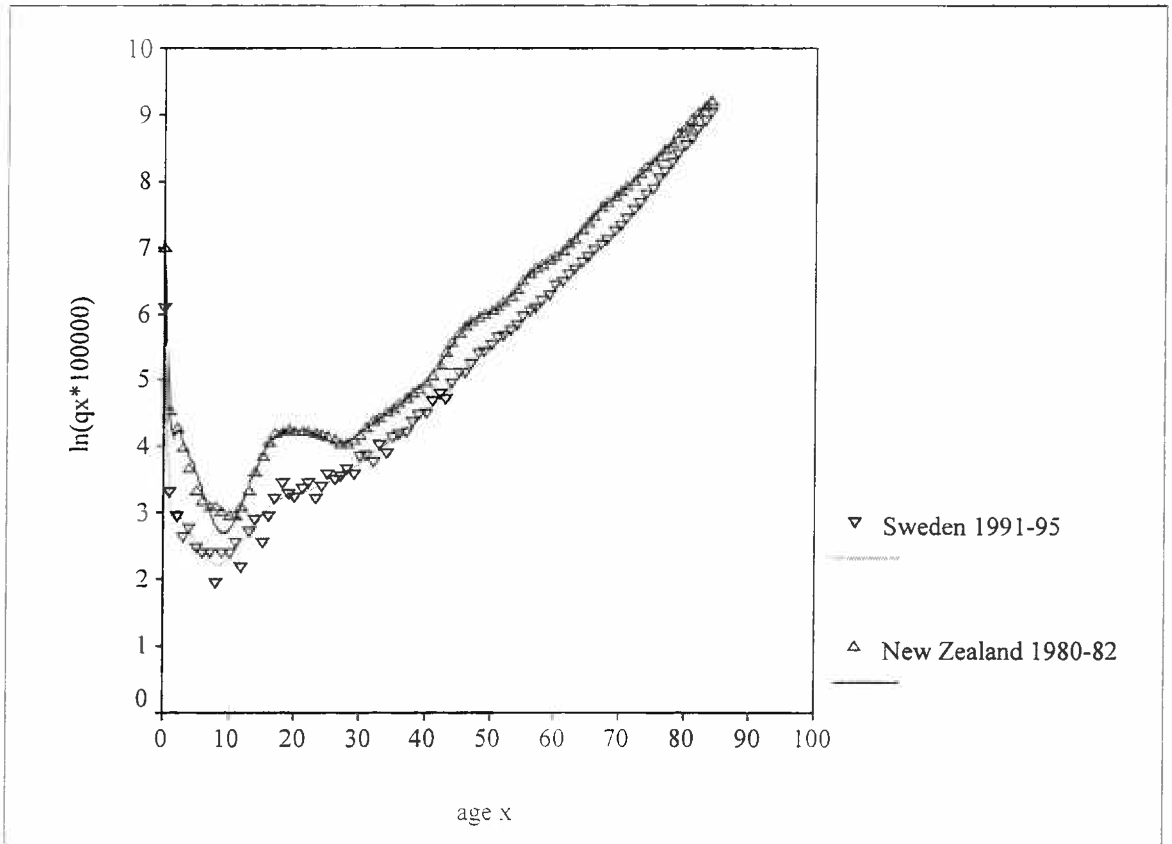

Figure A8.1: Empirical q_x - values (***points***) and estimations by complete cubic spline interpolation (***solid line***s) for New-Zealand's 1980-82 and Sweden's 1991-95 females life table.

COMPLETE CUBIC SPLINE INTERPOLATION

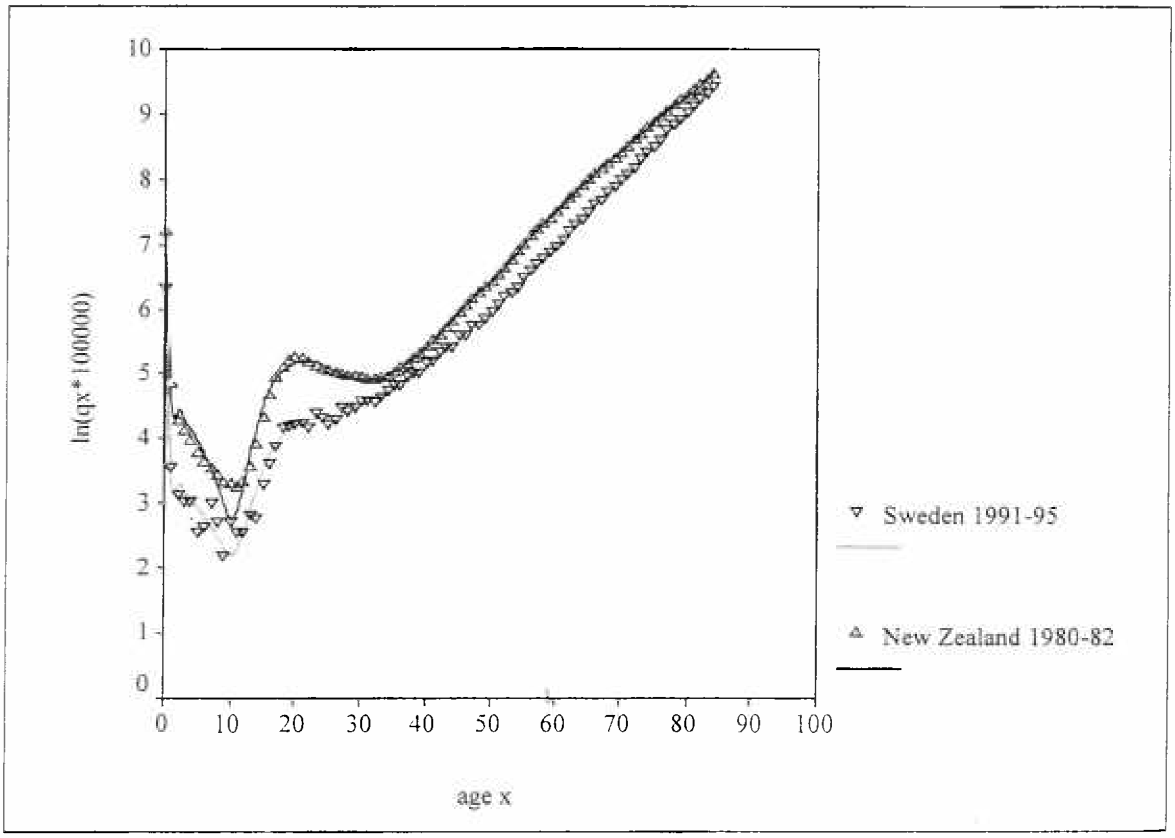

Figure A8.2: Empirical q_x - values (***points***) and estimations by complete cubic spline interpolation (***solid line***s) for New-Zealand's 1980-82 and Sweden's 1991-95 males life table.

COMPLETE CUBIC SPLINE INTERPOLATION

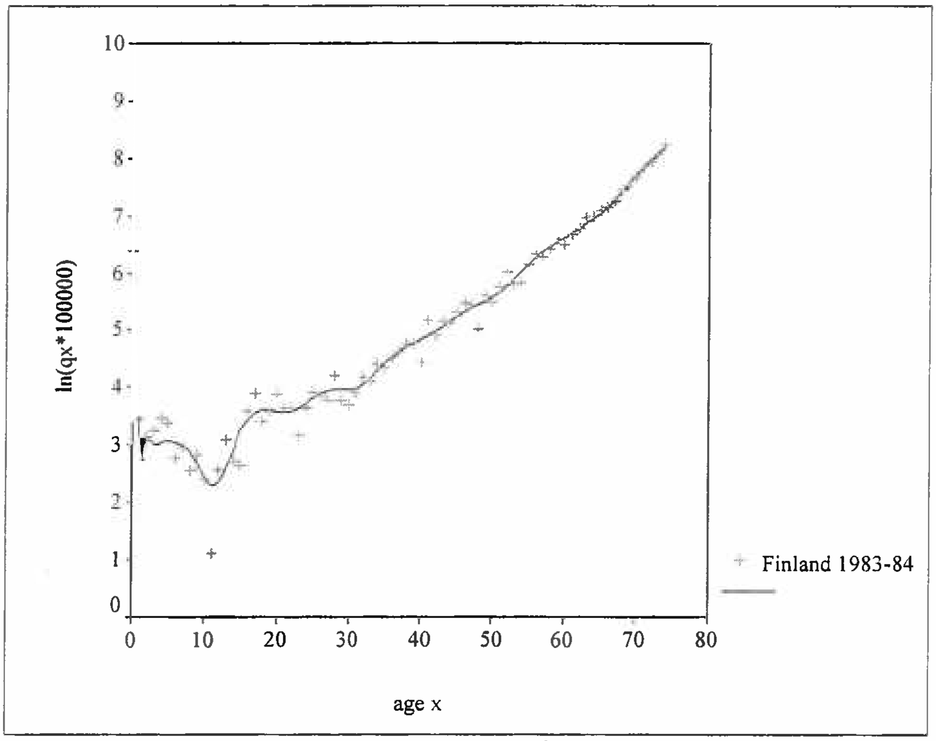

Figure A8.3: Empirical q_x - values (***points***) and estimations by complete cubic spline interpolation (***solid line***) for Finland's 1983-84 females life table.

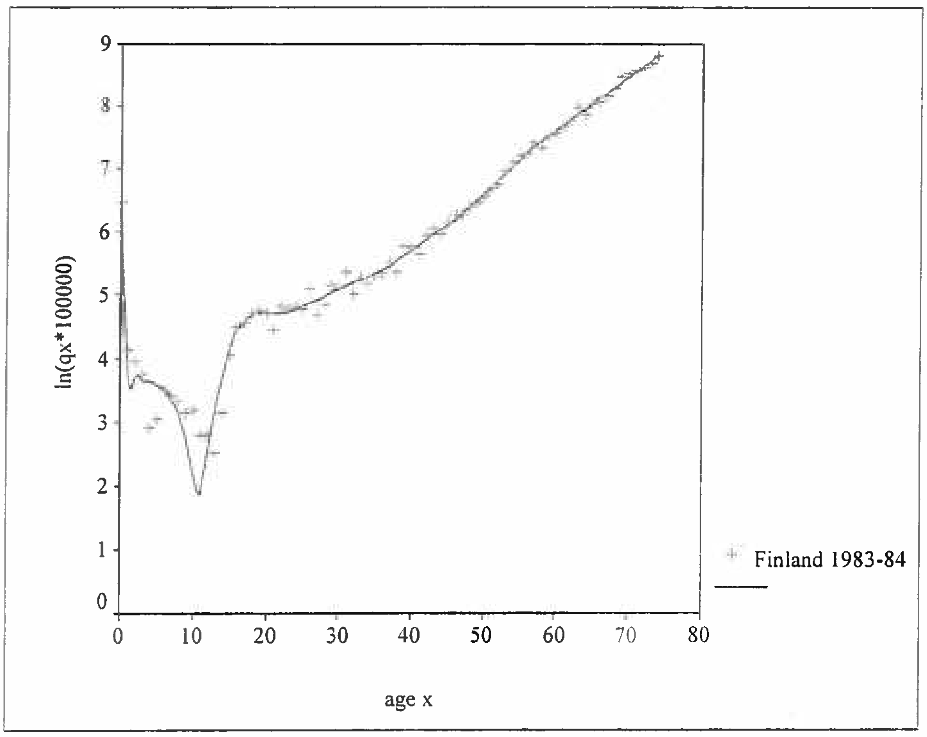

Figure A8.4: Empirical q_x - values (***points***) and estimations by complete cubic spline interpolation (***solid line***) for Finland's 1983-84 males life table.

COMPLETE CUBIC SPLINE INTERPOLATION

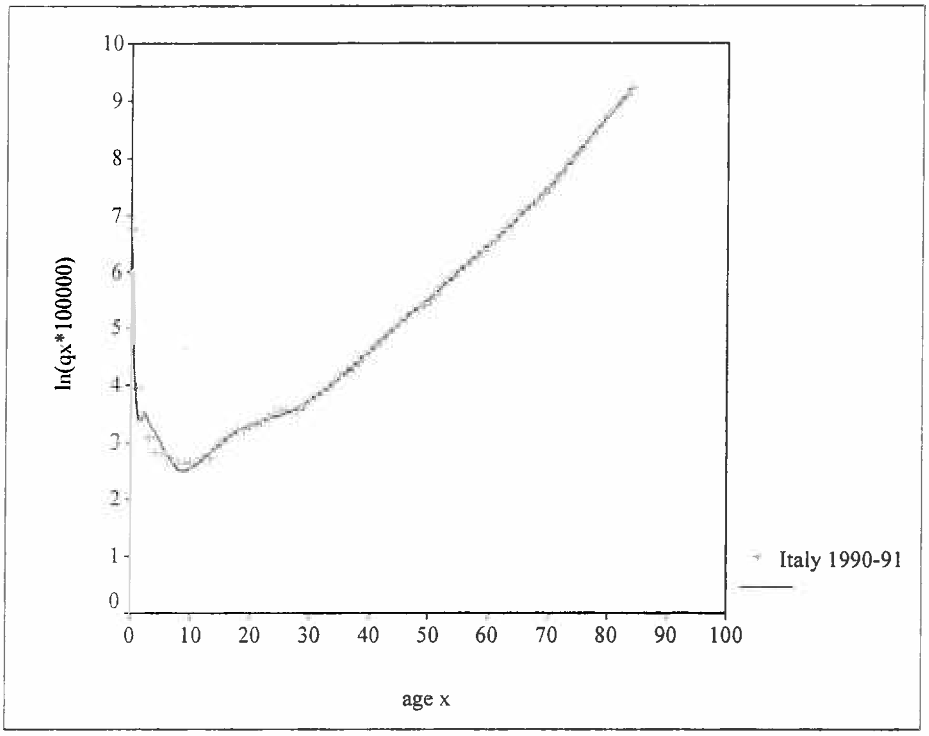

Figure A8.5: Empirical q_x - values (***points***) and estimations by complete cubic spline interpolation (***solid line***) for Italy's 1990-91 females life table.

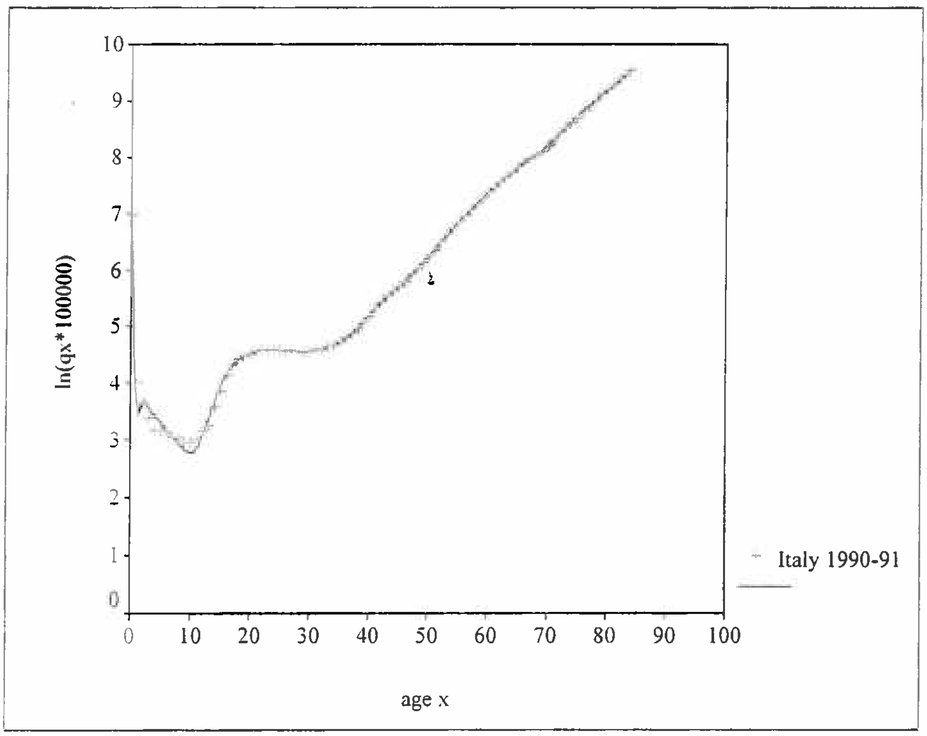

Figure A8.6: Empirical q_x - values (***points***) and estimations by complete cubic spline interpolation (***solid line***) for Italy's 1990-91 males life table.

PART 9

REED'S TECHNIQUE

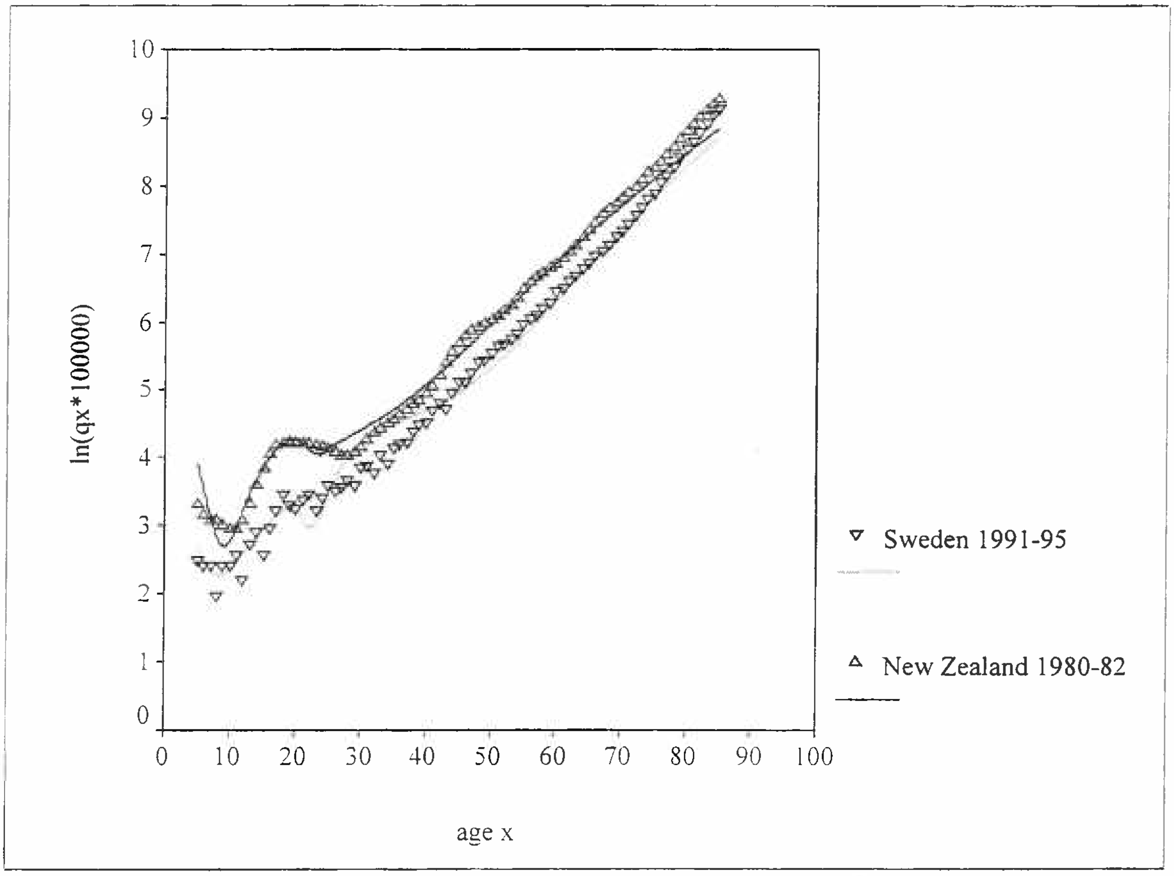

Figure A9.1: Empirical q_x - values (***points***) and estimations by Reed's expanding technique (***solid line**s*) for New-Zealand's 1980-82 and Sweden's 1991-95 females life table.

Figure A9.2: Empirical q_x - values (***points***) and estimations Reed's expanding technique (***solid line***s) for New-Zealand's 1980-82 and Sweden's 1991-95 males life

REED'S TECHNIQUE

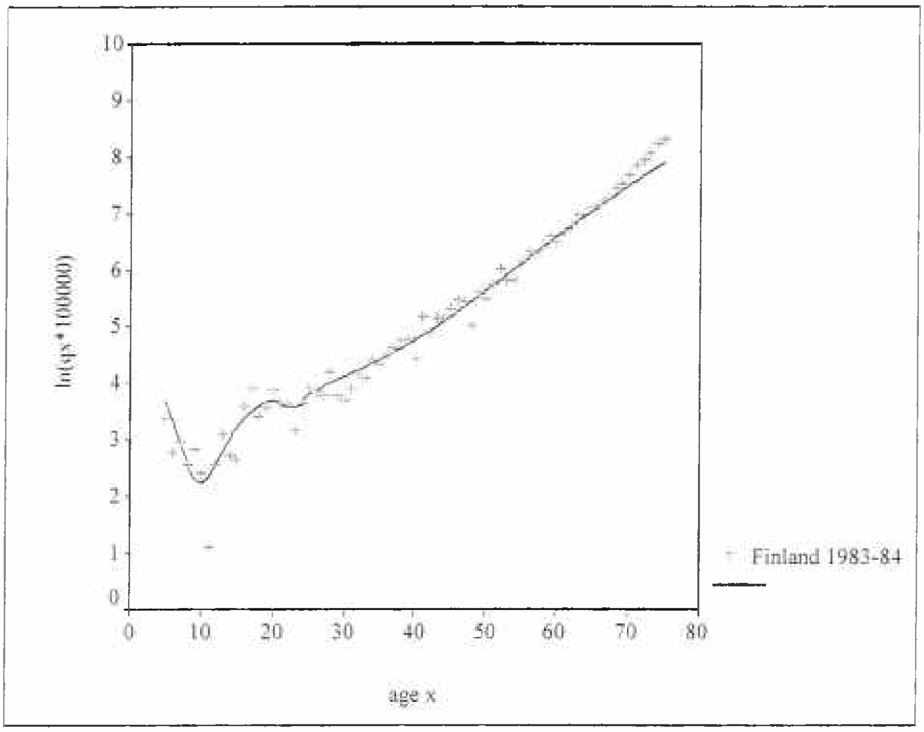

Figure A9.3: Empirical q_x - values (***points***) and estimations by Reed's expanding technique (***solid line***) for Finland's 1983-84 females life table.

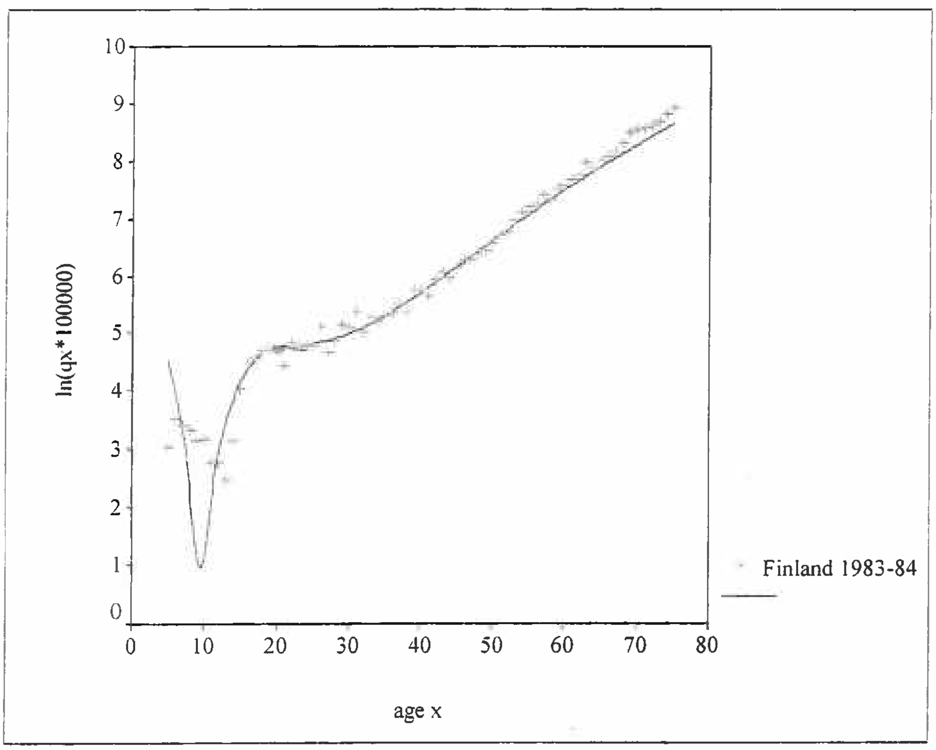

Figure A9.4: Empirical q_x - values (***points***) and estimations by Reed's expanding technique (***solid line***) for Finland's 1983-84 males life table.

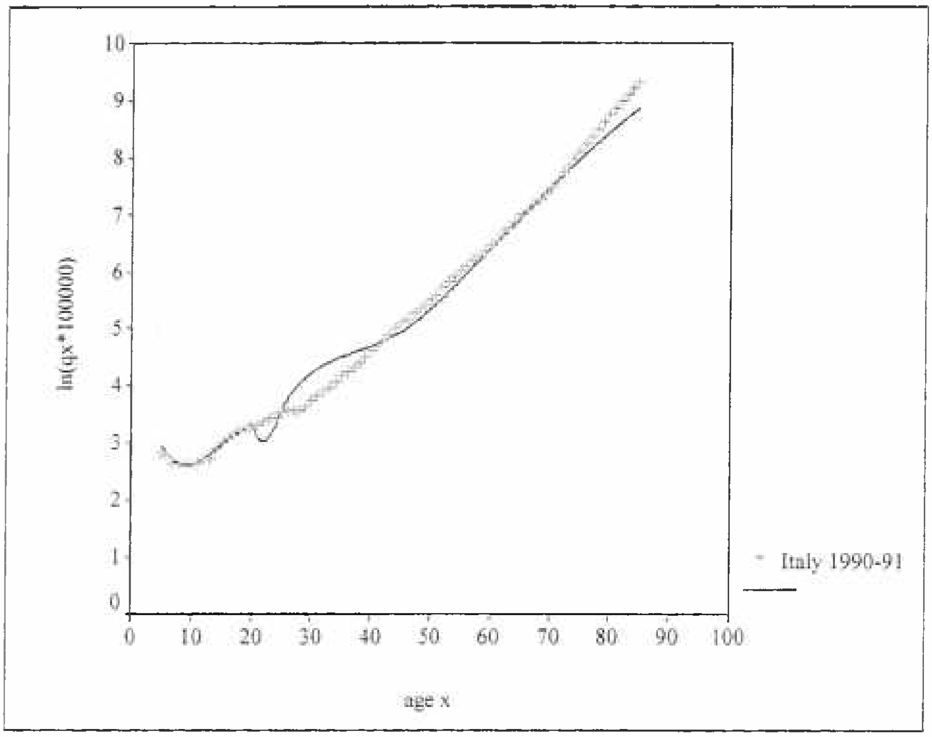

Figure A9.5: Empirical q_x - values (***points***) and estimations by Reed's expanding technique (***solid line***) for Italy's 1990-91 females life table.

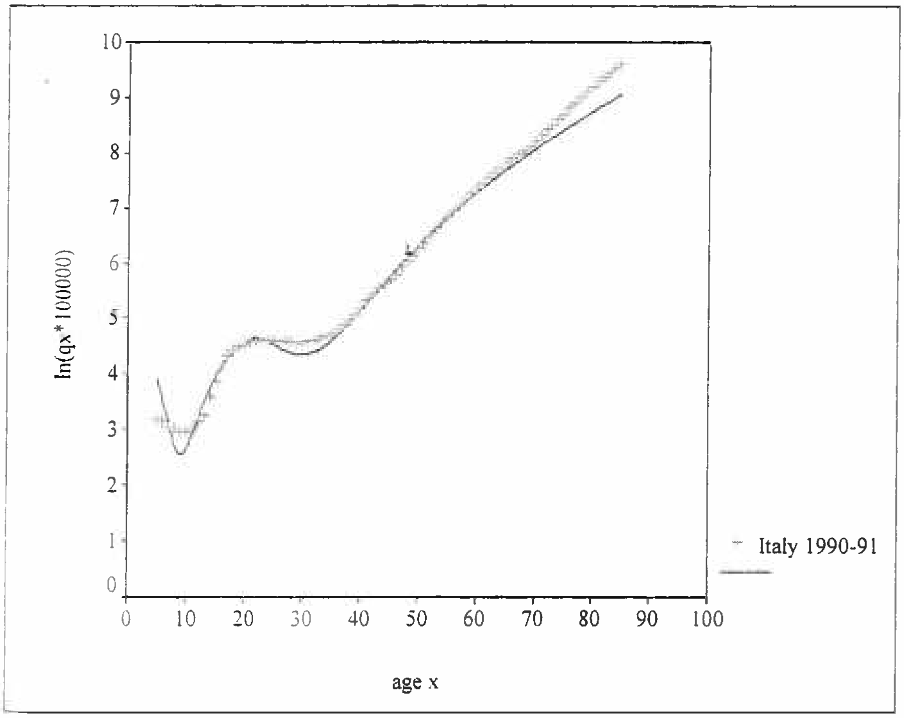

Figure A9.6: Empirical q_x- values (***points***) and estimations by Reed's expanding technique (***solid line***) Italy's 1990-91 males life table.

APPENDIX B:

COMPARISON FIGURES

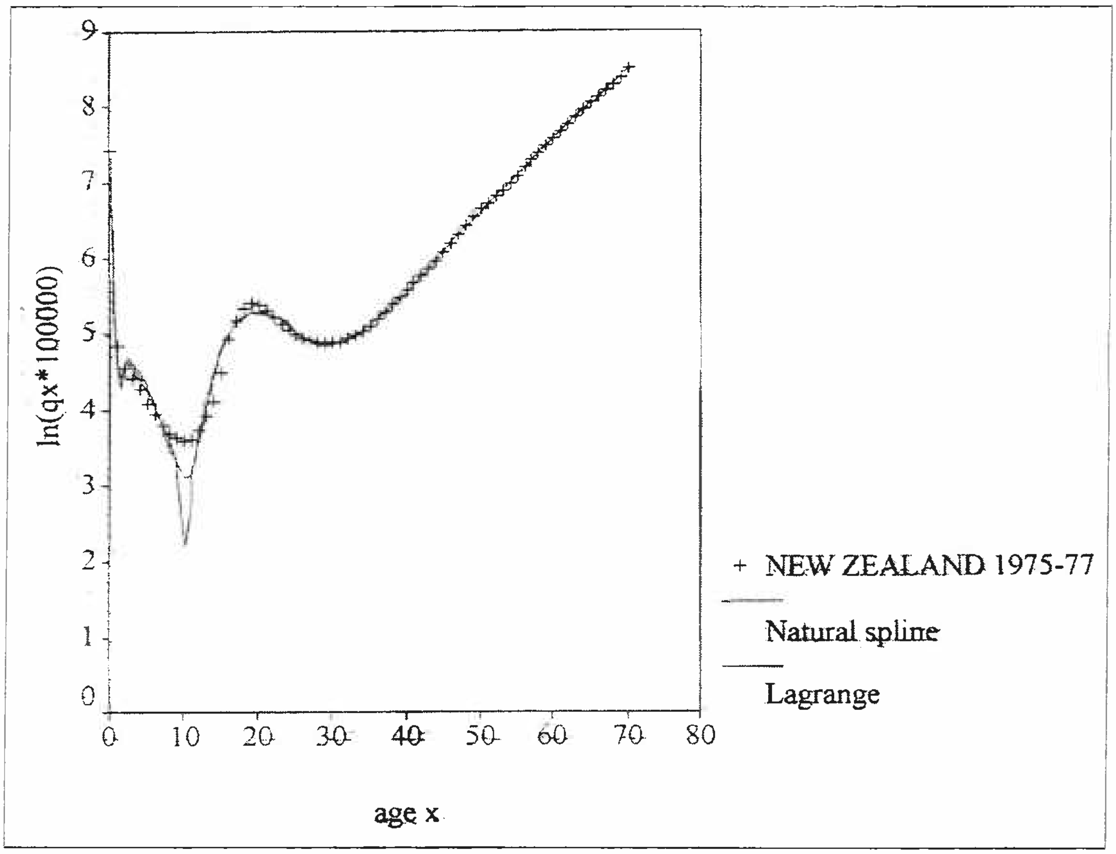

Figure B2.1: Empirical q_x - values (*crosses*) and estimations by Lagrangean and natural cubic spline interpolation (*solid lines*) for New Zealand's 1975-77 males life table.

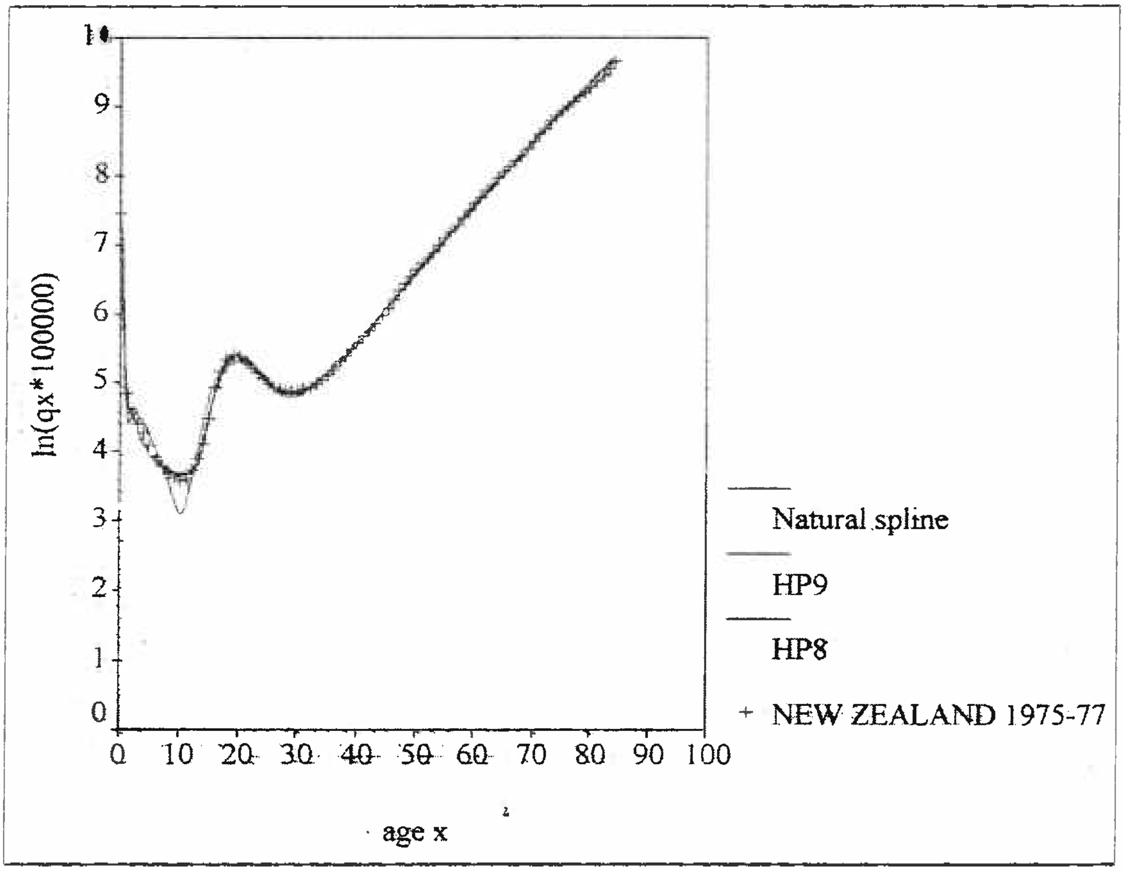

Figure B2.2: Empirical q_x - values (*crosses*) and estimations by the HP8 model, HP9 model and natural cubic spline interpolation (*solid lines*) for New Zealand's 1975-77 males life table.

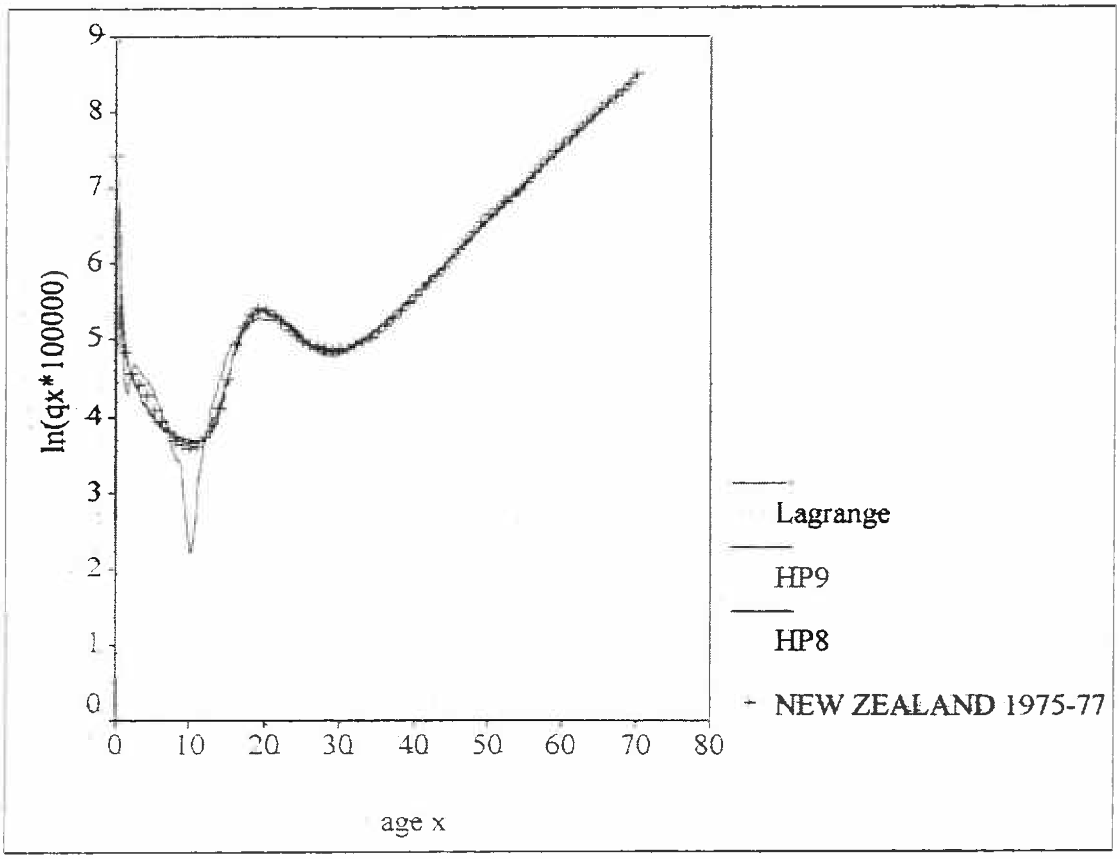

Figure B2.3: Empirical q_x - values (*crosses*) and estimations by the HP8 and HP9 adjusted models and Lagrangean interpolation (*solid lines*) for New Zealand's 1975-77 females life table.

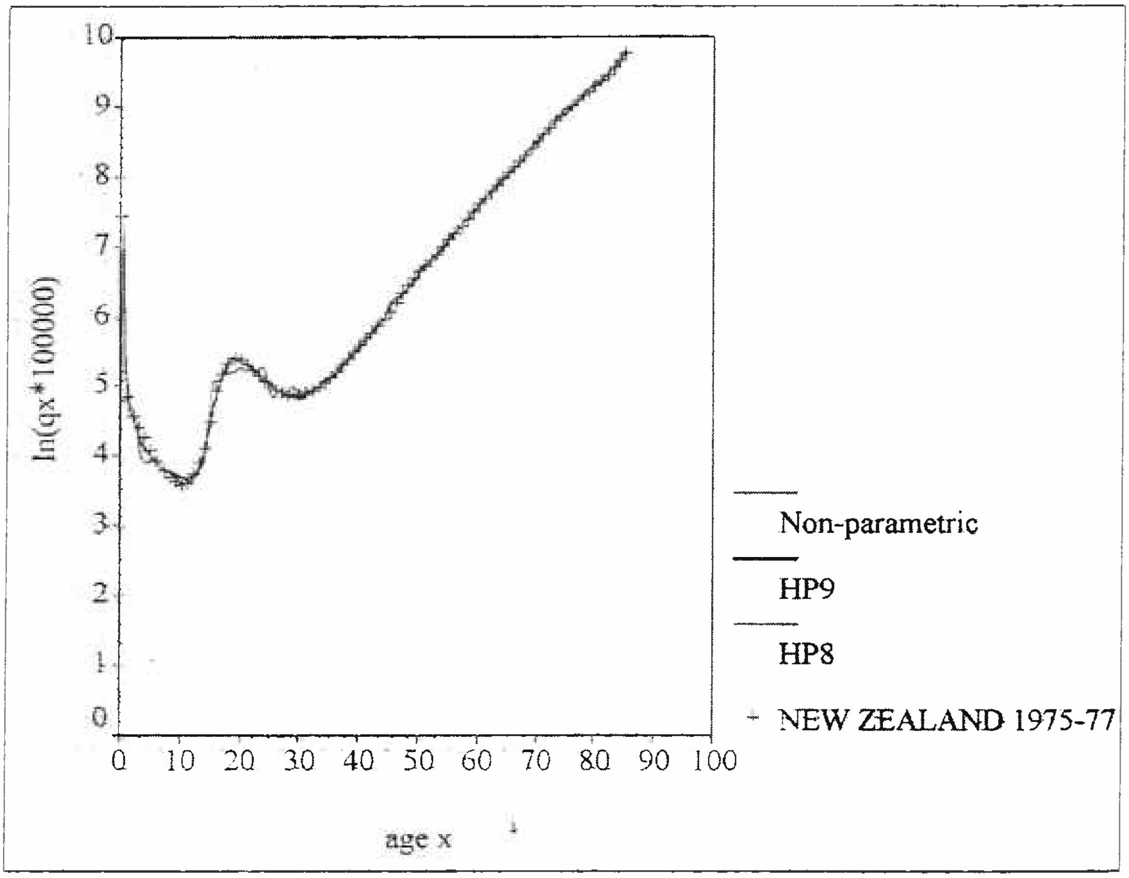

Figure B2.4: Empirical q_x - values (*crosses*) and estimations by the HP8 and HP9 adjusted models and Kostaki's non-parametric technique (*solid lines*) for New Zealand's 1975-77 females life table.

APPENDIX C:

TABLES

Table C.1a: Values of the sum of squares of the relative deviations between the resulting and the empirical q_x - values for the age interval, $x \leq 85$, of Kostaki's non-parametric expanding technique for several choices of a table of reference.

Expanding Method		
	Non-Parametric	*table of reference*
Life Table		
Italy females 1990-91	0.392	Italy males 1990-91
	0.980	Sweden females 1991-95
	1.291	Sweden males 1991-95
	0.395	NewZealand females 1980-82
	0.691	NewZealand males 1980-82
Italy males 1990-91	0.406*	Italy females 1990-91
	0.335*	NewZealand females 1975-77
	0.282*	NewZealand males 1975-77
	0.687*	Norway females 1951-55
	0.316*	Norway males 1951-55

*estimations were produced for the ages $x \leq 75$.

Table C.1b: Values** of the sum of squares of the absolut deviations between the resulting and the empirical q_x - values for the age interval, $x \leq 85$, of Kostaki's non-parametric expanding technique for several choices of a table of reference.

Expanding Method		
	Non-Parametric	*table of reference*
Life Table		
Italy females 1990-91	163	Italy males 1990-91
	1.07	Sweden females 1991-95
	63.99	Sweden males 1991-95
	76.71	NewZealand females 1980-82
	9.89	NewZealand males 1980-82
Italy males 1990-91	39.95*	Italy females 1990-91
	8.15*	NewZealand females 1975-77
	6.53*	NewZealand males 1975-77
	37.23*	Norway females 1951-55
	3.41*	Norway males 1951-55

*estimations were produced for the ages $x \leq 75$.

**multiplied by 10^6.

Table C2a: Estimates of parameters *A, B, C, D* and their standard errors for the eight parameter Heligman&Pollard model (**HP8**).

Parameter	$A*10^4$	$B*10^3$	$C*10^3$	$D*10^4$
Parameter's standard error	**(s.e.)$*10^4$**	(s.e.)$*10^3$	(s.e.)$*10^3$	(s.e.)$*10^4$
Life table				
Italy males 1990-91	5.175000	1.724500	79.879300	6.970000
	0.000242	0.002508	0.092283	0.000024
Italy females 1990-91	5.291563	13.635015	108.198617	9.589142
	0.000175	2.015401	5.730777	1.417471
Sweden males 1991-95	3.826000	13.212700	96.209200	5.100000
	0.000033	0.147631	0.263948	0.000030
Sweden females 1991-95	2.850000	8.146800	85.361500	1.230000
	0.000017	0.073080	0.228730	0.000004
New Zealand males 1975-77	14.861000	12.425600	107.181700	16.980000
	0.000165	0.042959	0.095129	0.000127
New Zealand females 1975-77	13.159000	33.572400	121.665300	3.410000
	0.000221	0.277155	0.155992	0.000020
New Zealand males 1980-82	13.362000	27.606400	119.323300	15.280000
	0.000196	0.180967	0.133164	0.000113
New Zealand females 1980-82	11.782000	51.456200	134.944300	4.830000
	0.000142	0.342130	0.119434	0.000018
Finland males 1983-84	7.115000	71.19370	137.80820	8.210000
	0.001003	14.453431	4.2910840	0.000490
Finland females 1983-84	2.489000	0.0001230	0.1230000	2.680000
	0.000052	0.000418	0.7598410	0.013887
Norway males 1951-55	19.251000	2.186200	53.330500	7.480000
	0.000199	0.0022486	0.0013887	0.000040
Norway females 1951-55	16.896000	9.736200	106.418300	5.240000
	0.000161	0.0229782	0.074751	0.000800

Table C2b: Estimates of parameters *E, F, G, H* and their standard errors for the eight parameter Heligman&Pollard model (**HP8**).

Parameter	**E**	**F**	**G*10^5**	**H**
Parameter's standard error	(s.e.)	(s.e.)	(s.e.)*10^5	(s.e.)
Life table				
Italy males 1990-91	9,612824	21,057105	2,995000	1,108290
	2,486814	0,148118	0,000001	0,000003
Italy females 1990-91	0,174214	516,759772	0,691841	1,120085
	1,74214	139306977,9	0,000001	0,000066
Sweden males 1991-95	6,698210	23,696873	2,555000	1,106880
	3,313204	0,994838	0,000002	0,000009
Sweden females 1991-95	3,535758	21,415017	1,643000	1,104530
	3,556470	4,663070	0,000001	0,000008
New Zealand males 1975-77	14,962399	19,804751	5,509000	1,102400
	4,208808	0,054534	0,000002	0,000002
New Zealand females 1975-77	10,137363	18,520898	4,319000	1,097020
	10,882659	0,371871	0,000002	0,000003
New Zealand males 1980-82	11,555139	21.015098	4.185000	1.105560
	2,642458	0.093418	0.000002	0.000004
New Zealand females 1980-82	12,678668	18.897960	3.559000	1.099120
	6,044556	0.122801	0.000001	0.000002
Finland males 1983-84	10.420834	23.012591	5.494000	1.103120
	64.617362	3.582706	0.000042	0.000048
Finland females 1983-84	47.997283	19.139034	2.322000	1.102370
	232933.38	5.525640	0.0000034	0.000024
Norway males 1951-55	7.592427	22.790478	5.175000	1.097610
	3.966309	0.359116	0.000004	0.000006
Norway females 1951-55	1.346263	43.762779	1.316000	1.115110
	0.994403	399.88801	0.000004	0.000054

Table C3a: Estimates of parameters *A, B, C, D* and their standard errors for the nine parameter Heligman&Pollard model (**HP9**).

Parameter	$A*10^4$	$B*10^3$	$C*10^3$	$D*10^4$
Parameter's standard error	$(s.e.)*10^4$	$(s.e.)*10^3$	$(s.e.)*10^3$	$(s.e.)*10^4$
Life table				
Italy males 1990-91	5.108550	1.373323	77.380726	6.928700
	0.000025	0.001999	0.103737	0.000026
Sweden males 1991-95	3.510260	2.702618	72.289118	4.899540
	0.000001	0.000596	0.008872	0.000001
New Zealand males 1975-77	14.43426	9.184438	101.234675	17.024360
	0.000115	0.021632	0.071483	0.000103
New Zealand females 1975-77	13.14150	33.008103	121.120640	3.397950
	0.000242	0.3000942	0.173210	0.000022
New Zealand males 1980-82	12.763260	19.514239	110.580101	15.216590
	0.000110	0.073369	0.082470	0.000076
New Zealand females 1980-82	11.798980	51.685770	135.074475	4.826220
	0.000161	0.392557	0.137814	0.000021
Finland males 1983-84	6.130000	7.390000	78.710000	7.480000
	0.000080	0.024173	0.680684	0.001075

Table C3b: Estimates of parameters *E1, E2, F, G, H* and their standard errors for the nine parameter Heligman&Pollard model (**HP9**).

Parameter	**E1**	**E2**	**F**	**G*10^5**	**H**
Parameter's standard error	**(s.e.)**	**(s.e.)**	**(s.e.)**	**(s.e.)*10^5**	**(s.e.)**
Life table					
Italy males 1990-91	12.262554	7.333303	20.27788	2.943580	1.108568
	18.12840	8.465538	0.871415	0.000001	0.000003
Sweden males 1991-95	21.706321	0.088856	19.22429	1.583930	1.114135
	13.460977	0.049685	0.102028	0.0000001	0.000002
New Zealand males 1975-77	25.285255	9.840330	18.71712	5.320900	1.102966
	69.129798	5.043978	0.288252	0.000001	0.000001
New Zealand females 1975-77	12:128468	8.123339	18.02112	4.277700	1.097177
	94.222600	61.788296	4.488211	0.000002	0.000003
New Zealand males 1980-82	21.920184	6.563069	19.35785	3.967390	1.106421
	51.186613	2.224358	0.343381	0.000001	0.000002
New Zealand females 1980-82	12.620102	12.754379	18.91032	3.560690	1.099114
	23.524682	45.981210	1.072624	0.000001	0.000002
Finland males 1983-84	132.082000	1.165000	16.070000	0.889000	1.128800
	424.168134	2.202468	17.769900	0.000126	0.004252

Table C4: Coefficients for six - point Lagrangean interpolation on values of a function u_x (e.g., l_x).

Coefficients			$x < 10$			
	$x = 0$	$x = 5$	$x = 10$	$x = 15$	$x = 20$	$x = 25$
u_1	0.612864	0.76608	-0.680690	0.437760	-0.161280	0.025536
u_2	0.344448	1.14816	-0.861120	0.52992	-0.19136	0.029952
u_3	0.167552	1.25664	-0.71808	0.41888	-0.14784	0.022848
u_4	0.059136	1.18272	-0.39424	0.21504	-0.07392	0.011264
u_6	-0.025536	0.76608	0.38304	-0.17024	0.05472	-0.008064
u_7	-0.029952	0.524160	0.69888	-0.26208	0.08064	-0.011648
u_8	-0.022848	0.30464	0.91392	-0.26112	0.07616	-0.010752
	-0.011264	0.12672	1.01376	-0.168896	0.04608	-0.006336
u_1 is used	$x = 1$	$x = 5$	$x = 10$	$x = 15$	$x = 20$	$x = 25$
u_2	0.56203	0.7176	-0.4784	0.283886	-0.10072	0.0156
u_3	0.273393	1.047199	-0.53191	0.2992	-0.10375	0.015867
u_4	0.096491	1.1088	-0.32853	0.1728	-0.05836	0.0088
u_6	-0.04167	0.798	0.354667	-0.152	0.048	-0.007
u_7	-0.04887	0.5616	0.6656	-0.24069	0.072758	-0.0104
u_8	-0.03728	0.3332	0.888533	-0.2448	0.070147	-0.0098
u_9	-0.01838	0.1408	1.001244	-0.16091	0.043116	-0.00587
Coefficients			$x > 10$			
	$x = 5m - 10$	$x = 5m - 5$	$x = 5m$	$x = 5m + 5$	$x = 5m + 10$	$x = 5m + 15$
u_{5m+1}	0.008064	-0.07392	0.88704	0.22176	-0.04928	0.006336
u_{5m+2}	0.011648	-0.09984	0.69888	0.46592	-0.08736	0.010752
u_{5m+3}	0.010752	-0.08736	0.49592	0.69888	-0.09984	0.011648
u_{5m+4}	0.006336	-0.04928	0.22176	0.88704	-0.07392	0.008064

REFERENCES

Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions with Formulas*, and Mathematical Tables, Dover ed., New York.

Barkalov, N. B. (1988). Interpolation of demographic data using rational spline functions. *Demograficheskie Issledovaniya*, 44-64. (*In Rus.*)

Barndorff, Ole E. and Cox, D. C. (1995). *Inference and Asymptotics*, Chapman&Hall ed., London.

Beers, S.H. (1944). Six - term formulae for Routine Actuarial Interpolation, *Record of the American Institute of Actuaries*, 33(2), 245-260.

Benjamin, B. and Pollard, J.H. (1980). *The Analysis of Mortality and other Actuarial Statistics*, second edition, Heinemann, London.

Bergström, T. and Lamm, D.(1989). Recovering Event Histories by Cubic Spline Interpolation, *Mathematical Population Studies*, 1(4), 327-355.

Biller, C. and Fährmeir, L. (1997). Bayesian Spline - type smoothing in generalized regression models, *Computational Statistics*, 12, 135-151.

Carlin, B. P. (1992). A simple Monte Carlo Approach to Bayesian Graduation, *Transactions of the Society of Actuaries.*, 44, 55-75.

Chambers, J.M. and Hastie, T.J. (1992). *Statistical Models in S*, Chapman&Hall ed., New York.

Congdon, P. (1993). Statistical Graduation in Local Demographic Analysis and Projection, *Journal of the Royal Statistical Society A*, 156, 237-270.

Dellaportas, P., Smith, A.F. and Stavropoulos, P. (1995). Bayesian Analysis of Mortality Data, *forthcoming to the Journal of the Royal Statistical Society A*.

Elandt-Johnson, R. and Johnson, N. (1980). *Survival Models and Data Analysis*, New York, John Wiley.

Forfar, D.O. and Smith, D.M. (1987). The changing shape of the English life tables, *Transactions of the Faculty of Actuaries*, 40(1), 98-134.

Forfar, D.O., McCutcheon, J.J. and Wilkie, A.D. (1988). *On Graduation by Mathematical Formula*, Institute of Actuaries, London.

Green, P.J. and Silverman, B.W. (1995). *Nonparametric regression and generalized linear models: A roughness penalty approach*, Chapman&Hall ed., London.

Greville, T.N.E. (1969). *Theory and Applications of Spline Functions*, New York Academic Press.

Haberman, S. and Rensaw, A.E. (1999). A Simple Graphical Method for the Comparison of two Mortality Experiences, *Applied Stochastic Models in Business and Industry,* 15, 333-352.

Härdle, W. (1992). Applied Non Parametric Methods, CORE Discussion Paper 9203, *Universite Catholique de Louvain*, Louvain - la - Neuve, Belgium.

Heligman, L. and Pollard, J.H. (1980). The Age pattern of mortality, *Journal of the Institute of Actuaries*, 107, 49-77.

Hsieh, J.J. (1991). Construction of Expanded Continuous Life Tables - A Generalization of Abridged and Complete Life Tables, *Mathematical Biosciences*, 103, 287-302.

Hsieh, J.J. (1991). A general theory of life table construction and a precise life table method, *Biometrical Journal*, 33(2), 143-162.

Karlis, D. and Kostaki, A. (2000), Bootstrap techniques for mortality models, Technical Report, No 116, *Athens University of Economics and Business*, Department of Statistics, Greece.

Keyfitz, N. (1985). *Applied Mathematical Demography*, second edition, Springer - Verlag, New York Inc.

King, G. (1914). On a Short Method of constructing an Abridged life table, *Journal of the Institute of Actuaries*, 48, 294-303.

Kostaki, A. (1991). The Heligman - Pollard Formula as a Tool for Expanding an Abridged Life Table, *Journal of Official Statistics*, 7(3), 311-323.

Kostaki A. (1992a). Methodology and Applications of the Heligman - Pollard Formula, Doctoral Dissertation, *University of Lund*, Department of Statistics, Sweden.

Kostaki A. (1992b). A Nine-Parameter Version of the Heligman - Pollard Formula. *Mathematical Population Studies*, 3(4), 277-288

Kostaki, A. (1999b). *Demometrics*, Publications of the Athens University of Economics and Business (*In Greek*).

Kostaki, A. and Lanke, J. (2000a). Degrouping Mortality Data for the Elderly, *Mathematical Population Studies*, 7(4), 1-11.

Kostaki, A. (2000b). A Relational Technique for Estimating the age - specific mortality pattern from grouped data, *Mathematical Population Studies*, 9(1), 83-95.

Lytrokapi, A. (1998). Parametric Models of Mortality: Their use in Demography and Actuarial Science, M.Sc. Dissertation, *Athens University of Economics and Business*, Department of Statistics, Greece.

McCutcheon, J.J. (1981). Some Remarks on Splines, *Transactions of the Faculty of Actuaries*, 37, 421-438.

McCutcheon, J.J. and Eilbeck, J.C. (1977). Experiments in the Graduation of the English Life Tables (No. 13) Data, *Transactions of the Faculty of Actuaries*, 35, 281-296.

McNeil, D.R., Trussel, T.J. and Turner, J.C. (1977). Spline Interpolation of Demographic Data, *Demography*, 14(2), 245-252.

NAG Fortran Library, Mark 17 (1995), Chapter E04: Minimizing or Maximizing a Function, Chapter E01: Interpolation.

Namboodiri, N.K. and Suchindran, C.M. (1987). *Life Tables Techniques and their applications*, Academic Press, San Diego.

Nanjo, Z. (1986). Practical Use of Interpolatory Cubic and Rational Spline Functions for Fertility Data, NUPRI Research Paper Series No. 34, vi, 13pp, *Nihon University*, Population research Institute, Tokyo, Japan.

Nanjo, Z. (1987) Demographic data and spline interpolation, *Journal of Population Studies*, No. 10, 43-53 pp. (*In Jpn. with sum. In Eng.*)

Nanjo, Z. (1993). New relational models of age - specific mortality - all and selected causes of death, NUPRI Research Paper Series No. 63 vi, 21pp, *Nihon University*, Population research Institute, Tokyo, Japan.

National Statistical Service of Greece (1964). Life Tables for Greece I:1926-30, II:1940, III:1955-59, IV:1960-62, *Publications of the National Statistical Service of Greece*, Athens, Greece.

Pollard, J.H. (1991). Fun with Gompertz, *Genus*, 47, 1-20.

Pollard, J.H. (1989). On the derivation of a full life table from mortality data recorded in five-year age groups, *Mathematical Population Studies*, 2(1), 1-14.

Renshaw, A.E. (1995). Graduation and Generalized Linear Models: An Overview, Actuarial Research Paper No. 73, *City University*, Department of Actuarial Science and Statistics, London, UK.

Rogers, A. (1986). Parameterized Multistate Population Dynamics and Projections, *Journal of the American Statistical Association*, 81, 48-61.

Speirs, R. (1986). The use of Cubic Spline functions in the derivation of the 1980-82 New Zealand Life Tables, *Demographic Trends Bulletin*, 28, 12-18.

Thiele, P.N. (1872). On a Mathematical Formula to Express the Rate of Mortality throughout the Whole of Life, *Journal of the Institute of Actuaries*, 29(1), 97-107.

Tyrtyshnikov, E.E. (1997). *A Brief Introduction to Numerical Analysis*, Birkhauser ed., Boston.

United Nations (1988a). *MORTPAK. The United Nations Software Package for Mortality Measurement* (Batch oriented Software for the Mainframe Computer), New York: United Nations, 114-118.

United Nations (1988b). *MORTPAK-LITE. The United Nations Software Package for Mortality Measurement* (Interactive Software for the IBM-PC and Compatibles), New York: United Nations, 111-114.

Valaoras , V.G. (1972). Population Analysis of Egypt 1935-1970 - With special reference to mortality, *Cairo Demographic Centre Occasional Paper No.* 1, 1-63.

Valaoras, V.G. (1984). The 1980 Life Tables for Greece, *Publications of the Academy of Athens*, 405-436.

Wegman, J.E. and Wright, W.I. (1983). Splines in Statistics, *Journal of the American Statistical Association*, 78, 351-365.

Wahba, G. (1975). Interpolating Spline Methods for Density Estimation: Equi - spaced Knots, *Annals of Statistics*, 3, 30-48.