Agnieszka Rossa
Lesław Socha
Andrzej Szymański

gerontologia

Hybrid Dynamic and Fuzzy Models of Mortality

Hybrid Dynamic and Fuzzy Models of Mortality

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

gerontologia

Agnieszka Rossa
Lesław Socha
Andrzej Szymański

Hybrid Dynamic and Fuzzy Models of Mortality

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO
Łódź 2018

Agnieszka Rossa, Andrzej Szymański – University of Łódź, Faculty of Economics and Sociology
Department of Demography and Social Gerontology, 90-214 Łódź, 41/43 Rewolucji 1905 r. St.

Lesław Socha – Cardinal Stefan Wyszyński University in Warsaw
Faculty of Mathematics and Natural Sciences. School of Exact Sciences, Institute of Informatics
01-938 Warszawa, 1/3 Wóycickiego St.

TRANSLATOR AND PROOFREADING
Janusz Kwitecki

REVIEWER
Jan Paradysz

INITIATING EDITOR
Iwona Gos

TYPESETTING
Agnieszka Rossa

TECHNICAL EDITOR
Leonora Wojciechowska

COVER DESIGN
Katarzyna Turkowska

Cover Image: © Depositphotos.com/hobbitart

Printed directly from camera-ready materials provided to the Łódź University Press

The research was supported by a grant from the National Science Centre, Poland, under contract UMO-2015/17/B/HS4/00927

Published by Łódź University Press
First edition. W.07707.16.0.K

Printing sheets 18.875

ISBN 978-83-8088-926-2
e-ISBN 978-83-8088-927-9
https://doi.org/10.18778/8088-926-2
Łódź University Press
90-131 Łódź, 8 Lindleya St.
www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
phone. (42) 665 58 63

Contents

Preface

Mortality is generally considered relatively easy to forecast, particularly when the forecasting horizon is short. In longer periods however, its course may be affected by various changes brought about by all kinds of disturbances and events. A case in point is the health crisis in Poland of the 1970s and 1980s [Okólski 2003]. In such cases, it is of key importance that appropriate assumptions and an adequate model are selected.

Mortality forecasting is usually supported by extrapolative models, making use of the regularity found in age patterns and trends of death rates or probabilities over time.

There are several reasons why one should learn more about mortality models. Forecasting of mortality has a wide range of applications outside the field of statistics and mathematics. It is of fundamental importance in such areas as funding of public or private pensions and life insurance. Annuity providers and policy makers use mortality projections to determine appropriate pension benefits, to assess retirement income or life insurance products, to hold additional reserving capital or to manage the long term demographic risk. Thus, one of the important question arises: What is the best way to forecast future mortality rates and to model the uncertainty of such forecasts? A key input to address this question is the development of advanced mortality modeling methodology.

These notes are an attempt to capture the stochastic nature of mortality by approaching the subject of mortality modeling and forecasting from a new theoretical point of view, using theory of stochastic differential equations, theory of fuzzy numbers and complex numbers.

The book is addressed to tertiary students, doctoral students and specialists in the fields of demography, life insurance, statistics and economics. This research project was funded by the National Science Center pursuant to its decision no. 2015/17/B/HS4/00927.

Abbreviation and notation

Throughout this book, the following abbreviations for mortality models have been adopted:

SLC	Standard Lee–Carter
LCH	Lee–Carter hybrid
DLC	Dynamic Lee–Carter
DDLC	Discrete Dynamic Lee–Carter model
LCH	Lee–Carter hybrid
DLCH	discrete Lee–Carter hybrid
FLC	Fuzzy Lee–Carter
EFLC	Extended Fuzzy Lee–Carter
MFLC	Modified Fuzzy Lee–Carter
CFLC	Complex-Function Lee–Carter
QVLC	Quaternion-Valued Lee–Carter
V	Vasiček
DV	Discrete Vasiček
VH	Vasiček hybrid
DVH	Discrete Vasiček hybrid
VHM	Vasiček hybrid moment
DVHM	Discrete Vasiček hybrid moment
MV	Modified Vasiček
DMV	Discrete Modified Vasiček
MVH	Modified Vasiček hybrid
DMVH	Discrete Modified Vasiček hybrid
MVHM	Modified Vasiček hybrid moment
DMVHM	Discrete Modified Vasiček hybrid moment
CIR	Cox–Ingersoll–Ross
DCIR	Discrete Cox–Ingersoll–Ross
CIRH	Cox–Ingersoll–Ross hybrid
DCIRH	Discrete Cox–Ingersoll–Ross hybrid
CIRHM	Cox–Ingersoll–Ross hybrid moment
DCIRHM	Discrete Cox–Ingersoll–Ross hybrid moment
MCIR	Modified Cox–Ingersoll–Ross
DMCIR	Discrete Modified Cox–Ingersoll–Ross
MCIRH	Modified Cox–Ingersoll–Ross hybrid
DMCIRH	Discrete Modified Cox–Ingersoll–Ross hybrid
MCIRHM	Modified Cox–Ingersoll–Ross hybrid moment
DMCIRHM	Discrete Modified Cox–Ingersoll–Ross hybrid moment
GOB	Giacometti–Ortobelli–Bertocchi
DGOB	Discrete Giacometti–Ortobelli–Bertocchi
GOBH	Giacometti–Ortobelli–Bertocchi hybrid
DGOBH	Discrete Giacometti–Ortobelli–Bertocchi hybrid
GOBHM	Giacometti–Ortobelli–Bertocchi hybrid moment
DGOBHM	Discrete Giacometti–Ortobelli–Bertocchi hybrid moment

MP	Milevsky–Promislow
DMP	Discrete Milevsky–Promislow
MMP	Modified Milevsky–Promislow
DMMP	Discrete Modified Milevsky–Promislow
DMPH	Discrete Milevsky–Promislow hybrid
MPHM	Milevsky–Promislow hybrid moment
DMPHM	Discrete Milevsky–Promislow hybrid moment
MMPH	Modified Milevsky–Promislow hybrid
DMMPH	Discrete Modified Milevsky–Promislow hybrid
MMPHM	Modified Milevsky–Promislow hybrid moment
DMMPHM	Discrete Modified Milevsky–Promislow hybrid moment
MP-2DF	Milevsky–Promislow, with 2 dependent filters
MPH-2DF	Milevsky–Promislow hybrid, with 2 dependent filters
MPHM-2DF	Milevsky–Promislow hybrid moment, with 2 dependent filters
DMPHM-2DF	Discrete Milevsky–Promislow hybrid moment with 2 dependent filters
MP-2IF	Milevsky–Promislow, with 2 independent filters
MPH-2IF	Milevsky–Promislow hybrid, with 2 independent filters
MPHM-2IF	Milevsky–Promislow hybrid moment, with 2 independent filters
DMPHM-2IF	Discrete Milevsky–Promislow hybrid moment with 2 independent filters
MP-VLF	Milevsky–Promislow with vector linear filter
MPH	Milevsky–Promislow hybrid
MPH-VLF	Milevsky–Promislow hybrid, with a vector linear filter
MPHM-VLF	Milevsky–Promislow hybrid moment, with a vector linear filter
DMPH-VLF	Discrete Milevsky–Promislow hybrid, with a vector linear filter

Introduction

The phenomenon of mortality has been studied for many centuries. In the early 3rd c., a Roman jurist, Domitius Ulpianus, created for fiscal purposes the so-called Ulpian table containing life expectancies for the citizens of the Roman Empire. As historical sources do not mention what calculation method and source materials he had used, the Ulpian table is mainly of historical value [Rosset 1979, pp. 102–103].

It is recognized that the father of the mortality table methodology is John Graunt (1620–1674), since his work [Graunt, 1662] where mortality of generations of London residents was examined. Graunt based his analysis on the records of London parishes, but did not specify which periods they concerned. Graunt's research was continued by an English astronomer Edmond Halley (1656–1742), who proposed mortality tables for the Wrocław population [Halley 1693].

The modern methodology for constructing mortality tables, also known as "life-tables", is credited to Chin L. Chiang (1914–2014) and his book [Chiang 1968]. The more works on life-tables and mortality models come from 19th c. [Gompertz 1825, Thiele, Sprague 1871], but it is only during the last decades that the mortality modeling methodology started to develop, as evidenced by numerous books on this subject [Rosset 1979, Keilman 1990, Okólski 1990, Benjamin, Pollard 1993, Kannisto 1994, Tabeau *et al.* 2001, Keilman 2005, Alho, Spencer 2005, Girosi, King 2006, Kędelski, Paradysz 2006, Rossa *et al.* 2011].

Since the introduction of the Lee–Carter model [Lee, Carter 1992] proposed to forecast the trend of age-specific mortality rates, a range of mortality models have been proposed with modeling the probability of death, the age-specific mortality rate or the force of mortality.

Among mortality models three main approaches can be identified: extrapolation, expectation and explanation [Pitacco 2004, Booth 2006, Tabeau *et al.* 2001]. The most common one is an extrapolative approach which uses a real or fuzzy variable functions of age and time to describe patterns and trends in death probabilities, mortality rates (or their transformations) and other measures [Heligman, Pollard 1980, Brouhns *et al.* 2002, Lee, Miller 2001, Renshaw, Haberman 2003a, 2003b, 2003c, 2006, 2008, Koissi, Shapiro 2006, Cairns *et al.* 2006, 2008a, 2008b, 2009, 2011, Denuit 2007, Debon *et al.* 2008, Haberman, Renshaw 2008, 2009, 2011, Hatzopoulos, Haberman 2011, Fung *et al.* 2017].

Mortality models can be divided also into two main categories: static and dynamic models. Models in the first group are based on some algebraic equations, while in dynamic models of the second group the force of mortality (the intensity process) is expressed as a solution of stochastic differential equations [Vasiček 1977, Cox *et al.* 1985a, 1985b, Janssen,Skiadas 1995, Milevsky, Promislow 2001, Dahl 2004, Biffis 2005, Biffis, Denuit 2006, Schrager 2006, Bravo, Braumann 2007, Yashin 2007, Hainaut, Devolder 2007, 2008, Luciano *et al.* 2008, Luciano, Vigna 2008, Plat 2009, Bayraktar *et al.* 2009, Biffis *et al.* 2010, Coelho *et al.* 2010, Giacometti *et al.* 2011, Russo*et al.* 2011, Wang*et al.* 2011, Hainaut 2012, Rossa, Socha 2013].

Unfortunately, the simple dynamic models based on stochastic differential equations can be inadequate to describe demographic processes. In particular, they may fail to explain evolution of the phenomena, meaning that their behavior changes in continuous time or discrete time intervals. To make up for this disadvantage, researchers put forward a new type of models, called hybrid models, which account for interactions between continuous and discrete dynamics.

Hybrid models, or switching models [Boukas 2005], are constructed as the generalizations of the models with switching points that have been already used for automatic control and for random structure models [Kazakov, Artemiev 1980] describing phenomena within mechanics, biology, economics or empirical sciences. The authors of some studies have proposed complex mortality models sharing characteristics with the hybrid models [Biffis, Denuit 2006, Biffis *et al.* 2010, Hainaut 2012, Rossa, Socha 2013].

For the purposes of this study, a hybrid system will henceforth be understood as a family of static or dynamic models where the switchings take place according to some switching rule. The dynamic models will be described using stochastic differential equations. There exists a class of equations for which analytical solutions of relatively complex structure can be found, therefore a new group of hybrid models will be proposed called the moment hybrid models. The idea underlying their construction involves the replacement of the stochastic models by equivalent differential equations for moments.

Another promising approach to mortality modeling offers theory of fuzzy numbers. It is well-known that the main difficulty in the applications of the Lee–Carter model is due to the assumed homogeneity of random terms. However, this property is not confirmed by the real-life data. The problem prompted search for solutions that could do without this assumption. One of the possible options is to set research in the framework of the fuzzy number theory. This line of thinking was adopted by [Koissi, Shapiro 2006], where empirical observations and parameters of the Lee–Carter model were converted into fuzzy symmetric triangular numbers.

Unfortunately, the Koissi–Shapiro model involves some difficulties, which arise from the necessity to find the minimum of a criterion function containing a max-type operator and cannot be solved using standard optimization algorithms. One approach to such a problem can be applying the Banach algebra of oriented fuzzy numbers (OFN) developed by [Kosiński *et al.* 2003]. The results of using this algebra to the Koissi–Shapiro model have been published in [Szymański, Rossa 2014].

A more sophisticated modification of the Koissi–Shapiro model involves the replacement of the Banach OFN algebra by the Banach C^*–algebra to allow the use of the Gelfand–Mazur theorem about isometric isomorphism between the C^*–algebra and the Banach algebra of complex functions and, consequently, to move the optimization problem into the framework of complex analysis. To our best knowledge, this is an innovative approach to mortality modeling.

This book has the following structure. In Chapter 1, basic mortality characteristics and some static and dynamic mortality models are discussed, especially the oldest historical mortality models (the so-called ”mortality laws”), the well-known Lee–Carter model with its extensions and generalizations, the Vasiček and Cox–Ingersoll-Ross models, the Giacometti–Ortobelli–Bertocchi model and some variants of the Milevsky–Promislow model. Chapter 2 introduces theoretical backgrounds of hybrid modeling. In Chapter 3, hybrid counterparts of the dynamic models presented in Chapter 1 are provided and some estimation procedures are proposed. Chapter 4 discusses the theoretical underpinnings of the fuzzy mortality modeling based on the algebra of Oriented Fuzzy Numbers (OFN), whereas Chapter 5 presents mortality models from the perspective of the so-called modified fuzzy numbers (MFN) and complex functions. Chapter 6 illustrates results of estimations of some proposed models, the parameters of which were estimated using empirical mortality data sets. The comparative analysis of the models' prediction accuracy is also performed.

Chapter 1

Basic mortality characteristics and models

1.1. Introduction

Demographic models are an attempt to generalize and simplify real demographic processes by means of mathematical functions or a set of mathematical relations in order to approximate possible variations observed in the real data and to support demographic forecasting.

In this chapter basic notions, relations and some discrete-time as well as continuous-time extrapolative mortality models are introduced.

The main attention is focused on the well-known Lee–Carter model, its generalizations, the Vasiček and Cox–Ingersoll–Ross models as well as the Milevsky–Promislow and Giacometti–Ortobelli–Bertocchi models. They will be converted to hybrid models in Chapter 3.

1.2. Discrete-time mortality frameworks

1.2.1. Age-specific rates and probabilities of death

The definition of a mortality rate used in this book draws on the general definition of a cohort (or period) demographic rate defined as a ratio of the number of demographic events occurring in some defined cohort (or in a real population within some defined time period) to the time-to-exposure, understood as the number of time units lived by the cohort (or by the population during the given time period) [Preston *et al.* 2001, pp. 5–32].

If person-years are used in the denominator, a demographic rate is termed ”an annualized rate”. Below the definitions of both a cohort and a period annualized age-specific mortality rates are provided [Rossa *et al.* 2011, pp. 229–231].

An important notion used in the Definition 1.1 is "a cohort", defined as a real or hypothetical aggregate of individuals that experience a specific demographic event, e.g. births, during a specific time interval. The cohort is identified by the event itself and by its time frame.

For the purposes of this discussion, let index t indicate a calendar year from the given set $\{1, 2, \ldots, T\}$, and index x the attained age, meaning that it takes values from the set $\{0, 1, \ldots, X\}$, where X is the fixed upper age limit.

Definition 1.1. A cohort age-specific mortality rate $m_x^{(s)}$ in the s-th cohort is a ratio of the number of deaths, $D_x^{(s)}$, among individuals aged x years last birthday to the number of person-years, $K_x^{(s)}$, lived in the age range $[x, x+1)$

$$m_x^{(s)} = \frac{D_x^{(s)}}{K_x^{(s)}}. \tag{1.2.1}$$

Definition 1.2. A period age-specific mortality rate $m_{x,t}$ is a ratio of the number of deaths, $D_{x,t}$, among individuals in the age range $[x, x+1)$ years during the calendar year t to the number of person-years, $K_{x,t}$, lived in th age interval $[x, x+1)$ during this year

$$m_{x,t} = \frac{D_{x,t}}{K_{x,t}}. \tag{1.2.2}$$

It is worth noting that the denominators $K_x^{(s)}$ in (1.2.1) and $K_{x,t}$ in (1.2.2) can be treated as the number of individuals exposed to the risk of death in the given age interval or in the age-time interval, respectively. In the case of (1.2.2) the denominator is usually replaced by the midyear population $\bar{L}_{x,t}$, lived in the age range $[x, x+1)$ during the given year t. Therefore, period mortality rates (1.2.2) are often described as central death rates because of a midyear population used in the denominator.

For convenience (1.2.1), (1.2.2) are often expressed in thousands as

$$m_x^{(s)} = \frac{D_x^{(s)}}{K_x^{(s)}} \cdot 1\,000, \quad m_{x,t} = \frac{D_{x,t}}{K_{x,t}} \cdot 1\,000. \tag{1.2.3}$$

In a more general discrete approach, it is possible to consider an age interval $[x, x+n)$, where $n \in \mathbb{N}$ and $n > 1$. The cohort age-specific mortality rates (1.2.1) are then denoted as ${}_n m_x^{(s)}$ and the period age-specific mortality rates (1.2.2) as ${}_n m_{x,t}$.

In addition to demographic rates, the set of measures used in demographic analysis also includes the probability of death.

Definition 1.3. The probability of death, $q_x^{(s)}$, in the s-th cohort is a ratio of the number of deaths, $D_x^{(s)}$, among individuals aged x years last birthday to the number of individuals, $L_x^{(s)}$, surviving to this age, i.e.

$$q_x^{(s)} = \frac{D_x^{(s)}}{L_x^{(s)}}. \tag{1.2.4}$$

When the age interval under consideration is $[x, x+n)$ and $n > 1$, the cohort death probability is denoted as ${}_nq_x^{(s)}$.

For simplicity, the index of s-th cohort (s) will be omitted from further notations.

1.2.2. The relationship between mortality rates and death probabilities

Let ${}_nD_x$ and ${}_nK_x$ be, respectively, the number of deaths and the time of exposure to the risk of death (time-to-exposure) for a cohort in the age range $[x, x+n)$ years. Let ${}_na_x$ represent the average number of years lived by individuals in that age range who died before their $(x+n)$–th birthday. Additionally, let L_x be the number of individuals surviving to the age x.

The characteristics ${}_nD_x$, ${}_nK_x$, ${}_na_x$ and L_x are linked by the following relations, defined by analogy to the balancing equations for a closed population [Preston *et al.* 2001, p. 2]

$${}_nK_x = n \cdot L_x - n \cdot {}_nD_x + {}_na_x \cdot {}_nD_x \tag{1.2.5}$$

and

$${}_nK_x = n \cdot L_{x+n} + {}_na_x \cdot {}_nD_x. \tag{1.2.6}$$

Additionally, we have

$$L_x = L_{x+n} + {}_nD_x. \tag{1.2.7}$$

From (1.2.5) we obtain

$$L_x = \frac{{}_nK_x}{n} + {}_nD_x - \frac{{}_nD_x \cdot {}_na_x}{n} = \frac{{}_nK_x + (n - {}_na_x)\,{}_nD_x}{n}, \tag{1.2.8}$$

hence

$$
{}_nq_x = \frac{{}_nD_x}{L_x} = \frac{n \cdot {}_nD_x}{{}_nK_x + (n - {}_na_x) \cdot {}_nD_x} =
$$
$$
= \frac{n \cdot \frac{{}_nD_x}{{}_nK_x}}{\frac{{}_nK_x}{{}_nK_x} + (n - {}_na_x) \cdot \frac{{}_nD_x}{{}_nK_x}} = \frac{n \cdot {}_nm_x}{1 + (n - {}_na_x) \cdot {}_nm_x}. \tag{1.2.9}
$$

Thus, the following relation is obtained

$$
{}_nq_x = \frac{n \cdot {}_nm_x}{1 + (n - {}_na_x) \cdot {}_nm_x}. \tag{1.2.10}
$$

Formula (1.2.10) shows how the cohort age-specific death rate and the probability of death are related to each other. In the special case of $n = 1$, we obtain

$$
q_x = \frac{m_x}{1 + (1 - a_x) \cdot m_x}. \tag{1.2.11}
$$

1.2.3. Interpolation models

Let L_{x+y} for a fixed age x and for $y \in [0, n]$ be the number of individuals surviving the age $x + y$. In this section we will assume that L_{x+y} is a continuous function of $y \in [0, n]$.

A linear interpolation model

Let us assume that L_{x+y} is a linear function of variable y, i.e.

$$
L_{x+y} = a + by \quad \text{for} \quad y \in [0, n]. \tag{1.2.12}
$$

The parameters a, b of this function are determined so that it takes values L_x for $y = 0$ and L_{x+n} for $y = n$, where $L_x \geq L_{x+n} > 0$ are fixed in advance. These two conditions can be written as

$$
L_{x+0} = a \quad \text{and} \quad L_{x+n} = a + bn. \tag{1.2.13}
$$

It follows from (1.2.13) that

$$
a = L_x \quad \text{and} \quad b = \frac{L_{x+n} - L_x}{n} = -\frac{{}_nD_x}{n}, \tag{1.2.14}
$$

where ${}_nD_x$ is the number of deaths observed in age interval $[x, x + n)$.

Therefore, (1.2.12) can be written as

$$L_{x+y} = a + by = L_x - \frac{{}_nD_x}{n} y. \tag{1.2.15}$$

Let us now calculate the time-to-exposure ${}_nK_x$ in the age interval $[x, x+n)$. If L_{x+n} is a integrable function, then ${}_nK_x$ can be calculated as the integral of L_{x+n} on the interval $[0, n]$. In this case, the integrability of L_{x+y} arises from its linear form (1.2.12). Hence, we have

$$\begin{aligned} {}_nK_x &= \int_0^n L_{x+y} dy = \int_0^n (L_x - \frac{{}_nD_x}{n} y) dy = \\ &= n \cdot L_x - \frac{{}_nD_x}{n} \cdot \frac{1}{2} y^2 \Big|_0^n = n \cdot L_x - \frac{n}{2} \cdot {}_nD_x. \end{aligned} \tag{1.2.16}$$

Since L_x, L_{x+n} are connected by relation (1.2.7), we obtain

$${}_nK_x = n \cdot L_{x+n} + \frac{n}{2} \cdot {}_nD_x. \tag{1.2.17}$$

A comparison of the above result with the general formula (1.2.6) for the time-to-exposure

$${}_nK_x = n \cdot L_{x+n} + {}_na_x \cdot {}_nD_x, \tag{1.2.18}$$

leads us to the conclusion that

$${}_na_x = \frac{n}{2}. \tag{1.2.19}$$

Thus in the linear interpolation framework, formula (1.2.10) identifying the relation between probability ${}_nq_x$ and the age-specific death rate ${}_nm_x$ can be reduced to

$${}_nq_x = \frac{n \cdot {}_nm_x}{1 + (n - \frac{n}{2}) \cdot {}_nm_x} = \frac{2n \cdot {}_nm_x}{2 + n \cdot {}_nm_x}. \tag{1.2.20}$$

For the special case of $n = 1$, we get

$$q_x = \frac{m_x}{1 + (1 - \frac{1}{2}) \cdot m_x} = \frac{2m_x}{2 + m_x}. \tag{1.2.21}$$

An exponential interpolation model

Let us assume now that L_{x+y} for $y \in [0, n]$ is an exponential function of variable y expressed by the formula

$$L_{x+y} = ab^y \quad \text{for} \quad y \in [0, n], \tag{1.2.22}$$

given the following constraints

$$L_{x+0} = ab^0 = a \quad \text{and} \quad L_{x+n} = ab^n, \tag{1.2.23}$$

where $L_x \geq L_{x+n} > 0$ are known values.
From (1.2.23) it follows that

$$a = L_x \quad \text{and} \quad b = \left(\frac{L_{x+n}}{L_x} \right)^{\frac{1}{n}}. \tag{1.2.24}$$

Hence, L_{x+y} defined in (1.2.22) is of the form

$$L_{x+y} = L_x \left(\frac{L_{x+n}}{L_x} \right)^{\frac{y}{n}}. \tag{1.2.25}$$

Let us denote

$$_n p_x = 1 - {_n q_x} \quad \text{for} \quad {_n q_x} \in (0, 1). \tag{1.2.26}$$

Since

$$\frac{L_{x+n}}{L_x} = {_n p_x}, \tag{1.2.27}$$

from (1.2.25) we get

$$L_{x+y} = L_x \left({_n p_x} \right)^{\frac{y}{n}}. \tag{1.2.28}$$

Let us calculate the time-to-exposure $_n K_x$ as an integral of L_{x+y} on the interval $[0, n]$. We have

$$\begin{aligned} K_x &= \int_0^n L_{x+y} dy = \int_0^n L_x \left({_n p_x} \right)^{\frac{y}{n}} dy = \\ &= L_x \int_0^n \exp\left\{ \frac{y}{n} \ln {_n p_x} \right\} dy. \end{aligned} \tag{1.2.29}$$

The last transformation under the integral is due to the fact that

$$a^z \equiv e^{z \ln a} \quad \text{for any positive constant } a. \tag{1.2.30}$$

By applying the variable substitution

$$z = \frac{y}{n} \ln {}_n p_x \quad \text{and} \quad dy = \frac{n}{\ln {}_n p_x} dz, \tag{1.2.31}$$

we get

$$\begin{aligned}
{}_n K_x &= n \cdot L_x \int_0^{\ln {}_n p_x} \frac{e^z}{\ln {}_n p_x} dz = \\
&= \frac{n \cdot L_x}{\ln {}_n p_x} \int_0^{\ln {}_n p_x} e^z dz = \frac{n \cdot L_x}{\ln {}_n p_x} e^z \Big|_0^{\ln {}_n p_x} = \\
&= \frac{n \cdot L_x}{\ln {}_n p_x} \left(e^{\ln {}_n p_x} - e^0 \right) = \frac{n \cdot L_x}{\ln {}_n p_x} \left({}_n p_x - 1 \right) = \\
&= -\frac{n \cdot L_x}{\ln {}_n p_x} {}_n q_x.
\end{aligned} \tag{1.2.32}$$

Because of the definition of probability of death (see Definition 1.3), the following equality holds

$$L_x \cdot {}_n q_x = {}_n D_x. \tag{1.2.33}$$

We obtain

$${}_n K_x = -n \frac{{}_n D_x}{\ln {}_n p_x}. \tag{1.2.34}$$

It follows from the comparison of the above result with the general formula (1.2.6) for the time-to-exposure ${}_n K_x$ that

$${}_n K_x = n \cdot L_{x+n} + {}_n a_x \cdot {}_n D_x. \tag{1.2.35}$$

We receive equality

$$-n \frac{{}_n D_x}{\ln {}_n p_x} = n \cdot L_{x+n} + {}_n a_x \cdot {}_n D_x. \tag{1.2.36}$$

Hence, the formula for ${}_n a_x$ takes the following form

$$\begin{aligned}
{}_n a_x &= -\frac{n L_{x+n}}{{}_n D_x} - \frac{n}{\ln {}_n p_x} = -n \frac{{}_n p_x}{{}_n q_x} - \frac{n}{\ln {}_n p_x} = \\
&= n - \frac{n}{{}_n q_x} - \frac{n}{\ln(1 - {}_n q_x)}.
\end{aligned} \tag{1.2.37}$$

Let us determine now how mortality rate ${}_nm_x$ and probability ${}_nq_x$ in the exponential interpolation model are related to each other. Based on the cohort age-specific mortality rate (see Definition 1.1), we have

$$ {}_nm_x = \frac{{}_nD_x}{{}_nK_x}. \tag{1.2.38} $$

Then, using (1.2.34) we can reduce (1.2.38) to

$$ {}_nm_x = \frac{{}_nD_x}{{}_nK_x} = \frac{{}_nD_x}{-n\frac{{}_nD_x}{\ln {}_np_x}} = -\frac{1}{n}\ln {}_np_x = -\frac{1}{n}\ln(1 - {}_nq_x). \tag{1.2.39} $$

Thus, we get

$$ {}_nm_x = -\frac{1}{n}\ln(1 - {}_nq_x). \tag{1.2.40} $$

The relation between mortality rates and death probabilities can be equivalently written as

$$ {}_nq_x = 1 - e^{-n\cdot {}_nm_x}. \tag{1.2.41} $$

In the special case of $n = 1$, the above formula reduces to

$$ q_x = 1 - e^{-m_x}. \tag{1.2.42} $$

1.2.4. Other life-table measures

Since John Graunt life-tables has been constructed by demographers, actuaries, statisticians and others to present mortality over the whole lifespan of a real or hypothetical cohort. Cohort age-specific mortality rates ${}_nm_x$ and death probabilities ${}_nq_x$ represent the main life-table parameters. They are also called tabular mortality parameters, because they are calculated for arbitrarily defined age intervals $[x, x+n)$, $n \in \mathbb{N}$.

Other major life-table measures of mortality are derived from death probabilities q_x. These are, for instance, person-years lived above age x, T_x or life expectancy, e_x (see e.g. [Balicki 2006]). The definitions for these characteristics are respectively:

T_x – person-years lived above age x, the remaining lifetime for all individuals surviving to the age of x

$$ T_x = \sum_{y=x}^{\infty} K_y, \tag{1.2.43} $$

e_x – remaining life expectancy, the average number of additional years that a survivor to age x will live beyond that age

$$e_x = \frac{T_x}{L_x}. \tag{1.2.44}$$

In practice, calculation of the above values from the observations of real cohorts (generations) causes some difficulties, because cohort data might be unavailable, outdated, or incomplete. Therefore, demographers have developed a concept of "a period life-table". In this approach, it is usually assumed that there is a hypothetical cohort, which is subjected throughout its life to a set of mortality conditions of the given period, which is usually a fixed calendar year. This solution allows calculating period mortality rates for all age groups and for an individual calendar year. Life-table characteristics that are calculated from period data will be marked with the symbol of the year t, e.g. $m_{x,t}, q_{x,t}, e_{x,t}$.

1.3. Continuous-time mortality frameworks

In the analysis of discrete life-table models, arbitrary, discrete age intervals $[x, x+n)$ are assumed, where x, n are non-negative integers. However, for some uses it is important that some mortality functions are calculated for any non-negative real ages x or for intervals $[x, x+y)$ of any small length $y > 0$. For this purpose, the lifetime of an individual is treated as a random variable of some continuous probability distribution, what is a natural extension, since lifetime and mortality evolve continuously.

1.3.1. Survival distributions

Definition 1.4. Let X be a non-negative and continuous scalar random variable representing the lifetime of a new-born (age-at-death) and let F_X be a cumulative distribution function of X, i.e.

$$F_X(x) = P(X < x), \tag{1.3.1}$$

for which $F_X(0) = 0$.

The survival function S_X of variable X is a complementary function to F_X of the form

$$S_X(x) = 1 - F_X(x). \tag{1.3.2}$$

To keep things simple, let functions F_X, S_X be denoted as F, S, respectively. It is by definition that for given $x \geq 0$ the value of $S(x)$ stands for the probability of a new-born surviving until the age x.

Definition 1.5. Let $Y(x)$ be a scalar random variable defined as

$$Y(x) = X - x \quad \text{for} \quad X \geq x, \tag{1.3.3}$$

where $x \geq 0$ is a known real number.

Random variable $Y(x)$ represents the residual lifetime of an individual aged x. The cumulative distribution function $F_{Y(x)}(y)$ of $Y(x)$ for a given $y \geq 0$ is as follows

$$\begin{aligned} F_{Y(x)}(y) &= P(Y(x) < y) = \\ &= P(X - x < y \mid X \geq x) = P(X < x + y \mid X \geq x) = \\ &= 1 - P(X \geq x + y \mid X \geq x) = \\ &= 1 - \frac{P(X \geq y + x)}{P(X \geq x)} = 1 - \frac{S(x+y)}{S(x)}. \end{aligned} \tag{1.3.4}$$

Let us notice that for $x = 0$ we have $Y(0) = X$ and $F_{Y(0)} = F$, which shows that variable X is a special case of $Y(x)$.

Definition 1.6. Let F' be a derivative of cumulative distribution function F, i.e. the density function f of variable X. The force of mortality is defined as the ratio

$$\mu(x) = \frac{f(x)}{S(x)}, \quad x \geq 0. \tag{1.3.5}$$

Expression $\mu(x)dx$ approximates the probability of dying in the age range $[x, x + dx)$ given that the person is surviving until the age x.

Integrating both sides of the expression (1.3.5) over interval $[0, x]$, we get

$$\int_0^x \mu(z)dz = \int_0^x \frac{f(z)}{S(z)} dz. \tag{1.3.6}$$

Let us apply the variable substitution $v = S(z)$. Then we have $dv = S'(z)dz$. Since the derivative of $S(z)$ is $-f(z)$, so $dv = -f(z)dz$. From this, we get

$$\int_0^x \mu(z)dz = -\int_1^{S(x)} \frac{1}{v}dv = -\ln v\Big|_1^{S(x)} = -\ln S(x). \tag{1.3.7}$$

Therefore, the following relations are true

$$F(x) = 1 - \exp\left\{-\int_0^x \mu(z)dz\right\} \tag{1.3.8}$$

$$S(x) = \exp\left\{-\int_0^x \mu(z)dz\right\}. \tag{1.3.9}$$

We also have

$$\begin{aligned} F_{Y(x)}(y) &= 1 - \frac{S(x+y)}{S(x)} = \\ &= 1 - \frac{\exp\left\{-\int_0^{x+y} \mu(z)dz\right\}}{\exp\left\{-\int_0^x \mu(z)dz\right\}} = \\ &= 1 - \exp\left\{-\int_x^{x+y} \mu(z)dz\right\} \end{aligned} \tag{1.3.10}$$

and

$$\begin{aligned} S_{Y(x)}(y) &= 1 - F_{Y(x)}(y) = \frac{S(x+y)}{S(x)} = \\ &= \exp\left\{-\int_x^{x+y} \mu(z)dz\right\}. \end{aligned} \tag{1.3.11}$$

Let us notice that the force of mortality $\mu(x)$ identifies the distributions of both random variables X and $Y(x)$.

Using the actuarial notation, the distribution functions (1.3.8) and (1.3.10) will be denoted by ${}_xq_0$ and ${}_yq_x$ and the survival functions (1.3.9) and (1.3.11) will be denoted by ${}_xp_0$ and ${}_yp_x$, respectively.

With the above formula (1.3.10), it is also easy to calculate density $f_{Y(x)}$ of random variable $Y(x)$. Since

$$F_{Y(x)}(y) = 1 - \frac{S(x+y)}{S(x)}, \quad x \geq 0, \tag{1.3.12}$$

thus for any $y \geq 0$, we have

$$\begin{aligned} f_{Y(x)}(y) &= \frac{dF_{Y(x)}(y)}{dy} = -\frac{1}{S(x)} \frac{dS(x+y)}{dy} = \\ &= -\frac{1}{S(x)} (-f(x+y)) = \frac{f(x+y)}{S(x)} = \\ &= \frac{f(x+y)}{S(x+y)} \frac{S(x+y)}{S(x)} = \mu(x+y)\, {}_y p_x = \\ &= \mu(x+y) \exp\left\{ -\int_x^{x+y} \mu(z) dz \right\}, \end{aligned} \tag{1.3.13}$$

where ${}_y p_x = 1 - {}_y q_x$.

It follows from (1.3.11) and (1.3.13) that the force of mortality $\mu_x(y)$ of the variable $Y(x)$ equals to

$$\mu_x(y) = \frac{f_{Y(x)}(y)}{S_{Y(x)}(y)} = \frac{\mu(x+y) \exp\left\{ -\int_x^{x+y} \mu(z) dz \right\}}{\exp\left\{ -\int_x^{x+y} \mu(z) dz \right\}} = \mu(x+y), \tag{1.3.14}$$

for $x, y \geq 0$.

1.3.2. The relationship between the mortality rate and the force of mortality

It is worth reminding here the relationship between the cohort mortality rate (1.2.2) and the force of mortality (1.3.5). The cohort mortality rate ${}_n m_x$ is defined as the following ratio

$${}_n m_x = \frac{{}_n D_x}{{}_n K_x}. \tag{1.3.15}$$

The number of deaths ${}_nD_x$ in the age range $[x, x+n)$ can be written as

$$ {}_nD_x = L_x - L_{x+n}, \tag{1.3.16} $$

where L_x the number of survivors until the age x.

The time-to-exposure for this interval is given as

$$ {}_nK_x = nL_{x+n} + {}_na_x \cdot {}_nD_x. \tag{1.3.17} $$

We now let the age interval range n to be positive real number, not necessarily an integer, tending to the limit $n \to 0$. Let us investigate $\lim_{n\to 0} {}_nm_x$.

For a small value of n, the time-to-exposure ${}_nK_x$ can be approximated as

$$ {}_nK_x \approx nL_x, \tag{1.3.18} $$

The approximation is the more accurate, the shorter length n of an interval $[x, x+n)$. For $n \to 0$ we have

$$ \lim_{n\to 0} {}_nm_x = \lim_{n\to 0} \frac{L_x - L_{x+n}}{nL_x}. \tag{1.3.19} $$

From the definition of a function derivative, we have

$$ \lim_{n\to 0} \frac{L_x - L_{x+n}}{n} = -L'_x. \tag{1.3.20} $$

Hence, we obtain

$$ \lim_{n\to 0} {}_nm_x = -\frac{L'_x}{L_x}. \tag{1.3.21} $$

Let us notice that the following equality holds

$$ L_x = L_0 S(x), \tag{1.3.22} $$

where L_0 is the number of births (real or hypothetical) and $S(x)$ is the probability of survival until age x. Finally, we have

$$ \lim_{n\to 0} {}_nm_x = -\frac{L_0 S'(x)}{L_0 S(x)} = \frac{f(x)}{S(x)} = \mu(x). \tag{1.3.23} $$

It follows that the force of mortality $\mu(x)$ is the limit to which the mortality rate ${}_nm_x$ tends with $n \to 0$.

1.4. Laws of mortality

Many models were proposed in the literature with the intention of defining "a simple law of mortality" expressing the way in which mortality changes age by age.

Different authors have considered mathematical functions depending on age x, among others $\mu(x)$, $S(x)$, ${}_xp_0$ or q_x as most suitable to represent the law of mortality.

In the simplest exponential model the force of mortality is constant

$$\mu(x) \equiv \mu = \text{const}, \quad x \geq 0. \tag{1.4.1}$$

However, the exponential model is not appropriate for human populations, as it assumes that the force of mortality is the same for all ages x. Other demographic models use a more realistic assumption that, for instance, $\mu(x)$ is piecewise constant, which in the case of sufficiently narrow age intervals offers a good approximation of the force of mortality as it is. Making an assumption about a piecewise constant force of mortality reduces, in fact, the problem to the approach considered in Section 1.2.3.

Among the better known laws of mortality are the historically oldest de Moivre model, the Lambert model and the Gompertz–Makeham or Weibull models [Frątczak 1997, Ostasiewicz 2011].

The assumption on which de Moivre built his model [Moivre 1725] states that there is a limit age X. Then the force of mortality is given by the formula

$$\mu(x) = \frac{1}{X - x} \quad \text{for} \quad 0 \leq x < X \quad \text{and} \quad \mu(x) = 0 \quad \text{for} \quad x \geq X. \tag{1.4.2}$$

This model is equivalent to an assumption that ${}_xp_0$ declines linearly with x. It follows from (1.4.2)

$${}_xp_0 = 1 - \frac{x}{X} \quad \text{for} \quad 0 \leq x < X \quad \text{and} \quad {}_xp_0 = 0 \quad \text{for} \quad x \geq X. \tag{1.4.3}$$

Lambert proposed a model for ${}_xp_0$ with four parameters [Lambert 1776]

$${}_xp_0 = \left[\frac{a - x}{x}\right]^2 - b\left[e^{-\frac{x}{c}} - e^{-\frac{x}{d}}\right]. \tag{1.4.4}$$

In the model of [Gompertz 1825] the force of mortality is defined as

$$\mu(x) = Bc^x, \quad B > 0, c \geq 1, \; x \geq 0, \tag{1.4.5}$$

what leads to the formula for ${}_xp_0$ of the form

$$ {}_xp_0 = e^{-k(c^x-1)}, \tag{1.4.6} $$

where $k = -B/\ln c$.

The Gompertz model is based on the assumption that the force of mortality increases exponentially with age. In 1867 Makeham modified this assumption by putting forward the following 3-parameter formula [Makeham 1867]

$$ \mu(x) = A + Bc^x, \quad A, B > 0, c \geq 1, x \geq 0, \tag{1.4.7} $$

or equivalently

$$ {}_xp_0 = e^{-k(bx+c^x-1)}, \tag{1.4.8} $$

where $b = A\ln c/B$ and $k = -B/\ln c$.

This model is an extended version of the Gompertz model by including a term A independent of age, representing constant level of mortality caused, for example, by accidents. Now (1.4.7) or (1.4.8) is called the Gompertz–Makeham mortality law and is frequently used by actuaries.

In 1939 Weibull proposed a 2-parameter formula [Weibull 1939]

$$ \mu(x) = ax^{b-1}, \quad a, b > 0, \ x \geq 0 \tag{1.4.9} $$

or equivalently

$$ {}_xp_0 = e^{-\frac{a}{b}x^b}. \tag{1.4.10} $$

It is worth noting that the Weibull force of mortality (1.4.9) is defined as a monotonic function (i.e. increasing, decreasing or constant), therefore it is inadequate as a mortality model of a human population with an unimodal or multi-modal force of mortality.

There are many studies in the contemporary literature that deal with models of the force of mortality, death probability etc. A review of historical mortality laws is provided in the books [Tabeau *et al.* 2001] and [Wunsch *et al.* 2002], among others (see Table 1.1).

Table 1.1. Mortality laws (overview of historical parametric models)

Author and year	Force of mortality μ_x, survival function $S(x)$ or probability of death q_x
[Moivre 1725]	$\mu(x) = \frac{1}{X-x}$
[Lambert 1776]	${}_xp_0 = \left[\frac{a-x}{x}\right]^2 - b\left[e^{-\frac{x}{c}} - e^{-\frac{x}{d}}\right]$
[Gompertz 1825]	$\mu(x) = Bc^x$
[Makeham 1867]	$\mu(x) = A + Bc^x$
[Opperman 1870]	$\mu(x) = \frac{a}{\sqrt{x}} + b + c\sqrt[3]{x}$ for young ages $x \in [0, 20]$
[Thiele, Sprague 1871]	$\mu(x) = a_1 e^{-b_1 x} + a_2 e^{-\frac{1}{2}b_2(x-c)^2} + a_3 e^{b_3 x}$
[Wittstein, Bumsted 1883]	$q_x = \frac{1}{m} a^{-(mx)^n} + a^{-(M-x)^n}$
[Steffensen 1930]	$\log_{10} S(x) = 10^{-A\sqrt{x}-B} + C$
[Perks 1932]	$\mu(x) = \frac{A+Bc^x}{1+Dc^x}$ $\mu(x) = \frac{A+Bc^x}{Kc^{-x}+1+Dc^x}$
[Harper 1936]	$\log_{10} S(x) = A + 10^{B\sqrt{x}+Cx+D}$
[Weibull 1939]	$\mu(x) = ax^{b-1}$
[Van der Maen 1943]	$\mu(x) = A + Bx + Cx^2 + \frac{1}{N-x}$ $\mu(x) = A + Bc^x + \frac{c}{N-x}$

Source: [Tabeau *et al.* 2001, p. 7] and [Wunsch *et al.* 2002, pp. 144–146]

[Thatcher *et al.* 1998] performed studies to fit different mathematical models to the mortality data of adult ages. They discovered that the logistic model was the best mathematical model of human adult mortality, even better then the popular Makeham–Gompertz model. The logistic force of mortality is a function of age x and can be written as

$$\mu(x) = \frac{ae^{bx}}{1 + ae^{bx}}, \quad x \geq 0. \tag{1.4.11}$$

Many studies deal with modeling of survival distributions, including the force of mortality, be means of parametric distributions, for instance the gamma, generalized gamma, log-gamma, log-normal distributions [Stacy 1962, Proschan 1963, Stacy, Mihram 1965, Harter 1967, Prentice 1974, DiCiccio 1987, Gupta *et al.* 1997] or the Pareto distribution [Quandt 1966, Malik 1970, Arnold 1983, Arnold, Press 1989, Brazauskas, Serfling 2000, 2001, Wu 2003].

Other classes of distributions that are considered are those with a non-monotonic force of mortality, including a quadratic function or, more generally, a polynomial or a bathtub function [Krane 1963, Kodlin 1967, Polovko 1968, Bain 1974, Hjorth 1980, Chen 2000].

In the 1960s, parametric regression models of the force of mortality were proposed, which assumed a multiplicative dependence of this function on the so-called base hazard and a set of risk factors [Feigl, Zelen 1965, Galsser 1967, Zippin, Armitage 1966].

Such models paved the way for the popular Cox semi-parametric model. Hazard estimation in these models boils down to the estimation of the base hazard and the regression coefficients of explanatory variables.

The Cox model is semi-parametric model, because it accepts any form of the base hazard function as a function of age and the parametric specification of the function of risk factors. If the explanatory variables are not time-dependent and the base hazard function is the same for all individuals in the population, the Cox model is called a model of proportional hazards. Its parameters are estimated by maximizing the so-called partial likelihood function [Cox 1972, Cox 1975].

An alternative to the multiplicative regression models of the force of mortality is additive regression models and proportional odds models [Bennett 1983, Pettitt 1984, Huffer, McKeague 1991, Lin 1991].

Other well-known models are of non-parametric kind, such as the Kaplan–Meier survival model or the Nelson–Aalen model of the cumulative force of mortality [Kaplan, Meier 1958, Nelson 1969, Aalen 1978].

Models described in this section belong to the group of models, i.e. the parameters are assumed to be constant over time. However, the actual mortality evolves continuously, therefore it seems justified to refit parameters in such models periodically to accommodate changes in mortality patterns.

1.5. The Lee–Carter model and its extensions

Long-term observations have shown significant improvements in mortality caused by different driving factors (e.g. medical advances, nutrition improvement, lifestyle changes), especially for older ages. For instance, it follows from the observation of mortality rates in developed countries that they are declining in time, while expected lifetimes are rising. Other general characteristics also change in time, e.g. the dominant lifetime (mode), the average age at death, or the maximal age at death. Thus, mortality profiles evolve in two main dimensions: age and time. However, some accidental shocks caused by war or natural disasters etc. can also appear.

The common features of mortality behavior that have been observed in developed countries since the second half of the 20th c. can be summed up as follows [Wilmoth, Horiuchi 1999, Vaupel *et al.* 2011]:

- mortality rates are falling at all ages,
- rate of decrease in mortality varies over time and by age group,
- the dominant lifetime (mode), the average age at death as well as the maximal age at death shift to the right, toward older ages,
- ages at death concentrating around the mode,
- the survival curve undergoing expansion and rectangularization (because of the aforementioned trends),
- life expectancy increases and life disparity decreases,
- levels of accidental deaths from external causes (injuries, accidents, poisoning) in the young population are rising, especially among young males aged 20+ years, with corresponding larger dispersion.

The observed two-dimensional evolution implies the use of models exhibiting both the age-period and stochastic nature of mortality. Some works on this subject have used for instance the time-series analysis to capture the general trend of mortality and its stochastic uncertainty. Nowadays, one of the most popular models of this type is the Lee–Carter model [Lee, Carter 1992]. Further, it will be termed as the Standard Lee–Carter model (SLC model).

In addition to static approaches presented in the previous sections, in the remainder of this chapter we review basic mortality models, both in discrete and continuous time.

Some sophisticated stochastic models can be performed in a framework of stochastic differential equations. Models based on such equations will be termed dynamic models or dynamic systems. In the dynamic context, force of mortality $\mu_x(t)$ is treated as a mortality intensity process and is derived as a solution of stochastic differential equations.

The representatives of this category of models are the Vasiček and Cox–Ingersoll–Ross models [Vasiček 1977, Cox *et al.* 1985a, 1985b], the dynamic Lee–Carter model [Rossa, Socha 2013], the Milevsky–Promislow model [Milevsky, Promislow 2001] as well as the Giacometti–Ortobelli–Bertocchi model [Giacometti *et al.* 2011]. The models are introduced in Sections 1.6–1.11.

1.5.1. The Lee–Carter model

Let $m_{x,t}$ be a central age-specific death rate for exact ages between x and $x+1$ registered for calendar year t

$$m_{x,t} = \frac{D_{x,t}}{\bar{L}_{x,t}}, \tag{1.5.1}$$

where

$D_{x,t}$ – number of deaths between ages x and $x+1$ in year t,
$\bar{L}_{x,t}$ – midyear population alive at the age x in year t,
$x = 0, 1, \ldots, X$ – subscripts denoting one-year age groups,
$t = 1, 2, \ldots, T$ – subscripts denoting calendar years.

The age-specific mortality rate $m_{x,t}$ is constructed as a ratio of deaths between ages x and $x+1$ to the midyear population alive at age x in year t, which is also referred to as the mean population in year t. Because of the midyear population being used in the denominator, (1.5.1) is described also as the central death rate.

The Lee–Carter model [Lee, Carter 1992] for the log-central death rates can be written as

$$\ln m_{x,t} = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}, \tag{1.5.2}$$

or, equivalently, as

$$m_{x,t} = \exp\{\alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}\}, \tag{1.5.3}$$

where α_x, β_x $(x = 0, 1, \ldots, X)$ and κ_t $(t = 1, 2, \ldots, T)$ are the unknown parameters and the double-indexed terms $\varepsilon_{x,t}$ are independent random variables, which are assumed to have the same normal distribution with mean $\mathrm{E}[\varepsilon_{x,t}] = 0$ and constant variance $\mathrm{Var}[\varepsilon_{x,t}] = \sigma^2$.

The system of equations (1.5.2) or (1.5.3) cannot be explicitly solved unless normalizing constraints are imposed. Indeed, let us assume for instance that the model (1.5.2) is valid for a set of parameters

$$\{\alpha_x,\ \beta_x,\ \kappa_t\}, \tag{1.5.4}$$

It is easy to see that the model holds true also for any constant c and parameters

$$\{\alpha_x - c\beta_x,\ \beta_x,\ \kappa_t + c\}, \quad \text{or} \quad \{\alpha_x,\ c\beta_x,\ \kappa_t/c\}. \tag{1.5.5}$$

Thus for full model identification, additional constraints are imposed. Lee and Carter assumed that the sum of parameters β_x is 1 and the sum of parameters κ_t is 0, i.e.

$$\sum_{x=0}^{X} \beta_x = 1 \tag{1.5.6}$$

and

$$\sum_{t=1}^{T} \kappa_t = 0. \tag{1.5.7}$$

The age-related effects α_x indicate the age profile of mortality, the time-related effects κ_t describe the general mortality trend, whereas β_x describe patterns of deviations from the age profile in response to change of the general trend.

It is worth noting that β_x could be negative at some ages, indicating that log-central mortality rates $\ln m_{x,t}$ at those ages tend to rise when falling at other ages. In other words, parameters β_x tell which rates decline rapidly and which slowly over time in response to change of κ_t.

Parameters α_x and β_x do not depend on time t, meaning that after they have been derived they can also be used to forecast $m_{x,t}$ for future periods $t > T$.

The time-related effects are κ_t. They can be modeled and predicted using, for instance, the time-series analysis. Lee and Carter applied a random walk model with a drift to find predicted $\tilde{\kappa}_t$ for $t > T$, but the range of proposals discussed in the literature is wider [Nielsen, Nielsen 2010].

A random walk process with a drift is given by the formula

$$\kappa_t = \delta + \kappa_{t-1} + \xi_t, \quad t \in \mathbb{N}, \tag{1.5.8}$$

where δ is a constant (a drift) and ξ_t are random terms.

Parameter δ in (1.5.8) usually takes negative values showing that mortality is declining. Random fluctuations are represented by independent random terms ξ_t, each having the normal distribution with mean 0 and finite variance $\sigma^2_{\xi_t}$.

With the predicted values $\tilde{\kappa}_t$ obtained from (1.5.8) for $t > T$ and the estimates a_x, b_x of parameters α_x, β_x, respectively, mortality rates $m_{x,t}$ can be predicted according to the formula

$$\tilde{m}_{x,t} = \exp\{\hat{\alpha}_x + \hat{\beta}_x \tilde{\kappa}_t\}, \tag{1.5.9}$$

where $\tilde{m}_{x,t}$ denote forecasts of $m_{x,t}$ for $t > T$.

The original approach used by Lee and Carter to estimate model's parameters β_x and κ_t is via Singular Value Decomposition (SVD) (a review of the SVD history was given by [Steward 1993]). This method was further developed by [Wilmoth 1993] as weighted SVD.

The SVD method allows decomposing any $m \times n$ matrix $\mathbf{W}$ into a matrix of singular values $\mathbf{D}$ and two matrices $\mathbf{U}$ and $\mathbf{V}$ of the left and right singular vectors, i.e.

$$\mathbf{W} = \mathbf{UDV}^T. \tag{1.5.10}$$

Matrix $\mathbf{D}$ takes the form

$$\mathbf{D} = \begin{bmatrix} \mathbf{\Sigma}_{r \times r} & \mathbf{0}_{r \times (n-r)} \\ \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{bmatrix}, \tag{1.5.11}$$

where

$$\mathbf{\Sigma}_{r \times r} = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d_r \end{bmatrix} \tag{1.5.12}$$

and r denotes the number of positive singular values $d_1, d_2, \dots, d_r$.

Singular values d_i are calculated as the square roots of the eigenvalues of matrix $\mathbf{W}^T\mathbf{W}$, and the orthogonal matrix $\mathbf{V}$ consists of the right singular column vectors $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$, derived as the eigenvectors of the square matrix $\mathbf{Y}^T\mathbf{Y}$. The orthogonal matrix $\mathbf{U}$ consists of the left singular vectors $[\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]$, where $\mathbf{u}_i = \frac{1}{d_i}\mathbf{W}\mathbf{v}_i$ for $i = 1, 2, \dots, r$.

It follows from (1.5.10) that each element $w_{x,t}$ of $\mathbf{W}$ can be represented by the following sum

$$w_{x,t} = \sum_{i=1}^{r} d_i u_{x,i} v_{t,i}, \tag{1.5.13}$$

where

$u_{x,i}$ – x-th element of the i-th left column vector of $\mathbf{U}$,
$v_{t,i}$ – t-th element of the i-th right column vector of $\mathbf{V}$,
d_i – i-th singular value of $\mathbf{W}$,

and

$$\sum_t v_{t,i} = 0, \quad \text{for} \quad i = 1, 2, \dots, r. \tag{1.5.14}$$

To express (1.5.13) in terms of the Lee-Carter model, let elements $w_{x,t}$ be defined as

$$w_{x,t} = \ln m_{x,t} - \alpha_x. \tag{1.5.15}$$

From (1.5.13) we have

$$\ln m_{x,t} - \alpha_x = \sum_{i=1}^{r} d_i u_{x,i} v_{t,i}, \tag{1.5.16}$$

or equivalently

$$\ln m_{x,t} = \alpha_x + \sum_{i=1}^{r} d_i u_{x,i} v_{t,i}. \tag{1.5.17}$$

Denoting

$$\beta_x^{(i)} = \frac{u_{x,i}}{\sum_{x=0}^{X} u_{x,i}}, \qquad \kappa_t^{(i)} = d_i v_{t,i} \sum_{x=0}^{X} u_{x,i}, \tag{1.5.18}$$

equality (1.5.13) reduces to the following one

$$\ln m_{x,t} = \alpha_x + \sum_{i=1}^{r} \beta_x^{(i)} \kappa_t^{(i)}, \tag{1.5.19}$$

where superscript (i) refers to the i-th singular value and to the i-th left and right singular vectors.

Let us reduce the number of components of the sum in (1.5.19) to the first one and let us replace the remaining components by $\epsilon_{x,t}$. For simplicity, we will denote $\beta_x^{(1)}, \kappa_t^{(1)}$ as β_x, κ_t, respectively. Thus, we have

$$\beta_x = \frac{u_{x,1}}{\sum_{x=0}^{X} u_{x,1}}, \quad \kappa_t = d_1 v_{t,1} \sum_{x=0}^{X} u_{x,1}. \tag{1.5.20}$$

We get from (1.5.19)

$$\ln m_{x,t} = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \tag{1.5.21}$$

where

$$\sum_{x=0}^{X} \beta_x = 1, \quad \sum_{t=1}^{T} \kappa_t = 0. \tag{1.5.22}$$

It is worth noting that such an approach assumes homoscedasticity of residuals $\epsilon_{x,t}$, i.e. it is assumed that the variance of $\epsilon_{x,t}$ is constant across age x and time t.

In order to determine α_x, we refer to the assumption that random terms $\epsilon_{x,t}$ have expectation 0, i.e.

$$\mathrm{E}[\epsilon_{x,t}] = 0. \tag{1.5.23}$$

It means that the following equality holds, i.e.

$$\sum_{t=1}^{T} [\ln m_{x,t} - (\alpha + \beta_x \kappa_t)] = 0. \tag{1.5.24}$$

After simple transformations we get

$$T\alpha_x + \beta_x \sum_{t=1}^{T} \kappa_t = \sum_{t=1}^{T} \ln m_{x,t}. \tag{1.5.25}$$

Because of the constraint (1.5.7), equality (1.5.25) can be reduced to

$$\alpha_x = \frac{1}{T} \sum_{t=1}^{T} \ln m_{x,t}. \tag{1.5.26}$$

Expressions (1.5.20) and (1.5.26) can be used to define estimators a_x, b_x, k_t of parameters $\alpha_x \beta_x, \kappa_t$ of the SLC model by taking the first singular value as well as the first right and left singular vectors of the matrix $\mathbf{W} = [\ln m_{x,t} - a_x]$, where a_x for $x = 0, 1, \ldots, X$ represent arithmetic averages of log-central death rates in rows of the sample matrix $\mathbf{M} = [\ln m_{x,t}]_{X+1 \times T}$.

Lee and Carter suggested, that after the parameters κ_t have been estimated, they can be re-estimated using a different criterion. However, we will skip this re-estimation stage, since it is not a defining feature of this method.

Another well-known method of parameters' estimation is via maximum likelihood, assuming the Poisson distribution of the number of deaths [Brouhns *et al.* 2002].

For constant mortality rates in age intervals $[x, x+1)$, number of deaths $D_{x,t}$ can be treated as independent random variables with the Poisson distribution, i.e.

$$D_{x,t} \sim \text{Poisson}(\lambda_{x,t}), \quad x = 0, 1, \ldots, X, \ t = 1, 2, \ldots, T, \qquad (1.5.27)$$

where $\lambda_{x,t} = K_{x,t} m_{x,t}$ represents one parameter of the Poisson distribution, $K_{x,t}$ is time-to-exposure in age group $[x, x+1)$ and $m_{x,t}$ is the age-specific central mortality rate.

The estimators of $\alpha_x, \beta_x, \kappa_t$ are then derived by means of the maximum likelihood method, where the likelihood function is defined as

$$L(\alpha_x, \beta_x, \kappa_t | D_{x,t}, K_{x,t}) = \prod_{x=0}^{X} \prod_{t=1}^{T} e^{-K_{x,t} m_{x,t}} \frac{(K_{x,t} m_{x,t})^{D_{x,t}}}{D_{x,t}!}. \qquad (1.5.28)$$

The logarithm of (1.5.28) can be expressed as

$$\ln L(\alpha_x, \beta_x, \kappa_t | D_{x,t}, K_{x,t}) = \sum_{x=0}^{X} \sum_{t=1}^{T} D_{x,t} \ln m_{x,t} - K_{x,t} m_{x,t} + C, \quad (1.5.29)$$

where

$$C = D_{x,t} \ln K_{x,t} - \ln(D_{x,t}!). \qquad (1.5.30)$$

Let us assume that (1.5.2) and (1.5.3) hold. Then (1.5.29) can be written as

$$\ln L(\alpha_x, \beta_x, \kappa_t | D_{x,t}, K_{x,t}) =$$

$$= \sum_{x=0}^{X} \sum_{t=1}^{T} D_{x,t} (\alpha_x + \beta_x \kappa_t) - K_{x,t} \exp\{\alpha_x + \beta_x \kappa_t\} + C, \qquad (1.5.31)$$

where C is defined in (1.5.30). Note that C is independent on the estimated parameters.

Estimators are defined as such values of $\alpha_x, \beta_x, \kappa_t$ for which function (1.5.31) reaches maximum. The maximum is found by means of iterative algorithms [Brouhns *et al.* 2002].

1.5.2. Age-period-cohort modifications

The SLC model (1.5.2) has been frequently modified by different authors mainly by allowing successive extensions of the right-hand side of (1.5.17) [Renshaw, Haberman 2003a, 2003b, 2006, Currie 2006].

Some authors have considered extrapolative models by substituting the so-called logits of death probabilities $q_{x,t}$ for the log-central death rates $\ln m_{x,t}$ [Cairns *et al.* 2006, 2009].

The $q_{x,t}$ logit can be defined as

$$\eta_{x,t} \equiv \text{logit } q_{x,t} = \ln \frac{q_{x,t}}{1 - q_{x,t}}, \tag{1.5.32}$$

where $q_{x,t}$ is the probability of death during one year for individuals aged x last birthday in the calendar year t.

Between $m_{x,t}$ and $q_{x,t}$ there is a general relation (1.2.11). Depending on the assumed interpolation model (Section 1.2.3), the relation can be reduced to (1.2.21) or (1.2.42), i.e.

$$q_{x,t} = \frac{2m_{x,t}}{2 + m_{x,t}}, \tag{1.5.33}$$

or

$$q_{x,t} = 1 - \exp\{-m_{x,t}\}. \tag{1.5.34}$$

Some extensions of the SLC model take also into account of the so-called cohort-related effects γ_{t-x}, which are functions of the year of birth of persons aged x.

The reason for adding γ_{t-x} is that cohorts may differ in terms of the course and pace of changes in mortality. The sub-index $t - x$ of the cohort parameter γ_{t-x} stands for the year in which the cohort was born.

Several models for $\log m_{x,t}$ or $\eta_{x,t}$, incorporating age- and time-related effects as well as cohort effects, are discussed in [Cairns *et al.* 2006, 2009, Van Berkum *et al.* 2013] as listed below in (1.5.35).

$$\begin{aligned}
M1:&\ \log m_{x,t} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)},\\
M2:&\ \log m_{x,t} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\gamma_{t-x},\\
M3:&\ \log m_{x,t} = \alpha_x + \kappa_t^{(1)} + \gamma_{t-x},\\
M4:&\ \log m_{x,t} = \sum_{i,j}\theta_{ij}B_{ij}^{ay}(x,t),\\
M5:&\ \eta_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}),\\
M6:&\ \eta_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \gamma_{t-x},\\
M7:&\ \eta_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}((x-\bar{x})^2-\sigma_x^2) + \gamma_{t-x},\\
M8:&\ \eta_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \gamma_{t-x}(x_c - x),
\end{aligned}\tag{1.5.35}$$

where

α_x and $\beta_x^{(i)}$ represent the age-related effects,
$\kappa_t^{(i)}$ are the time-related effects,
γ_c serve for the cohort-related effects, with $c = t - x$,
$\bar{x}$ and σ_x^2 are, respectively, average age and its variance in the age group under consideration, i.e.

$$\bar{x} = \frac{1}{n}\sum_{x=x_1}^{x_n} x, \quad \sigma_x^2 = \frac{1}{n}\sum_{x=x_1}^{x_n}(x-\bar{x})^2, \tag{1.5.36}$$

x_1 is the youngest and x_n the oldest age included in the data set,
x_c is a given constant adjusted to the age range $[x_1, x_n]$,
$B_{ij}^{ay}(x,t)$ denote the splines and θ_{ij} are their weights.

Which model is selected depends on our knowledge and beliefs about the mortality behavior in a given population. Let us notice that M1 represents simply the SLC model and M2 is its generalization [Renshaw, Haberman 2003a, 2003b, 2006], as it additionally takes account of cohort effects. Both models are equivalent for $\gamma_{t-x} = 0$.

Since M2 shares the identifiability problem like M1 does, additional constraints are imposed

$$\sum_x \beta_x^{(i)} = 1,\ i = 1,2, \quad \sum_t \kappa_t^{(1)} = 0, \quad \sum_{c\in C}\gamma_c = 0, \tag{1.5.37}$$

where $c = t - x$ and C is the set of years in which the analyzed generations were born.

From the second and third constraints it follows that α_x present in the model M2 are the arithmetic averages of the log-central death rates. The other parameters can be estimated using an iterative method [Renshaw 2006].

Model M3 is a special case of M2, when $\beta_x^{(1)} = \beta_x^{(2)} = 1$. In this model additional constraints are imposed, i.e.

$$\sum_t \kappa_t^{(1)} = 0, \quad \sum_{c \in C} \gamma_c = 0. \tag{1.5.38}$$

Model M4 assumes that there exists some surface reflecting smooth arrangement of log-central death rates in two dimensions: age x and time t. This approach is basically different from that used in models M1–M3, which do not assume smooth transition of mortality rates between age groups and calendar years, but rather jumping changes arising from casual factors.

A different class of models is represented by models M5–M8, which have logits (1.5.32) on the left-hand sides instead of the log-central death rates. We will call them the logit mortality models. In such models analogous parameters are used as in M1–M4 to represent the effects of age, time and cohort. The simplest one is M5 with two parameters $\kappa_t^{(1)}, \kappa_t^{(2)}$ and without any additional constraints.

The other three models, M6–M8, are the extended versions of M5 that incorporate cohort-related effects. However, because of the identifiability problem additional constraints have to be imposed on these parameters. In the case of M6, the constraints have the following form

$$\sum_{c \in C} \gamma_c = 0, \quad \sum_{c \in C} c\gamma_c = 0, \tag{1.5.39}$$

where $c = t - x$ and C is the set of years in which the analyzed generations were born.

Constraints (1.5.39) follow from the following reasoning. If we use the least squares method to fit linear function $\phi_1 + \phi_2 c$ to γ_c, then the fitted function should by identically equal to zero, what means that both scalars ϕ_1, ϕ_2 should satisfy equalities

$$\phi_1 = \phi_2 = 0. \tag{1.5.40}$$

This requires constraints (1.5.39) to ensure that equalities (1.5.40) hold. It follows from these constraints that estimates of γ_c will be centered around zero and there will be no constant trend up or down.

In M7 a quadratic term to the age effect is added inspired by the possible curvature identified in the plot of the $q_{x,t}$ logits. Therefore, three constraints are imposed on the cohort effects

$$\sum_{c\in C}\gamma_c = 0, \quad \sum_{c\in C} c\gamma_c = 0, \quad \sum_{c\in C} c^2\gamma_c = 0. \tag{1.5.41}$$

Constraints (1.5.41) enforce a quadratic function $\phi_1 + \phi_2 c + \phi_3 c^2$ fitted to γ_c to be identically equal to zero, what ensures that scalars ϕ_1, ϕ_2, ϕ_3 satisfy equalities

$$\phi_1 = \phi_2 = \phi_3 = 0, \tag{1.5.42}$$

what allows to obtain estimates of γ_c fluctuating around zero with no trend up or down and no systematic curvature [Cairns *et al.* 2006, 2009].

In $M8$ one simple constraint imposed on cohort effects is assumed

$$\sum_{c\in C}\gamma_c = 0. \tag{1.5.43}$$

The comparative study of the logit mortality models is presented more at length in [Haberman, Renshaw 2008, 2011], where the following models are investigated:

$$\begin{aligned}
&LC: \ \eta_{x,t} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)},\\
&H_1: \ \eta_{x,t} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \gamma_{t-x},\\
&M: \ \eta_{x,t} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\gamma_{t-x},\\
&LC2: \ \eta_{x,t} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)},\\
&M5: \ \eta_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}),\\
&M6: \ \eta_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \gamma_{t-x},\\
&M7: \ \eta_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}\upsilon_x + \gamma_{t-x},\\
&M8: \ \eta_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \gamma_{t-x}(x_c - x),\\
&M5^*: \ \eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}(\bar{x}-x)^+,\\
&M6^*: \ \eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}(\bar{x}-x)^+ + \gamma_{t-x},\\
&M7^*: \ \eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}(\bar{x}-x)^+ + \kappa_t^{(4)}\upsilon_x + \gamma_{t-x},\\
&M8^*: \ \eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}(\bar{x}-x)^+ + \gamma_{t-x}(x_c - x),
\end{aligned} \tag{1.5.44}$$

where

$\alpha_x, \beta_x^{(i)}$ are the age-related effects,
$\kappa_t^{(i)}$ are the time-related effects,
γ_{t-x} are the cohort-related effects,
x_c is a constant parameter,
$\bar{x}$ and σ_x^2 are, respectively, average age and age variance in the analyzed data set defined in (1.5.36),
v_x are coefficients expressed as $v_x = (x-\bar{x})^2 - \sigma_x^2$.

Components

$$(\bar{x}-x)^+ = \max(\bar{x}-x, 0) \tag{1.5.45}$$

are included in the group of models M5*–M8* because of additional parameters $\kappa_t^{(3)}$ representing higher level of mortality in young age groups, i.e. for $x < \bar{x}$.

Parameters of models (1.5.44) are estimated by means of iterative methods [Haberman, Renshaw 2008, 2011].

The generalized family of the above models can be given by the following one

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + (x_c - x)^+\kappa_t^{(3)} + v_x\kappa_t^{(4)} + (x_c - x)\gamma_{t-x}, \tag{1.5.46}$$

which serves as the generalization of M7* and M8*.

1.5.3. The fuzzy Lee–Carter model

One of the most interesting extensions of the SLC model was proposed by [Koissi, Shapiro 2006]. In their version of the SLC model – called the fuzzy Lee–Carter model (FLC model) or the Koissi–Shapiro model – fuzzy representation of the mortality data is assumed. This concept allows taking account of uncertainty involved in mortality and including random terms into the fuzzy structure of the model. Their approach builds on the assumption that exact mortality rates are usually not known, therefore fuzzified mortality data should be used.

It is well-known that death statistics are subject to reporting errors of several kinds. They may be reported for incorrect year, area, or assigned statistics that are incorrect, e.g. age. Moreover, the midyear population data that often serve as the denominators of mortality rates are also the subject of errors. It is regarded as the population at July 1 and is assumed to be the point at which half of the deaths in the population during the year have occurred. Such an estimate can be actually underestimated or overestimated. For these reasons, mortality rates cannot be exactly determined and their fuzzy representation seems to be justified.

[Koissi, Shapiro 2006] created fuzzy death rates $Y_{x,t}$ by converting log-central mortality rates $y_{x,t} = \ln m_{x,t}$ into symmetric, triangular fuzzy numbers $Y_{x,t}$ expressed by ordered pairs $(y_{x,t}, e_{x,t})$, i.e.

$$Y_{x,t} = (y_{x,t}, e_{x,t}), \quad x = 0, 1, \ldots, X, \; t = 1, 2, \ldots, T, \tag{1.5.47}$$

where $y_{x,t} = \ln m_{x,t}$ and $e_{x,t}$ are the so-called central values and spreads, respectively (basic notions relating to fuzzy numbers are provided in Chapter 4, see especially Definition 4.3 and Definition 4.4).

This approach supports the structure of the FLC model, where the role of the explanatory variable is played by the fuzzy log-central mortality rates (1.5.47).

The FLC model is then defined as

$$Y_{x,t} = A_x \oplus (B_x \odot K_t), \quad x = 0, 1, \ldots, X, \; t = 1, 2, \ldots, T, \tag{1.5.48}$$

where $A_x = (\alpha_x, s_{A_x}), B_x = (\beta_x, s_{B_x}), K_t = (\kappa_x, s_{K_t})$ are fuzzy triangular symmetric numbers with central values $\alpha_x, \beta_x, \kappa_x$ and spreads $s_{A_x}, s_{B_x}, s_{K_x}$, respectively, representing unknown parameters and $\oplus$, $\odot$ are the addition and multiplication operators of fuzzy numbers (Definition 4.6).

Koissi and Shapiro assumed that the parameters of the FLC model be estimated minimizing a criterion function based on the Diamond distance (Definition 4.8). The components S_1, S_2 of the criterion $S = S_1 + S_2$ can be written as follows

$$\begin{aligned} S_1(\alpha_x, \beta_x, \kappa_t) = \sum_{x=0}^{X} \sum_{t=1}^{T} \big[& 3\alpha_x^2 + 3(\beta_x \kappa_t)^2 + 3y_{x,t}^2 + 6\alpha_x \beta_x \kappa_t \\ & - 4\alpha_x y_{x,t} - 4\beta_x \kappa_t y_{x,t} + 2e_{x,t}^2 \big], \\ S_2(\beta_x, \kappa_t, s_{A_x}, s_{B_x}, s_{K_t}) = 2\sum_{x=0}^{X} \sum_{t=1}^{T} \big[& (\max\{s_{A_x}, |\beta_x| s_{K_t}, |\kappa_t| s_{B_x}\})^2 \\ & - 2e_{x,t} \max\{s_{A_x}, |\beta_x| s_{K_t}, |\kappa_t| s_{B_x}\} \big]. \end{aligned} \tag{1.5.49}$$

However, the FLC model poses a major problem in estimation of parameters, since the expression $\max\{s_{A_x}, |\beta_x| s_{K_t}, |\kappa_t| s_{B_x}\}$ appearing in S_2 prevents the use of standard non-linear optimization methods. In Chapters 4 and 5, new fuzzy mortality models simplifying the estimation procedure are proposed.

1.6. The dynamic Lee–Carter model

In contrast to the SLC model, [Rossa, Socha 2013] proposed the dynamic dynamic Lee–Carter models described by means of the Itô stochastic differential equations. Both models are discussed in this section.

1.6.1. Dynamic LC model

In the Dynamic Lee–Carter model (DLC model) the force of mortality $\mu_x(t)$ is expressed as

$$d\mu_x(t) = \left(\gamma_x(t) + \frac{1}{2}\sigma_x^2\right)\mu_x(t)dt + \sigma_x \mu_x(t) dw(t), \quad t \in \mathbb{R}^+, \tag{1.6.1}$$

$$\gamma_x(t) = \beta_x \kappa'(t), \quad \mu_x(t_0) = e^{\alpha_x + \beta_x \kappa(t_0)}, \tag{1.6.2}$$

where

α_x, β_x are age-related scalar coefficients,

$\kappa(t)$ is a scalar, differentiable, deterministic function of time t, with an initial value $\kappa(t_0)$,

$\sigma_x > 0$ are age-specific volatility parameters,

$w(t)$ is a standard Wiener process.

The solution of (1.6.1)–(1.6.2) follows from the Itô formula (Theorem A.7, formula (A.2.21) in Appendix A). It takes the form

$$\ln \mu_x(t) = \alpha_x + \beta_x \kappa(t) + \sigma_x w(t), \tag{1.6.3}$$

or, equivalently

$$\mu_x(t) = \exp\left\{\alpha_x + \beta_x \kappa(t) + \sigma_x w(t)\right\}. \tag{1.6.4}$$

This model structurally resembles the SLC model (see (1.5.2) or (1.5.3)), but differs from it in the properties of random terms.

Let us assume a linear form of function $\kappa(t)$

$$\kappa(t) = \chi + \delta t, \tag{1.6.5}$$

such that

$$\int_{t_0}^{t_1} \kappa(t) dt = 0, \tag{1.6.6}$$

where $[t_0, t_1]$ represents the interval of observation period.

Additionally, for full model identification the following constraint is imposed

$$\sum_{x=0}^{X} \beta_x = 1. \tag{1.6.7}$$

Then (1.6.3) is written as

$$\ln \mu_x(t) = \alpha_x + \beta_x(\chi + \delta t) + \sigma_x w(t). \tag{1.6.8}$$

1.6.2. Discrete dynamic LC model

The Discrete Dynamic Lee–Carter model (DDLC model) can be derived for $t \in \mathbb{N}$ by subtracting $\ln \mu_x(t)$ from $\ln \mu_x(t+1)$, where $\ln \mu_x(t)$ is defined in (1.6.8). As a result, we have

$$\ln \mu_x(t+1) = \ln \mu_x(t) + \beta_x \delta + \sigma_x \epsilon_{x,t+1}, \quad t \in \mathbb{N}, \tag{1.6.9}$$

or, substituting $\xi_{x,t+1}$ for $\sigma_x \epsilon_{x,t+1}$, we also have

$$\ln \mu_x(t+1) = \ln \mu_x(t) + \beta_x \delta + \xi_{x,t+1}, \quad t \in \mathbb{N}, \tag{1.6.10}$$

subject to constraints (1.6.5)–(1.6.7).

Terms $\xi_{x,t+1}$ are Gaussian random variables with mean $\mathrm{E}[\xi_{x,t+1}]$ and variance $\mathrm{Var}[\xi_{x,t+1}]$ equal, respectively,

$$\mathrm{E}[\xi_{x,t+1}] = 0, \quad \mathrm{Var}[\xi_{x,t+1}] = \mathrm{E}[\xi_{x,t+1}^2] = \sigma_x^2. \tag{1.6.11}$$

Note that it follows from (1.6.10) that

$$\xi_{x,t+1} = \ln \mu_x(t+1) - \ln \mu_x(t) - \beta_x \delta. \tag{1.6.12}$$

1.6.3. Parameters' estimation of the dynamic LC model

Let us consider the discrete dynamic LC model (1.6.10) subject to constraints (1.6.5)–(1.6.7). Let us denote by a_x, b_x, s_x^2, d, c estimators of parameters $\alpha_x, \beta_x, \sigma_x^2, \delta, \chi$, respectively.

The model's parameters can be estimated using the method of moments. The first and the second raw moments of random term (1.6.12) are given in (1.6.11). We will consider the following two moments

$$\begin{cases} \mathrm{E}\left[\ln \mu_x(t+1) - \ln \mu_x(t) - \beta_x\delta\right], \\ \\ \mathrm{E}\left[\left(\ln \mu_x(t+1) - \ln \mu_x(t) - \beta_x\delta\right)^2 - \sigma_x^2\right]. \end{cases} \tag{1.6.13}$$

Note that the moments are defined so that both are equal 0. Equaling analogous sample moments to 0 we obtain moment equations from which estimators b_x and s_x^2 can be determined. The sample moment equations are as follows

$$\begin{cases} \frac{1}{t_1-t_0}\sum_{t=t_0}^{t_1-1}\left[\ln m_{x,t+1} - \ln m_{x,t} - b_x d\right] = 0, \\ \\ \frac{1}{t_1-t_0}\sum_{t=t_0}^{t_1-1}\left[\left(\ln m_{x,t+1} - \ln m_{x,t} - b_x d\right)^2 - s_x^2\right] = 0, \end{cases} \tag{1.6.14}$$

where $\ln m_{x,t}$ are log-central death rates from a sample time series $\{\ln m_{x,t},\ t = t_0, t_0+1, \ldots, t_1\}$.

From the first moment equation of (1.6.14) we receive

$$b_x = \frac{\ln m_{x,t_1} - \ln m_{x,t_0}}{d(t_1 - t_0)}, \quad x = 0, 1, \ldots, X, \tag{1.6.15}$$

while from the second moment equation there is

$$s_x^2 = \frac{1}{t_1 - t_0}\sum_{t=t_0}^{t_1-1}\left(\ln m_{x,t+1} - \ln m_{x,t} - b_x d\right)^2,\ x = 0, 1, \ldots, X. \tag{1.6.16}$$

Moreover, the following equality holds

$$\sum_{x=0}^{X}\sum_{t=t_0}^{t_1-1}\left(\ln m_{x,t+1} - \ln m_{x,t} - b_x d\right) = 0. \tag{1.6.17}$$

Thus, allowing for condition (1.6.7) we get from (1.6.17)

$$d = \frac{1}{t_1 - t_0} \sum_{x=0}^{X} (\ln m_{x,t_1} - \ln m_{x,t_0}). \tag{1.6.18}$$

Let us denote $v_{x,t} = \ln m_{x,t+1} - \ln m_{x,t}$, then estimators d, b_x and s_x^2 can be expressed as

$$d = \sum_{x=0}^{X} \bar{v}_x, \tag{1.6.19}$$

$$b_x = \frac{\bar{v}_x}{d} = \frac{\bar{v}_x}{\sum_{x=0}^{X} \bar{v}_x}, \tag{1.6.20}$$

$$s_x^2 = \frac{1}{t_1 - t_0} \sum_{t=t_0}^{t_1-1} \left(v_{x,t} - \bar{v}_x\right)^2, \tag{1.6.21}$$

where $x = 0, 1, \ldots, X$ and $\bar{v}_x$ is an average value of $v_{x,t_0}, \ldots, v_{x,t_1-1}$.

Parameter χ is determined by relations (1.6.5)–(1.6.6). We have

$$\chi = -\frac{\delta(t_0 + t_1)}{2}, \tag{1.6.22}$$

thus estimator c of χ takes the form

$$c = -\frac{d(t_0 + t_1)}{2}. \tag{1.6.23}$$

In order to find estimator a_x of α_x, the method of moments is used again. Since from (1.6.8) we get

$$\mathrm{E}\left[\ln \mu_x(t) - \alpha_x - \beta_x(\chi + t\delta)\right] = 0, \tag{1.6.24}$$

therefore a_x will be derived from the equation

$$\sum_{t=t_0}^{t_1} \left[\ln m_{x,t} - a_x - b_x(c + td)\right] = 0. \tag{1.6.25}$$

After simple transformations and using formula (1.6.23), we arrive at

$$a_x = \frac{1}{t_1 - t_0 + 1} \sum_{t=t_0}^{t_1} \ln m_{x,t}, \quad x = 0, 1, \ldots, X. \tag{1.6.26}$$

If the observation period is $[1, T]$ then $t_0 = 1$ and $t_1 = T$ in the above formulas.

1.7. The Vasiček and Cox–Ingersoll–Ross models

Vasiček [Vasiček 1977] and Cox, Ingersoll, Ross [Cox *et al.* 1985a, 1985b] considered dynamic models for the risk-free spot interest rate. However, it is recognized that the spot interest rate and the force of mortality have nearly identical representations, although there are also some important differences (see e.g. [Milevsky, Promislow 2001, Dahl 2004]). For instance, mortality rates are assumed to be positive and show non-mean reversion property.

In this section we present the Vasiček and Cox–Ingersoll–Ross models defined for the force of mortality $\mu_x(t)$.

1.7.1. V and CIR models

The Vasiček model (V model) takes the form of the Itô scalar stochastic differential equation

$$d\mu_x(t) = \kappa_x\left(\theta_x - \mu_x(t)\right)dt + \sigma_x dw(t), \quad t \in \mathbb{R}^+, \tag{1.7.1}$$

and the Cox–Ingersoll–Ross model (CIR model) can be written as

$$d\mu_x(t) = \kappa_x\left(\theta_x - \mu_x(t)\right)dt + \sigma_x\sqrt{\mu_x(t)}dw(t), \quad t \in \mathbb{R}^+, \tag{1.7.2}$$

where $\sigma_x, \theta_x, \kappa_x > 0$ are constant parameters and $w(t)$ is a standard Wiener process.

In the case of the V model one can find the analytical solution. Let us introduce the following function

$$K(t, \mu_x(t)) = e^{\kappa_x t}(\mu_x(t) - \theta_x). \tag{1.7.3}$$

Using the Itô formula (see Theorem A.7, formula (A.2.21) in Appendix A), we have

$$dK = \frac{\partial K}{\partial t}dt + \frac{\partial K}{\partial \mu_x}d\mu_x + \frac{1}{2}\frac{\partial^2 K}{\partial \mu_x^2}d\mu_x^2 + \dots \tag{1.7.4}$$

If we substitute (1.7.1) for $d\mu_x(t)$ in (1.7.4), then we receive

$$\begin{aligned} dK = {} & \frac{\partial K}{\partial t}dt + \frac{\partial K}{\partial \mu_x}\left[\kappa_x\left(\theta_x - \mu_x(t)\right)dt + \right. \\ & \left. + \sigma_x dw(t)\right] + \frac{1}{2}\frac{\partial^2 K}{\partial \mu_x^2}d\mu_x^2 + \dots \end{aligned} \tag{1.7.5}$$

From (1.7.3) we have

$$\frac{\partial K}{\partial t} = \kappa_x e^{\kappa_x t}(\mu_x(t) - \theta_x), \quad \frac{\partial K}{\partial \mu_x} = e^{\kappa_x t}, \quad \frac{\partial^2 K}{\partial \mu_x^2} = 0. \tag{1.7.6}$$

Then (1.7.4) takes the form

$$\begin{aligned} dK = e^{\kappa_x t}\kappa_x(\mu_x(t) - \theta_x)dt + e^{\kappa_x t}\kappa_x\left(\theta_x - \mu_x(t)\right)dt + \\ + e^{\kappa_x t}\sigma_x dw(t) = e^{\kappa_x t}\sigma_x dw(t). \end{aligned} \tag{1.7.7}$$

After integration of (1.7.7) between times t_0 and t, we arrive at

$$K(t, \mu_x(t)) = K(t_0, \mu_x(t_0)) + \sigma_x \int_{t_0}^{t} e^{\kappa_x s} dw(s), \tag{1.7.8}$$

what leads to the following solution of the stochastic differential equation (1.7.1)

$$\begin{aligned} \mu_x(t) = \mu_x(t_0)e^{-\kappa_x(t-t_0)} + \theta_x\left(1 - e^{-\kappa_x(t-t_0)}\right) + \\ + \sigma_x e^{-\kappa_x t}\int_{t_0}^{t} e^{\kappa_x s} dw(s). \end{aligned} \tag{1.7.9}$$

In the case of the Cox–Ingersoll–Ross model, application of the Itô formula leads to the equality

$$\begin{aligned} \mu_x(t) = & \mu_x(t_0)e^{-\kappa_x(t-t_0)} + \theta_x\left(1 - e^{-\kappa_x(t-t_0)}\right) + \\ & + \sigma_x e^{-\kappa_x t}\int_{t_0}^{t} \sqrt{\mu_x(s)}e^{\kappa_x s} dw(s). \end{aligned} \tag{1.7.10}$$

It indicates that the CIR model is described by a non-linear stochastic differential equation and the solution cannot be found in an explicit form as it is possible for the Vasiček model.

The drawback of the Vasiček model (1.7.1) is that it can yield the negative values of $\mu_x(t)$. In the CIR model (1.7.2) the diffusion function $\sigma_x^2\mu_x(t)$ is proportional to $\mu_x(t)$ what ensures that the process stays on a positive domain.

A general form of both models can be written as

$$d\mu_x(t) = \kappa_x\left(\theta_x - \mu_x(t)\right)dt + \sigma_x\mu_x^{\gamma_x}(t)dw(t), \quad t \in \mathbb{R}^+. \tag{1.7.11}$$

For the Vasiček model we have $\gamma_x = 0$ and for the CIR model $\gamma_x = \frac{1}{2}$.

1.7.2. Discrete V and CIR models

The Discrete Vasiček model (DV model) comes down to the following approximation

$$\mu_x(t+1) = \kappa_x\theta_x + (1-\kappa_x)\mu_x(t) + \sigma_x\epsilon_{x,t+1}, \quad t \in \mathbb{N}. \qquad (1.7.12)$$

From (1.7.12) it follows that the value of $\mu_x(t+1)$ is the weighted average of $\mu_x(t)$ in the period preceding $t+1$ and of the long-term average θ_x. From (1.7.12) we have also

$$\xi_{x,t+1} = \mu_x(t+1) - \mu_x(t) - \kappa_x\left(\theta_x - \mu_x(t)\right), \quad t \in \mathbb{N}, \qquad (1.7.13)$$

where $\xi_{x,t+1} = \sigma_x\epsilon_{x,t+1}$ are Gaussian random variables with means and variances equal $\mathrm{E}[\xi_{x,t+1}] = 0$ and $\mathrm{Var}[\xi_{x,t+1}] = \sigma_x^2$.

By analogy, the Discrete Cox–Ingersoll–Ross model (DCIR model) takes the form

$$\mu_x(t+1) = \kappa_x\theta_x + (1-\kappa_x)\mu_x(t) + \sigma_x\sqrt{\mu_x(t)}\epsilon_{x,t+1}, \quad t \in \mathbb{N}. \quad (1.7.14)$$

It follows from (1.7.14) that

$$\xi_{x,t+1} = \mu_x(t+1) - \mu_x(t) - \kappa_x\left(\theta_x - \mu_x(t)\right), \quad t \in \mathbb{N}, \qquad (1.7.15)$$

where $\xi_{x,t+1} = \sigma_x\sqrt{\mu_x(t)}\epsilon_{x,t+1}$ are random variables with means and conditional variances equal $\mathrm{E}[\xi_{x,t+1}]=0$ and $\mathrm{Var}[\xi_{x,t+1}|\mu_x(t)]=\sigma_x^2\mu_x(t)$, respectively.

With the discrete-time version of (1.7.11), the following expression is obtained

$$\mu_x(t+1) = \kappa_x\theta_x + (1-\kappa_x)\mu_x(t) + \sigma_x\mu_x^{\gamma_x}(t)\epsilon_{x,t+1}, \ t \in \mathbb{N}, \quad (1.7.16)$$

where $\xi_{x,t+1} = \sigma_x\mu_x^{\gamma_x}(t)\epsilon_{x,t+1}$ are random variables with with means and conditional variances equal $\mathrm{E}[\xi_{x,t+1}] = 0$ and $\mathrm{Var}[\xi_{x,t+1}|\mu_x(t)] = \sigma_x^2\mu_x^{2\gamma_x}(t)$.

1.7.3. Modified V and CIR models

If we substitute a non-linear differentiable function $f_x(t)$ for the scalar parameter σ_x in the Vasiček and Cox–Ingersoll–Ross models (1.7.1) and (1.7.2), respectively, then we obtain modified dynamic models, hereafter termed as the Modified Vasiček model (MV model) and the Modified Cox–Ingersoll–Ross model (MCIR model).

Thus, the MV model can be expressed by means of the following Itô stochastic differential equation

$$d\mu_x(t) = \kappa_x \left(\theta_x - \mu_x(t)\right) dt + f_x(t) dw(t), \quad t \in \mathbb{R}^+ \tag{1.7.17}$$

and the MCIR model can be written as

$$d\mu_x(t) = \kappa_x \left(\theta_x - \mu_x(t)\right) dt + f_x(t)\sqrt{\mu_x(t)} dw(t) \quad t \in \mathbb{R}^+, \tag{1.7.18}$$

where $\theta_x, \kappa_x > 0$ are constant parameters, $f_x(t) > 0$ is a time-depending diffusion function and $w(t)$ represents a standard Wiener process.

We will further assume that the diffusion function takes the form

$$f_x(t) = e^{\zeta_x t}, \quad \zeta_x \in \mathbb{R}. \tag{1.7.19}$$

Large positive values of the scalar ζ_x imply that the volatility of the diffusion term can explode. On the contrary, negative values of ζ_x indicate that volatility decreases exponentially.

Let us apply the Itô formula to the following function defined as in (1.7.3)

$$K(t, \mu_x(t)) = e^{\kappa_x t}(\mu_x(t) - \theta_x). \tag{1.7.20}$$

Similarly as solution (1.7.9) of the Vasiček model, the solution of (1.7.17) is as follows

$$\begin{aligned} \mu_x(t) =& \mu_x(t_0) e^{-\kappa_x (t-t_0)} + \theta_x \left(1 - e^{-\kappa_x (t-t_0)}\right) + \\ &+ e^{-\kappa_x t} \int_{t_0}^{t} e^{(\zeta_x + \kappa_x) s} dw(s) \end{aligned} \tag{1.7.21}$$

and for the MCIR model (1.7.18) we receive the following equation

$$\begin{aligned} \mu_x(t) =& \mu_x(t_0) e^{-\kappa_x (t-t_0)} + \theta_x \left(1 - e^{-\kappa_x (t-t_0)}\right) + \\ &+ e^{-\kappa_x t} \int_{t_0}^{t} \sqrt{\mu_x(s)} e^{(\zeta_x + \kappa_x) s} dw(s). \end{aligned} \tag{1.7.22}$$

1.7.4. Discrete modified V and CIR models

The Discrete Modified Vasiček model (DMV model) is of the form

$$\mu_x(t+1) = \kappa_x\theta_x + (1-\kappa_x)\mu_x(t) + \xi_{x,t+1}, \quad t \in \mathbb{N}, \tag{1.7.23}$$

where $\xi_{x,t+1}$ are Gaussian random variables with means $\mathrm{E}[\xi_{x,t+1}] = 0$ and variances $\mathrm{Var}[\xi_{x,t+1}] = e^{2\zeta_x t}$.

By analogy to (1.7.23), the Discrete Modified Cox–Ingersoll–Ross model (DMCIR model) can be expressed as

$$\mu_x(t+1) = \kappa_x\theta_x + (1-\kappa_x)\mu_x(t) + \xi_{x,t+1},\ t \in \mathbb{N}, \tag{1.7.24}$$

where $\xi_{x,t+1}$ are random variables with means $\mathrm{E}[\xi_{x,t+1}] = 0$ and conditional variances $\mathrm{Var}[\xi_{x,t+1}|\mu_x(t)] = e^{2\zeta_x t}\mu_x(t)$.

1.7.5. Parameters' estimation of the V and CIR models

Parameters $\theta_x, \kappa_x, \sigma_x$ of the DV model (1.7.12) or the DCIR model (1.7.14) can be estimated using the generalized method of moments (GMM) [Hansen 1982].

Let us note that both for the DV and DCIR model there is

$$\mathrm{E}[\xi_{x,t+1}] = 0. \tag{1.7.25}$$

Moreover, for the DV model we have

$$\mathrm{E}[\xi^2_{x,t+1}] = \sigma^2_x. \tag{1.7.26}$$

In the case of the DCIR model there is

$$\mathrm{E}[\xi^2_{x,t+1}|\mu_x(t)] = \sigma^2_x\mu_x(t). \tag{1.7.27}$$

Thus, employing properties (1.7.25)–(1.7.27) of respective random terms (1.7.13) and (1.7.15) and assuming the orthogonality condition

$$\mathrm{E}[\xi_{x,t+1}\mu_x(t)] = 0, \tag{1.7.28}$$

we can consider a set of three moments.

For the DV model the set of moments is as follows

$$\begin{cases} \mathrm{E}\left[\mu_x(t+1) - \mu_x(t) - \kappa_x\left(\theta_x - \mu_x(t)\right)\right], \\ \\ \mathrm{E}\left[\left(\mu_x(t+1) - \mu_x(t) - \kappa_x\left(\theta_x - \mu_x(t)\right)\right)^2 - \sigma^2_x\right], \\ \\ \mathrm{E}\left[\left(\mu_x(t+1) - \mu_x(t) - \kappa_x\left(\theta_x - \mu_x(t)\right)\right)\mu_x(t)\right]. \end{cases} \tag{1.7.29}$$

For the DCIR model the analogous set of moments takes the form

$$\begin{cases} \mathrm{E}\left[\mu_x(t+1)-\mu_x(t)-\kappa_x\left(\theta_x-\mu_x(t)\right)\right], \\ \\ \mathrm{E}\left[\left(\mu_x(t+1)-\mu_x(t)-\kappa_x(\theta_x-\mu_x(t))\right)^2-\sigma_x^2\mu_x(t)\right], \\ \\ \mathrm{E}\left[\left(\mu_x(t+1)-\mu_x(t)-\kappa_x\left(\theta_x-\mu_x(t)\right)\right)\mu_x(t)\right]. \end{cases} \quad (1.7.30)$$

Note that the moments are defined so that they all are equal 0.

The sample moments corresponding with (1.7.29) can be written as

$$\mathbf{g}(\kappa_x,\theta_x,\sigma_x^2) = \\ = \begin{bmatrix} \frac{1}{T-1}\sum_{t=1}^{T-1}\left[m_{x,t+1}-m_{x,t}-\kappa_x\left(\theta_x-m_{x,t}\right)\right] \\ \\ \frac{1}{T-1}\sum_{t=1}^{T-1}\left[\left(m_{x,t+1}-m_{x,t}-\kappa_x(\theta_x-m_{x,t})\right)^2-\sigma_x^2\right] \\ \\ \frac{1}{T-1}\sum_{t=1}^{T-1}\left[\left(m_{x,t+1}-m_{x,t}-\kappa_x\left(\theta_x-m_{x,t}\right)\right)m_{x,t}\right] \end{bmatrix} \quad (1.7.31)$$

and the sample moments corresponding with (1.7.30) are

$$\mathbf{g}(\kappa_x,\theta_x,\sigma_x^2) = \\ = \begin{bmatrix} \frac{1}{T-1}\sum_{t=1}^{T-1}\left[m_{x,t+1}-m_{x,t}-\kappa_x\left(\theta_x-m_{x,t}\right)\right] \\ \\ \frac{1}{T-1}\sum_{t=1}^{T-1}\left[\left(m_{x,t+1}-m_{x,t}-\kappa_x(\theta_x-m_{x,t})\right)^2-\sigma_x^2 m_{x,t}\right] \\ \\ \frac{1}{T-1}\sum_{t=1}^{T-1}\left[\left(m_{x,t+1}-m_{x,t}-\kappa_x\left(\theta_x-m_{x,t}\right)\right)m_{x,t}\right] \end{bmatrix}, \quad (1.7.32)$$

where $m_{x,t}$ are age-specific central death rates from a sample time series $\{m_{x,t}, t=1,2,\ldots,T\}$.

The GMM estimators of $\kappa_x,\theta_x,\sigma_x^2$ can be found by minimizing the sum of squared sample moments, i.e. by solving the following optimization problem

$$\text{minimize } S(\kappa_x,\theta_x,\sigma_x^2) = \mathbf{g}^T(\kappa_x,\theta_x,\sigma_x^2)\mathbf{g}(\kappa_x,\theta_x,\sigma_x^2) \quad (1.7.33)$$

with respect to $\kappa_x,\theta_x,\sigma_x^2$.

In the case of the DMV model (1.7.23) or DMCIR model (1.7.24) the unknown parameters are $\kappa_x,\theta_x,\zeta_x$. They can be estimated in a similar way, i.e. by means of the generalized method of moments.

1.8. The Milevsky–Promislow model

The range of mortality models based on the Itô stochastic differential equation includes a model proposed by [Milevsky, Promislow 2001].

1.8.1. MP model

In the Milevsky–Promislow approach, the force of mortality $\mu_x(t)$ is defined as a stochastic process

$$\mu_x(t) = \mu_{x0} \exp\{\gamma_x t + q_x z(t)\}, \quad t \in \mathbb{R}^+, \quad \gamma_x, q_x, \mu_{x0} > 0, \tag{1.8.1}$$

where $z(t)$ is expressed via the Itô stochastic differential equation

$$dz(t) = -\beta_x z(t)dt + dw(t), \quad z(0) = 0, \quad \beta_x \geq 0, \tag{1.8.2}$$

with a scalar Wiener process $w(t)$, $t \in \mathbb{R}^+$.

The model (1.8.1)–(1.8.2) will be called the Milevsky–Promislow model (MP model).

It is worth noting that if $\beta_x = 0$, then $z(t)$ in (1.8.2) collapses to $w(t)$ and the process becomes a geometric Brownian motion. If $\beta_x > 0$ then it is the Ornstein–Uhlenbeck process.

Similarly to the derivation presented in [Giacometti *et al.* 2011], we introduce the twice-differentiable function

$$K(t, \ln \mu_x(t)) = \ln \mu_x(t). \tag{1.8.3}$$

Thus, taking under account equation (1.8.1) we have also

$$K(t, \ln \mu_x(t)) = \ln \mu_x(t) = \ln \mu_{x0} + \gamma_x t + q_x z(t). \tag{1.8.4}$$

By applying the Itô formula (see (A.2.21) in Appendix A), we arrive at

$$\begin{aligned} dK &= \frac{\partial K}{\partial t}dt + \frac{\partial K}{\partial z}dz + \frac{1}{2}\frac{\partial^2 K}{\partial z^2}dz^2 = \\ &= \frac{\partial K}{\partial t}dt + \frac{\partial K}{\partial z}\left[-\beta_x z(t)dt + dw(t)\right] + \frac{1}{2}\frac{\partial^2 K}{\partial z^2}dt = \\ &= \left[\frac{\partial K}{\partial t} - \beta_x z(t)\frac{\partial K}{\partial z} + \frac{1}{2}\frac{\partial^2 K}{\partial z^2}\right]dt + \frac{\partial K}{\partial z}dw(t). \end{aligned} \tag{1.8.5}$$

From (1.8.4) we have

$$\frac{\partial K}{\partial t} = \gamma_x, \quad \frac{\partial K}{\partial z} = q_x, \quad \frac{\partial^2 K}{\partial z^2} = 0. \tag{1.8.6}$$

It follows that (1.8.5) can be transformed to

$$\begin{aligned} dK &= \left[\gamma_x - q_x \frac{\beta_x}{q_x} \left(K(t, \ln \mu_x(t)) - \ln \mu_{x0} - \gamma_x t\right)\right] dt + q_x dw(t) = \\ &= \left[\gamma_x - \beta_x K(t, \ln \mu_x(t)) + \beta_x \ln \mu_{x0} + \gamma_x \beta_x t\right] dt + q_x dw(t). \end{aligned} \tag{1.8.7}$$

Since $K(t, \ln \mu_x(t)) = \ln \mu_x(t)$, thus we have also an equivalent stochastic differential equation

$$d \ln \mu_x(t) = \left[\gamma_x - \beta_x \ln \mu_x(t) + \beta_x \ln \mu_{x0} + \gamma_x \beta_x t\right] dt + q_x dw(t). \tag{1.8.8}$$

Equation (1.8.8) can be solved explicitly by applying the Itô formula to function $e^{\beta_x} K(t, \ln \mu_x(t))$ with $K(t, \ln \mu_x(t))$ defined in (1.8.4), obtaining the following solution

$$\begin{aligned} \ln \mu_x(t) = & e^{-\beta_x t} \ln \mu_{x0} + \int_0^t e^{-\beta_x (t-s)} [\gamma_x + \beta_x \ln \mu_{x0} + \beta_x \gamma_x s] ds + \\ & + q_x \int_0^t e^{-\beta_x (t-s)} dw(s). \end{aligned} \tag{1.8.9}$$

1.8.2. Discrete MP model

The respective Discrete Milevsky–Promislow model (DMP model) is derived from equality (1.8.9) by subtracting from $\ln \mu_x(t)$ the following product

$$e^{-\beta_x} \ln \mu_x(t-1). \tag{1.8.10}$$

Hence, we have

$$\ln \mu_x(t) - e^{-\beta_x} \ln \mu_x(t-1) = \psi_{x,t} + q_x \epsilon_{x,t}, \quad t \in \mathbb{N}, \tag{1.8.11}$$

or equivalently

$$\ln \mu_x(t) = e^{-\beta_x} \ln \mu_x(t-1) + \psi_{x,t} + q_x \epsilon_{x,t}, \quad t \in \mathbb{N}, \tag{1.8.12}$$

where

$$\psi_{x,t} = \int_0^1 e^{-\beta_x u}[\gamma_x + \beta_x \ln \mu_{x0} + \beta_x \gamma_x (t-u)]du,$$
$$\epsilon_{x,t} = -\int_0^1 e^{-\beta_x u} dw(t-u). \qquad (1.8.13)$$

After integration, $\psi_{x,t}$ reduces to

$$\psi_{x,t} = (1 - e^{-\beta_x})\ln \mu_{x0} + \gamma_x e^{-\beta_x} + \gamma_x t(1 - e^{-\beta_x}). \qquad (1.8.14)$$

Relationship (1.8.12) can be written also as a difference equation

$$y_x(t) = b_{x,0}(t) + b_{x,1} y_x(t-1) + \xi_{x,t}, \quad t \in \mathbb{N}, \qquad (1.8.15)$$

where

$$\begin{aligned} y_x(t) &= K(t, \ln \mu_x(t)) = \ln \mu_x(t), \\ b_{x,0}(t) &= \psi_{x,t} = (1 - e^{-\beta_x}) \ln \mu_{x0} + \gamma_x e^{-\beta_x} + \gamma_x t(1 - e^{-\beta_x}), \\ b_{x,1} &= e^{-\beta_x}, \\ \xi_{x,t} &= q_x \epsilon_{x,t}. \end{aligned} \qquad (1.8.16)$$

Formula (1.8.15) can be also transformed to the following one

$$y_x(t) = a_{x,0} + a_{x,1} t + a_{x,2} y_x(t-1) + \xi_{x,t}, \quad t \in \mathbb{N}, \qquad (1.8.17)$$

where

$$\begin{aligned} y_x(t) &= \ln \mu_x(t), \\ a_{x,0} &= (1 - e^{-\beta_x}) \ln \mu_{x0} + \gamma_x e^{-\beta_x}, \\ a_{x,1} &= \gamma_x (1 - e^{-\beta_x}), \\ a_{x,2} &= e^{-\beta_x}, \\ \xi_{x,t} &= q_x \epsilon_{x,t}, \end{aligned} \qquad (1.8.18)$$

After substituting $a_{x,1} a_{x,2}/(1 - a_{x,2})$ for $\gamma_x e^{-\beta_x}$, a 2-factor model can be derived

$$y_x(t) = (1 - a_{x,2}) \ln \mu_{x0} + \frac{a_{x,1} a_{x,2}}{1 - a_{x,2}} + a_{x,1} t +$$
$$+ a_{x,2} y_x(t-1) + \xi_{x,t}, \quad t \in \mathbb{N}, \qquad (1.8.19)$$

where

$$
\begin{aligned}
y_x(t) &= \ln \mu_x(t), \\
a_{x,1} &= \gamma_x(1 - e^{-\beta_x}), \\
a_{x,2} &= e^{-\beta_x}, \\
\xi_{x,t} &= q_x \epsilon_{x,t}.
\end{aligned}
\tag{1.8.20}
$$

It follows from properties of the Itô integral and from the Itô isometry [Oksendal 2003, p. 25 and 29] that

$$
\mathrm{E}\left[\int_0^t g(u)dw(u)\right] = 0, \quad \mathrm{E}\left[\int_0^t g(u)dw(u)\right]^2 = \int_0^t \mathrm{E}[g(u)]^2 du. \tag{1.8.21}
$$

Thus, random terms $\epsilon_{x,t}$ in (1.8.13) as well as $\xi_{x,t} = q_x \epsilon_{x,t}$ in (1.8.15), (1.8.17) and (1.8.19) are Gaussian random variables with means and variances equal, respectively,

$$
\begin{aligned}
&\mathrm{E}[\epsilon_{x,t}] = \mathrm{E}[\xi_{x,t}] = 0, \\
&\mathrm{Var}[\epsilon_{x,t}] = \mathrm{E}[\epsilon_{x,t}^2] = \frac{1 - e^{-2\beta_x}}{2\beta_x} \approx 1, \\
&\mathrm{Var}[\xi_{x,t}] = q_x^2 \mathrm{E}[\epsilon_{x,t}^2] = q_x^2 \frac{1 - e^{-2\beta_x}}{2\beta_x} \approx q_x^2.
\end{aligned}
\tag{1.8.22}
$$

1.8.3. Parameters' estimation of the MP model

Parameters of the DMP model (1.8.19) can be estimated by means of the least squares method, i.e. by minimizing the following sum of squared errors with respect to $a_{x,1}, a_{x,2}$

$$
\sum_{t=2}^{T} \left[y_{x,t} - \left((1 - a_{x,2}) \ln \mu_{x0} + \frac{a_{x,1} a_{x,2}}{1 - a_{x,2}} + a_{x,1} t + a_{x,2} y_{x,t-1} \right) \right]^2, \tag{1.8.23}
$$

where $y_{x,t} = \ln m_{x,t}$ represent log-central death rates from a sample time series $\{\ln m_{x,t},\ t = 1, 2, \ldots, T\}$.

Subsequently, estimates of β_x, γ_x are obtained from (1.8.20), while parameters q_x^2 representing the variance of residuals $\xi_{x,t}$ are estimated by determining errors $\hat{\xi}_{x,t}$ from the optimization problem (1.8.23) and by calculating the second sample moment of the residuals obtained.

1.9. The Giacometti–Ortobelli–Bertocchi model

Giacometti, Ortobelli and Bertocchi [Giacometti *et al.* 2011] considered an analogous model as given in [Milevsky, Promislow 2001], by extending the stochastic differential equation of the filter.

1.9.1. GOB model

Let us consider the equation (1.8.1)

$$\mu_x(t) = \mu_{x0} \exp\{\gamma_x t + q_x z(t)\}, \quad t \in \mathbb{R}^+, \quad \gamma_x, q_x, \mu_{x0} > 0, \tag{1.9.1}$$

where $z(t)$ is defined by means of the following Itô stochastic differential equation

$$dz(t) = -\beta_x z(t)dt + f_x(t)dw(t), \quad z(0) = 0, \quad \beta_x > 0, \tag{1.9.2}$$

with $f_x(t)$ being a non-linear differentiable function of time and $w(t)$ representing a standard Wiener process.

The model (1.9.1)–(1.9.2) is termed the Giacometti–Ortobelli–Bertocchi model (GOB model).

Similarly to the derivation presented in [Giacometti *et al.* 2011], let us consider a function

$$K(t, \ln \mu_x(t)) = \frac{\ln \mu_x(t)}{f_x(t)}. \tag{1.9.3}$$

Taking into account equation (1.9.1), we have

$$K(t, \ln \mu_x(t)) = \frac{\ln \mu_x(t)}{f_x(t)} = \frac{\ln \mu_{x0} + \gamma_x t + q_x z(t)}{f_x(t)}. \tag{1.9.4}$$

The following equation follows from the Itô formula (see (A.2.21), Theorem A.7 in Appendix A)

$$\begin{aligned} dK(t, \ln \mu_x(t)) &= \frac{\partial K}{\partial t}dt + \frac{\partial K}{\partial z}dz + \frac{1}{2}\frac{\partial^2 K}{\partial z^2}dz^2 = \\ &= \frac{\partial K}{\partial t}dt + \frac{\partial K}{\partial z}\left[-\beta_x z(t)dt + f_x(t)dw(t)\right] + \frac{1}{2}\frac{\partial^2 K}{\partial z^2}f_x^2(t)dt. \end{aligned} \tag{1.9.5}$$

From (1.9.4) we have

$$\frac{\partial K}{\partial t} = \frac{\gamma_x}{f_x(t)} - \frac{f_x'(t)}{f_x^2(t)}\left(\ln \mu_{x0} + \gamma_x t + q_x z\right),$$
$$\frac{\partial K}{\partial z} = \frac{q_x}{f_x(t)}, \quad \frac{\partial^2 K}{\partial z^2} = 0, \tag{1.9.6}$$

where $f_x'(t)$ denotes the first derivative of $f_x(t)$.

Equation (1.9.5) can then be expressed as

$$dK(t, \ln \mu_x(t)) = \left[\frac{\gamma_x}{f_x(t)} - \frac{f_x'(t)}{f_x(t)} K(t, \ln \mu_x(t))\right] dt +$$
$$+ \frac{q_x}{f_x(t)}\left[-\beta_x z(t)dt + f_x(t)dw(t)\right]. \tag{1.9.7}$$

After simple transformations of (1.9.7), we receive

$$dK(t, \ln \mu_x(t)) = \left[\frac{\gamma_x + \beta_x \ln \mu_{x0} + \gamma_x \beta_x t}{f_x(t)} + \right.$$
$$\left. -K(t, \ln \mu_x(t))\left(\frac{f_x'(t)}{f_x(t)} + \beta_x\right)\right] dt + q_x dw(t), \quad t \in \mathbb{R}^+. \tag{1.9.8}$$

Let us assume that the diffusion function $f_x(t)$ takes the form

$$f_x(t) = \exp\{\zeta_x t\}, \quad \zeta_x \in \mathbb{R}. \tag{1.9.9}$$

For (1.9.9) equation (1.9.8) can be written as

$$dK(t, \ln \mu_x(t)) = (\gamma_x + \beta_x \ln \mu_{x0} + \gamma_x \beta_x t) e^{-\zeta_x t} dt +$$
$$- (\zeta_x + \beta_x) K(t, \mu_x(t)) dt + q_x dw(t), \ t \in \mathbb{R}^+. \tag{1.9.10}$$

Equation (1.9.10) can be solved by applying again the Itô formula to the expression $e^{(\zeta_x + \beta_x)t} K(t, \ln \mu_x(t))$.

It leads to the following solution

$$K(t,\ln\mu_x(t)) = e^{-(\zeta_x+\beta_x)t}K(0,\ln\mu_{x0}) + \\ + \int_0^t e^{-(\zeta_x+\beta_x)(t-s)}[\gamma_x+\beta_x\ln\mu_{x0}+\beta_x\gamma_x s]e^{-\zeta_x s}ds + \\ + q_x\int_0^t e^{-(\zeta_x+\beta_x)(t-s)}dw(s). \tag{1.9.11}$$

For $K(t,\ln\mu_x(t))$ defined in (1.9.4) an equation can be also derived for $\ln\mu_x(t)$ from equality (1.9.11)

$$\ln\mu_x(t) = e^{-\beta_x t}\ln\mu_{x0} + \int_0^t e^{-\beta_x(t-s)}[\gamma_x+\beta_x\ln\mu_{x0}+\beta_x\gamma_x s]ds + \\ + e^{-\beta_x t}q_x\int_0^t e^{(\zeta_x+\beta_x)s}dw(s). \tag{1.9.12}$$

It is worth noting that analogous solution as in (1.9.12) can be obtained by using the Itô formula with respect to $e^{\beta_x t}\ln\mu_x(t)$.

1.9.2. Discrete GOB model

The discrete-time form of (1.9.8) is as follows

$$K(t,\ln\mu_x(t)) = \\ = K(t-1,\ln\mu_x(t-1)) + \left[\frac{\gamma_x+\beta_x\ln\mu_{x0}+\gamma_x\beta_x(t-1)}{f_x(t-1)} + \right. \\ \left. - K(t-1,\ln\mu_x(t-1))\left(\frac{f_x(t)-f_x(t-1)}{f_x(t-1)}+\beta_x\right)\right] + q_x\epsilon_{x,t}, \quad t\in\mathbb{N} \tag{1.9.13}$$

and the Discrete Giacometti–Ortobelli–Bertocchi model (DGOB model) is obtained from (1.9.11) by subtracting from $K(t,\ln\mu_x(t))$ the following product

$$e^{-(\zeta_x+\beta_x)}K(t-1,\ln\mu_x(t-1)). \tag{1.9.14}$$

Thus, we have for $t \in \mathbb{N}$

$$K(t, \ln \mu_x(t)) - e^{-(\zeta_x+\beta_x)} K(t-1, \ln \mu_x(t-1)) + \psi_{x,t} + q_x \epsilon_{x,t}, \quad (1.9.15)$$

or equivalently

$$K(t, \ln \mu_x(t)) = e^{-(\zeta_x+\beta_x)} K(t-1, \ln \mu_x(t-1)) = \psi_{x,t} + q_x \epsilon_{x,t}, \quad (1.9.16)$$

where

$$\begin{aligned} \psi_{x,t} &= \int_0^1 e^{-(\zeta_x+\beta_x)u}[\gamma_x + \beta_x \ln \mu_{x0} + \beta_x \gamma_x (t-u)] e^{-\zeta_x (t-u)} du, \\ \epsilon_{x,t} &= -\int_0^1 e^{-(\zeta_x+\beta_x)u} dw(t-u). \end{aligned} \quad (1.9.17)$$

After integration, $\psi_{x,t}$ reduces to

$$\psi_{x,t} = e^{-\zeta_x t}\left[(1 - e^{-\beta_x}) \ln \mu_{x0} + \gamma_x e^{-\beta_x} + \gamma_x t(1 - e^{-\beta_x})\right]. \quad (1.9.18)$$

Let us notice that random terms $\epsilon_{x,t}$ in (1.9.15) or (1.9.16) are Gaussian random variables with means and variances equal, respectively,

$$\begin{aligned} \mathrm{E}[\epsilon_{x,t}] &= 0, \\ \mathrm{Var}[\epsilon_{x,t}] = \mathrm{E}[\epsilon_{x,t}^2] &= \frac{1 - e^{-2(\zeta_x+\beta_x)}}{2(\zeta_x + \beta_x)} \approx 1. \end{aligned} \quad (1.9.19)$$

Moreover, the p-th raw absolute moments of random terms $\epsilon_{x,t}$ satisfy the following condition [Winkelbauer 2014]

$$\mathrm{E}\left[|\epsilon_{x,t}|^p\right] = 2^{\frac{p}{2}} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}} \left[\mathrm{Var}[\epsilon_{x,t}]\right]^{\frac{p}{2}}, \quad p > -1, \quad (1.9.20)$$

where $\Gamma(z) = \int_0^\infty v^{z-1} e^{-v} dv$ for $z \in \mathbb{R}^+$. Thus, from (1.9.20) it follows that the first raw absolute moments are

$$\mathrm{E}\left[|\epsilon_{x,t}|\right] = \sqrt{\frac{2}{\pi}} \cdot \sqrt{\frac{1 - e^{-2(\zeta_x+\beta_x)}}{2(\zeta_x + \beta_x)}} \approx \sqrt{\frac{2}{\pi}}, \quad (1.9.21)$$

since $|\epsilon_{x,t}|$ follows a half-normal distribution.

Relationship (1.9.16) can be written as a difference equation

$$K_x(t) = b_{x,0}(t) + b_{x,1}K_x(t-1) + \xi_{x,t}, \quad t \in \mathbb{N}, \tag{1.9.22}$$

where

$$\begin{aligned} K_x(t) &= K(t, \ln\mu_x(t)) = e^{-\zeta_x t}\ln\mu_x(t), \\ b_{x,0}(t) &= \psi_{x,t} = e^{-\zeta_x t}\big[(1-e^{-\beta_x})\ln\mu_{x0} + \gamma_x e^{-\beta_x} + \gamma_x t(1-e^{-\beta_x})\big], \\ b_{x,1} &= e^{-(\zeta_x+\beta_x)}, \\ \xi_{x,t} &= q_x\epsilon_{x,t}. \end{aligned} \tag{1.9.23}$$

By variable substitution, (1.9.22) can be transformed to

$$y_x(t) = a_{x,0} + a_{x,1}t + a_{x,2}y_x(t-1) + \varepsilon_{x,t}, \;\; t \in \mathbb{N}, \tag{1.9.24}$$

where

$$\begin{aligned} y_x(t) &= \ln\mu_x(t), \\ a_{x,0} &= (1-e^{-\beta_x})\ln\mu_{x0} + \gamma_x e^{-\beta_x}, \\ a_{x,1} &= \gamma_x(1-e^{-\beta_x}), \\ a_{x,2} &= e^{-\beta_x}, \\ \varepsilon_{x,t} &= e^{\zeta_x t}\xi_{x,t}. \end{aligned} \tag{1.9.25}$$

The 2-factor DGOB model derived from (1.9.24) is as follows

$$\begin{aligned} y_x(t) = (1-a_{x,2})\ln\mu_{x0} + \frac{a_{x,1}a_{x,2}}{1-a_{x,2}} + a_{x,1}t + \\ + a_{x,2}y_x(t-1) + \varepsilon_{x,t}, \quad t \in \mathbb{N}, \end{aligned} \tag{1.9.26}$$

where

$$\begin{aligned} y_x(t) &= \ln\mu_x(t), \\ a_{x,1} &= \gamma_x(1-e^{-\beta_x}), \\ a_{x,2} &= e^{-\beta_x}, \\ \varepsilon_{x,t} &= e^{\zeta_x t}\xi_{x,t}. \end{aligned} \tag{1.9.27}$$

Random terms $\xi_{x,t}$ appearing in (1.9.22) and $\varepsilon_{x,t}$ in (1.9.24), (1.9.26) are Gaussian random variables with means and variances

$$\begin{aligned} &\mathrm{E}[\xi_{x,t}] = \mathrm{E}[\varepsilon_{x,t}] = 0, \\ &\mathrm{Var}[\xi_{x,t}] = \mathrm{E}[\xi_{x,t}^2] = q_x^2, \quad \mathrm{Var}[\varepsilon_{x,t}] = \mathrm{E}[\varepsilon_{x,t}^2] = q_x^2 e^{2\zeta_x t}. \end{aligned} \tag{1.9.28}$$

The first absolute raw moments approximated according to (1.9.20) are as follows

$$\mathrm{E}\left[|\xi_{x,t}|\right] = q_x\sqrt{\frac{2}{\pi}}, \quad \mathrm{E}\left[|\varepsilon_{x,t}|\right] = q_x e^{\zeta_x t}\sqrt{\frac{2}{\pi}} \tag{1.9.29}$$

and the following equalities hold

$$\mathrm{E}\left[\frac{\varepsilon_{x,t}^2}{e^{2\zeta_x t}}\right] = q_x^2, \qquad \mathrm{E}\left[\left|\frac{\varepsilon_{x,t}}{e^{\zeta_x t}}\right|\right] = q_x\sqrt{\frac{2}{\pi}}. \tag{1.9.30}$$

1.9.3. Parameters' estimation of the GOB model

Parameters $a_{x,1}$, $a_{x,2}$ of the model (1.9.26) are estimated using the least squares method, i.e. by minimizing the following sum of squared errors with respect to $a_{x,1}$, $a_{x,2}$

$$\sum_{t=2}^{T}\left[y_{x,t} - \left((1-a_{x,2})\ln\mu_{x0} + \frac{a_{x,1}a_{x,2}}{1-a_{x,2}} + a_{x,1}t + a_{x,2}y_{x,t-1}\right)\right]^2, \tag{1.9.31}$$

where $y_{x,t} = \ln m_{x,t}$ are log-central death rates from a sample time series $\{\ln m_{x,t},\ t = 1, 2, \ldots, T\}$.

Subsequently, estimators of β_x, γ_x are obtained from (1.9.27) and parameters q_x, ζ_x are estimated by calculating errors $\hat{\varepsilon}_{x,t}$ from the optimization problem (1.9.31) and by finding estimates $\hat{q}_x$ and $\hat{\zeta}_x$ from sample moment equations determined by analogy to (1.9.30).

1.10. The modified Milevsky–Promislow model

In this section a mortality model analogous to (1.8.1)–(1.8.2) is proposed. It differs from the MP model with the definition of the filter equation.

1.10.1. Modified MP model

The Modified Milevsky–Promislow model (MMP model) is defined by means of the following equations

$$\mu_x(t) = \mu_{x0}\exp\{\gamma_x t + z(t)\}, \quad t \in \mathbb{R}^+, \quad \gamma_x, \mu_{x0} > 0, \tag{1.10.1}$$

where $z(t)$ is the Ornstein–Uhlenbeck stochastic process satisfying the following stochastic differential equation

$$dz(t)=\beta_x(\alpha_x-z(t))dt+\sigma_x dw(t), \quad z(0)=0, \ \alpha_x\in\mathbb{R}, \ \beta_x,\sigma_x>0, \quad (1.10.2)$$

with a scalar Wiener process $w(t)$.

To find an explicit solution of (1.10.1)–(1.10.2), let us consider function

$$K(t,\ln\mu_x(t)) = e^{\beta_x t}\ln\mu_x(t). \qquad (1.10.3)$$

Taking under account relation (1.10.1), function $K(t,\ln\mu_x(t))$ can be written as

$$K(t,\ln\mu_x(t)) = e^{\beta_x t}\ln\mu_x(t) = e^{\beta_x t}\left(\ln\mu_{x0}+\gamma_x t+z(t)\right). \qquad (1.10.4)$$

Application of the Itô formula results in the following equation

$$dK(t,\ln\mu_x(t)) = \left[\frac{\partial K}{\partial t}+\beta_x(\alpha_x-z(t))\frac{\partial K}{\partial z}+\frac{1}{2}\frac{\partial^2 K}{\partial z^2}\sigma_x^2\right]dt+ \qquad (1.10.5)$$

$$+\frac{\partial K}{\partial z}\sigma_x dw(t),$$

where

$$\frac{\partial K}{\partial t}=\beta e^{\beta_x t}\left(\ln\mu_{x0}+\gamma_x t+z\right)+\gamma_x e^{\beta_x t}, \ \frac{\partial K}{\partial z}=e^{\beta_x t}, \ \frac{\partial^2 K}{\partial z^2}=0. \qquad (1.10.6)$$

Thus, we have

$$dK(t,\ln\mu_x(t)) =$$

$$=\left[\beta_x e^{\beta_x t}\left(\ln\mu_{x0}+\gamma_x t+z(t)\right)+\gamma_x e^{\beta_x t}+\beta_x(\alpha_x-z(t))e^{\beta_x t}\right]dt+ \qquad (1.10.7)$$

$$+\sigma_x e^{\beta_x t}dw(t).$$

Finally, the following stochastic differential equation is obtained

$$dK(t,\ln\mu_x(t)) = e^{\beta_x t}\left[\beta_x\ln\mu_{x0}+\beta_x\gamma_x t+\gamma_x+\beta_x\alpha_x\right]dt+ \qquad (1.10.8)$$

$$+\sigma_x e^{\beta_x t}dw(t).$$

Equation (1.10.8) leads to the following solution

$$K(t,\ln\mu_x(t)) = K(0,\ln\mu_{x0}) + \int_0^t e^{\beta_x s}[\beta_x \ln\mu_{x0} + \beta_x\gamma_x s + \gamma_x + \beta_x\alpha_x]\,ds + \sigma_x \int_0^t e^{\beta_x s} dw(s). \tag{1.10.9}$$

Since $K(t,\ln\mu_x(t)) = e^{\beta_x t}\ln\mu_x(t)$, therefore (1.10.9) can be also rewritten as

$$\ln\mu_x(t) = e^{-\beta_x t}\ln\mu_{x0} + \int_0^t e^{-\beta_x(t-s)}\,[\beta_x\ln\mu_{x0} + \beta_x\gamma_x s + \gamma_x + \beta_x\alpha_x]\,ds + \sigma_x\int_0^t e^{-\beta_x(t-s)} dw(s). \tag{1.10.10}$$

1.10.2. Discrete modified MP model

The Discrete Modified Milevsky–Promislow model (DMMP model) is derived from (1.10.10) by multiplying $\ln\mu_x(t-1)$ by $e^{-\beta_x}$ and subtracting from $\ln\mu_x(t)$. Hence, we have

$$\ln\mu_x(t) - e^{-\beta_x}\ln\mu_x(t-1) = \psi_{x,t} + \sigma_x\epsilon_{x,t}, \quad t\in\mathbb{N}, \tag{1.10.11}$$

or equivalently

$$\ln\mu_x(t) = e^{-\beta_x}\ln\mu_x(t-1) + \psi_{x,t} + \sigma_x\epsilon_{x,t}, \quad t\in\mathbb{N}, \tag{1.10.12}$$

where

$$\psi_{x,t} = \int_0^1 e^{-\beta_x u}\,[\beta_x\ln\mu_{x0} + \beta_x\gamma_x(t-u) + \gamma_x + \beta_x\alpha_x]\,du, \tag{1.10.13}$$

$$\epsilon_{x,t} = -\int_0^1 e^{-\beta_x u} dw(t-u). \tag{1.10.14}$$

Integrating right-hand side of (1.10.13), we receive

$$\psi_{x,t} = (\ln\mu_{x0} + \alpha_x)\left(1 - e^{-\beta_x}\right) + \gamma_x e^{-\beta_x} + \gamma_x\left(1 - e^{-\beta_x}\right) t. \tag{1.10.15}$$

Relationship (1.10.12) can be written as a difference equation

$$\ln \mu_x(t) = b_{x,0}(t) + b_{x,1} \ln \mu_x(t-1) + \xi_{x,t}, \tag{1.10.16}$$

where

$$b_{x,0}(t) = \psi_{x,t} = (\ln \mu_{x0} + \alpha_x)\big(1 - e^{-\beta_x}\big) + \gamma_x e^{-\beta_x} + \gamma_x \big(1 - e^{-\beta_x}\big) t,$$

$$b_{x,1} = e^{-\beta_x}, \tag{1.10.17}$$

$$\xi_{x,t} = \sigma_x \epsilon_{x,t}.$$

The model (1.10.16) can be also expressed as

$$y_x(t) = a_{x,0} + a_{x,1} t + a_{x,2} y_x(t-1) + \xi_{x,t}, \tag{1.10.18}$$

where

$$y_x(t) = \ln \mu_x(t),$$

$$a_{x,0} = (\ln \mu_{x0} + \alpha_x)\left(1 - e^{-\beta_x}\right) + \gamma_x e^{-\beta_x},$$

$$a_{x,1} = \gamma_x (1 - e^{-\beta_x}), \tag{1.10.19}$$

$$a_{x,2} = e^{-\beta_x},$$

$$\xi_{x,t} = \sigma_x \epsilon_{x,t}.$$

Let us remark that the random terms $\epsilon_{x,t}$ in (1.10.11) and $\xi_{x,t}$ in (1.10.16) or (1.10.18) are Gaussian random variables with means and variances equal, respectively,

$$\mathrm{E}[\epsilon_{x,t}] = \mathrm{E}[\xi_{x,t}] = 0,$$

$$\mathrm{Var}[\epsilon_{x,t}] = \frac{1 - e^{-2\beta_x}}{2\beta_x} \approx 1, \tag{1.10.20}$$

$$\mathrm{Var}[\xi_{x,t}] = \sigma_x^2 \ \mathrm{Var}[\epsilon_{x,t}] = \sigma_x^2 \frac{1 - e^{-2\beta_x}}{2\beta_x} \approx \sigma_x^2.$$

1.10.3. Parameters' estimation of the modified MP model

Parameters of the DMMP model (1.10.18) are estimated using the least squares method, i.e. by minimizing the following sum of squared errors with respect to $a_{x,0}$, $a_{x,1}$, $a_{x,2}$

$$\sum_{t=2}^{T} \left[y_{x,t} - (a_{x,0} + a_{x,1}t + a_{x,2}y_{x,t-1})\right]^2, \tag{1.10.21}$$

where $y_{x,t} = \ln m_{x,t}$ are log-central death rates from a sample time series $\{\ln m_{x,t},\ t = 1, 2, \ldots, T\}$.

The estimates of $\alpha_x, \beta_x, \gamma_x$ are obtained from relations (1.10.19) and σ_x^2 are estimated by determining errors $\hat{\xi}_{x,t}$ from the optimization problem (1.10.21) and by calculating the second raw moment of the residuals obtained.

1.11. The Milevsky–Promislow models with two or more linear scalar filters

1.11.1. MP model with two dependent filters

Let us consider the MP model with two linear scalar filters, i.e.

$$\mu_x(t) = \mu_{x0} \exp\{\gamma_x t + q_{x1}z_1(t) + q_{x2}z_2(t)\}, \quad t \in \mathbb{R}^+, \tag{1.11.1}$$

where $\gamma_x, q_{x1}, q_{x2}, \mu_{x0} > 0$ and z_1, z_2 are two dependent filters defined by means of the following stochastic differential equations

$$dz_1(t) = -\beta_{x1}z_1(t)dt + \sigma_{x1}dw(t), \quad \beta_{x1}, \sigma_{x1} > 0, \tag{1.11.2}$$

$$dz_2(t) = -\beta_{x2}z_2(t)dt + \sigma_{x2}dw(t), \quad \beta_{x2}, \sigma_{x2} > 0, \tag{1.11.3}$$

with $w(t)$ standing for a standard Wiener process.

Equations (1.11.1)–(1.11.3) define the Milevsky–Promislow model with 2 Dependent Filters (MP-2DF model).

By applying the Itô formula to the logarithm of (1.11.1) we obtain the stochastic differential equation

$$\begin{aligned} d\ln\mu_x(t) = {} & [\gamma_x - \beta_{x1}q_{x1}z_1(t) - \beta_{x2}q_{x2}z_2(t)]dt + \\ & + [\sigma_{x1}q_{x1} + \sigma_{x2}q_{x2}]dw(t), \quad t \in \mathbb{R}^+. \end{aligned} \tag{1.11.4}$$

Let us assume that $\beta_{x1} \neq \beta_{x2}$ and introduce a new state vector

$$\mathbf{h}_x(t) = [h_{x1}(t), h_{x2}(t), h_{x3}(t)]^T = [\ln \mu_x(t), z_1(t), z_2(t)]^T. \qquad (1.11.5)$$

Then equations (1.11.4) and (1.11.2)–(1.11.3) can be then written down as a vector equation

$$d\mathbf{h}_x(t) = \left(\begin{bmatrix} 0 & -\beta_{x1}q_{x1} & -\beta_{x2}q_{x2} \\ 0 & -\beta_{x1} & 0 \\ 0 & 0 & -\beta_{x2} \end{bmatrix} \mathbf{h}_x(t) + \begin{bmatrix} \gamma_x \\ 0 \\ 0 \end{bmatrix} \right) dt + \qquad (1.11.6)$$

$$+ \begin{bmatrix} \sigma_{x_1}q_{x1} + \sigma_{x2}q_{x2} \\ \sigma_{x1} \\ \sigma_{x2} \end{bmatrix} dw(t).$$

1.11.2. MP model with two independent filters

By analogy to equations (1.11.1)–(1.11.3) the Milevsky–Promislow model with 2 Independent Filters (MP-2IF model) can be written as

$$\mu_x(t) = \mu_{x0} \exp\{\gamma_x t + q_{x1}z_1(t) + q_{x2}z_2(t)\}, \quad t \in \mathbb{R}^+, \qquad (1.11.7)$$

$$dz_1(t) = -\beta_{x1}z_1(t)dt + \sigma_{x1}dw_1(t), \quad \sigma_{x1}, \beta_{x1} > 0, \qquad (1.11.8)$$

$$dz_2(t) = -\beta_{x2}z_2(t)dt + \sigma_{x2}dw_2(t), \quad \sigma_{x2}, \beta_{x2} > 0, \qquad (1.11.9)$$

where $\gamma_x, q_{x1}, q_{x2}, \mu_{x0}, \sigma_{x1}, \sigma_{x2}, \beta_{x1}, \beta_{x2} > 0$ are model's parameters and $w_1(t), w_2(t)$ are two independent standard Wiener processes.

Let us take logarithms of both sides of equality (1.11.7) and apply the Itô formula. As a result, we receive the following representation

$$d \ln \mu_x(t) = [\gamma_x - \beta_{x1}q_{x1}z_1(t) - \beta_{x2}q_{x2}z_2(t)]dt + \qquad (1.11.10)$$

$$+ \sigma_{x1}q_{x1}dw_1(t) + \sigma_{x2}q_{x2}dw_2(t), \quad t \in \mathbb{R}^+.$$

We define a new state vector, assuming that $\beta_{x1} \neq \beta_{x2}$

$$\mathbf{h}_x(t) = [h_{x1}(t), h_{x2}(t), h_{x3}(t)]^T = [\ln \mu_x(t), z_1(t), z_2(t)]^T. \quad (1.11.11)$$

Then equations (1.11.10) and (1.11.8)–(1.11.9) can be replaced by a vector equation

$$d\mathbf{h}_x(t) = \left(\begin{bmatrix} 0 & -\beta_{x1} q_{x1} & -\beta_{x2} q_{x2} \\ 0 & -\beta_{x1} & 0 \\ 0 & 0 & -\beta_{x2} \end{bmatrix} \mathbf{h}_x(t) + \begin{bmatrix} \gamma_x \\ 0 \\ 0 \end{bmatrix} \right) dt + $$
$$+ \begin{bmatrix} \sigma_{x1} q_{x1} \\ \sigma_{x1} \\ 0 \end{bmatrix} dw_1(t) + \begin{bmatrix} \sigma_{x2} q_{x2} \\ 0 \\ \sigma_{x2} \end{bmatrix} dw_2(t). \quad (1.11.12)$$

1.11.3. MP model with a vector filter

Substituting a vector filter for one-dimensional filter equation in (1.8.1)–(1.8.2), we obtain the generalized version of the MP model

$$\mu_x(t) = \mu_{x0} \exp\{\gamma_x t + \mathbf{q}_x^T \mathbf{z}(t)\}, \quad t \in \mathbb{R}^+, \quad (1.11.13)$$

$$d\mathbf{z}(t) = \mathbf{A}_x \mathbf{z}(t) dt + \mathbf{G}_x(t) dw(t), \quad (1.11.14)$$

where

$\gamma_x, \mu_{x0} > 0$ are constant parameters and $\mathbf{q}_x = [q_x^1, \ldots, q_x^n]^T$ is a constant vector,

$\mathbf{z} \in \mathbb{R}^n$ is a filter vector,

$\mathbf{A}_x$ is an $n \times n$ constant matrix with $\mathbf{A}_x^i$ as an i-th row of the matrix,

$\mathbf{G}_x(t) = [G_x^1(t), \ldots, G_x^n(t)]^T$ is a vector with coordinates $G_x^i(t) > 0$ representing deterministic, differentiable functions of time,

$w(t)$ is a scalar standard Wiener process.

Model (1.11.13)–(1.11.14) will be called the Milevsky–Promislow model with a Vector Linear Filter (MP-VLF model).

By taking logarithms of both sides of (1.11.13) and using the Itô formula, the MP-VLF model can be transformed to

$$d\ln\mu_x(t) = [\gamma_x + \sum_{i=1}^{n} q_x^i \mathbf{A}_x^i \mathbf{z}(t)]dt + \sum_{i=1}^{n} q_x^i G_x^i(t)dw(t), \tag{1.11.15}$$

$$d\mathbf{z}(t) = \mathbf{A}_x\mathbf{z}(t)dt + \mathbf{G}_x(t)dw(t), \quad t \in \mathbb{R}^+.$$

1.12. Final remarks

In the beginning of this chapter basic mortality notions and characteristics are introduced, among others some life-table measures as well as continuous-time survival functions. An overview of static mortality models is also provided, including historical laws of mortality. In greater details the well-known Lee–Carter model and some of its extensions, i.e. a fuzzy version, are discussed.

In the second part of the chapter several dynamic mortality models described by Itô's stochastic differential equations are presented, both in the discrete-time and continuous-time framework, in particular the Vasiček, Cox–Ingersoll–Ross, dynamic Lee–Carter, Milevsky–Promislow and Giacometti–Ortobelli–Bertocchi models. A few modifications of these models are also proposed, i.e. with some constant coefficients replaced by time depending functions, with two or more linear filter equations, or with a particular combination of the Milevsky–Promislow and Vasiček models.

This methodology suggests possibilities to create and study a new class of advanced dynamic mortality models. For instance, one can apply the Milevsky–Promislow model with linear filters described by Itô's stochastic differential equations with states depending diffusion terms. Also the Milevsky–Promislow model with some non-linear filters or continuous non-Gaussian excitation can be considered.

Unfortunately, models of dynamic systems expressed by means of simple stochastic differential equations can be insufficient to represent adequately evolving demographic processes. As a result, a family of models called hybrid models can be employed, i.e. models which account for interactions between continuous and discrete dynamics. Such hybrid mortality models are developed in Chapter 3 and the general theoretical introduction to hybrid modeling is provided in Chapter 2.

Chapter 2

Static and dynamic hybrid models

2.1. Introduction

As it was mentioned at the beginning of the book, the idea of constructing generalized mathematical models can be successfully realized by applying hybrid (or switching) systems. These models are usually described by algebraic equations or differential equations, deterministic or stochastic. In successive switching points the structure of the models changes according to the given switching rule thereupon creates the hybrid model. The switching rule can be random or dependent on some state variables. When subsystems (structures) are described by algebraic equations the underlying models are called static hybrid models.

Similar switching rules are used when subsystems are described by stochastic differential equations. The models are then called dynamic hybrid models. Such models with a switching rule defined as a right-continuous Markov chain are called Markov jump processes.

In this chapter we consider hybrid models described by linear Itô's stochastic differential equations for all subsystems with a given set of switching time points.

When the diffusion part of equations do not depend on a vector state it is possible to find the solution for each subsystem analytically. Assuming that the final value of the solution for the first subsystem is equal to the initial value of the solution for the second subsystem etc., we obtain the continuous solution for the whole hybrid system. When the diffusion part of equations depends on a vector state and it is not possible to find the solution for each subsystem in an analytical way, we will find moment equations constituting a new deterministic hybrid system. The mathematical tools introduced in this chapter will be used in Chapter 3.

2.2. Static hybrid models

Let us consider a family of static random systems represented by non-linear random vector equations such as

$$\mathbf{y}(t,l,\omega) = \mathbf{f}(\mathbf{x}(t,\omega),l), \quad \mathbf{x}(t_0,\omega) = \mathbf{x}_0, \quad l \in \mathbb{S}, \tag{2.2.1}$$

where

$\mathbb{S} = \{1,\ldots,N\}$ is a set of states of the system,
ω is an element of probabilistic space Ω,
$\mathbf{f}(\mathbf{x}_0,l) = \mathbf{0}$, $l \in \mathbb{S}$,
$\mathbf{x}(t) \in \mathbb{R}^n$ is a continuous input process with the initial condition $\mathbf{x}_0 \in \mathbb{R}^n$,
$\mathbf{y}(t,l)$ is an output process of the l-th subsystem.

Let us assume that there are of non-negative constants K_l meeting the conditions

$$|\mathbf{f}(\mathbf{x}(t),l)| \le K_l|\mathbf{x}| \quad \forall \mathbf{x} \in \mathbb{R}^n, \ \forall l \in \mathbb{S}, \ \forall \omega \in \Omega. \tag{2.2.2}$$

The system of equations (2.2.1) can also be written as

$$\mathbf{y}(t,\sigma(t),\omega) = \mathbf{f}(\mathbf{x}(t,\omega),\sigma(t)), \quad \mathbf{x}(t_0,\omega) = \mathbf{x}_0, \ \sigma(t_0) = \sigma_0, \tag{2.2.3}$$

where $\sigma(t) : \mathbb{R}^+ \to \mathbb{S}$ is a switching rule. We assume that $\sigma(t)$ is independent of the initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$.

In the analysis of switching systems, an important role is played by switching processes. The three main types of switchings considered in the literature are the following:

- an arbitrary switching,
- a switching dependent on the value of $\mathbf{x}(t)$, i.e. $\sigma(\mathbf{x}(t)) : \mathbb{R}^n \to \mathbb{S}$,
- a random switching usually represented by the Markov chain, i.e. $\sigma(t)=r(t)$ is a right-continuous Markov chain defined on the probabilistic space Ω and taking values in a finite space of states $\mathbb{S} = \{1,\ldots,N\}$ with generator $\Gamma = [\gamma_{ij}]_{N\times N}$, i.e.

$$\mathbf{P}\{r(t+\delta)=j|r(t)=i\}=\begin{cases} \gamma_{ij}\delta + o(\delta) & \text{for} \quad i \ne j, \\ 1 + \gamma_{ii}\delta + o(\delta) & \text{for} \quad i = j, \end{cases} \tag{2.2.4}$$

where $\delta > 0$ and $\gamma_{ij} \ge 0$ is the probability of transition from state i to state j if $i \ne j$, $\gamma_{ii} = -\sum_{i\ne j} \gamma_{ij}$.

Since the Markov chain is assumed to be irreducible, rank$(\Gamma) = N-1$ and it has only one stationary solution $\mathcal{P} = [\pi_1, \pi_2, \ldots, \pi_N]^T \in \mathbb{R}^N$, which can be found by solving a system of equations

$$\begin{cases} \mathcal{P}\Gamma = \mathbf{0}, \\ \text{where } \sum_{i=1}^{N} p_i = 1 \ \text{ and } \ p_i > 0 \ \forall i \in \mathbb{S}. \end{cases} \tag{2.2.5}$$

The time points when changes in the discrete system take place, i.e. when a model given, for instance, by function $\mathbf{f}(\mathbf{x}(t), i)$ becomes a model described by $\mathbf{f}(\mathbf{x}(t), j)$, will be called switching time points or switchings and will be denoted by $\{\tau_j\}_{j\in\mathbb{N}}$, so

$$0 = \tau_0 < \tau_1 < \tau_2 < \ldots < \tau_j \tag{2.2.6}$$

Let us assume that at moment $t = \tau_j$ current discrete state $l_{current} = l(\tau_{j-1})$ changes into future state $l_{future} = l(\tau_j)$; at the same time, an abrupt change in the continuous state can also take place, i.e.

$$\mathbf{x}(\tau_j) \neq \mathbf{x}(\tau_j-). \tag{2.2.7}$$

The discrete state $l(t)$ remains constant between successive switching times

$$l(t) = l_{current} \in \mathbb{S} \ \text{ for } \ t \in [\tau_{j-1}, \tau_j), \tag{2.2.8}$$

so

$$\mathbf{f}(\mathbf{x}(t), l(t)) = \mathbf{f}(\mathbf{x}(t), l_{current}) \ \text{ for } \ t \in [\tau_{j-1}, \tau_j), \ \ j \in \mathbb{N}. \tag{2.2.9}$$

Example 2.1. Let us consider a deterministic scalar hybrid system with two states, defined by functions

$$\begin{aligned} f(x,1) &= a_1 \exp\{-\alpha_1 x\} + b_1, \\ f(x,2) &= a_2 \exp\{-\alpha_2 x\} + b_2, \end{aligned} \tag{2.2.10}$$

where $x \in \mathbb{R}$ and a_i, b_i, α_i for $i = 1, 2$ are constant parameters.

Let us assume that state 1 changes into state 2 when $x = \bar{x}$ and that the final value of the first function and the initial value of the second function are equal, i.e.

$$f(\bar{x},1) = a_1 \exp\{-\alpha_1 \bar{x}\} + b_1 = f(\bar{x},2) = a_2 \exp\{-\alpha_2 \bar{x}\} + b_2. \tag{2.2.11}$$

For $b_1 = b_2$, (2.2.11) is equivalent to condition

$$\ln\left(\frac{a_1}{a_2}\right) = -\bar{x}(\alpha_2 - \alpha_1) \tag{2.2.12}$$

and the output of the hybrid system is represented by

$$y(x) = \begin{cases} a_1 \exp\{-\alpha_1 x\} + b_1 & \text{for} \quad x < \bar{x}, \\ \\ a_2 \exp\{-\alpha_2 x\} + b_2 & \text{for} \quad x \geq \bar{x}. \end{cases} \tag{2.2.13}$$

2.3. Dynamic hybrid models

There are two classes of models that can be identified within dynamic hybrid models. The first class contains stationary and non-stationary models with known analytical solutions of the Itô stochastic differential equations, with known analytical solutions of the moment equations, or with known probability densities. The second class is represented by models which are solved through difference schemes.

In the approach applied here, the hybrid dynamic stochastic system will be presented as a family of the Itô stochastic differential vector equations describing their dynamics in the time intervals between the switching time points.

We will assume that a vector stochastic process $\mathbf{x}(t)$ solving a stochastic vector equation and starting at time t_0 from initial state $\mathbf{x}_0$ switches at times $\tau_1, \tau_2, \ldots, \tau_M$. We also assume that $\tau_0 = t_0$ and that in time intervals $[\tau_i, \tau_{i+1})$ the hybrid system remains in states $l_i \in \mathbb{S}$, $i = 0, 1, \ldots, M$, where $l_0, l_1, \ldots, l_M$ is any subsequence of N states (subsystems). The continuity of solutions is also assumed, meaning that the value of the process in state l_i at time τ_i, i.e. $\mathbf{x}(\tau_i, l_i)$, is equal to the value of the process in state l_{i-1} at time τ_i, i.e. $\mathbf{x}(\tau_i, l_{i-1})$.

For the sake of illustration, let us use an example.

Example 2.2. Let us consider a hybrid system with a three-element set of states, i.e. $\mathbb{S} = \{1, 2, 3\}$. Let the initial state be $l_0 = 2$ and the successive states are $l_1 = 3, l_2 = 2, l_3 = 1, l_4 = 3, l_5 = 1$. The times when switchings take place are $\tau_1, \tau_2, \ldots, \tau_5$. These switchings are assumed to be significant, meaning that two successive states are different from each other. In other words, a switching from state l_2 to state l_2 is not considered as a switching.

Let functions $\mathbf{f} : \mathbb{R}^n \times \mathbb{S} \to \mathbb{R}^n$ and $\mathbf{g}_k : \mathbb{R}^n \times \mathbb{S} \to \mathbb{R}^n$ define the initial conditions for the subsystems. The Itô vector stochastic differential equation for the l-th subsystem ($l \in \mathbb{S}$) is then of the form

$$d\mathbf{x}(t,l) = \mathbf{f}(\mathbf{x}(t,l),l)dt + \sum_{k=1}^{m} \mathbf{g}_k(\mathbf{x}(t,l),l)dw_k(t), \quad \mathbf{x}(t_{0l},l) = \mathbf{x}_{0l}, \quad (2.3.1)$$

where

$l \in \mathbb{S}$, $\mathbf{x}_{0l} \in \mathbb{R}^n$, $t_{0l} \in \mathbb{R}^+$,

$\mathbf{f}(\mathbf{0},l) = \mathbf{0}$, $\mathbf{f}(\mathbf{x},l) = [f_1(\mathbf{x},l), \ldots, f_n(\mathbf{x},l)]^T$,

$\mathbf{g}_k(\mathbf{0},l) = \mathbf{0}$, $\mathbf{g}_k(\mathbf{x},l) = [\sigma_{k1}(\mathbf{x},l), \ldots, \sigma_{kn}(\mathbf{x},l)]^T$, $k = 1, \ldots, m$, are such that there exist non-negative constants K_l satisfying the following conditions

$$\begin{aligned} &|\mathbf{f}(\mathbf{x},l)|^2 + \sum_{k=1}^{m} |\mathbf{g}_k(\mathbf{x},l)|^2 \leq K_l(1 + |\mathbf{x}|^2), \ \forall \mathbf{x} \in \mathbf{U} \subset \mathbb{R}^n, \\ &|\mathbf{f}(\mathbf{x},l) - \mathbf{f}(\mathbf{y},l)| + \sum_{k=1}^{m} |\mathbf{g}_k(\mathbf{x},l) - \mathbf{g}_k(\mathbf{y},l)| \leq K_l|\mathbf{x} - \mathbf{y}|, \ \forall \mathbf{x}, \mathbf{y} \in \mathbf{U}. \end{aligned} \quad (2.3.2)$$

The above conditions ensure the existence of a solution of (2.3.1).

Equations (2.3.1) can be presented as a hybrid system

$$\begin{aligned} &d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t),\sigma(t))dt + \sum_{k=1}^{m} \mathbf{g}_k(\mathbf{x}(t),\sigma(t))dw_k(t), \\ &\mathbf{x}(t_0) = \mathbf{x}_0, \quad \sigma(t_0) = \sigma_0, \end{aligned} \quad (2.3.3)$$

where $\sigma(t)$ is a switching rule defined in the same manner as for the static hybrid models.

For a special case of linear systems with additive noise, let us consider a hybrid system given by a vector stochastic differential equation

$$\begin{aligned} &d\mathbf{x}(t) = [\mathbf{A}_0(t,\sigma(t)) + \mathbf{A}(t,\sigma(t))\mathbf{x}(t)]dt + \sum_{k=1}^{m} \mathbf{G}_{k0}(t,\sigma(t))dw_k(t), \\ &\mathbf{x}(t_0) = \mathbf{x}_0, \ \sigma(t_0) = \sigma_0, \end{aligned} \quad (2.3.4)$$

where
$\mathbf{x}(t) = [x_1(t), ..., x_n(t)]^T$ and $\mathbf{x}_0 = [x_{01}, ..., x_{0n}]^T$ is an initial condition,
$\sigma(t) : \mathbf{T} \to \mathbb{S}$ is a switching rule and $\sigma_0 \in \mathbb{S}$,
$\mathbf{A}(t, l) = [a_{ij}(t, l)]$, $l \in \mathbb{S}$, $i, j = 1, ..., n$, $k = 1, \dots, m$,
$\mathbf{A}_0(t, l) = [a_0^1(t, l), ..., a_0^n(t, l)]^T$,
$\mathbf{G}_{k0}(t, l) = [g_{k0}^1(t, l), ..., g_{k0}^n(t, l)]^T$ are n-dimensional vectors,
$a_0^i(t, l)$, $a_{ij}(t, l)$ and $g_{k0}^i(t, l)$, are limited measurable deterministic functions of $t \in \mathbb{R}^+$,
$w_k(t)$, $k = 1, \dots, m$ are independent standard Wiener processes.

The solution is then given by

$$\mathbf{x}(t) = \mathbf{\Psi}(t, t_0, \sigma(t))\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{\Psi}(t, s, \sigma(s))\mathbf{A}_0(s, \sigma(s))ds + \\ + \int_{t_0}^{t} \mathbf{\Psi}(t, s, \sigma(s)) \sum_{k=1}^{m} \mathbf{G}_{k0}(s, \sigma(s))dw_k(s), \tag{2.3.5}$$

where the $n \times n$ fundamental matrix $\mathbf{\Psi}(t, t_0, \sigma(t))$ is defined by appropriate $n \times n$ fundamental matrices for subsystems $\mathbf{\Psi}(t, t_{0l}, l)$, $l \in \mathbb{S}$

$$\frac{d\mathbf{x}(t, l)}{dt} = \mathbf{A}(t, l)\mathbf{x}(t, l), \quad \mathbf{x}(t_{0l}, l) = \mathbf{x}_{0l}. \tag{2.3.6}$$

In particular, when $\mathbf{A}(t, l) = \mathbf{A}(l)$ are constant matrices, the following relation takes place in time $t - t_{0l}$ for l-th subsystem

$$\mathbf{\Psi}(t, t_{0l}, l) = \mathbf{\Psi}(t - t_{0l}, l) = \exp\{\mathbf{A}(l)(t - t_{0l})\} = \\ = \sum_{j=0}^{\infty} \frac{1}{j!} \mathbf{A}^j(l)(t - t_{0l})^j. \tag{2.3.7}$$

Relation (2.3.5) is then reduced to

$$\mathbf{x}(t, l) = \exp\{\mathbf{A}(l)(t - t_{0l})\}\mathbf{x}_{0l} + \int_{t_{0l}}^{t} \exp\{\mathbf{A}(l)(t - s)\}\mathbf{A}_0(s, l)ds + \\ + \int_{t_{0l}}^{t} \exp\{\mathbf{A}(l)(t - s)\} \sum_{k=1}^{m} \mathbf{G}_{k0}(s, l)d\boldsymbol{w}(s). \tag{2.3.8}$$

Let us assume that the switching time points are $\tau_1, \tau_2, \ldots, \tau_M$, that $\tau_0 = t_0$ and that in time intervals $[\tau_i, \tau_{i+1})$ the hybrid system is in states $l_i \in \mathbb{S}$, $i = 0, 1, \ldots, M$, where $l_0, l_1, \ldots, l_M$ is any subsequence of N states (subsystems). The continuity of solutions is also assumed, i.e. $\mathbf{x}(\tau_i, l_i) = \mathbf{x}(\tau_i, l_{i-1})$.

In this case, for $\mathbf{A}_0(s, l_i) = \mathbf{0}$ for all $i = 0, 1, \ldots, k$ the following solution of the hybrid system (2.3.4) reduced to (2.3.8) is obtained

$$\begin{aligned}
\mathbf{x}(t, l_0) &= \exp\{\mathbf{A}(l_0)(t - t_0)\}\mathbf{x}_0 + \\
&\quad + \int_{t_0}^{t} \exp\{\mathbf{A}(l_0)(t-s)\} \sum_{k=1}^{m} \mathbf{G}_{k0}(s, l_0) d\boldsymbol{w}(s), \quad t \in [\tau_0, \tau_1), \\
\mathbf{x}(t, l_1) &= \exp\{\mathbf{A}(l_1)(t - \tau_1)\}\mathbf{x}(\tau_1, l_0) + \\
&\quad + \int_{\tau_1}^{t} \exp\{\mathbf{A}(l_2)(t-s)\} \sum_{k=1}^{m} \mathbf{G}_{k0}(s, l_1) d\boldsymbol{w}(s), \quad t \in [\tau_1, \tau_2), \\
&\;\vdots \\
\mathbf{x}(t, l_M) &= \exp\{\mathbf{A}(l_M)(t - \tau_M)\}\mathbf{x}(\tau_M, l_{M-1}) + \\
&\quad + \int_{\tau_M}^{t} \exp\{\mathbf{A}(l_M)(t-s)\} \sum_{k=1}^{m} \mathbf{G}_{k0}(s, l_M) d\boldsymbol{w}(s), \quad t \geq \tau_M.
\end{aligned} \tag{2.3.9}$$

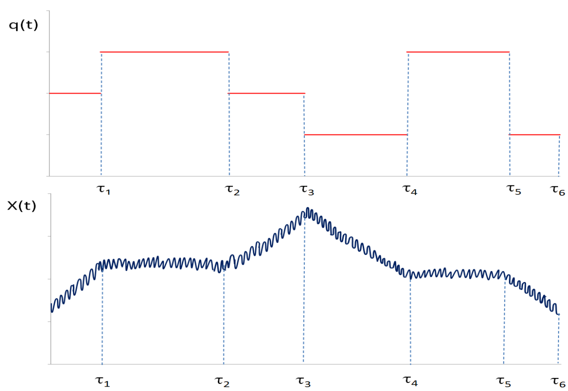

Figure 2.1. Trajectories of a switching process and of a dynamic system
Source: Developed by the authors

A trajectory of a switching rule $q(t)$ described in Example 2.2 and a corresponding trajectory of a dynamic switching system solution $x(t)$ are plotted in Figure 2.1.

Linear systems with parametric noise do not have an explicit solution, excluding a scalar case.

Scalar homogeneous hybrid system

Let us consider a scalar linear hybrid system with M switchings that constitutes a family of the Itô homogenous stochastic equations

$$dx(t,l_i) = a(t,l_i)x(t,l_i)dt + g(t,l_i)x(t,l_i)dw(t), \tag{2.3.10}$$

with an initial condition $x(\tau_0, l_0) = x_0$, where $t \in [t_0, \infty)$, $a(t, l_i)$ and $g(t, l_i)$, $i = 0, 1, \ldots, M$ are some linear functions of variable t,

We assume that switching time points are $\tau_1, \ldots, \tau_M$, that $\tau_0 = t_0$ and that in time intervals $[\tau_i, \tau_{i+1})$ the hybrid system is in states $l_i \in \mathbb{S}$, $i = 0, 1, \ldots, M$, where $l_0, l_1, \ldots, l_M$ is any subsequence of N states. The continuity of the solutions of (2.3.10) is assumed, i.e. $x(\tau_i, l_i) = x(\tau_i, l_{i-1})$. Initial conditions $x(\tau_i, l_i)$ are random variables independent of a standard Wiener process $w(t)$.

Based on the Itô formula, we can prove that the solution of equation (2.3.10) is a stochastic process

$$x(t,l_i) = \psi(t,\tau_i,l_i)\, x(\tau_i, l_{i-1}),$$

where

$$\psi(t,\tau_i,l_i) = \exp\left\{\int_{\tau_i}^{t}\left[a(s,l_i) - \frac{g^2(s,l_i)}{2}\right]ds + \int_{\tau_i}^{t} g(s,l_i)dw(s)\right\}. \tag{2.3.11}$$

From equalities (2.3.11) it follows that the value of the solution of equation (2.3.10) after M switchings is given by

$$\begin{aligned} x(t) &= \psi(t,\tau_M,l_M)\ \ldots\ \psi(\tau_2,\tau_1,l_1)\,\psi(\tau_1,\tau_0,l_0)\,x(\tau_0) = \\ &= \exp\left\{\int_{\tau_M}^{t}\left[a(s,l_M) - \frac{g^2(s,l_M)}{2}\right]ds + \int_{\tau_M}^{t} g(s,l_M)dw(s) + \right. \\ &\left. + \sum_{i=1}^{M}\int_{\tau_{i-1}}^{\tau_i}\left[a(s,l_i) - \frac{g^2(s,l_i)}{2}\right]ds + \sum_{i=1}^{M}\int_{\tau_{i-1}}^{\tau_i} g(s,l_i)dw(s)\right\} x(\tau_0). \end{aligned} \tag{2.3.12}$$

Scalar heterogeneous hybrid system

Let us consider a scalar linear hybrid system with M switchings, constituting a family of the Itô stochastic equations

$$
\begin{aligned}
dx(t,l_i) &= [a(t,l_i)x(t,l_i) + b(t,l_i)]dt + \\
&\quad + [g(t,l_i)x(t,l_i) + q(t,l_i)]dw(t),
\end{aligned} \tag{2.3.13}
$$

with an initial condition $x(\tau_0, l_0) = x_0$, where $t \in [t_0, \infty)$, $b(t, l_i)$ and $q(t, l_i)$, $i = 0, 1, \ldots, M$ are some non-linear functions of time t; all other symbols are the same as in equation (2.3.10). The continuity of solutions is also assumed, i.e. $x(\tau_i, l_i) = x(\tau_i, l_{i-1})$.

The solution of (2.3.13) is then given by

$$
\begin{aligned}
x(t,l_i) &= \psi(t,\tau_i,l_i)\Big\{x(\tau_i,l_{i-1}) + \int_{\tau_i}^{t}\psi^{-1}(s,\tau_i,l_i)q(s,l_i)dw(s) \\
&+ \int_{\tau_i}^{t}\psi^{-1}(s,\tau_i,l_i)[b(s,l_i) - q(s,l_i)g(s,l_i)]ds\Big\}, \quad t \in [\tau_i, \tau_{i+1}).
\end{aligned} \tag{2.3.14}
$$

By dividing (2.3.14) into subintervals we have

$$
\begin{aligned}
x(t,l_0) &= \Big\{x(\tau_0,l_0) + \int_{\tau_0}^{t}\psi^{-1}(s,\tau_0,l_0)[b(s,l_0) - q(s,l_0)g(s,l_0)]ds \\
&+ \int_{\tau_0}^{t}\psi^{-1}(s,\tau_0,l_0)q(s,l_0)dw(s)\Big\}\psi(t,\tau_0,l_0) \text{ for } t \in [\tau_0,\tau_1), \\[2ex]
x(t,l_1) &= \Big\{x(\tau_1,l_0) + \int_{\tau_1}^{t}\psi^{-1}(s,\tau_1,l_1)[b(s,l_1) - q(s,l_1)g(s,l_1)]ds \\
&+ \int_{\tau_1}^{t}\psi^{-1}(s,\tau_1,l_1)q(s,l_1)dw(s)\Big\}\psi(t,\tau_1,l_1) \text{ for } t \in [\tau_1,\tau_2), \\
&\vdots \\
x(t,l_M) &= \Big\{x(\tau_M,l_{M-1}) + \int_{\tau_M}^{t}[\psi^{-1}(s,\tau_M,l_M)[b(s,l_M) - q(s,l_M)g(s,l_M)]ds \\
&+ \int_{\tau_M}^{t}\psi^{-1}(s,\tau_M,l_M)q(s,l_M)dw(s)\Big\}\psi(t,\tau_M,l_M) \text{ for } t \geq \tau_M.
\end{aligned} \tag{2.3.15}
$$

2.4. Moment equations for the hybrid models

When equations (2.3.1) cannot be explicitly solved, the moment equations come in handy in analyzing a stochastic hybrid system. We shall consider them with respect to the linear systems with additive and parametric excitations for any switchings.

Let us analyze the hybrid vector linear Itô stochastic equation with additive and parametric excitation and any switchings

$$d\mathbf{x}(t) = [\mathbf{A}_0(t,\sigma(t)) + \mathbf{A}(t,\sigma(t))\mathbf{x}(t)]dt +$$

$$+\sum_{k=1}^{m}[\mathbf{G}_{k0}(t,\sigma(t)) + \mathbf{G}_k(t,\sigma(t))\mathbf{x}(t)]dw_k(t), \qquad (2.4.1)$$

$$\mathbf{x}(t_0) = \mathbf{x}_0, \ \sigma(t_0) = \sigma_0,$$

where

$\mathbf{x}(t) = [x_1(t), ..., x_n(t)]^T$ and $\mathbf{x}_0 = [x_{01}, ..., x_{0n}]^T$ is an initial condition,

$\mathbf{A}_0(t,\sigma(t)) = [a_0^1(t,\sigma(t)), ..., a_0^n(t,\sigma(t))]^T$,

$\mathbf{G}_{k0}(t) = [\sigma_{k0}^1(t,\sigma(t)), ..., \sigma_{k0}^n(t,\sigma(t))]^T$ are n-dimensional vectors, $k = 1, \ldots, m$,

$\mathbf{A}(t,\sigma(t)) = [a_{pj}(t,\sigma(t))]$, $\mathbf{G}_k(t,\sigma(t)) = [\sigma_{kj}^p(t,\sigma(t))]$ are $n \times n$ matrices, $p, j = 1, \ldots, n$,

a_0^i, a_{ij} and σ_{k0}^i are limited measurable and deterministic functions of $t \in [0, \infty)$,

$w_k(t)$, $k = 1, \ldots, m$ are independent standard Wiener processes,

For simplicity, let us assume that $\mathbf{x}_0 = [x_{01}, \ldots, x_{0n}]^T$ is a random variable independent of $w_k(t)$, $k = 1, \ldots, m$.

Using the Itô formula and averaging, we arrive at equations for the first- and second-order moments of the l_i-th subsystem. Thus, for $t \in [\tau_i, \tau_{i+1})$ we have

$$\frac{d\mathbf{m}(t,l_i)}{dt} = \mathbf{A}_0(t,l_i) + \mathbf{A}(t,l_i)\mathbf{m}(t,l_i),$$

$$\mathbf{m}(\tau_i,l_i) = \mathbf{m}(\tau_i,l_{i-1}), \qquad (2.4.2)$$

$$\frac{d\mathbf{\Gamma}(t,l_i)}{dt} = \mathbf{m}(t,l_i)\mathbf{A}_0^T(t,l_i) + \mathbf{A}_0(t,l_i)\mathbf{m}^T(t,l_i) +$$
$$+\mathbf{\Gamma}(t,l_i)\mathbf{A}^T(t,l_i) + \mathbf{A}(t,l_i))\mathbf{\Gamma}(t,l_i) +$$
$$+\sum_{k=1}^{m}[\mathbf{G}_{k0}(t,l_i)\mathbf{G}_{k0}^T(t,l_i) + \mathbf{G}_k(t,l_i)\mathbf{m}(t,l_i)\mathbf{G}_{k0}(t,l_i) + \tag{2.4.3}$$
$$+\mathbf{G}_{k0}(t,l_i)\mathbf{m}^T(t,l_i)\mathbf{G}_k^T(t,l_i) + \mathbf{G}_k(t,l_i)\mathbf{\Gamma}(t,l_i)\mathbf{G}_k^T(t,l_i)],$$
$$\mathbf{\Gamma}(\tau_i,l_i) = \mathbf{\Gamma}(\tau_i,l_{i-1}),$$

where

$$\mathbf{m}(t,l_i) = \mathrm{E}[\mathbf{x}(t,l_i)], \quad \mathbf{\Gamma}(t,l_i) = \mathrm{E}[\mathbf{x}(t,l_i)\mathbf{x}^T(t,l_i)],$$
$$\mathbf{m}(\tau_0,l_0) = \mathrm{E}[\mathbf{x}(t_0,l_0)], \; \mathbf{\Gamma}(\tau_0,l_0) = \mathrm{E}[\mathbf{x}(\tau_0,l_0)\mathbf{x}^T(\tau_0,l_0)]. \tag{2.4.4}$$

The coordinate equations (2.4.2) and (2.4.3) for l_i-th subsystem and $t \in [\tau_i, \tau_{i+1})$ are of the following form

$$\frac{dm_p(t,l_i)}{dt} = a_0^p(t,l_i) + \sum_{j=1}^{n} a_{pj}(t,l_i)m_j(t,l_i),$$
$$m_p(\tau_i,l_i) = m_p(\tau_i,l_{i-1}),$$
$$\frac{d\Gamma_{pj}(t,l_i)}{dt} = a_0^p(t,l_i)m_j(t,l_i) + a_0^j(t,l_i)m_p(t,l_i) +$$
$$+\sum_{q=1}^{n}[a_{pq}(t,l_i))\Gamma_{qj}(t,l_i) + a_{jq}(t,l_i)\Gamma_{qp}(t,l_i)] +$$
$$+\sum_{k=1}^{m}\sigma_{k0}^p(t,l_i)\sigma_{k0}^j(t,l_i) + \tag{2.4.5}$$
$$+\sum_{k=1}^{m}\sum_{\alpha=1}^{n}\sigma_{k\alpha}^i(t,l_i)\sigma_{k0}^j(t,l_i)m_\alpha(t,l_i) +$$
$$+\sum_{k=1}^{m}\sum_{\alpha=1}^{n}\sigma_{k\alpha}^j(t,l_i)\sigma_{k0}^p(t,l_i)m_\alpha(t,l_i) +$$
$$+\sum_{k=1}^{m}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\sigma_{k\alpha}^p(t,l_i)\sigma_{k\alpha}^j(t,l_i)\Gamma_{\alpha\beta}(t,l_i),$$
$$\Gamma_{pj}(\tau_0,l_0) = \Gamma_{pj0}, \; \Gamma_{pj}(\tau_i,l_i) = \Gamma_{pj}(\tau_i,l_{i-1}),$$

where

$$m_p(t,l_i) = \mathrm{E}[x_p(t,l_i)],\ \Gamma_{pj}(t,l_i) = \mathrm{E}[x_p(t,l_i)x_j(t,l_i)],$$
$$m_{p0} = \mathrm{E}[x_p(t_0,l_0)],\ \Gamma_{pj0} = \mathrm{E}[x_p(t_0,l_0)x_j(t_0,l_0)]. \tag{2.4.6}$$

It is noteworthy that the moment equations obtained are closed, i.e. their right-hand side and left-hand side moments are of the same order and the second-order moments only depend on variable $t \in \mathbb{R}^+$.

Example 2.3. Let us consider a scalar hybrid system with parametric and additive stochastic excitation and a random initial condition that constitutes a family of the Itô scalar linear equations

$$dx(t,l_i) = [a_0(l_i)+a(l_i)x(t,l_i)]dt+[b_0(l_i)+b(l_i)x(t,l_i)]dw(t), \tag{2.4.7}$$

with an initial condition $x(\tau_0,l_0) = x_0$, where $t \in [t_0,\infty)$, $a_0(l_i)$, $a(l_i)$, $b_0(l_i)$ and $b(l_i)$, $i = 0,1,...,M$ are constants and initial condition $x(\tau_0,l_0)$ is a random variable independent of a standard Wiener process $w(t)$. Let us denote $\mathrm{E}[x(\tau_0,l_0)]=m_0$, $\mathrm{E}[x^2(\tau_0,l_0)] = v_0$.

We assume that switchings take place at times $\tau_1,...,\tau_M$, that $\tau_0=t_0$ and that in time intervals $[\tau_i,\tau_{i+1})$ the hybrid system is in states $l_i \in \mathbb{S}$, $i = 0,1,\dots,M$, where $l_0,l_1,\dots,l_M$ is any subsequence of N states. The continuity of solutions is also assumed, i.e. $x(\tau_i,l_i) = x(\tau_i,l_{i-1})$.

Based on the Itô formula we can prove that the first- and second-order moment equations $m(t,l_i) = \mathrm{E}[x(t,l_i)]$, $\Gamma(t,l_i) = \mathrm{E}[x^2(t,l_i)]$, satisfy, respectively, the following differential equations for l_i-th subsystem and $t \in [\tau_i,\tau_{i+1})$

$$\frac{dm(t,l_i)}{dt} = a_0(l_i) + a(l_i)m(t,l_i),$$
$$m(\tau_i,l_i) = m(\tau_i,l_{i-1}), \tag{2.4.8}$$

$$\frac{d\Gamma(t,l_i)}{dt} = 2a_0(l_i)m(t,l_i) + 2a(l_i)\Gamma(t,l_i) + b_0^2(l_i)+$$
$$+ 2b_0(l_i)b(l_i)m + b_0(l_i)\Gamma(t,l_i), \tag{2.4.9}$$
$$\Gamma(\tau_i,l_i)=\Gamma(\tau_i,l_{i-1}).$$

By solving the system of equations (2.4.8), (2.4.9), we obtain

$$m(t,l_i) = -\frac{a_0(l_i)}{a(l_i)} + \left(m(\tau_i,l_i) + \frac{a_0(l_i)}{a(l_i)}\right)\exp\{a(l_i)(t-\tau_i)\}, \quad (2.4.10)$$

$$\Gamma(t,l_i) = -A_1 - A_2\exp\{a(l_i)(t-\tau_i)\} + \\ + \left[\Gamma^2(\tau_i,l_i)+A_1+A_2\right]\exp\left\{(2a(l_i)+b^2(l_i))(t-\tau_i)\right\}, \quad (2.4.11)$$

where

$$A_1 = \frac{b_0^2(l_i) - \frac{a_0(l_i)}{a(l_i)}(2a_0(l_i)+2b_0(l_i)b(l_i))}{2a(l_i)+b^2(l_i)}, \\ A_2 = \frac{\left(m(\tau_i,l_i)+\frac{a_0(l_i)}{a(l_i)}\right)(2a_0(l_i)+2b_0(l_i)b(l_i))}{a(l_i)+b^2(l_i)}. \quad (2.4.12)$$

Example 2.4. Let us consider a linear hybrid oscillator equation with deterministic and stochastic coefficients and excitation and with a deterministic initial condition, represented by the Itô vector linear hybrid equation

$$d\mathbf{x}(t) = \mathbf{A}_0(\sigma(t)) + \mathbf{A}(\sigma(t))\mathbf{x}(t)]dt + \\ + \sum_{k=0}^{2}[\mathbf{G}_{k0}(\sigma(t)) + \mathbf{G}_k(\sigma(t))\mathbf{x}(t)]dw_k(t), \quad (2.4.13) \\ \mathbf{x}(t_0) = \mathbf{x}_0, \;\; \sigma(t_0) = \sigma_0,$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \;\; \mathbf{x}_0 = \begin{bmatrix} x_{10} \\ x_{20} \end{bmatrix}, \\ \mathbf{A}_0(l) = \begin{bmatrix} 0 \\ -a_{0l} \end{bmatrix}, \;\; \mathbf{A}(l) = \begin{bmatrix} 0 & 1 \\ -\lambda_{0l}^2 & -2\zeta_l\lambda_{0l} \end{bmatrix}, \quad (2.4.14)$$

$$\mathbf{G}_{10}(l)=\mathbf{G}_{20}(l)=\mathbf{G}_{0}(l)=\mathbf{0},$$

$$\mathbf{G}_{00}(l)=\begin{bmatrix}0\\ \sigma_{0l}\end{bmatrix},\ \mathbf{G}_{1}(l)=\begin{bmatrix}0 & 0\\ \sigma_{1l} & 0\end{bmatrix},\ \mathbf{G}_{2}(l)=\begin{bmatrix}0 & 0\\ 0 & \sigma_{2l}\end{bmatrix}, \tag{2.4.15}$$

λ_{0l}, ζ_l, a_{0l}, σ_{0l}, σ_{1l} and σ_{2l}, $l=1,2$ are constant parameters and $w_k(t)$, $k=0,1,2$ are independent Wiener processes. Initial conditions x_{10} and x_{20} are random variables independent of $w_k(t)$.

The first- and second-order moment equations for the l_i-th subsystem have the following form

$$\frac{dm_1(t,l_i)}{dt}=m_2(t,l_i),$$

$$\frac{dm_2(t,l_i)}{dt}=-\lambda_{0l_i}^2 m_1(t,l_i)-2\zeta_{l_i}\lambda_{0l_i}m_2(t,l_i)-a_{0l_i}, \tag{2.4.16}$$

$$\frac{d\Gamma_{11}(t,l_i)}{dt}=2\Gamma_{12}(t,l_i),$$

$$\frac{d\Gamma_{12}(t,l_i)}{dt}=\Gamma_{22}(t,l_i)-a_{0l_i}m_1(t,l_i)-2\zeta_{l_i}\lambda_{0l_i}\Gamma_{12}(t,l_i)+$$

$$-\lambda_{0l_i}^2\Gamma_{11}(t,l_i), \tag{2.4.17}$$

$$\frac{d\Gamma_{22}(t,l_i)}{dt}=-2a_{0l_i}m_2(t,l_i)-4\zeta_{l_i}\lambda_{0l_i}\Gamma_{22}(t,l_i)+\sigma_{0l_i}^2+$$

$$-2\lambda_{0l_i}^2\Gamma_{12}(t,l_i)+\sigma_{1l_i}^2\Gamma_{11}(t,l_i)+\sigma_{2l_i}^2\Gamma_{22}(t,l_i),$$

where

$$m_p(t,l_i)=\mathrm{E}[x_p(t,l_i)],$$

$$\Gamma_{pj}(t,l_i)=E[x_p(t,l_i)x_j(t,l_i)],\ \ p,j=1,2,\ \ i=1,2,\ldots \tag{2.4.18}$$

Since moment equations (2.4.2) constitute a system of linear, deterministic equations, an analytical solution of this system is possible and it is of the following form

$$\mathbf{m}(t,l_i) = \exp\left\{\int_{\tau_i}^{t} \mathbf{A}(s,l_i)ds\right\}\mathbf{m}(\tau_i,l_{i-1})+$$

$$+\int_{\tau_i}^{t} \exp\left\{\int_{s}^{t} \mathbf{A}(u,l_i)du\right\}\mathbf{A}_0(s,l_i)ds, \tag{2.4.19}$$

$$\mathbf{m}(\tau_i,l_i) = \mathbf{m}(\tau_i,l_{i-1}) \quad \text{for} \quad t\in[\tau_i,\tau_{i+1}).$$

To solve the system of the moment equations (2.4.3) in a similar manner, equation (2.4.3) must first be replaced by the corresponding deterministic vector equation. To perform this, we will use $sc\mathbf{A}$, a notation used in the literature, which represents a column vector consisting of the column vectors of matrix $\mathbf{A}$

$$sc\mathbf{A} = ||[\mathbf{A}^{(\alpha)}]||. \tag{2.4.20}$$

The Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$ is a matrix $\mathbf{A}\otimes\mathbf{B}$, defined as follows

$$\mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \dots & A_{1n}\mathbf{B} \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & \dots & A_{2n}\mathbf{B} \\ \dots & \dots & \dots & \dots \\ A_{m1}\mathbf{B} & A_{m2}\mathbf{B} & \dots & A_{mn}\mathbf{B} \end{bmatrix}, \tag{2.4.21}$$

We have

$$sc\mathbf{\Gamma}(t,l_i) = \exp\left\{\int_{\tau_i}^{t} \mathcal{A}(s,l_i)ds\right\} sc\mathbf{\Gamma}(\tau_i,l_{i-1})+$$

$$+\int_{\tau_i}^{t} \exp\left\{\int_{s}^{t} \mathcal{A}(u,l_i)du\right\} sc\mathcal{Q}_0(s,l_i)ds, \tag{2.4.22}$$

$$sc\mathbf{\Gamma}(\tau_i,l_i) = sc\mathbf{\Gamma}(\tau_i,l_{i-1}) \quad \text{for} \quad t\in[\tau_i,\tau_{i+1}),$$

where

$$\mathcal{A}(s,l_i) = \mathbf{I}\otimes\mathbf{A}(s,l_i)+\mathbf{A}(s,l_i)\otimes\mathbf{I}+\sum_{k=1}^{m}\mathbf{G}_k(s,l_i)\otimes\mathbf{G}_k(s,l_i), \tag{2.4.23}$$

$$scQ_0(s,l_i) = sc\left((\mathbf{A}_0(s,l_i) + \bar{\mathbf{G}}(\mathbf{m}))(\mathbf{A}_0(s,l_i) + \bar{\mathbf{G}}(\mathbf{m}))^T\right), \quad (2.4.24)$$

$$\bar{\mathbf{G}}(\mathbf{m}) = \left[\sum_{j=1}^{n} \sigma_{kj}^{p}(s,l_i) m_j(s,l_i)\right]. \qquad (2.4.25)$$

2.5. Final remarks

In this chapter some linear hybrid systems are presented. Especially, the explicit solutions of linear hybrid vector systems with an additive Gaussian excitation are provided, both for realizations and for moments.

For linear hybrid vector systems with additive and multiplicative Gaussian excitations the explicit solutions can be obtained only for a special class of scalar systems. For non-linear hybrid vector systems with additive and multiplicative Gaussian excitations the exact solutions for realizations and for moments cannot be found analytically, since on the right-hand sides of the differential moment equations appear functions of moments of higher order than those on the left-hand sides of the respective equations. To omit these difficulties some closure techniques can be applied but in such cases only approximate solutions can be found. The wide analysis of these problems is provided for instance in the following books [Evlanov, Konstantinov 1976, Ibrahim 1985, Pugachev, Sinitsyn 1987, Socha 2008].

The methodology of constructing dynamic hybrid models presented in this chapter will be utilized in Chapter 3 on the hybrid mortality models.

Chapter 3

Dynamic hybrid mortality models

3.1. Introduction

In this chapter we will extend the mortality models presented in Chapter 1 to the corresponding hybrid mortality models. In the case of models (1.5.44) a natural approach to generalizing this family of models to the family of hybrid mortality models is defining them individually for separate subsystems $l \in \mathbb{S}$, e.g. for each time interval in which a given model "works" with the same set of parameters' values. For instance, by analogy to (1.5.44), the hybrid models for the log-central mortality rates $\ln m_{x,t}$ ($x = 0, 1, \ldots, X$, $t = 1, 2, \ldots, T$), can be written as

$$
\begin{aligned}
\ln m_{x,t}(l) &= \alpha_x(l) + \beta_x(l)\kappa_t(l), \\
\ln m_{x,t}(l) &= \alpha_x(l) + \beta_x(l)\kappa_t(l) + \gamma_{t-x}(l), \\
\ln m_{x,t}(l) &= \alpha_x(l) + \beta_x(l)\kappa_t(l) + \beta_x^{(1)}(l)\gamma_{t-x}(l), \\
\ln m_{x,t}(l) &= \alpha_x(l) + \kappa_t^{(1)}(l) + \kappa_t^{(2)}(l)(x-\bar{x}) + \kappa_t^{(3)}(l(x_c-x)^+), \\
\ln m_{x,t}(l) &= \alpha_x(l) + \kappa_t^{(1)}(l) + \kappa_t^{(2)}(l)(x-\bar{x}) + \kappa_t^{(3)}(l)(x_c-x)^+ + \\
&\quad + \gamma_{t-x}(l), \\
\ln m_{x,t}(l) &= \alpha_x(l) + \kappa_t^{(1)}(l) + \kappa_t^{(2)}(l)(x-\bar{x}) + \kappa_t^{(3)}(l)(x_c-x)^+ + \\
&\quad + \upsilon_x \kappa_t^{(4)}(l) + \gamma_{t-x}(l), \\
\ln m_{x,t}(l) &= \alpha_x(l) + \kappa_t^{(1)}(l) + \kappa_t^{(2)}(l)(x-\bar{x}) + \kappa_t^{(3)}(l)(x_c-x)^+ + \\
&\quad + \gamma_{t-x}(l)(x_c-x).
\end{aligned}
\tag{3.1.1}
$$

$$
\begin{aligned}
\eta_{x,t}(l) =& \kappa_t^{(1)}(l) + \kappa_t^{(2)}(l)(x-\bar{x}),\\
\eta_{x,t}(l) =& \kappa_t^{(1)}(l) + \kappa_t^{(2)}(l)(x-\bar{x}) + \gamma_{t-x}(l),\\
\eta_{x,t}(l) =& \kappa_t^{(1)}(l) + \kappa_t^{(2)}(l)(x-\bar{x}) + \kappa_t^{(3)}(l)v_x + \gamma_{t-x}(l),\\
\eta_{x,t}(l) =& \kappa_t^{(1)}(l)+\kappa_t^{(2)}(l)(x-\bar{x})+\gamma_{t-x}(l)(x_c-x),
\end{aligned}
\tag{3.1.2}
$$

where

$\eta_{x,t}(l)$ is a logit transformation of $q_{x,t}(l)$ for age x, year t and subsystem $l \in \mathbb{S}$, i.e.

$$\eta_{x,t}(l) = \ln \frac{q_{x,t}(l)}{1-q_{x,t}(l)}, \tag{3.1.3}$$

x_c is a fixed value,

$v_x = (x-\bar{x})^2 - \sigma_x^2$, where $\bar{x}$ and σ_x^2 denote, respectively, average age and age variance for the age groups,

$\alpha_x(l), \beta_x(l), \beta_x^{(1)}(l)$ are age-related parameters for l-th subsystem,

$\kappa_t(l)$ and $\kappa_t^{(i)}(l)$ for $i = 1, 2, 3, 4$ and $\gamma_{t-x}(l)$ are time- and cohort-related parameters for l-th subsystem, respectively.

The first equation in (3.1.1) will be called the Lee–Carter Hybrid model (LCH model).

In the remainder of this chapter we will consider some hybrid dynamic mortality models, which are hybrid versions of the dynamic models developed in Chapter 1.

Since the dynamic models are described by the Itô stochastic differential equations, thus the corresponding hybrid models will be also described by means of the Itô stochastic differential equations. Further, for simplicity, we will omit the term ”dynamic”.

For almost all the hybrid models we will consider their discrete-time representations, both for realizations and for moments.

To estimate the parameters of the proposed models, we shall use methods described in the literature [Ladde, Wu 2009, Yin *et al.* 2002, 2003]. In particular, we will consider estimation procedures which can be described according to the following two main steps.

Firstly, some switching points $\tau_1, \tau_2, \ldots, \tau_M$ are determined. Time intervals $I_0 = [t_0, \tau_1), I_1 = [\tau_1, \tau_2), \ldots, I_M = [\tau_M, T]$ within the observation period $[t_0, T]$ will be called "mortality regimes". Different subsystems $l \in \mathbb{S}$ of hybrid models will be identified by successive mortality regimes I_l.

In the second step, estimates of the model's parameters for each mortality regime are found, by solving the optimization problem or by using some iterative estimation algorithms.

Let us start our considerations from identification of switchings.

3.2. Identification of switchings

An important issue in defining hybrid models are the so-called switchings, used to identify sub-models of the hybrid system. In this approach switchings are recognized as different times at which the underlying mortality process switches from one state to another.

This section provides theoretical backgrounds of the self-adaptive statistical test which will be used for finding significant switchings in the time series of mortality rates. The test was originally proposed by [Janic-Wróblewska, Ledwina 2000].

3.2.1. An introductory example

Example 3.1. Let us consider the time series of log-central death rates $\{\ln m_{x,t},\ t = 1, 2, \ldots, T\}$ for a subpopulation of Polish women aged $x = 40$ years observed in the period 1958–2014 (Figure 3.1).

Estimates $\hat{c}_{1x}, \hat{c}_{0x}$ of the trend line parameters plotted in Figure 3.2 equal, respectively,

$$\hat{c}_{1x} \approx -0.0142,\ \hat{c}_{0x} \approx 0.8515 \tag{3.2.1}$$

and mean forecast error is $S_{\epsilon_x} \approx 0.1236$.

The way in which points in Figure 3.1 are arranged reveals a change observed in mortality pattern for the investigated population of women between years 1990 and 1991.

Better goodness-of-fit for trend lines for sub-periods 1958–1990 and 1991–2014 seems to confirm that this change is significant (Figure 3.3).

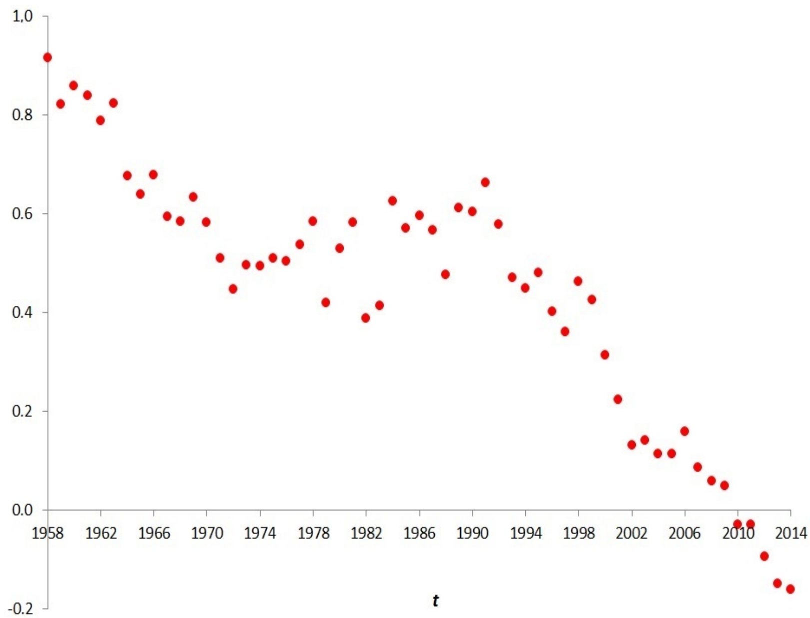

Figure 3.1. Log-central death rates for women aged $x = 40$ years (observation period 1958–2014)

Source: Developed by the authors

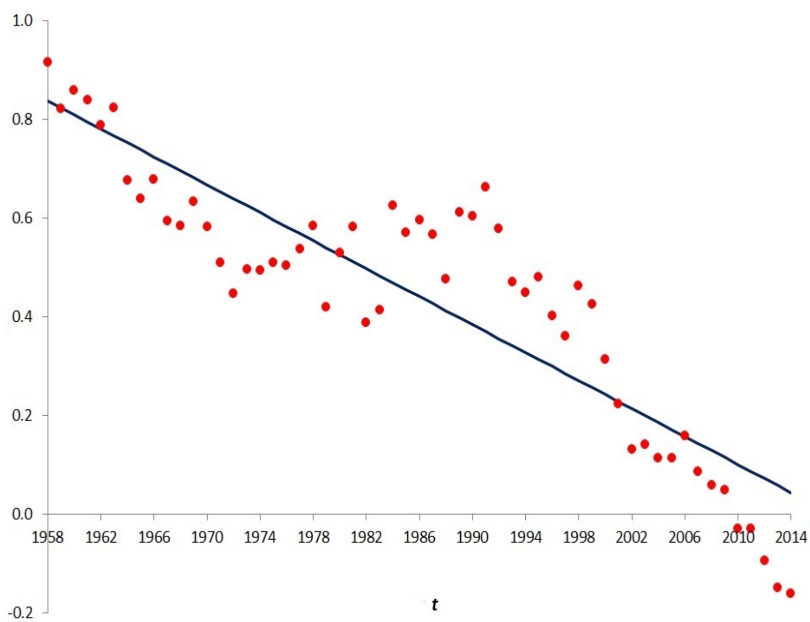

Figure 3.2. Log-central death rates for women aged $x = 40$ years and a fitted trend line

Source: Developed by the authors

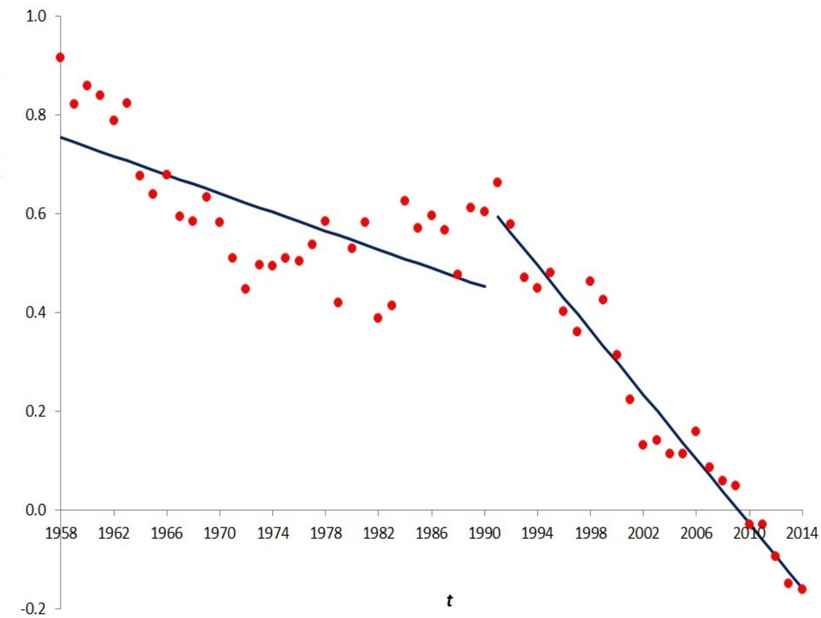

Figure 3.3. Log-central death rates for women aged $x = 40$ years and two fitted trend lines for years 1958–1990 and 1991–2014
Source: Developed by the authors

The 1958–1990 estimates of coefficients c_{1x}, c_{0x} are the following

$$\hat{c}_{1x}^{*} \approx -0.0095, \;\; \hat{c}_{0x}^{*} \approx 0.7636 \tag{3.2.2}$$

with mean forecast error $S_{\epsilon_x}^{*} \approx 0.1012$.

The values of the 1991–2014 estimates of coefficients c_{1x}, c_{0x} are

$$\hat{c}_{1x}^{**} \approx -0.0327, \hat{c}_{0x}^{**} \approx 0.6256 \tag{3.2.3}$$

with mean forecast error $S_{\epsilon_x}^{**} \approx 0.0514$.

Because of distinctively different direction and pace of mortality change, it seems also rational to break up the period of observation into three sub-periods or "mortality regimes" (see Figure 3.4).

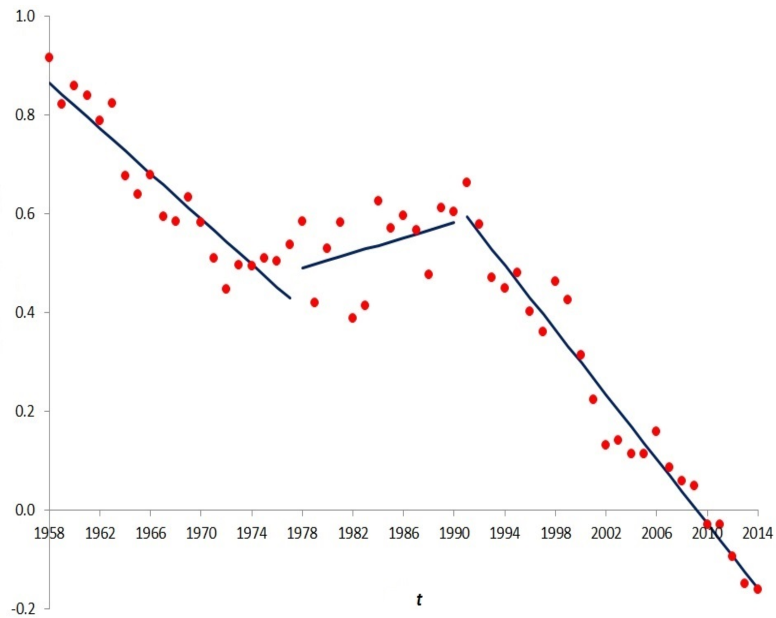

Figure 3.4. Log-central death rates for women aged $x = 40$ years
and three fitted trend lines for years 1958–1977, 1978–1990 and 1991–2014
Source: Developed by the authors

According to the above example, years 1978 and 1991 seem potentially significant as time points marking a change in the mortality trend for women aged $x = 40$ years in the period 1958–2014.

However, it is necessary to consider whether or not the potential switchings are accidental. If they are significant, several sub-models can be used, separately for each of the identified sub-periods.

The concept for finding statistically significant switching points used in determining sub-models of the hybrid models will be based here on the so-called self-adaptive statistical test (JL) introduced by [Janic-Wróblewska, Ledwina 2000, Antoch *et al.* 2008] and adopted to analyze mortality data series.

3.2.2. Theoretical backgrounds of the JL test

Let continuous random variables $U_1, U_2, \ldots, U_N$ be observed and let F_t denote the cumulative distribution function (CDF) of random variable U_t. We wish to verify the null hypothesis

$$H_0: \; F_1 = F_2 = \ldots = F_N, \tag{3.2.4}$$

against an alternative hypothesis

$$H_1 : \ \exists_{\eta \in (0.1)} \qquad F_1 = F_{[N\eta]} \neq F_{[N\eta+1]} = \ldots = F_N, \tag{3.2.5}$$

where $[N\eta]$ denotes an integer part of $N\eta$.

Statistics M_N of the JL test is of the form

$$M_N(e, p_N) = \max_{[eN] \leq m \leq [(1-e)N]} T\left(S(m, p_N); m\right), \tag{3.2.6}$$

where

N – sample size,

$e \in \left(0, \frac{1}{2}\right)$ – fixed value (hereafter $e = 0.1$),

$p_N = 1,5 \log N$ – positive number representing "penalty",

$S(m, p_N)$ – statistics defined by the formula

$$S(m, p_N) = \min\{k \in [1, d_N] : T(k, m) - kp_N \geq T(l, m) - lp_N; l = 1, \ldots, d_N\}, \tag{3.2.7}$$

d_N – natural number representing the complexity of the problem (hereafter $d_N = 10$),

$T(k, m)$ – statistics defined as

$$T(k, m) = \sum_{n=1}^{k} L^2(m; b_n), \tag{3.2.8}$$

$L(m, b_n)$ – statistics defined by formula

$$L(m, b_n) = \sum_{t=1}^{N} c_{mt} b_n \left(\frac{R_t - 0.5}{N}\right), \tag{3.2.9}$$

R_t – rank of U_t in a non-decreasing sequence $U_1, \ldots, U_N$,

c_{mt} – weights defined as

$$c_{mt} = \begin{cases} \sqrt{\frac{m(N-m)}{N}} \frac{1}{m}, & t = 1, 2, \ldots, m, \\ \\ -\sqrt{\frac{m(N-m)}{N}} \frac{1}{N-m}, & t = m+1, \ldots, N, \end{cases} \tag{3.2.10}$$

$b_n, n = 1, \ldots, k$ – Legendre's polynomials, orthonormal on interval $[0, 1]$, i.e.

$$b_n(z) = \frac{1}{n!\ 2^n} \frac{d^n\{(z^2 - 1)^n\}}{dz^n}. \tag{3.2.11}$$

The critical values of the JL test were derived by means of the Monte-Carlo method [Janic-Wróblewska, Ledwina 2000]. Since for $k = 1$ the statistics of the JL test and the statistics of Wilcoxon rank test are equivalent, the tabulated critical values of the latter test can be applied. A high value of M_N supports the rejection of H_0 hypothesis in favor of H_1 hypothesis.

3.2.3. Determining switching points from mortality rates

To find switching points in the mortality process $\ln \mu_x(t)$, we shall apply the JL test. According to the relation (1.3.23), we assume that $y_x(t) = \ln \mu_x(t)$ are approximated by log-central death rates $y_{x,t} = \ln m_{x,t}$ observed in sample time series $\{\ln m_{x,t},\ t = 1, 2, \ldots, T\}$ (see Section 1.3.2).

Let us consider the following sequence defined for fixed $x = 0, 1, \ldots, X$

$$\{U_{x,1}, U_{x,2}, \ldots, U_{x,N}\}, \tag{3.2.12}$$

where $U_{x,t} = y_{x,t+1} - y_{x,t}$, $y_{x,t} = \ln m_{x,t}$, $t = 1, 2, \ldots, T$.

Random variables $U_{x,t}$ represent here increments in the log-central death rates for the selected age group x. Because we have $T - 1$ of differences, hence $N = T - 1$.

The proposed application of the JL test to the analyzed problem is based on the following concept. Assume that for any t^* $(1 \leq t^* < T)$ variables

$$U_{x,1}, U_{x,2}, \ldots, U_{x,t^*} \tag{3.2.13}$$

have the same probability distributions as variables

$$U_{x,t^*+1}, U_{x,t^*+2} \ldots, U_{x,N}. \tag{3.2.14}$$

In such a case a switching point does not exist; otherwise, there is at least one switching point.

Example 3.2. Let us consider the data set from Example 3.1, i.e. the time series of log-central death rates recorded in years 1958–2000 for Polish women aged $x = 40$ (see Table 3.1).

Let us calculate differences $U_{x,t} = y_{x,t+1} - y_{x,t}$ and then statistics $L(m, b_n)$ from (3.2.9) for $m = 1, 2, \ldots, N-1$, $n = 1, 2, \ldots, k$, $k \in \mathbb{N}$.

For $m = 1$ expressions (3.2.9)–(3.2.10) defining statistics $L(1, b_n)$ and coefficients c_{1t} reduce to

$$L(1, b_n) = \sum_{t=1}^{N} c_{1t} b_n \left(\frac{R_t - 0.5}{N} \right), \tag{3.2.15}$$

where

$$c_{1t} = \begin{cases} \sqrt{\frac{N-1}{N}}, & t = 1, \\ \\ -\sqrt{\frac{N-1}{N}} \frac{1}{N-1}, & t = 2, 3, \ldots, N. \end{cases} \tag{3.2.16}$$

The Legendre polynomial b_n in (3.2.15) is determined from formula

$$b_n(z_t) = B_n(z_t)\sqrt{2n+1}. \tag{3.2.17}$$

For $n = 0, 1, 2,$ polynomials B_n take the following forms

$$\begin{aligned} B_0(z_t) &= 1, \\ B_1(z_t) &= 2z_t - 1, \\ B_2(z_t) &= \frac{1}{2}\left(3(2z_t - 1)^2 - 1\right). \end{aligned} \tag{3.2.18}$$

For $n = 3, 4, \ldots$, polynomials B_n are obtained from formula

$$B_{n+1}(z_t) = \frac{2n+1}{n+1}(2z_t - 1)B_n(z_t) - \frac{n}{n+1}B_{n-1}(z_t), \tag{3.2.19}$$

where arguments z_t are expressed as

$$z_t = \frac{R_t - 0.5}{N}, \quad t = 1, 2, \ldots, N \tag{3.2.20}$$

and R_t are the ranks of observations $U_{x,t}$ in the ordered sample.

Table 3.1. Log-central death rates for women aged $x = 40$ years (Poland, 1958–2000)

Year	$\ln m_{x,t}$	Year	$\ln m_{x,t}$
1958	0.9155	1979	0.4187
1959	0.8220	1980	0.5295
1960	0.8591	1981	0.5822
1961	0.8398	1982	0.3880
1962	0.7871	1983	0.4141
1963	0.8242	1984	0.6259
1964	0.6755	1985	0.5693
1965	0.6387	1986	0.5955
1966	0.6785	1987	0.5664
1967	0.5939	1988	0.4762
1968	0.5839	1989	0.6109
1969	0.6323	1990	0.6038
1970	0.5828	1991	0.6617
1971	0.5104	1992	0.5789
1972	0.4472	1993	0.4700
1973	0.4965	1994	0.4479
1974	0.4929	1995	0.4793
1975	0.5098	1996	0.4008
1976	0.5038	1997	0.3598
1977	0.5365	1998	0.4618
1978	0.5844	1999	0.4246
1979	0.4187	2000	0.3141

Source: Human Mortality Data Base.

Table 3.2 contains statistics $L(m, b_1)$ calculated for $m = 1$ and $m = 2$. Let us note that statistics $L(1, b_1)$ and $L(2, b_1)$ have different values only because of changes in the values of coefficients c_{mt}.

The sums of terms in columns 7 and 9 of Table 3.2 are the values of statistics $L(1; b_1)$ and $L(2; b_1)$, i.e.

$$L(1, b_1) = -1.2939, \quad L(2, b_1) = -0.2988. \tag{3.2.21}$$

Thus, we also get

$$L^2(1, b_1) = \ 1.6742, \quad L^2(2, b_1) = \ 0.0893. \tag{3.2.22}$$

Statistics $L^2(1, b_1)$ and $L^2(2, b_1)$ are the components of $T(k, 1)$ and $T(k, 2)$, respectively. The general definition of statistics $T(k, m)$ for given values of k and m comes from formula (3.2.8).

For $k = 1$ and $m = 1$, statistics $T(k, m)$ can be reduced to $T(1, 1) = L^2(1, b_1)$. For $k = 1$ and $m = 2$, we have $T(1, 2) = L^2(2, b_1)$.

Table 3.2. Auxiliary calculations for statistics $L(1;b_1)$ and $L(2;b_1)$

Year	$U_{x,t}$	R_t	z_t	$b_1(z_t)$	c_{1t}	$c_{1t}b_1(z_t)$	c_{2t}	$c_{2t}b_1(z_t)$
1959	−0.0935	6	0.13	−1.2784	0.9880	−1.2631	0.6901	−0.8822
1960	0.0371	32	0.75	0.8660	−0.0241	−0.0209	0.6901	0.5976
1961	−0.0192	21	0.49	−0.0412	−0.0241	0.0010	−0.0345	0.0014
1962	−0.0527	14	0.32	−0.6186	−0.0241	0.0149	−0.0345	0.0213
1963	0.0371	31	0.73	0.7835	−0.0241	−0.0189	−0.0345	−0.0270
1964	−0.1487	3	0.06	−1.5259	−0.0241	0.0368	−0.0345	0.0526
1965	−0.0368	18	0.42	−0.2887	−0.0241	0.0070	−0.0345	0.0100
1966	0.0399	33	0.77	0.9485	−0.0241	−0.0229	−0.0345	−0.0327
1967	−0.0847	8	0.18	−1.1135	−0.0241	0.0268	−0.0345	0.0384
1968	−0.0100	22	0.51	0.0412	−0.0241	−0.0010	−0.0345	−0.0014
1969	0.0484	35	0.82	1.1135	−0.0241	−0.0268	−0.0345	−0.0384
1970	−0.0496	15	0.35	−0.5361	−0.0241	0.0129	−0.0345	0.0185
1971	−0.0723	11	0.25	−0.8660	−0.0241	0.0209	−0.0345	0.0299
1972	−0.0632	12	0.27	−0.7835	−0.0241	0.0189	−0.0345	0.0270
1973	0.0493	36	0.85	1.1959	−0.0241	−0.0288	−0.0345	−0.0413
1974	−0.0037	25	0.58	0.2887	−0.0241	−0.0070	−0.0345	−0.0100
1975	0.0170	26	0.61	0.3712	−0.0241	−0.0089	−0.0345	−0.0128
1976	−0.0060	24	0.56	0.2062	−0.0241	−0.0050	−0.0345	−0.0071
1977	0.0327	30	0.70	0.7011	−0.0241	−0.0169	−0.0345	−0.0242
1978	0.0480	34	0.80	1.0310	−0.0241	−0.0248	−0.0345	−0.0356
1979	−0.1657	2	0.04	−1.6083	−0.0241	0.0388	−0.0345	0.0555
1980	0.1107	40	0.94	1,5259	−0.0241	−0.0368	−0.0345	−0.0526
1981	0.0528	37	0.87	1,2784	−0.0241	−0.0308	−0.0345	−0.0441
1982	−0.1942	1	0.01	−1.6908	−0.0241	0.0407	−0.0345	0.0583
1983	0.0261	27	0.63	0.4536	−0.0241	−0.0109	−0.0345	−0.0157
1984	0.2118	42	0.99	1,6908	−0.0241	−0.0407	−0.0345	−0.0583
1985	−0.0567	13	0.30	−0.7011	−0.0241	0.0169	−0.0345	0.0242
1986	0.0263	28	0.65	0.5361	−0.0241	−0.0129	−0.0345	−0.0185
1987	−0.0291	19	0.44	−0.2062	−0.0241	0.0050	−0.0345	0.0071
1988	−0.0902	7	0.15	−1.1959	−0.0241	0.0288	−0.0345	0.0413
1989	0.1346	41	0.96	1.6083	−0.0241	−0.0388	−0.0345	−0.0555
1990	−0.0071	23	0.54	0.1237	−0.0241	−0.0030	−0.0345	−0.0043
1991	0.0579	38	0.89	1.3609	−0.0241	−0.0328	−0.0345	−0.0470
1992	−0.0828	9	0.20	−1.0310	−0.0241	0.0248	−0.0345	0.0356
1993	−0.1089	5	0.11	−1.3609	−0.0241	0.0328	−0.0345	0.0470
1994	−0.0221	20	0.46	−0.1237	−0.0241	0.0030	−0.0345	0.0043
1995	0.0314	29	0.68	0.6186	−0.0241	−0.0149	−0.0345	−0.0213
1996	−0.0785	10	0.23	−0.9485	−0.0241	0.0229	−0.0345	0.0327
1997	−0.0410	16	0.37	−0.4536	−0.0241	0.0109	−0.0345	0.0157
1998	0.1021	39	0.92	1,4434	−0.0241	−0.0348	−0.0345	−0.0498
1999	−0.0372	17	0.39	−0.3712	−0.0241	0.0089	−0.0345	0.0128
2000	−0.1105	4	0.08	−1.4434	−0.0241	0.0348	−0.0345	0.0498
				Total	×	−1.2939	×	−0.2988

Source: Own calculations.

Statistics $L(m, b_1)$ and $T(1, m)$ for $k = 1$ and $m = 1, 2, \ldots, N - 1$ can be calculated in a similar way.

Based on the definition of (3.2.6), we shall consider statistics $T(1,m)$ for m satisfying double inequality $[eN] \leq m \leq [(1-e)N]$, where $e = 0.1$. For $N = 42$, we have $5 \leq m \leq 37$.

The values of $T(1, m)$, $m = 5, 6, \ldots, 37$ are contained in Table 3.3.

Table 3.3. Values of $L(m, b_1)$, $T(1, m)$, $m = 5, \ldots, 37$

m	$L(m, b_1)$	$T(1, m)$
5	−0.1375	0.0189
6	−0.8001	0.6402
7	−0.8708	0.7583
8	−0.4537	0.2059
9	−0.8529	0.7275
10	−0.8068	0.6509
11	−0.3908	0.1527
12	−0.5634	0.3175
13	−0.8396	0.7050
14	−1.0799	1.1662
15	−0.6773	0.4587
16	−0.5766	0.3324
17	−0.4537	0.2059
18	−0.3858	0.1488
19	−0.1662	0.0276
20	0.1529	0.0234
21	−0.3436	0.1181
22	0.1274	0.0162
23	0.5242	0.2748
24	0.0000	0.0000
25	0.1426	0.0203
26	0.6814	0.4643
27	0.4648	0.2160
28	0.6479	0.4198
29	0.5919	0.3503
30	0.1972	0.0389
31	0.7671	0.5884
32	0.8367	0.7000
33	1.3802	1.9050
34	1.0371	1.0756
35	0.5293	0.2802
36	0.5092	0.2593
37	0.8449	0.7139

Source: Own calculations.

Statistics $T(2,m)$ for $m = 5, 6, \ldots, 37$ are defined as $T(2,m) = L^2(m, b_1) + L^2(m, b_2)$, where $L(m, b_n)$ is given in (3.2.9) and polynomials b_1, b_2 are as follows

$$b_1(z_t) = \sqrt{3}\,(2z_t - 1), \quad b_2(z_t) = \sqrt{5}\left(\frac{3}{2}(2z_t - 1)^2 - \frac{1}{2}\right), \qquad (3.2.23)$$

where $z_t = \frac{R_t - 0.5}{N}$, $t = 1, 2, \ldots, N$.

Table 3.4. Values of $L(m, b_1)$, $L(m, b_2)$, $T(2,m)$, $m = 5, \ldots, 37$

m	$L(m,b_1)$	$L(m,b_2)$	$T(2,m)$
5	−0.1375	−0.8601	0.7586
6	−0.8001	−0.1409	0.6601
7	−0.8708	−0.5563	1.0678
8	−0.4537	−0.5718	0.5329
9	−0.8529	−0.4462	0.9266
10	−0.8068	−0.8340	1.3464
11	−0.3908	−0.7136	0.6619
12	−0.5634	−0.9664	1.2514
13	−0.8396	−1.0374	1.7813
14	−1.0799	−1.1585	2.5082
15	−0.6773	−0.9846	1.4282
16	−0.5766	−1.2969	2.0145
17	−0.4537	−1.5860	2.7212
18	−0.3858	−1.9067	3.7842
19	−0.1662	−2.0718	4.3199
20	0.1529	−2.0428	4.1963
21	−0.3436	−1.4928	2.3465
22	0.1274	−1.0355	1.0885
23	0.5242	−0.8190	0.9455
24	0.0000	−0.1755	0.0308
25	0.1426	−0.4559	0.2282
26	0.6814	0.1998	0.5042
27	0.4648	0.0196	0.2164
28	0.6479	−0.2407	0.4777
29	0.5919	−0.6025	0.7133
30	0.1972	−0.4520	0.2432
31	0.7671	0.1584	0.6135
32	0.8367	−0.2351	0.7553
33	1.3802	0.1144	1.9181
34	1.0371	0.1474	1.0974
35	0.5293	0.5500	0.5827
36	0.5092	0.1006	0.2694
37	0.8449	−0.2199	0.7622

Source: Own calculations.

The values of $L(m, b_1), L(m, b_2), T(2, m)$ are presented in Table 3.4.

Table 3.5. Values of statistics $T(k, m)$, $k = 3, \ldots, 10$, $m = 5, \ldots, 37$

m	$T(k, m)$							
	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
5	0.9267	1.7483	2.0939	2.0939	2.9248	2.9277	5.8176	5.8231
6	1.3520	2.1503	2.1509	2.1509	4.5209	5.0748	6.1496	6.2405
7	1.3381	1.4499	1.4862	1.4862	5.0686	5.4595	7.5431	7.5773
8	1.3803	1.7862	1.8306	1.8306	6.8936	7.1247	10.038	10.4501
9	1.2658	2.6588	2.7375	2.7375	5.7573	5.7573	9.0131	9.7243
10	1.7007	2.3980	2.4151	2.4151	4.9796	5.1454	7.8567	9.2966
11	1.3934	2.5297	2.5325	2.5325	6.2929	6.9749	9.0355	10.9292
12	1.4816	2.3959	2.4137	2.4137	7.0958	7.3029	9.4206	10.3077
13	1.7881	3.3140	3.5862	3.5862	6.9205	7.0415	8.1072	9.5486
14	2.6020	4.7308	5.5466	5.5466	8.2464	8.2545	8.6556	10.2335
15	1.4426	5.2268	5.8376	5.8376	9.2229	9.4829	9.5565	10.9032
16	2.0212	4.4868	4.8554	4.8554	7.0228	7.3360	7.3447	8.4891
17	2.8288	4.4745	4.7713	4.7713	5.9781	6.1925	6.3603	7.0330
18	4.0043	4.9293	5.0142	5.0142	5.6768	6.0986	6.6259	7.4652
19	5.0063	6.2539	6.6442	6.6442	7.2939	7.3928	7.5989	8.2147
20	5.4609	7.5609	8.2483	8.2483	9.6650	9.9934	10.1295	11.4057
21	4.9719	6.1165	6.7412	6.7412	8.8118	8.8353	9.6520	12.5092
22	2.7919	3.7694	3.9520	3.9520	4.8603	4.9715	5.2511	8.6941
23	2.8865	4.9621	4.9864	4.9864	5.9325	5.9334	6.8463	9.1806
24	4.4906	6.4617	7.0780	7.0780	7.2847	7.4196	7.9331	10.0277
25	6.0775	7.1514	7.8739	7.8739	7.8906	7.9120	8.7812	10.0214
26	3.4641	3.5774	5.8252	5.8252	6.2664	6.5731	7.9923	9.0625
27	2.0868	2.3624	5.8498	5.8498	6.3075	6.3546	8.5621	9.4037
28	3.4637	3.5232	7.8305	7.8305	8.0092	8.0221	10.4742	10.7785
29	3.2992	3.7175	7.1394	7.1394	7.6665	7.6742	9.2476	9.6996
30	2.3554	2.3680	5.5344	5.5344	5.7905	6.0790	6.9078	7.2776
31	1.4747	1.8014	4.9894	4.9894	5.0449	5.0506	5.1573	6.7546
32	1.8784	2.8958	5.0879	5.0879	5.0891	5.2639	5.6174	8.1620
33	3.0548	3.5118	4.5960	4.5960	4.6490	4.9207	5.9439	7.1763
34	1.6361	1.6869	3.5109	3.5109	4.0481	4.8135	6.1667	8.7338
35	1.2399	1.6408	2.4105	2.4105	2.6407	3.6767	4.3030	5.4855
36	0.8162	1.1686	1.4118	1.4118	1.4774	3.7230	4.0036	6.2890
37	2.5332	3.6750	4.5559	4.5559	4.7973	5.9429	6.0956	7.5271

Source: Own calculations.

Proceeding in the same way for $k = 3, 4, \ldots$ and $m = 5, 6, \ldots, 37$, we obtain the values of statistics $T(k, m)$ (table 3.5). In the next step, statistics $S(m, p_N)$ given by formula (3.2.7) will be found. For a fixed m, $S(m, p_N)$ corresponds to the smallest value k^* of index k ($k = 1, 2, \ldots, d_N$) for which difference $T(k, m) - kp_N$ is the largest.

Parameter p_N is found from formula

$$p_N = 1.5\ln(N) = 1.5\ln(42) \approx 5.61, \qquad (3.2.24)$$

where 10 represents the value of d_N according to suggestion given by [Janic-Wróblewska, Ledwina 2000].

Table 3.6. Values of statistics $T(k,m) - kp_N$, $k^* = S(m, p_N)$

m	$T(k,m) - kp_N$										k^*
	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$	
5	–5.6	–10.5	–15.9	–20.9	–26.3	–31.5	–36.3	–41.9	–44.6	–50.2	1
6	–5.0	–10.6	–15.5	–20.8	–25.9	–31.5	–34.7	–39.8	–44.3	–49.8	1
7	–4.8	–10.1	–15.5	–21.1	–26.6	–32.2	–34.2	–39.4	–42.9	–48.5	1
8	–5.4	–10.7	–15.4	–20.7	–26.2	–31.8	–32.4	–37.7	–40.4	–45.6	1
9	–4.9	–10.3	–15.6	–20.1	–25.4	–30.9	–33.5	–39.1	–41.4	–46.3	1
10	–5.0	–9.9	–15.1	–20.4	–25.6	–31.2	–34.3	–39.7	–42.6	–46.8	1
11	–5.5	–10.6	–15.4	–20.0	–25.5	–31.1	–33.0	–37.9	–41.4	–45.1	1
12	–5.3	–10.0	–15.3	–20.0	–25.6	–31.2	–32.1	–37.5	–41.0	–45.8	1
13	–4.9	–9.4	–15.0	–19.2	–24.7	–30.1	–32.3	–37.8	–42.4	–46.5	1
14	–4.4	–8.7	–14.2	–17.9	–23.3	–28.1	–31.0	–36.6	–41.8	–45.8	1
15	–5.1	–9.8	–15.4	–17.8	–22.8	–27.8	–30.0	–35.4	–40.9	–45.2	1
16	–5.3	–9.2	–14.8	–18.2	–23.5	–28.8	–32.2	–37.5	–43.1	–47.6	1
17	–5.4	–8.5	–14.0	–18.0	–23.6	–28.9	–33.3	–38.7	–44.1	–49.0	1
18	–5.5	–7.4	–12.8	–17.5	–23.1	–28.6	–33.6	–38.8	–43.8	–48.6	1
19	–5.6	–6.9	–11.8	–16.3	–21.8	–27.0	–32.0	–37.5	–42.9	–47.9	1
20	–5.6	–7.0	–11.4	–14.9	–20.5	–25.4	–29.6	–34.9	–40.3	–44.7	1
21	–5.5	–8.9	–11.8	–16.3	–21.9	–26.9	–30.4	–36.0	–40.8	–43.6	1
22	–5.6	–10.1	–14.0	–18.7	–24.3	–29.7	–34.4	–39.9	–45.2	–47.4	1
23	–5.3	–10.3	–13.9	–17.8	–23.1	–28.7	–33.3	–38.9	–43.6	–46.9	1
24	–5.6	–11.2	–12.3	–17.6	–21.6	–26.6	–32.0	–37.4	–42.5	–46.0	1
25	–5.6	–11.0	–10.7	–16.1	–20.9	–25.8	–31.4	–36.9	–41.7	–46.0	1
26	–5.1	–10.7	–13.4	–18.9	–24.5	–27.8	–33.0	–38.3	–42.5	–47.0	1
27	–5.4	–11.0	–14.7	–20.3	–25.7	–27.8	–32.9	–38.5	–41.9	–46.7	1
28	–5.2	–10.7	–13.4	–18.9	–24.5	–25.8	–31.2	–36.8	–40.0	–45.3	1
29	–5.3	–10.5	–13.5	–18.8	–24.3	–26.5	–31.6	–37.2	–41.2	–46.4	1
30	–5.6	–11.0	–14.5	–20.1	–25.7	–28.1	–33.5	–38.8	–43.6	–48.8	1
31	–5.0	–10.6	–15.3	–20.7	–26.2	–28.6	–34.2	–39.8	–45.3	–49.3	1
32	–4.9	–10.5	–14.9	–19.7	–25.1	–28.6	–34.2	–39.6	–44.8	–47.9	1
33	–3.7	–9.3	–13.8	–18.9	–24.5	–29.0	–34.6	–39.9	–44.5	–48.9	1
34	–4.5	–10.1	–15.2	–20.7	–26.3	–30.1	–35.2	–40.0	–44.3	–47.3	1
35	–5.3	–10.6	–15.6	–21.2	–26.4	–31.2	–36.6	–41.2	–46.2	–50.6	1
36	–5.3	–10.9	–16.0	–21.5	–26.9	–32.2	–37.8	–41.1	–46.5	–49.8	1
37	–4.9	–10.5	–14.3	–19.8	–24.4	–29.1	–34.4	–38.9	–44.4	–48.5	1

Source: Own calculations.

Table 3.5 shows the values of $T(k,m)$ for $k = 3,4,\dots,d_N$, while Table 3.6 contains differences $T(k,m)-kp_N$ and the values of statistics $S(m,p_N)$ for $k = 1,2,\dots,d_N$ and $m = 5,6,\dots,37$.

Since for each m we have $k^* = 1$ (table 3.6), we shall calculate statistics M_N taking account of the terms corresponding to $k=1$ and representing the values of statistics $T(1,m)$ (the last column in Table 3.3). Thereafter, we shall review the values of $T(1,m)$ to find its greatest value. It is $T(1,33)$, meaning that the value of the JL test statistics is $M_N=1.905$ and the switching point $m = 33$ that corresponds to it occurs in the year 1991.

Comparing the statistics obtained in this example with the critical value of the JL test that for $k = 1$ equals the critical value of the Wilcoxon test we find that the switching point is statistically significant.

The calculations presented in Example 3.2 refer to switching points' identification for Polish women aged $x=40$ years. The JL test results obtained for all one-year age groups of males and females in Poland are provided in Chapter 6 (see Table 6.1).

The concept of switching time points is incorporated in the structure of hybrid mortality models proposed in the remainder of this chapter.

3.3. The dynamic Lee–Carter hybrid model

3.3.1. Dynamic LCH model

The family of the subsystems comprising the Lee–Carter Hybrid model (LCH model) is defined by analogy to (1.6.1)–(1.6.2), i.e.

$$d\mu_x(t,l)=\left[\gamma_x(t,l)+\frac{1}{2}\sigma_x^2(l)\right]\mu_x(t,l)dt+\sigma_x(l)\mu_x(t,l)dw(t),\ t\in\mathbb{R}^+, \quad (3.3.1)$$

$$\gamma_x(t,l)=\beta_x(l)\kappa'(t,l), \quad \mu_x(t_0,l)=\exp\{\alpha_x(l)+\beta_x(l)\kappa(t_0,l)\}, \quad (3.3.2)$$

where

$l \in \mathbb{S}$ – state of the hybrid system,

$\alpha_x(l),\beta_x(l)$ – age-related scalar coefficients,

$\kappa(t,l)$ – scalar differentiable and deterministic functions of time t,

$\kappa(t_0, l)$ – initial constant parameters,
$\sigma_x^2(l) > 0$ – age-specific volatility parameters,
$w(t)$ – standard Wiener process.

Let us assume that $\kappa(t, l)$ are linear functions of time t and can be different for different subsystems, i.e.

$$\kappa(t, l) = \chi(l) + \delta(l)t, \quad l \in \mathbb{S}, \tag{3.3.3}$$

such that

$$\int_{t_0}^{\tau_1} \kappa(t, l_1)dt + \int_{\tau_1}^{\tau_2} \kappa(t, l_2)dt + \ldots + \int_{\tau_M}^{T} \kappa(t, l_M)dt = 0 \tag{3.3.4}$$

and

$$\begin{aligned} &\kappa(\tau_1, l_0) = \kappa(\tau_1, l_1), \\ &\kappa(\tau_2, l_1) = \kappa(\tau_2, l_2), \\ &\ldots\ldots\ldots\ldots \\ &\kappa(\tau_M, l_{M-1}) = \kappa(\tau_M, l_M). \end{aligned} \tag{3.3.5}$$

where $I_0 = [t_0, \tau_1)$, $I_1 = [\tau_1, \tau_2), \ldots, I_M = [\tau_M, T]$ are time intervals (mortality regimes) and $\tau_1, \tau_2, \ldots, \tau_M$ are switching points.

Successive subsystems of the LCH model correspond to different mortality regimes $I_0, I_1, \ldots, I_M$. In the more complex approach, functions $\kappa(t, l)$ for separate subsystems can be represented by more sophisticated formulas then (3.3.3).

Let the following additional constraint be also imposed

$$\sum_{x=0}^{X} \beta_x(l) = 1, \quad l \in \mathbb{S}. \tag{3.3.6}$$

By applying the Itô formula (see (A.2.21), Theorem A.7 in Appendix A), equation (3.3.1) has the following solution

$$\ln \mu_x(t, l) = \alpha_x(l) + \beta_x(l)\kappa(t, l) + \sigma_x(l)w(t), \tag{3.3.7}$$

or, equivalently, in its exponential form

$$\mu_x(t, l) = \exp\left\{\alpha_x(l) + \beta_x(l)\kappa(t, l) + \sigma_x(l)w(t)\right\}. \tag{3.3.8}$$

The solutions of the subsystems will be used to create a solution for the hybrid model.

We assume the continuity of solutions of stochastic differential equations, it means that the value of the process in state l_i at time τ_i, i.e. $\mu_x(\tau_i, l_{i-1})$, and the value of the process in state l_{i-1} at switching time τ_i, i.e. $\mu_x(\tau_i, l_{i-1})$, are equal. Then the solution is of the following form

$$\mu_x(t, l_i) = \mu_x(\tau_i, l_{i-1}) \exp\{\beta_x(l_i)(\kappa(t, l_i) - \kappa(\tau_i, l_i))+ \\ + \sigma_x(l_i)(w(t) - w(\tau_i))\} \quad \text{for} \quad t \in [\tau_i, \tau_{i+1}). \tag{3.3.9}$$

3.3.2. Discrete LCH model

The Discrete Lee–Carter Hybrid model(DLCH model) can be derived from equation (3.3.7) by subtracting $\ln \mu_x(t-1, l)$ from $\ln \mu_x(t, l)$

$$\ln \mu_x(t, l) = \ln \mu_x(t-1, l) + \beta_x(l)[\kappa(t, l) - \kappa(t-1, l)]+ \\ + \sigma_x(l)\epsilon_{x,t}(l), \quad t \in \mathbb{N}, \tag{3.3.10}$$

or, equivalently

$$\ln \mu_x(t,l) = \ln \mu_x(t-1,l) + \beta_x(l)[\kappa(t,l) - \kappa(t-1,l)] + \xi_{x,t}(l), \;\; t \in \mathbb{N}, \tag{3.3.11}$$

where $\xi_{x,t}(l) = \sigma_x(l)\epsilon_{x,t}(l)$ are Gaussian random variables with means $\mathrm{E}[\xi_{x,t}(l)]$ and variances $\mathrm{Var}[\xi_{x,t}(l)]$ equal, respectively,

$$\mathrm{E}[\xi_{x,t}(l)] = 0, \quad \mathrm{Var}[\xi_{x,t}(l)] = \mathrm{E}[\xi^2_{x,t}(l)] = \sigma^2_x(l). \tag{3.3.12}$$

Coefficients $\alpha_x(l), \beta_x(l), \chi(l), \delta(l), \sigma^2_x(l)$ for $l \in \mathbb{S}$ constitute a set of unknown parameters. Parameters $\beta_x(l)$ and functions $\kappa(t, l)$ satisfy constraints (3.3.3)–(3.3.6).

3.3.3. Parameters' estimation of the dynamic LCH model

Let us consider the discrete LCH model as given in (3.3.11) and assume that there exist M switching time points $\tau_1, \tau_2, \ldots, \tau_M$ within the observation period $[t_0, T]$.

Parameters $\delta(l), \beta_x(l)$ and $\sigma^2_x(l)$ can be estimated by analogy to (1.6.19)–(1.6.21). Thus, the respective estimators are as follows

$$d(l) = \sum_{x=0}^{X} \bar{v}_x(l), \tag{3.3.13}$$

$$b_x(l) = \frac{\bar{v}_x(l)}{d(l)} = \frac{\bar{v}_x(l)}{\sum_{x=0}^{X} \bar{v}_x(l)} \tag{3.3.14}$$

and

$$s_x^2(l) = \overline{(v_{x,t}(l) - \bar{v}_x(l))^2}, \tag{3.3.15}$$

where

$v_{x,t}(l) = \ln m_{x,t+1} - \ln m_{x,t}$ for $t \in I_l$,

$\ln m_{x,t},\ t = t_0, t_0 + 1, \ldots, T$ are log-central death rates in a sample time series,

$\bar{v}_x(l)$, $\overline{(v_{x,t}(l) - \bar{v}_x(l))^2}$ are arithmetic averages of the respective expressions calculated for each mortality regime I_l.

Parameters $\chi(l)$ defined via (3.3.3) satisfy constraints (3.3.3)–(3.3.5). Their estimators $c(l)$ can be obtained in an analogous way as provided in Section 1.6.3. As a result, estimators $k(t,l)$ of $\kappa(t,l)$ take the form

$$k(t,l) = c(l) + d(l)t. \tag{3.3.16}$$

Remaining parameters $\alpha_x(l)$ can be estimated for each sub-period I_l using formula (1.6.26) with t_0, t_1 replaced by successive switching points τ_i, τ_{i+1}.

3.4. The Vasiček and Cox–Ingersoll–Ross hybrid models

In this section we propose some extensions of the Vasiček and Cox–Ingersoll–Ross models discussed in Section 1.7, termed, respectively, the hybrid Vasiček and hybrid Cox–Ingersoll–Ross models.

3.4.1. VH and CIRH models

The family of subsystems representing the Vasiček Hybrid model (VH model) is of the form

$$d\mu_x(t,l) = \kappa_x(l)\left[\theta_x(l) - \mu_x(t,l)\right] dt + \sigma_x(l) dw(t),\ \ t \in \mathbb{R}^+, \tag{3.4.1}$$

and the family of subsystems representing the Cox–Ingersoll–Ross Hybrid model (CIRH model) is as follows

$$d\mu_x(t,l) = \kappa_x(l)[\theta_x(l) - \mu_x(t,l)]dt + \sigma_x(l)\sqrt{\mu_x(t,l)}dw(t),\ t \in \mathbb{R}^+, \tag{3.4.2}$$

where $\sigma_x(l), \theta_x(l), \kappa_x(l) > 0$ for $l \in \mathbb{S}$ are constant parameters and $w(t)$ is a standard Wiener process.

For the Vasiček model one can find the exact analytical solution using the Itô formula (see (A.2.21), Theorem A.7 in Appendix A)

$$\mu_x(t,l) = \mu_x(t_0,l)e^{-\kappa_x(l)(t-t_0)} + \theta_x(l)\left(1 - e^{-\kappa_x(l)(t-t_0)}\right) + \\ + \sigma_x(l)e^{-\kappa_x(l)t}\int_{t_0}^{t} e^{\kappa_x(l)s}dw(s). \tag{3.4.3}$$

We assume that the scalar hybrid differential model starting at moment t_0 from some initial state μ_{x0} switches at time points $\tau_1, \ldots, \tau_M$. We also assume that $\tau_0 = t_0$ and that in time intervals $[\tau_i, \tau_{i+1})$ the hybrid system stays in state $l_i \in \mathbb{S}$, $i = 0, 1, \ldots, M$, where $l_0, l_1, \ldots, l_M$ is any subsequence of N states.

The continuity of solutions is also assumed, meaning that the initial value of the process $\mu_x(t,l)$ in state l_i at time τ_i is equal to the value of the process in state l_{i-1} at time τ_i respectively, i.e. $\mu_x(\tau_i, l_i) = \mu_x(\tau_i, l_{i-1})$. Then we have

$$\mu_x(t,l_i) = \mu_x(\tau_i,l_i)e^{-\kappa_x(l)(t-\tau_i)} + \theta_x(l)\left(1 - e^{-\kappa_x(l)(t-\tau_i)}\right) + \\ + \sigma_x(l)e^{-\kappa_x(l)t}\int_{\tau_i}^{t} e^{\kappa_x(l)s}dw(s), \quad t \in [\tau_i, \tau_{i+1}). \tag{3.4.4}$$

3.4.2. VH and CIRH moment models

Using the Itô formula one can find the first and second moment equations for the Vasiček Hybrid Moment model (VHM model). The equations are as follows

$$\frac{d\mathrm{E}[\mu_x(t,l)]}{dt} = -\kappa_x(l)\mathrm{E}[\mu_x(t,l)] + \kappa_x(l)\theta_x(l), \\ \frac{d\mathrm{E}[\mu_x^2(t,l)]}{dt} = -2\kappa_x(l)\mathrm{E}[\mu_x^2(t,l)] + 2\kappa_x(l)\theta_x(l)\mathrm{E}[\mu_x(t,l)] + \sigma_x^2(l) \tag{3.4.5}$$

and the stationary solutions are

$$\mathrm{E}[\mu_x(t,l)] = \theta_x(l), \quad \mathrm{E}[\mu_x^2(t,l)] = \theta_x^2(l) + \frac{\sigma_x^2(l)}{2\kappa_x(l)}. \tag{3.4.6}$$

Similarly, the moment equations for the Cox–Ingersoll–Ross Hybrid Moment model (CIRHM model) are as follows

$$\frac{d\mathrm{E}[\mu_x(t,t)]}{dt} = -\kappa_x(l)\mathrm{E}[\mu_x(t,l)] + \kappa_x(l)\theta_x(l),$$

$$\frac{d\mathrm{E}[\mu_x^2(t,l)]}{dt} = -2\kappa_x(l)\mathrm{E}[\mu_x^2(t,l)] + 2\kappa_x(l)\theta_x(l)\mathrm{E}[\mu_x(t,l)] + \tag{3.4.7}$$

$$+\sigma_x^2(l)\mathrm{E}[\mu_x(t,l)]$$

and the stationary solutions are

$$\mathrm{E}[\mu_x(t,l)] = \theta_x(l), \quad \mathrm{E}[\mu_x^2(t,l)] = \theta_x^2(l) + \frac{\sigma_x^2(l)\theta_x(l)}{2\kappa_x(l)}. \tag{3.4.8}$$

The continuity of the moment of the solutions of the stochastic differential equations can be assumed only in the case of non-stationary solutions, meaning that the initial value of the process $\mathrm{E}[\mu_x(t)]$ in state l_i at time τ_i, i.e. $\mathrm{E}[\mu_x(\tau_i, l_i)]$, is equal to the value of the process in state l_{i-1} at time τ_i, i.e. $\mathrm{E}[\mu_x(\tau_i, l_i)] = \mathrm{E}[\mu_x(\tau_i, l_{i-1})]$.

3.4.3. Discrete VH and CIRH models

The discrete-time approximation of the model (3.4.1) results in the Discrete Vasiček Hybrid model (DVH model) defined as

$$\mu_x(t+1,l) = \kappa_x(l)\theta_x(l) + (1-\kappa_x(l))\mu_x(t,l) + \tag{3.4.9}$$

$$+\sigma_x(l)\epsilon_{x,t+1}(l), \quad t \in \mathbb{N},$$

where $\xi_{x,t+1}(l) = \sigma_x(l)\epsilon_{x,t+1}(l)$ on the right-hand side of (3.4.9) are Gaussian random variables with means and variances equal, respectively,

$$\mathrm{E}[\xi_{x,t+1}(l)] = 0, \quad \mathrm{Var}[\xi_{x,t+1}(l)] = \mathrm{E}[\xi_{x,t+1}^2(l)] = \sigma_x^2(l). \tag{3.4.10}$$

By analogy, the discrete-time approximation of (3.4.2) leads to the Discrete Cox–Ingersoll–Ross Hybrid model (DCIRH model)

$$\mu_x(t+1,l) = \kappa_x(l)\theta_x(l) + (1-\kappa_x(l))\mu_x(t,l) + \tag{3.4.11}$$

$$+\sigma_x(l)\sqrt{\mu_x(t,l)}\epsilon_{x,t+1}(l), \quad t \in \mathbb{N},$$

where $\xi_{x,t+1}(l) = \sigma_x(l)\sqrt{\mu_x(t,l)}\epsilon_{x,t+1}(l)$ are random variables with means and conditional variances equal, respectively,

$$\mathrm{E}[\xi_{x,t+1}(l)] = 0, \quad \mathrm{Var}\left[\xi_{x,t+1}(l)|\mu_x(t,l)\right] = \sigma_x^2(l)\mu_x(t,l). \tag{3.4.12}$$

Note that in the case of the DVH model we also have

$$\begin{aligned}\xi_{x,t+1} &= \sigma_x(l)\epsilon_{x,t+1}(l) = \\ &= \mu_x(t+1,l) - \mu_x(t,l) - \kappa_x(l)\left(\theta_x(l) - \mu_x(t,l)\right),\end{aligned} \tag{3.4.13}$$

while for the DCIRH model there is

$$\begin{aligned}\xi_{x,t+1} &= \sigma_x(l)\sqrt{\mu_x(t,l)}\epsilon_{x,t+1}(l) = \\ &= \mu_x(t+1,l) - \mu_x(t,l) - \kappa_x(l)\left(\theta_x(l) - \mu_x(t,l)\right).\end{aligned} \tag{3.4.14}$$

Both expressions will be used in the estimation procedure described in Section 3.4.9.

3.4.4. Discrete VH and CIRH moment models

Using the moment equations (3.4.5) we find the Discrete Vasiček Hybrid Moment model (DVHM model)

$$\mathrm{E}[\mu_x]_{i+1}(l) = \mathrm{E}[\mu_x]_i(l) - (\kappa_x(l)\mathrm{E}[\mu_x]_i(l) - \kappa_x(l)\theta_x(l))\delta, \tag{3.4.15}$$

$$\begin{aligned}\mathrm{E}[\mu_x^2]_{i+1}(l) = \mathrm{E}[\mu_x^2]_i(l) - (2\kappa_x(l)\mathrm{E}[\mu_x^2]_i(l) + \\ - 2\kappa_x(l)\theta_x(l)\mathrm{E}[\mu_x]_i(l) - \sigma_x^2(l))\delta,\end{aligned} \tag{3.4.16}$$

Similarly, using the moment equations (3.4.7) we find the Discrete Cox–Ingersoll–Ross Hybrid Moment model (DCIRHM model)

$$\mathrm{E}[\mu_x]_{i+1}(l) = \mathrm{E}[\mu_x]_i(l) - (\kappa_x(l)\mathrm{E}[\mu_x]_i(l) - \kappa_x(l)\theta_x(l))\delta, \tag{3.4.17}$$

$$\begin{aligned}\mathrm{E}[\mu_x^2]_{i+1}(l) = \mathrm{E}[\mu_x^2]_i(l) - (2\kappa_x(l)\mathrm{E}[\mu_x^2]_i(l) + \\ - 2\kappa_x(l)\theta_x(l)\mathrm{E}[\mu_x]_i(l) - \sigma_x^2(l)\mathrm{E}[\mu_x]_i(l))\delta.\end{aligned} \tag{3.4.18}$$

3.4.5. Modified VH and CIRH models

The Modified Vasiček Hybrid model (MVH model) takes the form of the scalar stochastic Itô equation

$$d\mu_x(t,l) = \kappa_x(l)\left[\theta_x(l) - \mu_x(t,l)\right]dt + f_x(t,l)dw(t), \quad t \in \mathbb{R}^+, \quad (3.4.19)$$

whereas the Modified Cox–Ingersoll–Ross Hybrid model (MCIRH model) can be written as

$$\begin{aligned} d\mu_x(t,l) = \kappa_x(l)\left[\theta_x(l) - \mu_x(t,l)\right]dt + \\ f_x(t,l)\sqrt{\mu_x(t,l)}dw(t), \quad t \in \mathbb{R}^+, \end{aligned} \quad (3.4.20)$$

where $\theta_x(l), \kappa_x(l) > 0$ for $l \in \mathbb{S}$ are constant parameters, $f_x(t,l) > 0$ are time-depending diffusion functions and $w(t)$ is a standard Wiener process.

Similarly as in (1.9.9), we will assume in further considerations that the diffusion functions have the following parametric form

$$f_x(t,l) = e^{\zeta_x(l)t}, \quad \zeta_x(l) \in \mathbb{R}. \quad (3.4.21)$$

To find solution of (3.4.19), let us consider the following expression

$$K(t,\mu_x(t,l)) = \mu_x(t,l)e^{\kappa_x(l)t}. \quad (3.4.22)$$

The Itô formula applied to (3.4.22) leads to the following solution

$$\begin{aligned} \mu_x(t,l) = \mu_x(t_0,l)e^{-\kappa_x(l)(t-t_0)} + \theta_x(l)\left(1 - e^{-\kappa_x(l)(t-t_0)}\right) + \\ + e^{-\kappa_x(l)t}\int_{t_0}^{t} e^{(\zeta_x(l)+\kappa_x(l))s}dw(s). \end{aligned} \quad (3.4.23)$$

In the case of the MCIRH model (3.4.20) the following equation is received by means of the Itô formula

$$\begin{aligned} \mu_x(t,l) = \mu_x(t_0,l)e^{-\kappa_x(l)(t-t_0)} + \theta_x(l)\left(1 - e^{-\kappa_x(l)(t-t_0)}\right) + \\ + e^{-\kappa_x(l)t}\int_{t_0}^{t} \sqrt{\mu_x(s,l)}e^{(\zeta_x(l)+\kappa_x(l))s}dw(s). \end{aligned} \quad (3.4.24)$$

3.4.6. Modified VH and CIRH moment models

Using the Itô formula one can find the first and second moment equations of the Modified Vasiček Hybrid Moment model (MVHM model), i.e.

$$\begin{aligned}\frac{d\mathrm{E}[\mu_x(t,l)]}{dt} =& -\kappa_x(l)\mathrm{E}[\mu_x(t,l)] + \kappa_x(l)\theta_x(l),\\ \frac{d\mathrm{E}[\mu_x^2(t,l)]}{dt} =& -2\kappa_x(l)\mathrm{E}[\mu_x^2(t,l)] + 2\kappa_x(l)\theta_x(l)\mathrm{E}[\mu_x(t,l)]+\\ &+ f_x^2(t,l).\end{aligned} \tag{3.4.25}$$

The moment equations for the modified Cox–Ingersoll–Ross Hybrid Moment model (MCIRHM model) take the similar form

$$\begin{aligned}\frac{d\mathrm{E}[\mu_x(t,t)]}{dt} =& -\kappa_x(l)\mathrm{E}[\mu_x(t,l)] + \kappa_x(l)\theta_x(l),\\ \frac{d\mathrm{E}[\mu_x^2(t,l)]}{dt} =& -2\kappa_x(l)\mathrm{E}[\mu_x^2(t,l)] + 2\kappa_x(l)\theta_x(l)\mathrm{E}[\mu_x(t,l)]+\\ &+ f_x^2(t,l)\mathrm{E}[\mu_x(t,l)].\end{aligned} \tag{3.4.26}$$

3.4.7. Discrete modified VH and CIRH models

The discrete-time version of the MVH model (3.4.19) is represented by the following equation called the Discrete Modified Vasiček Hybrid model (DMVH model)

$$\begin{aligned}\mu_x(t+1,l) =& \kappa_x(l)\theta_x(l) + (1-\kappa_x(l))\mu_x(t,l)+\\ &+ f_x(t,l)\epsilon_{x,t+1}(l), \quad t\in\mathbb{N},\end{aligned} \tag{3.4.27}$$

where $f_x(t,l) = e^{\zeta_x(l)t}$ $(l\in\mathbb{S})$ are time-depending diffusion functions. The terms

$$\xi_{x,t+1}(l) = e^{\zeta_x(l)t}\epsilon_{x,t+1}(l) \tag{3.4.28}$$

are Gaussian random variables with means and variances equal, respectively,

$$\mathrm{E}[\xi_{x,t+1}(l)] = 0,$$
$$\mathrm{Var}[\xi_{x,t+1}(l)] = e^{2\zeta_x(l)t}. \tag{3.4.29}$$

By analogy, the Discrete Modified Cox–Ingersoll–Ross Hybrid model (DMCIRH model) based on (3.4.20) takes the form

$$\mu_x(t+1,l) = \kappa_x(l)\theta_x(l) + (1-\kappa_x(l))\mu_x(t,l)$$
$$+ f_x(t,l)\sqrt{\mu_x(t,l)}\epsilon_{x,t+1}(l), \quad t \in \mathbb{N}. \tag{3.4.30}$$

3.4.8. Discrete modified VH and CIRH moment models

Using the moment equations (3.4.25) we find also the Discrete Modified Vasiček Hybrid Moment model (DMVHM model)

$$\mathrm{E}[\mu_x]_{i+1}(l) = \mathrm{E}[\mu_x]_i(l) - (\kappa_x(l)\mathrm{E}[\mu_x]_i(l) - \kappa_x(l)\theta_x(l))\delta, \tag{3.4.31}$$

$$\mathrm{E}[\mu_x^2]_{i+1}(l) = \mathrm{E}[\mu_x^2]_i(l) - (2\kappa_x(l)\mathrm{E}[\mu_x^2]_i(l) +$$
$$- 2\kappa_x(l)\theta_x(l)\mathrm{E}[\mu_x]_i(l) - f_{xi}^2(l))\delta, \tag{3.4.32}$$

where $f_{xi}^2(l) = e^{2\zeta(l)i}$.

Similarly, from the moment equations (3.4.26) we find the Discrete Modified Cox–Ingersoll–Ross Hybrid Moment model (DMCIRHM model) expressed by equations

$$\mathrm{E}[\mu_x]_{i+1}(l) = \mathrm{E}[\mu_x]_i(l) - (\kappa_x(l)\mathrm{E}[\mu_x]_i(l) - \kappa_x(l)\theta_x(l))\delta, \tag{3.4.33}$$

$$\mathrm{E}[\mu_x^2]_{i+1}(l) = \mathrm{E}[\mu_x^2]_i(l) - (2\kappa_x(l)\mathrm{E}[\mu_x^2]_i(l) +$$
$$- 2\kappa_x(l)\theta_x(l)\mathrm{E}[\mu_x]_i(l) - f_x^2(t,l)\mathrm{E}[\mu_x]_i(l))\delta. \tag{3.4.34}$$

3.4.9. Parameters' estimation of the VH and CIRH models

In this section we present some procedures used in estimation of the Vasiček and Cox–Ingersoll–Ross hybrid models, concerning both models for realizations and models for moments.

Estimation of the DVH/DMVH and DCIRH/DMCIRH models for realizations

Let us consider the DVH and DCIRH models expressed by equations (3.4.9), (3.4.11), respectively. We apply the method of estimation termed the generalized method of moments (see Section 1.7.5).

Estimates of parameters $\kappa_x(l), \theta_x(l), \sigma_x^2(l)$ can be found for separate $l \in \mathbb{S}$ by solving the following optimization problem with respect to $\kappa_x(l), \theta_x(l), \sigma_x^2(l)$

$$\begin{aligned}&\text{minimize } S(\kappa_x(l), \theta_x(l), \sigma_x^2(l)) = \\ &\qquad = \mathbf{g}^T(\kappa_x(l), \theta_x(l), \sigma_x^2(l))\mathbf{g}(\kappa_x(l), \theta_x(l), \sigma_x^2(l)),\end{aligned} \tag{3.4.35}$$

where $\mathbf{g}(\kappa_x(l), \theta_x(l), \sigma_x^2(l))$ is defined by analogy to (1.7.31) or (1.7.32), respectively, i.e. by means of sample moments of random terms (3.4.13) in the DVH model or sample moments of random terms (3.4.14) in the DCIRH model.

Parameters $\kappa_x(l), \theta_x(l), \zeta_x(l)$ of the MDVH and MDCIRH models can be estimated in a similar way.

Iterative estimation of the DVH and DCIRH moment models

Let us consider estimation of the DVHM and DCIRHM models using the discrete-time moment equations.

The estimation procedure for the DVHM model (3.4.15)–(3.4.16) and for the DCIRHM model (3.4.17)–(3.4.18) is as follows:

1^o Take constant initial values for $\mathrm{E}[\mu_x]_0(l) = p_x(l)$, $\mathrm{E}[\mu_x^2]_0(l) = p_x^2(l)$, where $p_x(l)$ is a fixed parameter.

2^o Assume initial conditions for parameters $p_x(l)$, $\kappa_x(l)$, $\theta_x(l)$, $\sigma_x(l)$ e.g. $p_x(l) = 0.1$, $\kappa_x(l) = 0.1$, $\theta_x(l) = 0.1$, $\sigma_x(l) = 0.1$.

3^o Estimate the successive values of $\mathrm{E}[\mu_x]_i(l)$, $\mathrm{E}[\mu_x^2]_i(l)$ from equations (3.4.15)–(3.4.16) or from equations (3.4.17)–(3.4.18), for an i-th iteration ($i = 1, 2, \ldots$) and for the given values of $p_x(l)$, $\kappa_x(l)$, $\theta_x(l)$, $\sigma_x(l)$.

4^o Determine the values of $\hat{\mathrm{E}}[\mu_x]_i(l)$, $\hat{\mathrm{E}}[\mu_x^2]_i(l)$, i.e. the central mortality rates and their squares from a sample time series.

5^o Minimize the following estimation criterion

$$S_V = \sum_l \sum_i \Big((\hat{\mathrm{E}}[\mu_x]_{i+1}(l) - \mathrm{E}[\mu_x]_i(l)(1-\kappa_x(l)) - \kappa_x(l)\theta_x(l) \Big)^2 + \\ + \Big(\hat{\mathrm{E}}[\mu_x^2]_{i+1}(l) - \mathrm{E}[\mu_x^2]_i(l) + (2\kappa_x(l)\mathrm{E}[\mu_x^2]_i(l) + \\ -2\,\kappa_x(l)\theta_x(l)\mathrm{E}[\mu_x]_i(l) - \sigma_x^2(l)) \Big)^2 , \tag{3.4.36}$$

or

$$S_{CIR} = \sum_l \sum_i \Big((\hat{\mathrm{E}}[\mu_x]_{i+1}(l) - \mathrm{E}[\mu_x]_i(l)(1-\kappa_x(l)) - \kappa_x(l)\theta_x(l) \Big)^2 + \\ + \Big(\hat{\mathrm{E}}[\mu_x^2]_{i+1}(l) - \mathrm{E}[\mu_x^2]_i(l) + (2\kappa_x(l)\mathrm{E}[\mu_x^2]_i(l) + \\ -2\,\kappa_x(l)\theta_x(l)\mathrm{E}[\mu_x]_i(l) - \sigma_x^2(l)\mathrm{E}[\mu_x]_i(l)) \Big)^2 . \tag{3.4.37}$$

Estimation of the VH and CIRH moment models with stationary first order moments

Let us consider the VHM and CIRHM models using the moments' equations (3.4.6) and (3.4.8), respectively.

The Vasiček stationary hybrid moment model has the form

$$\mathrm{E}[\mu_x(t,l)] = \theta_x(l), \quad \mathrm{E}[\mu_x^2(t,l)] = \theta_x^2(l) + \frac{\sigma_x^2(l)}{2\kappa_x(l)}, \tag{3.4.38}$$

while the Cox–Ingersoll–Ross stationary hybrid moment model is as follows

$$\mathrm{E}[\mu_x(t,l)] = \theta_x(l), \quad \mathrm{E}[\mu_x^2(t,l)] = \theta_x^2(l) + \frac{\sigma_x^2(l)\theta_x(l)}{2\kappa_x(l)}. \tag{3.4.39}$$

In both models only $\theta_x(l)$ and $\frac{\sigma_x^2(l)}{2\kappa_x(l)}$ for each $l \in \mathbb{S}$ can be estimated. It means that estimates of $\sigma_x^2(l)$ and $\kappa_x(l)$ cannot be found. To this end, it is enough to determine the theoretical moments (3.4.38) and (3.4.39) and compare them with values of their respective counterparts obtained from sample central death rates $\{m_{x,t},\ t+1,2,\ldots,T\}$, separately for each mortality regime I_l. It reduces to solving the following equations

$$\hat{\mathrm{E}}[\mu_x(t,l)] = \theta_x(l), \quad \hat{\mathrm{E}}[\mu_x^2(t,l)] = \theta_x^2(l) + \frac{\sigma_x^2(l)}{2\kappa_x(l)}, \tag{3.4.40}$$

or

$$\hat{\mathrm{E}}[\mu_x(t,l)] = \theta_x(l), \quad \hat{\mathrm{E}}[\mu_x^2(t,l)] = \theta_x^2(l) + \frac{\sigma_x^2(l)\theta_x(l)}{2\kappa_x(l)}. \tag{3.4.41}$$

Iterative estimation of the DMVHM and DMCIRHM models

Let us consider the DMVHM model expressed by the discrete-time moment equations (3.4.31)–(3.4.32) and the DMCIRHM model defined by the discrete-time moment equations (3.4.33)–(3.4.34). The iterative estimation procedure for both models can be described as follows:

1^o Take constant initial values of $\mathrm{E}[\mu_x]_0(l) = p_x(l)$, $\mathrm{E}[\mu_x^2]_0(l) = p_x^2(l)$, where $p_x(l)$ is a fixed parameter.

2^o Assume initial values for parameters $p_x(l)$, $\kappa_x(l)$, $\theta_x(l)$, $\zeta_x(l)$, e.g. $p_x(l) = 0.1$, $\kappa_x(l) = 0.1$, $\theta_x(l) = 0.1$, $\zeta_x(l) = -0.1$.

3^o Estimate successive values of $\mathrm{E}[\mu_x]_i(l)$, $\mathrm{E}[\mu_x^2]_i(l)$ from expressions (3.4.31)–(3.4.32) or (3.4.33)–(3.4.34) for an i-th iteration ($i = 1, 2, \ldots$) and for the given values of $p_x(l)$, $\kappa_x(l)$, $\theta_x(l)$, $\zeta_x(l)$.

4^o Determine the values of $\hat{\mathrm{E}}[\mu_x]_i(l)$, $\hat{\mathrm{E}}[\mu_x^2]_i(l)$, i.e. the central mortality rates and their squares from a sample time series.

5^o Minimize the estimation criterion

$$\begin{aligned} S_V = \sum_l \sum_i & [(\hat{\mathrm{E}}[\mu_x]_{i+1}(l) - \mathrm{E}[\mu_x]_i(l)(1-\kappa_x(l)) - \kappa_x(l)\theta_x(l))^2 + \\ & + (\hat{\mathrm{E}}[\mu_x^2]_{i+1}(l) - \mathrm{E}[\mu_x^2]_i(l) - (2\kappa_x(l)\mathrm{E}[\mu_x^2]_i(l) + \\ & - 2\kappa_x(l)\theta_x(l)\mathrm{E}[\mu_x]_i(l) - f_{xi}^2(l)))^2, \end{aligned} \tag{3.4.42}$$

or

$$S_{CIR}=\sum_l\sum_i[(\hat{\mathrm{E}}[\mu_x]_{i+1}(l)-\mathrm{E}[\mu_x]_i(l)(1-\kappa_x(l))-\kappa_x(l)\theta_x(l))^2+$$

$$+(\hat{\mathrm{E}}[\mu_x^2]_{i+1}(l)-\mathrm{E}[\mu_x^2]_i(l)+(2\kappa_x(l)\mathrm{E}[\mu_x^2]_i(l)+ \qquad (3.4.43)$$

$$-2\kappa_x(l)\theta_x(l)\mathrm{E}[\mu_x]_i(l)-f_{xi}^2(l)\mathrm{E}[\mu_x]_i(l)))^2,$$

where $f_{xi}^2(l)=e^{2\zeta(l)i}$.

3.5. The Milevsky–Promislow hybrid models with one linear scalar filter

In this section dynamic hybrid mortality models are presented which are some generalizations of the Milevsky–Promislow model.

3.5.1. MPH model

Let us consider the Milevsky–Promislow Hybrid model (MPH model) with a linear scalar filter

$$\mu_x(t,l)=\mu_x(0,l)\exp\{\gamma_x(l)t+q_x(l)z(t,l)\},\quad t\in\mathbb{R}^+, \qquad (3.5.1)$$

$$dz(t,l)=-\beta_x(l)z(t,l)dt+dw(t), \qquad (3.5.2)$$

where $\mu_{x0},\gamma_x(l),\beta_x(l),q_x(l)>0$ for $l\in\mathbb{S}$ are constant parameters and $w(t)$ is a standard Wiener process.

By taking logarithms on both sides of (3.5.1) and using the Itô formula, we receive the following equation

$$d\ln\mu_x(t,l)=[\gamma_x(l)-\beta_x(l)q_x(l)z(t,l)]dt+q_x(l)dw(t). \qquad (3.5.3)$$

Let us introduce a new state vector

$$\mathbf{h}_x(t,l)=[h_{x1}(t,l),h_{x2}(t,l)]^T=[\ln\mu_x(t,l),z(t,l)]^T. \qquad (3.5.4)$$

Then equations (3.5.2) and (3.5.3) can be written as vector equations

$$d\mathbf{h}_x(t)=\left\{\begin{bmatrix}0 & -\beta_x(l)q_x(l)\\ 0 & -\beta_{x1}(l)\end{bmatrix}\mathbf{h}_x(t,l)+\begin{bmatrix}\gamma_x(l)\\ 0\end{bmatrix}\right\}dt+\begin{bmatrix}q_x(l)\\ 1\end{bmatrix}dw(t). \qquad (3.5.5)$$

For the simplified version, we can also assume that $q_x(l)=1$ for $l\in\mathbb{S}$.

3.5.2. MPH moment model

The moment equations of the Milevsky–Promislow Hybrid Moment model (MPHM model) are of the form

$$\frac{d\mathrm{E}[h_{x1}(t,l)]}{dt}=\gamma_x(l)-\beta_x(l)q_{x1}(l)\mathrm{E}[h_{x2}(t,l)], \tag{3.5.6}$$

$$\frac{d\mathrm{E}[h_{x2}(t,l)]}{dt}=-\beta_{x1}(l)\mathrm{E}[h_{x2}(t,l)], \tag{3.5.7}$$

$$\begin{aligned}\frac{d\mathrm{E}[h_{x1}^2(t,l)]}{dt}&=2\gamma_x(l)\mathrm{E}[h_{x1}(t,l)]+\\&-2\beta_x(l)q_x(l)\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)]+q_x^2(l),\end{aligned} \tag{3.5.8}$$

$$\frac{d\mathrm{E}[h_{x2}^2(t,l)]}{dt}=-2\beta_x(l)\mathrm{E}[h_{x2}^2(t,l)]+1, \tag{3.5.9}$$

$$\begin{aligned}\frac{d\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)]}{dt}&=\gamma_x(l)\mathrm{E}[h_{x2}(t,l)]-\beta_x(l)q_x(l)\mathrm{E}[h_{x2}^2(t,l)]+\\&-\beta_x(l)\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)]+q_x(l).\end{aligned} \tag{3.5.10}$$

MPHM model with a part of stationary moments

By equating to zero the derivatives in equations (3.5.7), we obtain the following condition for the stationary first order moment

$$\mathrm{E}[h_{x2}(l)]=0. \tag{3.5.11}$$

Thus, from equation (3.5.6) we find

$$\mathrm{E}[h_{x1}(t,l)]=\gamma_x(l)t+\gamma_{x0}(l), \tag{3.5.12}$$

where $\gamma_{x0}(l)$ is a constant of integration of equation (3.5.6). In the special case it can be assumed that $\gamma_{x0}(l)=\ln\mu_x(0,l)=\mathrm{E}[h_{x1}(0,l)]$ and $\ln\mu_x(0,l)$ is a constant parameter.

Next, by equating to zero the derivatives in (3.5.9) and (3.5.10) and taking into account conditions (3.5.11) we obtain

$$\mathrm{E}[h_{x2}^2(l)] = \frac{1}{2\beta_x(l)}, \quad \mathrm{E}[h_{x1}(l)h_{x2}(l)] = \frac{q_x(l)}{2\beta_x(l)}. \tag{3.5.13}$$

By introducing quantities $\mathrm{E}[h_{x1}(l)h_{x2}(l)]$ given in (3.5.13) to equation (3.5.8), we obtain

$$\frac{d\mathrm{E}[h_{x1}^2(t,l)]}{dt} = 2\gamma_x(l)\mathrm{E}[h_{x1}(t,l)]. \tag{3.5.14}$$

Hence, from (3.5.12) and (3.5.14) we find

$$\mathrm{E}[h_{x1}^2(t,l)] = \gamma_x^2(l)t^2 + 2\gamma_x(l)\gamma_{x0}(l)t + c_{x0}(l), \tag{3.5.15}$$

where $c_{x0}(l)$ is a constant of integration of equation (3.5.14).
In the particular case, we have $c_{x0}(l) = \ln^2\mu_x(0,l) = \mathrm{E}[h_{x1}^2(0,l)]$, where $\ln^2\mu_x(0,l)$ is a constant parameter.

3.5.3. Discrete MPH model

The derivation of the discrete-time form of the MP model (1.8.17) allow defining the Discrete Milevsky–Promislow Hybrid model (DMPH model)

$$y_x(t,l) = a_{x,0}(l) + a_{x,1}(l)t + a_{x,2}(l)y_x(t-1,l) + \xi_{x,t}(l), \quad t \in \mathbb{N}, \tag{3.5.16}$$

where

$$\begin{aligned}
y_x(t,l) &= \ln\mu_x(t,l), \quad y_x(0,l) = \ln\mu_x(0,l), \\
a_{x,0}(l) &= (1-e^{-\beta_x(l)})y_x(0,l) + \gamma_x(l)e^{-\beta_x(l)}, \\
a_{x,1}(l) &= \gamma_x(l)(1-e^{-\beta_x(l)}), \\
a_{x,2}(l) &= e^{-\beta_x(l)}, \\
\xi_{x,t}(l) &= -q_x(l)\int_0^1 e^{-\beta_x(l)u}dw(t-u).
\end{aligned} \tag{3.5.17}$$

The model can be equivalently expressed as

$$y_x(t,l) = (1 - a_{x,2}(l))y_x(0,l) + \frac{a_{x,1}(l)a_{x,2}(l)}{1-a_{x,2}}(l) + a_{x,1}(l)t + \\ + a_{x,2}(l)y_x(t-1,l) + \xi_{x,t}(l), \quad t \in \mathbb{N}, \tag{3.5.18}$$

where

$$y_x(t,l) = \ln \mu_x(t,l), \quad y_x(0,l) = \ln \mu_x(0,l),$$

$$a_{x,1}(l) = \gamma_x(l)(1 - e^{-\beta_x(l)}), \tag{3.5.19}$$

$$a_{x,2}(l) = e^{-\beta_x(l)}.$$

Random terms $\xi_{x,t}(l)$ in (3.5.16) or (3.5.18) are Gaussian random variables with means and variances equal, respectively,

$$\mathrm{E}[\xi_{x,t}(l)] = 0, \quad \mathrm{Var}[\xi_{x,t}(l)] = \mathrm{E}[\xi^2_{x,t}(l)] = q^2_x(l). \tag{3.5.20}$$

3.5.4. Discrete MPH moment model

From the moment equations (3.5.6)–(3.5.10) we receive the Discrete Milevsky–Promislow Hybrid Moment model (DMPHM model) as

$$\mathrm{E}[h_{x1}]_{i+1}(l) = \mathrm{E}[h_{x1}]_i(l) + \gamma_x(l)\delta, \tag{3.5.21}$$

$$\mathrm{E}[h^2_{x1}]_{i+1}(l) = \mathrm{E}[h^2_{x1}]_i(l) + (2\gamma_x(l)\mathrm{E}[h_{x1}]_i(l) + \\ - 2\beta_x(l)q_x(l)\mathrm{E}[h_{x1}h_{x2}]_i(l) + q^2_x(l))\delta, \tag{3.5.22}$$

$$\mathrm{E}[h^2_{x2}]_{i+1}(l) = \mathrm{E}[h^2_{x2}]_i(l) + (-2\beta_{x_1}(l)\mathrm{E}[h^2_{x2}]_i(l) + 1)\delta, \tag{3.5.23}$$

$$\mathrm{E}[h_{x1}h_{x2}]_{i+1}(l) = \mathrm{E}[h_{x1}h_{x2}]_i(l) + (-\beta_x(l)q_x(l)\mathrm{E}[h^2_{x2}]_i(l) + \\ - \beta_x(l)\mathrm{E}[h_{x1}h_{x2}]_i(l) + q_x(l))\delta, \tag{3.5.24}$$

where

$$\mathrm{E}[h_{x1}]_i(l) = \mathrm{E}[h_{x1}](t_i,l),\ \mathrm{E}[h^2_{xj}]_i(l) = \mathrm{E}[h^2_{xj}](t_i,l), j = 1,2, \\ \mathrm{E}[h_{x1}h_{x2}]_i = \mathrm{E}[h_{x1}h_{x2}](t_i),\ \delta = t_{i+1} - t_i = \mathrm{const}. \tag{3.5.25}$$

3.5.5. Parameters' estimation of the MPH models

In this section we will present iterative procedures of the parameters' estimation for the Milevsky–Promislow hybrid models, both for realizations and for moments.

Estimation of the DMPH model for realizations

Let us consider the discrete version of the MPH model expressed by (3.5.18). The parameters of the DMPH model are estimated by minimizing the sum of squared errors with respect to $a_{x,1}(l)$ and $a_{x,2}(l)$, i.e. by minimizing the following criterion

$$S = \sum_{l} \sum_{t \in I_l} \Big(y_{x,t} - (1 - a_{x,2}(l)) y_{x,0} - \frac{a_{x,1}(l) a_{x,2}(l)}{1 - a_{x,2}(l)} + \\ - a_{x,1}(l) t - a_{x,2}(l) y_{x,t-1} \Big)^2 , \tag{3.5.26}$$

where $y_{x,t} = \ln m_{x,t}$ are log-central death rates observed in a sample time series $\{\ln m_{x,t},\ t = 1, 2, \ldots, T\}$.

Iterative estimation of the DMPH moment model

The iterative estimation procedure for the DMPHM model given in (3.5.21)–(3.5.24) consists of the following steps:

1^o Take constant initial conditions, e.g. $\mathrm{E}[h_{x1}h_{x2}]_0(l) = 0$, $\mathrm{E}[h_{x2}^2]_0(l) = 1$, and initial conditions $\mathrm{E}[h_{x1}]_0(l) = p_x(l)$, $\mathrm{E}[h_{x1}^2]_0(l) = p_x^2(l)$, where $p_x(l)$ is a fixed parameter.

2^o Assume initial values for parameters $p_x(l)$, $\gamma_x(l)$, $\beta_x(l)$, $q_x(l)$, e.g. $p_x(l) = \ln \mu_x(0, l)$, $\gamma_x(l) = 0.1$, $\beta_x(l) = 0.1$, $q_x(l) = 1$.

3^o Estimate the successive values of $\mathrm{E}[h_{x1}]_i(l)$, $\mathrm{E}[h_{x1}^2]_i(l)$, $\mathrm{E}[h_{x1}h_{x2}]_i(l)$, $\mathrm{E}[h_{x2}^2]_i(l)$ from expressions (3.5.21)–(3.5.24) for an i-th iteration ($i = 1, 2, \ldots$) and for the given values of $p_x(l) = \ln \mu_x(0, l)$, $\gamma_x(l)$, $\beta_x(l)$, $q_x(l)$.

4^o Determine the values of $\hat{\mathrm{E}}[h_{x1}]_i(l)$, $\hat{\mathrm{E}}[h_{x1}^2]_i(l)$, i.e. the log-central mortality rates and their squares from a sample time series.

5^o Minimize the estimation criterion

$$S = \sum_l \sum_i \Big(\hat{\mathrm{E}}[h_{x1}]_{i+1}(l) - \mathrm{E}[h_{x1}]_i(l) - \gamma_x(l)\Big)^2 + \\ + \Big(\hat{\mathrm{E}}[h_{x1}^2]_{i+1}(l) - \mathrm{E}[h_{x1}^2]_i(l) - (2\gamma_x(l)\mathrm{E}[h_{x1}]_i(l) + \\ - 2\beta_x(l) q_x(l) \mathrm{E}[h_{x1}h_{x2}]_i(l) + q_x^2(l))\Big)^2 . \tag{3.5.27}$$

Estimation of the MPH moment model with stationary first order moments

Let us assume the MPHM model expressed by moment equations (3.5.12) and (3.5.15). In this case the estimation procedure reduces to the minimization of the following sum

$$S = \sum_l \sum_{t \in I_l} [(\hat{\mathrm{E}}[h_{x1}(t,l)] - \gamma_x(l)t - \gamma_{x0}(l))^2 + \\ + (\hat{\mathrm{E}}[h_{x1}^2(t,l)] - \gamma_x^2(l)t^2 - 2\gamma_x(l)\gamma_{x0}(l)t - c_{x0}(l))^2, \tag{3.5.28}$$

where $\gamma_x(l), \gamma_{x0}(l)$ and $c_{x0}(l)$ are some parameters.

In the general case, (3.5.28) is minimized with respect to three parameters $\gamma_x(l)$, $\gamma_{x0}(l)$ and $c_{x0}(l)$ for $l \in \mathbb{S}$. However, for $\gamma_{x0}(l) = \ln \mu_x(0,l)$ and $c_{x0}(l) = \ln^2 \mu_x(0,l)$ criterion (3.5.28) can be minimized with respect to one parameter $\gamma_x(l)$ for $l \in \mathbb{S}$.

3.6. The Giacometti–Ortobelli–Bertocchi hybrid models

3.6.1. GOBH model

The family of subsystems making up the Giacometti–Ortobelli–Bertocchi Hybrid model (GOBH model) with a scalar linear filter is given by equations

$$\mu_x(t,l) = \mu_x(0,l) \exp\{\gamma_x(l)t + q_x(l)z(t,l)\}, \quad t \in \mathbb{R}^+, \tag{3.6.1}$$

$$dz(t,l) = -\beta_x(l)z(t,l)dt + f_x(t,l)dw(t), \tag{3.6.2}$$

where $\gamma_x(l), \beta_x(l), q_x(l), \mu_x(0,l) > 0$ for $l \in \mathbb{S}$ are constant parameters and $f_x(t,l) = e^{\zeta(l)t}$ are time dependent differentiable functions of t.

Hence, the following stochastic differential equation results from the Itô formula

$$d\ln\mu_x(t,l) = [\gamma_x(l) - \beta_x(l)(\ln\mu_x(t,l) - \ln\mu_x(0,l) - \gamma_x(l)t)]dt + \\ + q_x(l)f_x(t,l)dw(t). \tag{3.6.3}$$

Let us introduce a new state vector

$$\mathbf{h}_x(t,l) = [h_{x1}(t,l), h_{x2}(t,l)]^T = [\ln\mu_x(t,l), z(t,l)]^T. \tag{3.6.4}$$

Then equations (3.6.3) and (3.6.2) can be written as a vector equation

$$d\mathbf{h}_x(t) = \left\{ \begin{bmatrix} -\beta_x(l) & 0 \\ 0 & -\beta_x(l) \end{bmatrix} \mathbf{h}_x(t,l) + \right. \\ \left. + \begin{bmatrix} \gamma_x(l) + \beta_x(l)(\ln\mu_x(0,l) + \gamma_x(l)t) \\ 0 \end{bmatrix} \right\} dt + \tag{3.6.5} \\ + \begin{bmatrix} q_x(l)f_x(t,l) \\ f_x(t,l) \end{bmatrix} dw(t).$$

For the simplified version, we can assume that $q_x(l) = 1$ for $l \in \mathbb{S}$.

3.6.2. GOBH moment model

The moment equations of the Giacometti–Ortobelli–Bertocchi Hybrid Moment model (GOBHM model) are as follows

$$\frac{d\mathrm{E}[h_{x1}(t,l)]}{dt} = \gamma_x(l) - \beta_x(l)q_{x1}(l)\mathrm{E}[h_{x2}(t,l)], \tag{3.6.6}$$

$$\frac{d\mathrm{E}[h_{x2}(t,l)]}{dt} = -\beta_{x1}(l)\mathrm{E}[h_{x2}(t,l)], \tag{3.6.7}$$

$$\frac{d\mathrm{E}[h_{x1}^2(t,l)]}{dt} = 2\gamma_x(l)\mathrm{E}[h_{x1}(t,l)]+$$
$$- 2\beta_x(l)q_x(l)\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)] + q_x^2(l), \tag{3.6.8}$$

$$\frac{d\mathrm{E}[h_{x2}^2(t,l)]}{dt} = -2\beta_x(l)\mathrm{E}[h_{x2}^2(t,l)] + f_x^2(t,l), \tag{3.6.9}$$

$$\frac{d\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)]}{dt} = \gamma_x(l)\mathrm{E}[h_{x2}(t,l)] - \beta_x(l)q_x(l)\mathrm{E}[h_{x2}^2(t,l)]+$$
$$- \beta_x(l)\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)] + f_x(t,l)q_x(l). \tag{3.6.10}$$

3.6.3. Discrete GOBH model

By analogy to the derivations of the discrete form (1.9.24) of the GOB model, we obtain the Discrete Giacometti–Ortobelli– Bertocchi Hybrid model (DGOBH model)

$$y_x(t,l) = a_{x,0}(l) + a_{x,1}(l)t + a_{x,2}(l)y_x(t-1,l) + \varepsilon_{x,t}(l), \quad t \in \mathbb{N}, \tag{3.6.11}$$

where

$$y_x(t,l) = \ln \mu_x(t,l), \quad y_x(0,l) = \ln \mu_x(0,l),$$

$$a_{x,0}(l) = (1 - e^{-\beta_x(l)} \ln \mu_{x0,l} + \gamma_x(l)e^{-\beta_x(l)},$$

$$a_{x,1}(l) = \gamma_x(l)(1 - e^{-\beta_x(l)}), \tag{3.6.12}$$

$$a_{x,2}(l) = e^{-\beta_x(l)},$$

$$\varepsilon_{x,t}(l) = -q_x(l)e^{\zeta_x(l)t} \int_0^1 e^{-(\zeta_x(l)+\beta_x(l))u} dw(t-u).$$

The model takes also the following equivalent form

$$y_x(t,l) = (1 - a_{x,2}(l))y_x(0,l) + \frac{a_{x,1}(l)a_{x,2}(l)}{1 - a_{x,2}(l)} + a_{x,1}(l)t+$$
$$+ a_{x,2}(l)y_x(t-1,l) + \varepsilon_{x,t}(l), \ t \in \mathbb{N}, \tag{3.6.13}$$

where

$$y_x(t,l) = \ln \mu_x(t,l), \quad y_x(0,l) = \ln \mu_x(0,l),$$

$$a_{x,1}(l) = \gamma_x(l)(1 - e^{-\beta_x(l)}), \tag{3.6.14}$$

$$a_{x,2}(l) = e^{-\beta_x(l)}.$$

Random terms $\varepsilon_{x,t}(l)$ in (3.6.11) or (3.6.13) are Gaussian random variables with means and variances equal, respectively,

$$\mathrm{E}[\varepsilon_{x,t}(l)] = 0, \quad \mathrm{Var}[\varepsilon_{x,t}(l)] = \mathrm{E}[\varepsilon^2_{x,t}(l)] = q^2_x(l)e^{2\zeta_x(l)t}. \tag{3.6.15}$$

3.6.4. Discrete GOBH moment model

From the moment equations (3.6.6)–(3.6.10) we find the discrete representation, called the Giacometti–Ortobelli–Bertocchi Hybrid Moment model (DGOBHM model), expressed by the following equations

$$\mathrm{E}[h_{x1}]_{i+1}(l) = \mathrm{E}[h_{x1}]_i(l) + \gamma_x(l)\delta, \tag{3.6.16}$$

$$\begin{aligned} \mathrm{E}[h^2_{x1}]_{i+1}(l) = \mathrm{E}[h^2_{x1}]_i(l) + (2\gamma_x(l)\mathrm{E}[h_{x1}]_i(l)+ \\ - 2\beta_x(l)q_x(l)\mathrm{E}[h_{x1}h_{x2}]_i(l) + q^2_x(l))\delta, \end{aligned} \tag{3.6.17}$$

$$\mathrm{E}[h^2_{x2}]_{i+1}(l) = \mathrm{E}[h^2_{x2}]_i(l) + (-2\beta_{x_1}(l)\mathrm{E}[h^2_{x2}]_i(l) + f^2_{xi}(l))\delta, \tag{3.6.18}$$

$$\begin{aligned} \mathrm{E}[h_{x1}h_{x2}]_{i+1}(l) = \mathrm{E}[h_{x1}h_{x2}]_i(l) + (-\beta_x(l)q_x(l)\mathrm{E}[h^2_{x2}]_i(l)+ \\ - \beta_{x1}(l)\mathrm{E}[h_{x1}h_{x2}]_i(l) + f_{xi}(l)q_x(l))\delta, \end{aligned} \tag{3.6.19}$$

where

$$\mathrm{E}[h_{x1}]_i(l) = \mathrm{E}[h_{x1}](t_i,l), \ \mathrm{E}[h^2_{xj}]_i(l) = \mathrm{E}[h^2_{xj}](t_i,l), \ j = 1,2,$$

$$\mathrm{E}[h_{x1}h_{x2}]_i = \mathrm{E}[h_{x1}h_{x2}](t_i), \tag{3.6.20}$$

$$f_{xi}(l) = e^{\zeta_x(l)i}, \ \delta = t_{i+1} - t_i = \mathrm{const}.$$

3.6.5. Parameters' estimation of the GOBH models

Estimation of the DGOBH model for realizations

Parameters' estimation of the DGOBH model (3.6.13) is performed by minimizing the sum of squared errors, i.e. by minimizing the following criterion with respect to $a_{x,1}(l)$ and $a_{x,2}(l)$

$$S = \sum_{l} \sum_{t \in I_l} \left(y_{x,t} - (1 - a_{x,2}(l)) y_{x,0} - \frac{a_{x,1}(l) a_{x,2}(l)}{1 - a_{x,2}(l)} + \right. \\ \left. - a_{x,1}(l) t - a_{x,2}(l) y_{x,t-1} \right)^2, \tag{3.6.21}$$

where $y_{x,t} = \ln m_{x,t}$ are log-central death rates from a sample.

Iterative estimation of the DGOBH moment model

The estimation procedure for the DGOBHM model expressed by moment equations (3.6.16)–(3.6.19) is as follows.

1^o Take constant initial conditions, e.g. $\mathrm{E}[h_{x1}h_{x2}]_0(l) = 0$, $\mathrm{E}[h_{x2}^2]_0(l) = 1$ and initial conditions $\mathrm{E}[h_{x1}]_0(l) = p_x(l)$, $\mathrm{E}[h_{x1}^2]_0(l) = p_x^2(l)$, where $p_x(l)$ is a fixed parameter.

2^o Assume initial values for $p_x(l)$, $\gamma_x(l)$, $\beta_x(l)$, $q_x(l)$, $\zeta_x(l)$, e.g. $p_x(l) = \ln \mu_x(0, l)$, $\gamma_x(l) = 0.1$, $\beta_x(l) = 0.1$, $q_x(l) = 1$.

3^o Estimate the successive values of $\mathrm{E}[h_{x1}]_i(l)$, $\mathrm{E}[h_{x1}^2]_i(l)$, $\mathrm{E}[h_{x1}h_{x2}]_i(l)$, $\mathrm{E}[h_{x2}^2]_i(l)$ from expressions (3.6.16)–(3.6.19) for an i-th iteration ($i = 1, 2, ...$) and for the given values of $p_x(l) = \ln \mu_x(0, l)$, $\gamma_x(l)$, $\beta_x(l)$, $q_x(l), \zeta_x(l)$.

4^o Determine the values of $\hat{\mathrm{E}}[h_{x1}]_i(l)$, $\hat{\mathrm{E}}[h_{x1}^2]_i(l)$, i.e. the log-central mortality rates and their squares from a sample time series.

5^o Minimize the following estimation criterion

$$S = \sum_{l} \sum_{i} \left(\hat{\mathrm{E}}[h_{x1}]_{i+1}(l) - \mathrm{E}[h_{x1}]_i(l) - \gamma_x(l) \right)^2 + \\ + \left(\hat{\mathrm{E}}[h_{x1}^2]_{i+1}(l) - \mathrm{E}[h_{x1}^2]_i(l) - (2\gamma_x(l)\mathrm{E}[h_{x1}]_i(l) + \right. \\ \left. - 2\beta_x(l) q_x(l) \mathrm{E}[h_{x1}h_{x2}]_i(l) + q_x^2(l)) \right)^2. \tag{3.6.22}$$

It is worth noting that although parameter $\zeta_x(l)$ does not appear explicit in (3.6.22), it is a part of $\mathrm{E}[h_{x1}h_{x2}]_i(l)$ and therefore criterion (3.6.22) depends also on $\zeta_x(l)$.

3.7. Modified Milevsky–Promislow hybrid models

3.7.1. Modified MPH model

We will formulate a model analogous to (3.5.1)–(3.5.2), further termed as the Modified Milevsky–Promislow Hybrid model (MMPH model)

$$\mu_x(t,l) = \mu_x(0,l)\exp\{\gamma_x(l)t + z(t,l)\}, \quad t \in \mathbb{R}^+, \tag{3.7.1}$$

where $z(t,l)$ for $l \in \mathbb{S}$ are the Ornstein–Uhlenbeck stochastic processes satisfying the following stochastic differential equations

$$dz(t,l) = \beta_x(l)(\alpha_x(l) - z(t,l))dt + \sigma_x(l)dw(t), \quad z(0,l) = 0, \tag{3.7.2}$$

where $\alpha_x(l) \in \mathbb{R}$ and $\mu_x(0,l), \gamma_x(l), \beta_x(l), \sigma_x(l) > 0$ are constant parameters.

Hence, the following stochastic differential equation results from the Itô formula

$$d\ln\mu_x(t,l) = [\gamma_x(l) + \beta_x(l)\alpha_x(l) - \beta_x(l)z(t,l)]dt + \sigma_x(l)dw(t). \tag{3.7.3}$$

Let us introduce a new state vector

$$\mathbf{h}_x(t,l) = [h_{x1}(t,l), h_{x2}(t,l)]^T = [\ln\mu_x(t,l), z(t,l)]^T. \tag{3.7.4}$$

Then equations (3.7.3) and (3.7.2) can be written as a vector equation

$$d\mathbf{h}_x(t) = \left\{\begin{bmatrix} 0 & -\beta_x(l) \\ 0 & -\beta_x(l) \end{bmatrix}\mathbf{h}_x(t,l) + \begin{bmatrix} \gamma_x(l) + \beta_x(l)\alpha_x(l) \\ \beta_x(l)\alpha_x(l) \end{bmatrix}\right\}dt + \begin{bmatrix} \sigma_x(l) \\ \sigma_x(l) \end{bmatrix}dw(t). \tag{3.7.5}$$

3.7.2. Modified MPH moment model

The moment equations of the Modified Milevsky–Promislow Hybrid Moment model (MMPHM model) are of the form

$$\frac{d\mathrm{E}[h_{x1}(t,l)]}{dt} = \gamma_x(l) + \beta_x(l)\alpha_x(l) - \beta_x(l)\mathrm{E}[h_{x2}(t,l)], \tag{3.7.6}$$

$$\frac{d\mathrm{E}[h_{x2}(t,l)]}{dt} = -\beta_x(l)\mathrm{E}[h_{x2}(t,l)] + \beta_x(l)\alpha_x(l), \tag{3.7.7}$$

$$\begin{aligned}\frac{d\mathrm{E}[h_{x1}^2(t,l)]}{dt} &= 2(\gamma_x(l)+\beta_x(l)\alpha_x(l))\mathrm{E}[h_{x1}(t,l)]+ \\ &- 2\beta_x(l)\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)] + \sigma_x^2(l),\end{aligned} \tag{3.7.8}$$

$$\begin{aligned}\frac{d\mathrm{E}[h_{x2}^2(t,l)]}{dt} &= 2\beta_x(l)\alpha_x(l)\mathrm{E}[h_{x2}(t,l)]+ \\ &- 2\beta_x(l)\mathrm{E}[h_{x2}^2(t,l)] + \sigma_x^2(l),\end{aligned} \tag{3.7.9}$$

$$\begin{aligned}\frac{d\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)]}{dt} &= (\gamma_x(l) + \beta_x(l)\alpha_x(l))\mathrm{E}[h_{x2}(t,l)]+ \\ &- \beta_x(l)\mathrm{E}[h_{x2}^2(t,l)] - \beta_x(l)\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)]+ \\ &+ \beta_x(l)\alpha_x(l)\mathrm{E}[h_{x1}(t,l)] + \sigma_x^2(l).\end{aligned} \tag{3.7.10}$$

MMPHM model with a part of stationary moments

By equating to zero the derivative in equation (3.7.7), we receive the condition for the stationary first order moment

$$\mathrm{E}[h_{x2}(l)] = \alpha_x(l). \tag{3.7.11}$$

Thus, from equation (3.7.6) we find

$$\mathrm{E}[h_{x1}(t,l)] = \gamma_x(l)t + \gamma_{x0}(l), \tag{3.7.12}$$

where $\gamma_{x0}(l)$ is a constant in integration of equation (3.7.6).

In the particular case, it can be assumed that $\gamma_{x0}(l) = \ln\mu_x(0,l) = \mathrm{E}[h_{x1}(0,l)]$ with $\ln\mu_x(0,l)$ as a constant parameter.

By equating to zero the derivatives in (3.7.9) and (3.7.10) and taking into account conditions (3.7.11), we obtain

$$\mathrm{E}[h_{x2}^2(l)] = \alpha_x^2(l)\frac{\sigma_x^2(l)}{2\beta_x(l)}, \tag{3.7.13}$$

$$\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)] = \frac{\sigma_x^2(l)}{2\beta_x(l)} + \alpha_x(l)\mathrm{E}[h_{x1}(t,l)] + \frac{\gamma_x(l)\alpha_x(l)}{\beta_x(l)}. \tag{3.7.14}$$

Let us replace $\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)]$ in equation (3.7.8) with an expression given on the right-hand side of (3.7.14). Then we obtain

$$\frac{d\mathrm{E}[h_{x1}^2(t,l)]}{dt} = 2\gamma_x(l)\mathrm{E}[h_{x1}(t,l)] - 2\gamma_x(l)\alpha_x(l). \tag{3.7.15}$$

Hence, from (3.7.12) and (3.7.15) we have

$$\mathrm{E}[h_{x1}^2(t,l)] = \gamma_x^2(l)t^2 + 2\gamma_x(l)(\gamma_{x0}(l) - \alpha_x(l))t + c_{x0}(l), \tag{3.7.16}$$

where $c_{x0}(l)$ is a constant of integration of equation (3.7.15).
In the particular case, we have $c_{x0}(l) = \ln^2\mu_x(0,l) = \mathrm{E}[h_{x1}^2(0,l)]$, where $\ln^2\mu_x(0,l)$ is a constant parameter.

3.7.3. Discrete modified MPH model

From the derivation of the discrete-time version of the modified Milevsky–Promislow model (1.10.1)–(1.10.2) we find analogous equations defining the Discrete Modified Milevsky–Promislow Hybrid model (DMMPH model) which corresponds with the model (3.7.1)–(3.7.2)

$$y_x(t,l) = a_{x,0}(l) + a_{x,1}(l)t + a_{x,2}(l)y_x(t-1,l) + \xi_{x,t}(l), \quad t\in\mathbb{N}, \tag{3.7.17}$$

where

$$\begin{aligned}
&y_x(t,l) = \ln\mu_x(t,l), \quad y_x(0,l) = \ln\mu_x(0,l),\\
&a_{x,0}(l) = (1-e^{-\beta_x(l)})(y_x(0,l) + \alpha_x(l)) + \gamma_x(l)e^{-\beta_x(l)},\\
&a_{x,1}(l) = \gamma_x(l)(1-e^{-\beta_x(l)}),\\
&a_{x,2}(l) = e^{-\beta_x(l)},\\
&\xi_{x,t}(l) = -\sigma_x(l)\int_0^1 e^{-\beta_x(l)u}dw(t-u).
\end{aligned} \tag{3.7.18}$$

Let us remark that $\xi_{x,t}(l)$ in (3.7.17) are Gaussian random variables with means and variances equal, respectively,

$$\mathrm{E}[\xi_{x,t}(l)] = 0, \quad \mathrm{Var}[\xi_{x,t}(l)] = \mathrm{E}[\xi_{x,t}^2(l)] = \sigma_x^2(l). \tag{3.7.19}$$

3.7.4. Discrete modified MPH moment model

From the moment equations (3.7.6)–(3.7.10) we find the Discrete Modified Milevsky–Promislow Hybrid Moment model (DMMPHM model) described by equations

$$\mathrm{E}[h_{x1}]_{i+1}(l) = \mathrm{E}[h_{x1}]_i(l) + (\gamma_x(l) + \beta_x(l)\alpha_x(l))\delta, \tag{3.7.20}$$

$$\begin{aligned}\mathrm{E}[h_{x1}^2]_{i+1}(l) = \mathrm{E}[h_{x1}^2]_i(l) + (2(\gamma_x(l) + \beta_x(l)\alpha_x(l))\mathrm{E}[h_{x1}]_i(l) + \\ - 2\beta_x(l)\mathrm{E}[h_{x1}h_{x2}]_i(l) + \sigma_x^2(l))\delta,\end{aligned} \tag{3.7.21}$$

$$\mathrm{E}[h_{x2}^2]_{i+1}(l) = \mathrm{E}[h_{x2}^2]_i(l) + (-2\beta_x(l)\mathrm{E}[h_{x2}^2]_i(l) + \sigma_x^2(l))\delta, \tag{3.7.22}$$

$$\begin{aligned}\mathrm{E}[h_{x1}h_{x2}]_{i+1}(l) = \mathrm{E}[h_{x1}h_{x2}]_i(l) + (-\beta_x(l)\mathrm{E}[h_{x2}^2]_i(l) + \\ - \beta_x(l)\mathrm{E}[h_{x1}h_{x2}]_i(l) + \beta_x(l)\alpha_x(l)\mathrm{E}[h_{x1}]_i(l) + \sigma_x^2(l))\delta,\end{aligned} \tag{3.7.23}$$

where

$$\begin{aligned}&\mathrm{E}[h_{x1}]_i(l) = \mathrm{E}[h_{x1}](t_i, l),\ \mathrm{E}[h_{xj}^2]_i(l) = \mathrm{E}[h_{xj}^2](t_i, l),\ j = 1, 2,\\ &\mathrm{E}[h_{x1}h_{x2}]_i(l) = \mathrm{E}[h_{x1}h_{x2}](t_i, l),\ \delta = t_{i+1} - t_i = \mathrm{const}.\end{aligned} \tag{3.7.24}$$

3.7.5. Parameters' estimation of the modified MPH models

Estimation of the DMMPH model for realizations

Let as consider the DMMPH model (3.7.17). One can find estimates of the model's parameters by minimizing the sum of squared errors, i.e. by minimizing the following criterion with respect to $a_{x,0}(l)$, $a_{x,1}(l)$ and $a_{x,2}(l)$

$$S = \sum_l \sum_{t \in I_l} [y_{x,t} - (a_{x,0}(l) + a_{x,1}(l)t + a_{x,2}(l)y_{x,t-1})]^2, \tag{3.7.25}$$

where $y_{x,t} = \ln m_{x,t}$ are log-central death rates from a sample time series.

Iterative estimation of the DMMPH moment model

For the DMMPHM model (3.7.20)–(3.7.23) the estimation procedure can be described in the following steps.

1^o Take constant initial conditions, e.g. $\mathrm{E}[h_{x1}h_{x2}]_0(l)=0$, $\mathrm{E}[h_{x2}^2]_0(l)=1$, and initial conditions $\mathrm{E}[h_{x1}]_0(l)=p_x(l)$, $\mathrm{E}[h_{x1}^2]_0(l)=p_x^2(l)$, where $p_x(l)$ is a fixed parameter.

2^o Assume initial values for $p_x(l)$, $\gamma_x(l)$, $\beta_x(l)$, $q_x(l)$, e.g. $p_x(l)=\ln\mu_x(0,l)$, $\gamma_x(l)=0.1$, $\beta_x(l)=0.1$, $q_x(l)=1$.

3^o Estimate the successive values of $\mathrm{E}[h_{x1}]_i(l)$, $\mathrm{E}[h_{x1}^2]_i(l)$, $\mathrm{E}[h_{x1}h_{x2}]_i(l)$, $\mathrm{E}[h_{x2}^2]_i(l)$ from expressions (3.7.20)–(3.7.23) for an i-th iteration ($i=1,2,...$) and for the given values of $p_x(l)=\ln\mu_x(0,l)$, $\gamma_x(l)$, $\beta_x(l)$, $q_x(l)$.

4^o Determine the values of $\hat{\mathrm{E}}[h_{x1}]_i(l)$, $\hat{\mathrm{E}}[h_{x1}^2]_i(l)$, i.e. the log-central mortality rates and their squares from a sample time series.

5^o Minimize the estimation criterion

$$S=\sum_l\sum_i\Big(\hat{\mathrm{E}}[h_{x1}]_{i+1}(l)-\mathrm{E}[h_{x1}]_i(l)-\gamma_x(l)-\beta_x(l)\alpha_x(l)\Big)^2+ \\ +\Big(\hat{\mathrm{E}}[h_{x1}^2]_{i+1}(l)-\mathrm{E}[h_{x1}^2]_i(l)-(2(\gamma_x(l)+\beta_x(l)\alpha_x(l))\mathrm{E}[h_{x1}]_i(l)+ \\ -2\beta_x(l)q_x(l)\mathrm{E}[h_{x1}h_{x2}]_i(l)+q_x^2(l))\Big)^2. \tag{3.7.26}$$

Estimation of the MMPH moment model with stationary first order moments

Let us consider the DMMPHM model using the discrete-time version of moment equations (3.7.11)–(3.7.16). The estimation procedure reduces here to the minimization of the following square criterion

$$S=\sum_l\sum_{t\in I_l}\Big(\hat{\mathrm{E}}[h_{x1}(t,l)]-\gamma_x(l)t-\gamma_0(l)\Big)^2+ \\ +\Big(\hat{\mathrm{E}}[h_{x1}^2(t,l)]-\gamma_x^2(l)t^2-2\gamma_x(l)(\gamma_{x0}(l)-\alpha_x(l))t-c_{x0}(l)\Big)^2, \tag{3.7.27}$$

where $\gamma_x(l),\gamma_{x0}(l)$ and $c_{x0}(l)$ are some parameters.

In the general case, criterion (3.7.27) is minimized with respect to three parameters $\gamma_x(l)$, $\gamma_{x0}(l)$ and $c_{x0}(l)$ for $l \in \mathbb{S}$. However, for $\gamma_{x0}(l) = \ln \mu_x(0, l)$ and $c_{x0}(l) = \ln^2 \mu_x(0, l)$ the criterion can be minimized with respect to one parameter $\gamma_x(l)$ for $l \in \mathbb{S}$.

3.8. The Milevsky–Promislow hybrid models with two or more linear filters

3.8.1. MPH model with two dependent filters

The family of subsystems making up the Milevsky–Promislow Hybrid Model with 2 Dependent Filters (MPH-2DF model) is described by the following equations

$$\mu_x(t, l) = \mu_x(0, l) \exp\{\gamma_x(l)t + q_{x1}(l)z_1(t, l) + q_{x2}(l)z_2(t, l)\}, \quad (3.8.1)$$

$$dz_1(t, l) = -\beta_{x1}(l)z_1(t, l)dt + \sigma_{x1}(l)dw(t), \quad (3.8.2)$$

$$dz_2(t, l) = -\beta_{x2}(l)z_2(t, l)dt + \sigma_{x2}(l)dw(t), \quad (3.8.3)$$

where $t \in \mathbb{R}^+$, $l \in \mathbb{S}$ and $\gamma_x(l), q_{x1}(l), q_{x2}(l), \mu_x(0, l), \beta_{x1}(l), \beta_{x2}(l), \sigma_{x1}(l), \sigma_{x2}(l) > 0$ are constant parameters, $w(t)$ is a standard Wiener process.

The Itô formula applied to the logarithm of (3.8.1) leads to the following equation

$$\begin{aligned} d\ln \mu_x(t,l) = {} & [\gamma_x(l) - \beta_{x1}(l)q_{x1}(l)z_1(t,l) - \beta_{x2}(l)q_{x2}(l)z_2(t,l)]dt + \\ & +[\sigma_{x1}(l)q_{x1}(l) + \sigma_{x2}(l)q_{x2}(l)]dw(t). \end{aligned} \quad (3.8.4)$$

We assume that $\beta_{x1}(l) \neq \beta_{x2}(l)$. To obtain the moment equations for system (3.8.4) and (3.8.2)–(3.8.3) we introduce a new state vector

$$\begin{aligned} \mathbf{h}_x(t, l) &= [h_{x1}(t, l), h_{x2}(t, l), h_{x3}(t, l)]^T = \\ &= [\ln \mu_x(t, l), z_1(t, l), z_2(t, l)]^T. \end{aligned} \quad (3.8.5)$$

Equations (3.8.4) and (3.8.2)–(3.8.3) can then be written as vector equation

$$d\mathbf{h}_x(t,l)=\left\{\begin{bmatrix}0 & -\beta_{x1}(l)q_{x1}(l) & -\beta_{x2}(l)q_{x2}(l)\\ 0 & -\beta_{x1}(l) & 0\\ 0 & 0 & -\beta_{x2}(l)\end{bmatrix}\mathbf{h}_x(t,l)+\begin{bmatrix}\gamma_x(l)\\ 0\\ 0\end{bmatrix}\right\}dt+ \tag{3.8.6}$$

$$+\begin{bmatrix}\sigma_{x_1}(l)q_{x1}(l)+\sigma_{x2}(l)q_{x2}(l)\\ \sigma_{x1}(l)\\ \sigma_{x2}(l)\end{bmatrix}dw(t).$$

For the simplified version, we can assume that $q_{x1}(l) = q_{x2}(l) = 1$.

3.8.2. MPH moment model with two dependent filters

The moment equations of the Milevsky–Promislow Hybrid Moment model with 2 Dependent Filters (MPHM-2DF model) are as follows

$$\frac{d\mathrm{E}[h_{x1}(t,l)]}{dt}=\gamma_x(l)-\beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x2}(t,l)]+ \tag{3.8.7}$$
$$-\beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x3}(t,l)],$$

$$\frac{d\mathrm{E}[h_{x2}(t,l)]}{dt}=-\beta_{x1}(l)\mathrm{E}[h_{x2}(t,l)], \tag{3.8.8}$$

$$\frac{d\mathrm{E}[h_{x3}(t,l)]}{dt}=-\beta_{x2}(l)\mathrm{E}[h_{x3}(t,l)], \tag{3.8.9}$$

$$\frac{d\mathrm{E}[h_{x1}^2(t,l)]}{dt}=2\gamma_x(l)\mathrm{E}[h_{x1}(t,l)]-2\beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)] \tag{3.8.10}$$
$$-2\beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x1}(t,l)h_{x3}(t,l)]+$$
$$+(\sigma_{x1}(l)q_{x1}(l)+\sigma_{x2}(l)q_{x2}(l))^2,$$

$$\frac{d\mathrm{E}[h_{x2}^2(t,l)]}{dt}=-2\beta_{x1}(l)\mathrm{E}[h_{x2}^2(t,l)]+\sigma_{x1}^2(l), \tag{3.8.11}$$

$$\frac{d\mathrm{E}[h_{x3}^2(t,l)]}{dt} = -2\beta_{x2}(l)\mathrm{E}[h_{x3}^2(t,l)] + \sigma_{x2}^2(l), \tag{3.8.12}$$

$$\begin{aligned}\frac{d\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)]}{dt} =& \gamma_x(l)\mathrm{E}[h_{x2}(t,l)] - \beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x2}^2(t,l)] +\\ &- \beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x2}(t,l)h_{x3}(t,l)] +\\ &- \beta_{x1}(l)\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)] +\\ &+ (\sigma_{x1}(l)q_{x1}(l) + \sigma_{x2}(l)q_{x2}(l))\sigma_{x1}(l),\end{aligned} \tag{3.8.13}$$

$$\begin{aligned}\frac{d\mathrm{E}[h_{x1}(t,l)h_{x3}(t,l)]}{dt} =& \gamma_x(l)\mathrm{E}[h_{x3}(t,l)] - \beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x3}^2(t,l)] +\\ &- \beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x2}(t,l)h_{x3}(t,l)] +\\ &- \beta_{x2}(l)\mathrm{E}[h_{x1}(t,l)h_{x3}(t,l)] +\\ &+ (\sigma_{x1}(l)q_{x1}(l) + \sigma_{x2}(l)q_{x2}(l))\sigma_{x2}(l),\end{aligned} \tag{3.8.14}$$

$$\begin{aligned}\frac{d\mathrm{E}[h_{x2}(t,l)h_{x3}(t,l)]}{dt} =& -\beta_{x1}(l)\mathrm{E}[h_{x2}(t,l)h_{x3}(t,l)] +\\ &- \beta_{x2}(l)\mathrm{E}[h_{x2}(t,l)h_{x3}(t,l)] + \sigma_{x1}(l)\sigma_{x2}(l).\end{aligned} \tag{3.8.15}$$

MPHM-2DF model with a part of stationary moments

By equating to zero the derivatives in equations (3.8.8)–(3.8.9), we find the following conditions for stationary first order moments

$$\mathrm{E}[h_{x2}(l)] = \mathrm{E}[h_{x3}(l)] = 0. \tag{3.8.16}$$

From equation (3.8.7) we have

$$\mathrm{E}[h_{x1}(t,l)] = \gamma_x(l)t + \gamma_{x0}(l), \tag{3.8.17}$$

where $\gamma_{x0}(l)$ is a constant in integration of equation (3.8.7).

In the particular case, it can be assumed that $\gamma_{x0}(l) = \ln \mu_x(0,l) = \mathrm{E}[h_{x1}(0,l)]$.

After equating to zero derivatives in (3.8.11)–(3.8.15) and taking into account conditions (3.8.16), we obtain the following equations

$$\mathrm{E}[h_{x2}^2(l)] = \frac{\sigma_{x1}^2(l)}{2\beta_{x1}(l)}, \tag{3.8.18}$$

$$\mathrm{E}[h_{x3}^2(l)] = \frac{\sigma_{x2}^2(l)}{2\beta_{x2}(l)}, \tag{3.8.19}$$

$$\mathrm{E}[h_{x2}(l)h_{x3}(l)] = \frac{\sigma_{x_1}(l)\sigma_{x2}(l)}{\beta_{x1}(l)+\beta_{x2}(l)}, \tag{3.8.20}$$

$$\begin{aligned}\mathrm{E}[h_{x1}(l)h_{x2}(l)] = &\frac{1}{\beta_{x1}(l)}\left(-\frac{q_{x1}\sigma_{x1}^2(l)}{2} - \beta_{x2}q_{x2}\frac{\sigma_{x1}(l)\sigma_{x2}(l)}{\beta_{x1}(l)+\beta_{x2}(l)}\right) + \\ &+ \frac{1}{\beta_{x1}(l)}\left(\sigma_{x1}(l)q_{x1}(l) + \sigma_{x2}(l)q_{x2}(l)\right)\sigma_{x1}(l),\end{aligned} \tag{3.8.21}$$

$$\begin{aligned}\mathrm{E}[h_{x1}(l)h_{x3}(l)] = &\frac{1}{\beta_{x2}(l)}\left(-\frac{q_{x2}\sigma_{x2}^2(l)}{2} - \beta_{x1}q_{x1}\frac{\sigma_{x1}(l)\sigma_{x2}(l)}{\beta_{x1}(l)+\beta_{x2}(l)}\right) + \\ &+ \frac{1}{\beta_{x2}(l)}\left(\sigma_{x1}(l)q_{x1}(l) + \sigma_{x2}(l)q_{x2}(l)\right)\sigma_{x2}(l).\end{aligned} \tag{3.8.22}$$

Let us introduce quantities (3.8.18)–(3.8.22) to equation (3.8.10). Then we obtain

$$\frac{d\mathrm{E}[h_{x1}^2(t,l)]}{dt} = 2\gamma_x(l)\mathrm{E}[h_{x1}(t,l)]. \tag{3.8.23}$$

Hence, from (3.8.17) and (3.8.23) we find

$$\mathrm{E}[h_{x1}^2(t,l)] = \gamma_x^2(l)t^2 + 2\gamma_x(l)\gamma_{x0}(l)t + c_{x0}(l), \tag{3.8.24}$$

where $c_{x0}(l)$ is a constant of integration of equation (3.8.23).
In the particular case, there is $c_{x0}(l) = \ln^2\mu_x(0,l) = \mathrm{E}[h_{x1}^2(0,l)]$.

3.8.3. MPH model with two independent filters

Let us consider the Milevsky–Promislow Hybrid Model with 2 Independent Linear Scalar Filters (MPH-2IF model) of the form

$$\mu_x(t,l) = \mu_x(0,l)\exp\{\gamma_x(l)t + q_{x1}(l)z_1(t,l) + q_{x2}(l)z_2(t,l)\}, \quad (3.8.25)$$

$$dz_1(t,l) = -\beta_{x1}(l)z_1(t,l)dt + \sigma_{x1}(l)dw_1(t), \quad (3.8.26)$$

$$dz_2(t,l) = -\beta_{x2}(l)z_2(t,l)dt + \sigma_{x2}(l)dw_2(t), \quad (3.8.27)$$

where $t \in \mathbb{R}^+$, $l \in \mathbb{S}$ and $\gamma_x(l)$, $q_{x1}(l), q_{x2}(l)$, $\mu_x(0,l)$, $\beta_{x1}(l)$, $\beta_{x2}(l)$, $\sigma_{x1}(l), \sigma_{x2}(l) > 0$ are constant parameters, and $w_1(t), w_2(t)$ are independent Wiener processes.

By taking logarithms on both sides of equality (3.8.25) and using the Itô formula, we receive the following equation

$$\begin{gathered} d\ln\mu_x(t) = [\gamma_x(l) - \beta_{x1}(l)q_{x1}(l)z_1(t) - \beta_{x2}(l)q_{x2}(l)z_2(t)]dt + \\ + \sigma_{x1}(l)q_{x1}(l)dw_1(t) + \sigma_{x2}(l)q_{x2}(l)dw_2(t). \end{gathered} \quad (3.8.28)$$

We assume that $\beta_{x1}(l) \neq \beta_{x2}(l)$. Let us introduce a new state vector

$$\begin{gathered} \mathbf{h}_x(t,l) = [h_{x1}(t,l), h_{x2}(t,l), h_{x3}(t,l)]^T = \\ = [\ln\mu_x(t,l), z_1(t,l), z_2(t,l)]^T. \end{gathered} \quad (3.8.29)$$

Then equations (3.8.26)–(3.8.28) can be written as vector equations

$$\begin{gathered} d\mathbf{h}_x(t,l) = \left\{ \begin{bmatrix} 0 & -\beta_{x1}(l)q_{x1}(l) & -\beta_{x2}(l)q_{x2}(l) \\ 0 & -\beta_{x1}(l) & 0 \\ 0 & 0 & -\beta_{x2}(l) \end{bmatrix} \mathbf{h}_x(t,l) + \begin{bmatrix} \gamma_x(l) \\ 0 \\ 0 \end{bmatrix} \right\} dt + \\ + \begin{bmatrix} \sigma_{x1}(l)q_{x1}(l) \\ \sigma_{x1}(l) \\ 0 \end{bmatrix} dw_1(t) + \begin{bmatrix} \sigma_{x2}(l)q_{x2}(l) \\ 0 \\ \sigma_{x2}(l) \end{bmatrix} dw_2(t), \end{gathered} \quad (3.8.30)$$

For the simplified version, we can assume that $q_{x1}(l) = q_{x2}(l) = 1$.

3.8.4. MPH moment model with two independent filters

The moment equations of system (3.8.30) lead to the Milevsky–Promislow Hybrid Moment model with 2 Independent Filters (MPHM-2IF model) of the form

$$\frac{d\mathrm{E}[h_{x1}(t,l)]}{dt} = \gamma_x(l) - \beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x2}(t,l)]+ \\ - \beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x3}(t,l)], \tag{3.8.31}$$

$$\frac{d\mathrm{E}[h_{x2}(t,l)]}{dt} = -\beta_{x1}(l)\mathrm{E}[h_{x2}(t,l)], \tag{3.8.32}$$

$$\frac{d\mathrm{E}[h_{x3}(t,l)]}{dt} = -\beta_{x2}(l)\mathrm{E}[h_{x3}(t,l)], \tag{3.8.33}$$

$$\frac{d\mathrm{E}[h_{x1}^2(t,l)]}{dt} = 2\gamma_x(l)\mathrm{E}[h_{x1}(t,l)] - 2\beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)] \\ - 2\beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x1}(t,l)h_{x3}(t,l)]+ \\ + \sigma_{x1}^2(l)q_{x1}^2(l) + \sigma_{x2}^2(l)q_{x2}^2(l), \tag{3.8.34}$$

$$\frac{d\mathrm{E}[h_{x2}^2(t,l)]}{dt} = -2\beta_{x1}(l)\mathrm{E}[h_{x2}^2(t,l)] + \sigma_{x1}^2(l), \tag{3.8.35}$$

$$\frac{d\mathrm{E}[h_{x3}^2(t,l)]}{dt} = -2\beta_{x2}(l)\mathrm{E}[h_{x3}^2(t,l)] + \sigma_{x2}^2(l), \tag{3.8.36}$$

$$\frac{d\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)]}{dt} = \gamma_x(l)\mathrm{E}[h_{x2}(t,l)] - \beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x2}^2(t,l)]+ \\ - \beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x2}(t,l)h_{x3}(t,l)]+ \\ - \beta_{x1}(l)\mathrm{E}[h_{x1}(t,l)h_{x2}(t,l)] + \sigma_{x1}^2(l)q_{x1}(l), \tag{3.8.37}$$

$$\frac{d\mathrm{E}[h_{x1}(t,l)h_{x3}(t,l)]}{dt} = \gamma_x(l)\mathrm{E}[h_{x_3}(t,l)] - \beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x3}^2(t,l)]+ \\ - \beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x2}(t,l)h_{x3}(t,l)]+ \\ - \beta_{x2}(l)\mathrm{E}[h_{x1}(t,l)h_{x3}(t,l)] + \sigma_{x2}^2(l)q_{x2}(l), \tag{3.8.38}$$

$$\frac{d\text{E}[h_{x2}(t,l)h_{x3}(t,l)]}{dt} = -\ (\beta_{x1}(l) + \beta_{x2}(l))\text{E}[h_{x2}(t,l)h_{x3}(t,l)]. \quad (3.8.39)$$

MPHM-2IF model with a part of stationary moments

By equating to zero the derivatives in (3.8.32) and (3.8.33), we receive the following conditions for stationary moments

$$\text{E}[h_{x2}(l)] = \text{E}[h_{x3}(l)] = 0. \quad (3.8.40)$$

From equation (3.8.31) we find

$$\text{E}[h_{x1}(t,l)] = \gamma_x(l)t + \gamma_{x0}(l), \quad (3.8.41)$$

where $\gamma_{x0}(l)$ is a constant in integration of equation (3.8.31).

In the particular case, it can be assumed that $\gamma_{x0}(l) = \ln \mu_x(0,l) = \text{E}[h_{x1}(0,l)]$ and $\ln \mu_x(0,l)$ is a constant parameter.

By equating to zero the derivatives in (3.8.35)–(3.8.39) and taking into account conditions (3.8.41), we obtain

$$\text{E}[h_{x2}^2(l)] = \frac{\sigma_{x1}^2(l)}{2\beta_{x1}(l)}, \quad (3.8.42)$$

$$\text{E}[h_{x3}^2(l)] = \frac{\sigma_{x2}^2(l)}{2\beta_{x2}(l)}, \quad (3.8.43)$$

$$\text{E}[h_{x2}(l)h_{x3}(l)] = 0, \quad (3.8.44)$$

$$\text{E}[h_{x1}(l)h_{x2}(l)] = \frac{q_{x1}(l)\sigma_{x1}^2(l)}{2\beta_{x1}(l)}, \quad (3.8.45)$$

$$\text{E}[h_{x1}(l)h_{x3}(l)] = \frac{q_{x2}(l)\sigma_{x2}^2(l)}{2\beta_{x2}(l)}. \quad (3.8.46)$$

By substituting quantities (3.8.45) and (3.8.46) for $\text{E}[h_{x1}(l)h_{x2}(l)]$ and $\text{E}[h_{x1}(l)h_{x3}(l)]$ in the expression (3.8.34), we obtain

$$\frac{d\text{E}[h_{x1}^2(t,l)]}{dt} = 2\gamma_x(l)\text{E}[h_{x1}(t,l)]. \quad (3.8.47)$$

Hence, from (3.8.41) and (3.8.47) we find

$$\text{E}[h_{x1}^2(t,l)] = \gamma_x^2(l)t^2 + 2\gamma_x(l)\gamma_{x0}(l)t \ + c_{x0}(l), \quad (3.8.48)$$

where $c_{x0}(l)$ is a constant of integration of equation (3.8.47).
In the particular case, we have $c_{x0}(l) = \ln^2\mu_x(0,l) = \text{E}[h_{x1}^2(0,l)]$.

3.8.5. MPH model with a vector linear filter

Let us replace the one-dimensional filter (3.5.2) in the MPH model with a vector linear filter. As a result, we obtain the Milevsky–Promislow Hybrid model with a Vector Linear Filter (MPH-VLF model) of the following form

$$\mu_x(t,l) = \mu_x(0,l)\exp\{\gamma_x(l)t + \mathbf{q}_x^T(l)\mathbf{z}_x(t,l)\}, \tag{3.8.49}$$

$$d\mathbf{z}_x(t,l) = \mathbf{A}_x(l)\mathbf{z}_x(t,l)dt + \mathbf{G}_x(t,l)dw(t),\ \ \mathbf{z}_x(t_0,l) = \mathbf{z}_{x0l}, \tag{3.8.50}$$

where

$t \in \mathbb{R}^+$, $l \in \mathbb{S}$,
$\mathbf{z}_x(t,l) \in \mathbb{R}^n$ is a filter vector,
$\mathbf{z}_{x0l} \in \mathbb{R}^n$ is an initial condition filter vector,
$\mathbf{A}_x(l)$ are $n \times n$ constant stable matrices, i.e. $\mathrm{Re}\,\lambda_i(\mathbf{A}_x(l)) < 0$, $i = 1, \ldots, n$,
$\mathbf{q}_x(l) \in \mathbb{R}^n$ are constant vectors,
$\mathbf{G}_x(t,l) = [G_x^1(t,l), \ldots, G_x^n(t,l)]^T$,
$G_x^i(t,l)$ are the deterministic, non-linear functions of time representing filter dynamics,
$\gamma_x(l)$, $\mu_x(0,l) > 0$ are constant parameters,
$w(t)$ is a standard Wiener process.

Model (3.8.49)–(3.8.50) represented by the Itô stochastic differential vector equation has the following form

$$\begin{aligned} &d\ln\mu_x(t,l) = \\ &= [\gamma_x(l) + \sum_{i=1}^{n} q_x^i(l)\mathbf{A}_x^i(l)\mathbf{z}_x(t,l)]dt + \sum_{i=1}^{n} q_x^i(l)G_x^i(t,l)dw(t), \end{aligned} \tag{3.8.51}$$

$$d\mathbf{z}_x(t,l) = \mathbf{A}_x(l)\mathbf{z}_x(t,l)dt + \mathbf{G}_x(t,l)dw(t), \quad \mathbf{z}_x(t_0,l) = \mathbf{z}_{x0l}, \tag{3.8.52}$$

where $\mathbf{A}_x^i(l)$ is the i-th row of matrix $\mathbf{A}_x(l)$, and $q_x^i(l)$, $G_x^i(t,l)$ are the i-th coordinates of vectors $\mathbf{q}_x(l)$, $\mathbf{G}_x(t,l)$, respectively.

To find an analytical solution of equation (3.8.52), we will use a special case of relation (2.3.8) for $\mathbf{A}_0 = \mathbf{0}$, $M = 1$ and one dimensional noise $w(t)$. The solution is the following

$$\begin{aligned} &\mathbf{z}_x(t,l) = \\ &= \exp\{\mathbf{A}_x(l)(t - t_{0l})\}\mathbf{z}_{x0l} + \int_{t_{0l}}^{t} \exp\{\mathbf{A}_x(l)(t-s)\}\mathbf{G}_x(s,l)dw(s). \end{aligned} \tag{3.8.53}$$

By introducing (3.8.53) to equality (3.8.49) we obtain

$$\mu_x(t,l)=\mu_x(0,l)\exp\left\{\gamma_x(l)t+\mathbf{q}_x^T(l)\exp\{\mathbf{A}_x(l)(t-t_{0l})\}\mathbf{z}_{0l}+\right.$$
$$\left.+\int_{t_{0l}}^{t}\mathbf{q}_x^T(l)\exp\{\mathbf{A}_x(l)(t-s)\}\mathbf{G}_x(s,l)dw(s)\right\}. \tag{3.8.54}$$

The above solutions for the subsystems will be used to construct the solution of a hybrid model.

We assume that scalar stochastic process $\mu_x(t)$, solving the scalar hybrid stochastic equation and starting at moment t_0, switches at times $\tau_1,\ldots,\tau_M$. We also assume that $\tau_0 = t_0$ and that in time intervals $[\tau_i,\tau_{i+1})$ the hybrid system is in states $l_i \in \mathbb{S}$, $i = 0,\ldots,M$, where $l_0, l_1,\ldots,l_M$ is any subsequence of N states.

The continuity of solutions is also assumed, meaning that the value of the process in state l_i at time τ_i and the value of the process in state l_{i-1} at time τ_i are the same, i.e. $\mu_x(\tau_i,l_i) = \mu_x(\tau_i,l_{i-1})$. The solution for $t \in [\tau_i,\tau_{i+1})$ is then written as

$$\mu_x(t,l_i) = \mu_x(\tau_i,l_{i-1})\exp\left\{\gamma_x(l_i)t+\right.$$
$$+\mathbf{q}_x^T(l_i)\exp\{\mathbf{A}_x(l_i)(t-\tau_i)\}\mathbf{z}(\tau_i,l_{i-1})+ \tag{3.8.55}$$
$$\left.+\int_{\tau_i}^{t}\mathbf{q}_x^T(l_i)\exp\{\mathbf{A}_x(l_i)(t-s)\}\mathbf{G}_x(s,l_i)dw(s)\right\}.$$

3.8.6. MPH moment model with a vector linear filter

The family of the first- and second-order moments of the subsystems describing the Milevsky–Promislow Hybrid Moment model with a Vector Linear Filter (MPHM-VLF model) is represented by the following formulas

$$\mathrm{E}[\mu_x(t,l)] = \mathrm{E}[\mu_x(0,l)]\exp\{\gamma_x(l)t+\mathbf{q}_x^T(l)\mathrm{E}[\mathbf{z}_x(t,l)]+$$
$$+\frac{1}{2}tr\{\mathbf{Q}_x(l)\mathrm{cov}[\mathbf{z}_x(t,l)]\}, \tag{3.8.56}$$

$$\mathrm{E}[\mu_x^2(t,l)] = \mathrm{E}[\mu_{x0}^2(l)] \exp\{2\gamma_x(l)t + 2\mathbf{q}_x^T(l)\mathrm{E}[\mathbf{z}_x(t,l)] + \\ + tr\{\mathbf{Q}_x(l)\mathrm{cov}[\mathbf{z}_x(t,l)]\}, \tag{3.8.57}$$

where

$$\mathbf{Q}_x(l) = \mathbf{q}_x(l)\mathbf{q}_x^T(l), \\ \mathrm{cov}[\mathbf{z}_x(t,l)] = \mathrm{E}[\mathbf{z}_x(t,l)\mathbf{z}_x^T(t,l)] - \mathrm{E}[\mathbf{z}_x(t,l)]\mathrm{E}[\mathbf{z}_x^T(t,l)] \tag{3.8.58}$$

and

$$\frac{d\mathrm{E}[\mathbf{z}_x(t,l)]}{dt} = \mathbf{A}_x(l)\mathrm{E}[\mathbf{z}_x(t,l)], \\ \frac{d\mathrm{E}[\mathbf{z}_x(t,l)\mathbf{z}_x^T(t,l)]}{dt} = \mathbf{A}_x(l)\mathrm{E}[\mathbf{z}_x(t,l)\mathbf{z}_x^T(t,l)] + \\ + \mathrm{E}[\mathbf{z}_x(t,l)\mathbf{z}_x^T(t,l)]\mathbf{A}_x^T(l) + \mathbf{G}_x(t,l)\mathbf{G}_x^T(t,l). \tag{3.8.59}$$

It follows from (3.8.59) that for stationary solutions, $\mathrm{E}[\mathbf{z}_x(t,l)] = \mathbf{0}$. Consequently, there is $\mathrm{cov}[\mathbf{z}_x(t,l)] = \mathrm{E}[\mathbf{z}_x(t,l)\mathbf{z}_x^T(t,l)]$ for $l \in \mathbb{S}$ and relationships (3.8.56) and (3.8.57) are then written, respectively, as

$$\mathrm{E}[\mu_x(t,l)] = \\ = \mathrm{E}[\mu_x(0,l)] \exp\{\gamma_x(l)t + \frac{1}{2}tr\{\mathbf{Q}_x(l)\mathrm{E}[\mathbf{z}_x(t,l)\mathbf{z}_x^T(t,l)]\}, \tag{3.8.60}$$

$$\mathrm{E}[\mu_x^2(t,l))] = \\ = \mathrm{E}[\mu_{x0}^2(l)] \exp\{2\gamma_x(l)t + tr\{\mathbf{Q}_x(l)\mathrm{E}[\mathbf{z}_x(t,l)\mathbf{z}_x^T(t,l)]\}. \tag{3.8.61}$$

As before, solutions obtained for the moments of the subsystems will be used to solve the moment hybrid model.

We assume that the first two raw moments of the scalar stochastic process $\mathrm{E}[\mu_x(t)]$, $\mathrm{E}[\mu_x^2(t)]$, constituting the solutions of the scalar moment hybrid differential equations and starting at time t_0, switch at times $\tau_1, \tau_2, \ldots, \tau_M$. We also assume that $\tau_0 = t_0$ and that in time intervals $[\tau_i, \tau_{i+1})$ the moment hybrid system is in states $l_i \in \mathbb{S}$, $i = 0, \ldots, M$, where $l_0, l_2, \ldots, l_M$ is any subsequence of N states.

The continuity of solutions is also assumed, meaning that the values of the first and second moments of the process in state l_i at time τ_i are equal to the respective values of the first and second moments of the process in state l_{i-1} at time τ_i, hence

$$\mathrm{E}[\mu_x(\tau_i, l_i)] = \mathrm{E}[\mu_x(\tau_i, l_{i-1})], \quad \mathrm{E}[\mu_x^2(\tau_i, l_i)] = \mathrm{E}[\mu_x^2(\tau_i, l_{i-1})]. \quad (3.8.62)$$

Then, the respective solutions for $t \in [\tau_i, \tau_{i+1})$ are written as follows

$$\begin{aligned} \mathrm{E}[\mu_x(t, l_i)] = \mathrm{E}[\mu_x(\tau_i, l_{i-1})] \exp\{\gamma_x(l_i)t + \\ + \frac{1}{2} tr\{\mathbf{Q}_x(l_i)\mathrm{E}[\mathbf{z}_x(t, l_i)\mathbf{z}_x^T(t, l_i)]\}\}, \end{aligned} \quad (3.8.63)$$

$$\begin{aligned} \mathrm{E}[\mu_x^2(t, l_i)] = \mathrm{E}[\mu_x^2(\tau_i, l_{i-1})] \exp\{2\gamma_x(l_i) + \\ + tr\{\mathbf{Q}_x(l_i)\mathrm{E}[\mathbf{z}_x(t, l_i)\mathbf{z}_x^T(t, l_i)]\}\}, \end{aligned} \quad (3.8.64)$$

where matrix $\mathbf{\Gamma}_x(t, l_i) = \mathrm{E}[\mathbf{z}_x(t, l_i)\mathbf{z}_x^T(t, l_i)]$ satisfies for $t \in [\tau_i, \tau_{i+1})$ the recurrence equation

$$\begin{aligned} \frac{d\mathbf{\Gamma}_x(t,l_i)}{dt} = \mathbf{A}_x(l_i)\mathbf{\Gamma}_x(t,l_i) + \mathbf{\Gamma}_x(t,l_i)\mathbf{A}_x^T(l_i) + \mathbf{G}_x(t,l_i)\mathbf{G}_x^T(t,l_i), \\ \mathbf{\Gamma}_x(\tau_i, l_i) = \mathbf{\Gamma}_x(\tau_i, l_{i-1}). \end{aligned} \quad (3.8.65)$$

Equation (3.8.65) for the elements of the matrix

$$\mathbf{\Gamma}_x(t, l_i) = [\Gamma_{xkj}(t, l_i)] = [\mathrm{E}[z_{xk}(t, l_i)z_{xj}(t, l_i)]] \quad (3.8.66)$$

can be written as

$$\begin{aligned} \frac{d\Gamma_{xkj}(t, l_i)}{dt} = \sum_{q=1}^{n} [a_{xkq}(t,l_i)\Gamma_{xqj}(t, l_i) + a_{xjq}(t,l_i)\Gamma_{xqk}(t,l_i)] + \\ + G_{xk}(t, l_i)G_{xj}(t, l_i), \\ \Gamma_{xkj}(\tau_i, l_i) = \Gamma_{xkj}(\tau_i, l_{i-1}), \quad k, j = 1, ..., n. \end{aligned} \quad (3.8.67)$$

In the special case of $G_{xk}(t,l_i) = g_{xk}(l_i)\exp\{a_{xk}(l_i)t\}$, equation (3.8.67) is written as

$$\frac{d\Gamma_{xkj}(t,l_i)}{dt} = \sum_{q=1}^{n}[a_{xkq}(t,l_i)\Gamma_{xqj}(t,l_i)+a_{xjq}(t,l_i)\Gamma_{xqk}(t,l_i)]+ \\ + g_{xk}(t,l_i)g_{xj}(t,l_i)\exp\{(a_{xk}(l_i)+a_{xj}(l_i))t\}, \qquad (3.8.68)$$

$$\Gamma_{xkj}(\tau_i,l_i) = \Gamma_{xkj}(\tau_i,l_{i-1}), \quad k,j = 1,...,n.$$

Example 3.3. Let us consider a model with a two-dimensional linear filter represented by equations (3.8.49) and (3.8.50), where matrices $\mathbf{A}_x(l)$ and vectors $\mathbf{G}_x(l)$, $\mathbf{q}_x(l)$, $l = 1,2$ are defined as follows

$$\mathbf{A}_x(1)=\begin{bmatrix} 0 & 1 \\ -\omega_{x0}^2(1) & 2\nu_x(1)\end{bmatrix}, \ \mathbf{G}_x(1)=\begin{bmatrix} 0 \\ g_{x1}e^{a_{x1}}\end{bmatrix}, \ \mathbf{q}_x(1)=\begin{bmatrix} q_{x1} \\ 0\end{bmatrix}, \qquad (3.8.69)$$

$$\mathbf{A}_x(2)=\begin{bmatrix} 0 & 1 \\ -\omega_{x0}^2(2) & 2\nu_x(2)\end{bmatrix}, \ \mathbf{G}_x(2)=\begin{bmatrix} 0 \\ g_{x2}e^{a_{x2}}\end{bmatrix}, \ \mathbf{q}_x(2)=\begin{bmatrix} 0 \\ q_{x2}\end{bmatrix}, \qquad (3.8.70)$$

where $\mu_x(0,l), \gamma_x(l), \omega_{x0}(l), \nu_x(l), g_{xi}(l), a_{xi}(l)$, $l = 1,2$, $i = 1,2$ are constant parameters.

In this case, two models of the subsystems are described by the following equalities

$$\mu_x(t,1) = \mu_{x0}(1)\exp\{\gamma_x(1)t + q_{x1}z_{x1}(t,1)\}, \qquad (3.8.71)$$

$$d\mathbf{z}_x(t,1) = \mathbf{A}_x(1)\mathbf{z}_x(t,1)dt + \mathbf{G}_x(t,1)dw(t) \qquad (3.8.72)$$

and

$$\mu_x(t,2) = \mu_{x0}(2)\exp\{\gamma_x(2)t + q_{x2}z_{x2}(t,2)\}, \qquad (3.8.73)$$

$$d\mathbf{z}_x(t,2) = \mathbf{A}_x(2)\mathbf{z}_x(t,2)dt + \mathbf{G}_x(t,2)dw(t). \qquad (3.8.74)$$

Let us assume that the system in a state $l = 1$ operates under initial conditions for $t_0 = 0$ $\mathbf{z}_x(0,1) = [z_{x1}(0,1), z_{x2}(0,1)]^T = [z_{x10}, z_{x20}]^T$ and $\mu_{x0}(1) = \mu_{x0}$, and that the state changes into a state $l = 2$ at time $t = \tau_1$.

Then stochastic equations (3.8.72) and (3.8.74) have the following solutions

$$\mathbf{z}_x(t,1) = \exp\{\mathbf{A}_x(1)(t-0)\}\mathbf{z}_x(0,1) + \\ + \int_0^t \exp\{\mathbf{A}_x(1)(t-s)\}\mathbf{G}_x(s,1)dw(s) \text{ for } t \in [0,\tau_1), \tag{3.8.75}$$

$$\mathbf{z}_x(t,2) = \exp\{\mathbf{A}_x(2)(t-\tau_1)\}\mathbf{z}_x(\tau_1,1) + \\ + \int_{\tau_1}^t \exp\{\mathbf{A}_x(2)(t-s)\}\mathbf{G}_x(s,2)dw(s) \text{ for } \ t \geq \tau_1. \tag{3.8.76}$$

The family of the first- and second-order moments of subsystems representing the MPH-VLF model has the following form

$$\mathrm{E}[\mu_x(t,1)] = \mathrm{E}[\mu_{x0}(1)] \exp\{\gamma_x(1)t + \mathbf{q}_x^T(1)\mathrm{E}[\mathbf{z}_x(t,1)] + \\ + \frac{1}{2} tr\{\mathbf{Q}_x(1)\mathrm{cov}[\mathbf{z}_x(t,1)]\}\},$$

$$\mathrm{E}[\mu_x(t,2)] = \mathrm{E}[\mu_{x0}(2)] \exp\{\gamma_x(2)t + \mathbf{q}_x^T(2)\mathrm{E}[\mathbf{z}_x(t,2)] + \\ + \frac{1}{2} tr\{\mathbf{Q}_x(2)\mathrm{cov}[\mathbf{z}_x(t,2)]\}\}, \tag{3.8.77}$$

$$\mathrm{E}[\mu_x^2(t,1)] = \mathrm{E}[\mu_{x0}^2(1)] \exp\{2\gamma_x(1)t + 2\mathbf{q}_x^T(1)\mathrm{E}[\mathbf{z}_x(t,1)] + \\ + tr\{\mathbf{Q}_x(1)\mathrm{cov}[\mathbf{z}_x(t,1)]\}\},$$

$$\mathrm{E}[\mu_x^2(t,2)] = \mathrm{E}[\mu_{x0}^2(2)] \exp\{2\gamma_x(2)t + 2\mathbf{q}_x^T(2)\mathrm{E}[\mathbf{z}_x(t,2)] + \\ + tr\{\mathbf{Q}_x(2)\mathrm{cov}[\mathbf{z}_x(t,2)]\}\}, \tag{3.8.78}$$

where

$$\mathbf{Q}_x(1) = \begin{bmatrix} q_{x1}^2 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{Q}_x(2) = \begin{bmatrix} 0 & 0 \\ 0 & q_{x2}^2 \end{bmatrix}. \tag{3.8.79}$$

The moments of processes $\mathbf{z}_x(t,l)$, $l=1,2$ satisfy equations

$$\frac{d\mathrm{E}[\mathbf{z}_x(t,l)]}{dt}=\mathbf{A}_x(l)\mathrm{E}[\mathbf{z}_x(t,l)], \tag{3.8.80}$$

$$\begin{gathered}\frac{d\mathrm{E}[\mathbf{z}_x(t,l)\mathbf{z}_x^T(t,l)]}{dt}=\mathbf{A}_x(l)\mathrm{E}[\mathbf{z}_x(t,l)\mathbf{z}_x^T(t,l)]+\\ +\mathrm{E}[\mathbf{z}_x(t,l)\mathbf{z}_x^T(t,l)]\mathbf{A}_x^T(l)+\mathbf{G}_x(t,l)\mathbf{G}_x^T(t,l).\end{gathered} \tag{3.8.81}$$

From equation (3.8.80) it follows that for the stationary solution of equation (3.8.80) we have $\mathrm{E}[\mathbf{z}_x(t,l)]=\mathbf{0}$. Therefore, we get

$$\begin{gathered}\mathrm{cov}[\mathbf{z}_x(t,l)]=\mathrm{E}[\mathbf{z}_x(t,l)\mathbf{z}_x^T(t,l)]=\\ =\begin{bmatrix}\mathrm{E}[z_{x1}^2(t,l)] & \mathrm{E}[z_{x1}(t,l)z_{x2}(t,l)]\\ \mathrm{E}[z_{x1}(t,l)z_{x2}(t,l)] & \mathrm{E}[z_{x2}^2(t,l)]\end{bmatrix}.\end{gathered} \tag{3.8.82}$$

For $l=1,2$ coordinate equation (3.8.81) takes the following form

$$\begin{gathered}\frac{d\mathrm{E}[z_{x1}^2(t,l)]}{dt}=2\mathrm{E}[z_{x1}(t,l)z_{x2}(t,l)],\quad \mathrm{E}[z_{x1}^2(0,l)]=\Gamma_{x11_0},\\ \frac{d\mathrm{E}[z_{x1}(t,l)z_{x2}(t,l)]}{dt}=\mathrm{E}[z_{x2}^2(t,l)]-\omega_{x0}(l)\mathrm{E}[z_{x1}^2(t,l)]+\\ +2\nu_x(l)\mathrm{E}[z_{x1}(t,l)z_{x2}(t,l)],\quad \mathrm{E}[z_{x1}(t,l)z_{x2}(0,l)]=\Gamma_{x12_0},\\ \frac{d\mathrm{E}[z_{x2}^2(t,l)]}{dt}=-2\omega_{x0}(l)\mathrm{E}[z_{x1}(t,l)z_{x2}(t,l)]+\\ +4\nu_x(l)\mathrm{E}[z_{x2}^2(t,l)]+g_{xl}^2e^{2a_{xl}t},\quad \mathrm{E}[z_{x2}^2(0,l)]=\Gamma_{x22_0},\end{gathered} \tag{3.8.83}$$

when initial conditions $\Gamma_{x11_0},\Gamma_{x12_0},\Gamma_{x22_0}>0$ are positive,

$$\Gamma_{x11_0}\Gamma_{x22_0}-\Gamma_{x12_0}^2>0. \tag{3.8.84}$$

From equalities (3.8.79) and (3.8.82) it follows that the respective relationships (3.8.77) can be expressed as

$$E[\mu_x(t,1)] = E[\mu_{x0}(1)]\exp\{\gamma_x(1)t + \frac{1}{2}q_{x1}^2 E[z_{x1}^2(t,1)]\},$$
$$E[\mu_x(t,2)] = E[\mu_{x0}(2)]\exp\{\gamma_x(2)t + \frac{1}{2}q_{x2}^2 E[z_{x2}^2(t,1)]\} \tag{3.8.85}$$

and

$$E[\mu_x^2(t,1)] = E[\mu_{x0}^2(1)]\exp\{2\gamma_x(1)t + q_{x1}^2 E[z_{x1}^2(t,1)]\},$$
$$E[\mu_x^2(t,2)] = E[\mu_{x0}^2(2)]\exp\{2\gamma_x(2)t + q_{x2}^2 E[z_{x2}^2(t,2)]\}. \tag{3.8.86}$$

When the state $l=1$ switches to $l=2$ at time $t=\tau_1$, the first and second order moments of $\mu_x(t,1)$ in the model (3.8.49)–(3.8.50) are written as follows

$$E[\mu_x(t,1)] = E[\mu_{x0}(1)]\exp\{\gamma_x(1)t + \frac{1}{2}q_{x1}^2 E[z_{x1}^2(t,1)]\}, \quad 0 \le t < \tau_1,$$
$$E[\mu_x(t,2)] = E[\mu_x(\tau_1,1)]\exp\{\gamma_x(2)(t-\tau_1) + \frac{1}{2}q_{x2}^2 E[z_{x2}^2(t,1)]\}, \quad t \ge \tau_1, \tag{3.8.87}$$

$$E[\mu_x^2(t,1)] = E[\mu_{x0}^2(1)]\exp\{2\gamma_x(1)t + q_{x1}^2 E[z_{x1}^2(t,1)]\}, \; 0 \le t < \tau_1,$$
$$E[\mu_x^2(t,2)] = E[\mu_x^2(\tau_1,1)]\exp\{2\gamma_x(2)(t-\tau_1) + q_{x2}^2 E[z_{x2}^2(t,2)]\}, \quad t \ge \tau_1, \tag{3.8.88}$$

$$\frac{d\mathrm{E}[z_{x1}^2(t,1)]}{dt} = 2\mathrm{E}[z_{x1}(t,1)z_{x2}(t,1)], \quad \mathrm{E}[z_{x1}^2(0,1)] = \Gamma_{x11_0},$$

$$\frac{d\mathrm{E}[z_{x1}^2(t,2)]}{dt} = 2\mathrm{E}[z_{x1}(t,1)z_{x2}(t,1)], \tag{3.8.89}$$

$$\mathrm{E}[z_{x1}^2(\tau_1,2)] = \mathrm{E}[z_{x1}^2(\tau,1)],$$

$$\frac{d\mathrm{E}[z_{x2}^2(t,1)]}{dt} = -2\omega_{x0}(1)\mathrm{E}[z_{x1}(t,1)z_{x2}(t,1)]+$$

$$+4\nu_x(1)\mathrm{E}[z_{x2}^2(t,1)] + g_{x1}^2 e^{2a_{x1}t}, \tag{3.8.90}$$

$$\mathrm{E}[z_{x2}^2(0,1)] = \Gamma_{x22_0}, \quad 0 \le t < \tau_1,$$

$$\frac{d\mathrm{E}[z_{x2}^2(t,2)]}{dt} = -2\omega_{x0}(2)\mathrm{E}[z_{x1}(t,2)z_{x2}(t,2)]+$$

$$+4\nu_x(2)\mathrm{E}[z_{x2}^2(t,2)] + g_{x1}^2 e^{2a_{x2}(t-\tau_1)}, \tag{3.8.91}$$

$$\mathrm{E}[z_{x2}^2(\tau_1,2)] = \mathrm{E}[z_{x2}^2(\tau_1,1)], \quad t \ge \tau_1.$$

$$\frac{d\mathrm{E}[z_{x1}(t,1)z_{x2}(t,1)]}{dt} = \mathrm{E}[z_{x2}^2(t,1)] - \omega_{x0}(1)\mathrm{E}[z_{x1}^2(t,1)]+$$

$$+2\nu_x(1)\mathrm{E}[z_{x1}(t,1)z_{x2}(t,1)], \quad \mathrm{E}[z_{x1}(t,1)z_{x2}(0,1)] = \Gamma_{x12_0},$$

$$\frac{d\mathrm{E}[z_{x1}(t,2)z_{x2}(t,2)]}{dt} = \mathrm{E}[z_{x2}^2(t,1)] - \omega_{x0}(2)\mathrm{E}[z_{x1}^2(t,1)]+ \tag{3.8.92}$$

$$+2\nu_x(2)\mathrm{E}[z_{x1}(t,1)z_{x2}(t,1)], \quad \mathrm{E}[z_{x1}(\tau,2)z_{x2}(\tau,2)] =$$

$$= \mathrm{E}[z_{x1}(\tau,1)z_{x2}(\tau_1,1)],$$

3.8.7. Discrete MPH moment model with two dependent filters

From the moment equations (3.8.7)–(3.8.15) we define the Discrete Milevsky–Promislow Hybrid Moment model with 2 Dependent Linear Scalar Filters (DMPHM-2DF model)

$$\mathrm{E}[h_{x1}]_{i+1}(l) = \mathrm{E}[h_{x1}]_i(l)+\gamma_x(l)\delta, \tag{3.8.93}$$

$$\begin{aligned}\mathrm{E}[h_{x1}^2]_{i+1}(l)=&\mathrm{E}[h_{x1}^2]_i(l)+(2\gamma_x(l)\mathrm{E}[h_{x1}]_i(l)+\\ &-2\beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x1}h_{x2}]_i(l)-2\beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x1}h_{x3}]_i(l)+\\ &+(\sigma_{x1}(l)q_{x1}(l)+\sigma_{x2}(l)q_{x2}(l))^2)\delta,\end{aligned} \tag{3.8.94}$$

$$\mathrm{E}[h_{x2}^2]_{i+1}(l) = \mathrm{E}[h_{x2}^2]_i(l) + (-2\beta_{x1}(l)\mathrm{E}[h_{x2}^2]_i(l) + \sigma_{x1}^2(l))\delta, \tag{3.8.95}$$

$$\mathrm{E}[h_{x3}^2]_{i+1}(l) = \mathrm{E}[h_{x3}^2]_i(l) + (-2\beta_{x2}(l)\mathrm{E}[h_{x3}^2]_i(l) + \sigma_{x2}^2(l))\delta, \tag{3.8.96}$$

$$\begin{aligned}\mathrm{E}[h_{x1}h_{x2}]_{i+1}(l) =&\mathrm{E}[h_{x1}h_{x2}]_i(l)+(-\beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x2}^2]_i(l)+\\ &-\beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x2}h_{x3}]_i(l)-\beta_{x1}(l\mathrm{E}[h_{x1}h_{x2}]_i(l)+\\ &+(\sigma_{x1}(l)q_{x1}(l)+\sigma_{x2}(l)q_{x2}(l))\sigma_{x1}(l))\delta,\end{aligned} \tag{3.8.97}$$

$$\begin{aligned}\mathrm{E}[h_{x1}h_{x3}]_{i+1}(l) =&\mathrm{E}[h_{x1}h_{x3}](l)+(-\beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x2}h_{x3}]_i(l)+\\ &-\beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x3}^2]_i(l)-\beta_{x2}(l)\mathrm{E}[h_{x1}h_{x3}]_i(l)+\\ &+(\sigma_{x1}(l)q_{x1}(l)+\sigma_{x2}(l)q_{x2}(l))\sigma_{x2}(l))\delta,\end{aligned} \tag{3.8.98}$$

$$\begin{aligned}\mathrm{E}[h_{x2}h_{x3}]_{i+1}(l) =& \mathrm{E}[h_{x2}h_{x3}]_i(l)+(-\beta_{x1}(l)\mathrm{E}[h_{x2}h_{x3}]_i(l)+\\ &-\beta_{x2}(l)\mathrm{E}[h_{x2}h_{x3}]_i(l)+\sigma_{x1}(l)\sigma_{x2}(l))\delta,\end{aligned} \tag{3.8.99}$$

where

$$\mathrm{E}[h_{x1}]_i(l)=\mathrm{E}[h_{x1}](t_i,l),\ \mathrm{E}[h_{xj}^2]_i(l)=\mathrm{E}[h_{xj}^2](t_i,l),\ j=1,2,3,$$
$$\mathrm{E}[h_{x1}h_{x2}]_i(l)=\mathrm{E}[h_{x1}h_{x2}](t_i,l),$$
$$\mathrm{E}[h_{x1}h_{x3}]_i(l)=\mathrm{E}[h_{x1}h_{x3}](t_i,l),$$
$$\mathrm{E}[h_{x2}h_{x3}]_i(l)=\mathrm{E}[h_{x2}h_{x3}](t_i,l),\ \delta=t_{i+1}-t_i=\mathrm{const.} \tag{3.8.100}$$

3.8.8. Discrete MPH moment model with two independent filters

From the moment equations (3.8.31)–(3.8.39) we obtain the Discrete Milevsky–Promislow Hybrid Moment model with 2 Independent Filters (DMPHM-2IF model) expressed as

$$\mathrm{E}[h_{x1}]_{i+1}(l) = \mathrm{E}[h_{x1}]_i(l)+(\gamma_x(l)-\beta_x(l)q_x(l)\mathrm{E}[h_{x2}]_i(l))\delta, \tag{3.8.101}$$

$$\begin{aligned}\mathrm{E}[h_{x1}^2]_{i+1}(l) &= \mathrm{E}[h_{x1}^2]_i(l) + (2\gamma_x(l)\mathrm{E}[h_{x1}]_i(l)+\\ &- 2\beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x1}h_{x2}]_i(l) - 2\beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x1}h_{x3}]_i(l)+\\ &+ (\sigma_{x1}^2(l)q_{x1}^2(l) + \sigma_{x2}^2(l)q_{x2}^2(l)))\delta,\end{aligned} \tag{3.8.102}$$

$$\mathrm{E}[h_{x2}^2]_{i+1}(l) = \mathrm{E}[h_{x2}^2]_i(l) + (-2\beta_{x_1}(l)\mathrm{E}[h_{x2}^2]_i(l) + \sigma_{x1}^2(l))\delta, \tag{3.8.103}$$

$$\mathrm{E}[h_{x3}^2]_{i+1}(l) = \mathrm{E}[h_{x3}^2]_i(l) + (-2\beta_{x2}(l)\mathrm{E}[h_{x3}^2]_i(l) + \sigma_{x2}^2(l))\delta, \tag{3.8.104}$$

$$\begin{aligned}\mathrm{E}[h_{x1}h_{x2}]_{i+1}(l)&=\mathrm{E}[h_{x1}h_{x2}]_i(l)+(-\beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x2}^2]_i(l)+\\ &- \beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x2}h_{x3}]_i(l) - \beta_{x1}(l)\mathrm{E}[h_{x1}h_{x2}]_i(l)+\\ &+ \sigma_{x1}^2(l)q_{x1}(l))\delta,\end{aligned} \tag{3.8.105}$$

$$
\begin{aligned}
\mathrm{E}[h_{x1}h_{x3}]_{i+1}(l) = {} & \mathrm{E}[h_{x1}h_{x3}]_i(l) + (-\beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x2}h_{x3}]_i(l) + \\
& - \beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x3}^2]_i(l) - \beta_{x2}(l)\mathrm{E}[h_{x1}h_{x3}]_i(l) + \\
& + \sigma_{x2}^2(l)q_{x2}(l))\delta,
\end{aligned} \tag{3.8.106}
$$

$$
\begin{aligned}
\mathrm{E}[h_{x2}h_{x3}]_{i+1}(l) = {} & \mathrm{E}[h_{x2}h_{x3}]_i(l) + (-\beta_{x1}(l)\mathrm{E}[h_{x2}h_{x3}]_i(l) + \\
& - \beta_{x2}(l)\mathrm{E}[h_{x2}h_{x3}]_i(l))\delta,
\end{aligned} \tag{3.8.107}
$$

where

$$
\begin{aligned}
& \mathrm{E}[h_{x1}]_i(l) = \mathrm{E}[h_{x1}](t_i, l), \ \mathrm{E}[h_{xj}^2]_i(l) = \mathrm{E}[h_{xj}^2](t_i, l), \ j = 1, 2, 3, \\
& \mathrm{E}[h_{x1}h_{x2}]_i = \mathrm{E}[h_{x1}h_{x2}](t_i), \\
& \mathrm{E}[h_{x1}h_{x3}]_i(l) = \mathrm{E}[h_{x1}h_{x3}](t_i, l), \\
& \mathrm{E}[h_{x2}h_{x3}]_i(l) = \mathrm{E}[h_{x2}h_{x3}](t_i, l), \ \delta = t_{i+1} - t_i = \mathrm{const}.
\end{aligned} \tag{3.8.108}
$$

3.8.9. Discrete MPH model with a vector linear filter

The discrete-time version of the MPH model (3.8.49)–(3.8.50), further termed as the Discrete Milevsky–Promislow Hybrid model with a Vector Linear Filter (DMPH-VLF model), can be written using the following equations

$$
\begin{aligned}
y_x(t, l) = {} & \alpha_{x0}(l) + \alpha_{x1}(l)t + \mathbf{q}_x^T(l)\mathbf{z}_x(t, l), \quad t \in \mathbb{N}, \\
\mathbf{z}_x(t, l) = {} & (\mathbf{A}_x(l) + \mathbf{I})\mathbf{z}_x(t-1, l) + \mathbf{G}_x(l)\Delta w_t.
\end{aligned} \tag{3.8.109}
$$

To create the discrete solution of the hybrid model, the recurrence relationships (3.8.109) of the equation solving subsystems will be employed.

We assume that a scalar stochastic process $y_x(t) = \ln \mu_x(t)$, solving a scalar hybrid stochastic equation and starting at moment t_0, switches at times $\tau_1, \ldots, \tau_M$. We assume that $\tau_0 = t_0$ and that in time intervals

$[\tau_i, \tau_{i+1})$ the hybrid system is in states $l_i \in \mathbb{S}$, $i = 0, \ldots, M$, where $l_0, l_1, \ldots, l_M$ is any subsequence of N states.

The continuity of the solutions is also assumed, meaning that the value of the process in state l_i at time τ_i and the value of the process in state l_{i-1} at time τ_i are equal, i.e. $y_x(\tau_i, l_i) = y_x(\tau_i, l_{i-1})$. Then for $t \in [\tau_i, \tau_{i+1})$ the hybrid discrete equations are written as

$$y_x(t, l_i) = \alpha_{x0}(l_i) + \alpha_{x1}(l_i)(t - \tau_i) + \mathbf{q}_x^T(l_i)\mathbf{z}_x(t).$$

$$\mathbf{z}_x(t, l_i) = (\mathbf{A}(l_i) + \mathbf{I})\mathbf{z}_x(t-1, l_i) + \mathbf{G}(l_i)\Delta w_t, \tag{3.8.110}$$

$$z_x(\tau_i, l_i) = z_x(\tau_i, l_{i-1}), \quad y_x(\tau_i, l_i) = y_x(\tau_i, l_{i-1}).$$

3.8.10. Parameters' estimation of the DMPH moment models with two filters

Iterative estimation of the DMPHM-2DF model

The iterative estimation procedure for the DMPHM-2DF model (3.8.93)–(3.8.99) can be described as follows:

1^o Take constant initial values, e.g. $\mathrm{E}[h_{x1}h_{x2}]_0(l)=0$, $\mathrm{E}[h_{x1}h_{x3}]_0(l)=0$, $\mathrm{E}[h_{x2}^2]_0(l) = 1$, $\mathrm{E}[h_{x2}h_{x3}]_0(l) = 0$, $\mathrm{E}[h_{x3}^2]_0(l) = 1$ and initial conditions $\mathrm{E}[h_{x1}]_0(l) = p_x(l)$, $\mathrm{E}[h_{x1}^2]_0(l)= p_x^2(l)$, where $p_x(l)$ is a fixed parameter.

2^o Assume initial values for $p_x(l)$, $\gamma_x(l)$, $\beta_{x1}(l)$, $\beta_{x2}(l)$, $q_{x1}(l), q_{x2}(l)$, $\sigma_{x1}(l), \sigma_{x2}(l)$, e.g. $p_x(l) = \ln \mu_x(0, l)$, $\gamma_x(l) = 0.1$, $\beta_{x1}(l) = 0.1$, $\beta_{x2}(l) = 0.1$, $q_{x1}(l) = 1$, $q_{x2}(l) = 1$, $\sigma_{x1}(l) = 0.01$, $\sigma_{x_2}(l) = 0.01$.

3^o Estimate the successive values of $\mathrm{E}[h_{x1}]_i(l)$, $\mathrm{E}[h_{x1}^2]_i(l)$, $\mathrm{E}[h_{x1}h_{x2}]_i(l)$, $\mathrm{E}[h_{x1}h_{x3}]_i(l)$, $\mathrm{E}[h_{x2}^2]_i(l)$, $\mathrm{E}[h_{x2}h_{x3}]_i(l)$, $\mathrm{E}[h_{x3}^2]_i(l)$ from (3.8.93)–(3.8.99) for an i-th iteration ($i = 1, 2, ...$) and for the given values of $p_x(l) = \ln \mu_x(0, l)$, $\gamma_x(l)$, $\beta_{x1}(l)$, $\beta_{x2}(l)$, $q_{x1}(l)$, $q_{x2}(l)$, $\sigma_{x1}(l)$, $\sigma_{x2}(l)$.

4^o Determine the values of $\hat{\mathrm{E}}[h_{x1}]_i(l)$, $\hat{\mathrm{E}}[h_{x1}^2]_i(l)$, i.e. the log-central mortality rates and their squares from a sample time series.

5^o Minimize the following sum with respect to parameters $p_x(l) = \ln \mu_x(0,l)$, $\gamma_x(l)$, $\beta_{x1}(l)$, $\beta_{x2}(l)$, $q_{x1}(l)$, $q_{x2}(l)$, $\sigma_{x1}(l)$, $\sigma_{x2}(l)$

$$\begin{aligned} S = \sum_l \sum_i \Big(&\hat{\mathrm{E}}[h_{x1}]_{i+1}(l) - \mathrm{E}[h_{x1}]_i(l) - \gamma_x(l)\Big)^2 + \\ &+ \Big(\hat{\mathrm{E}}[h_{x1}^2]_{i+1}(l) - \mathrm{E}[h_{x1}^2]_i(l) - (2\gamma_x(l)\mathrm{E}[h_{x1}]_i(l) + \\ &-2\beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x1}h_{x_2}]_i(l) - 2\beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x1}h_{x3}]_i(l) + \\ &+(\sigma_{x1}(l)q_{x1}(l) + \sigma_{x2}(l)q_{x2}(l))^2)\Big)^2, \end{aligned} \tag{3.8.111}$$

or the following sum defined for $q_{x1}(l) = q_{x2}(l) = 1$

$$\begin{aligned} S = \sum_l \sum_i \Big(&\hat{\mathrm{E}}[h_{x1}]_{i+1}(l) - \mathrm{E}[h_{x1}]_i(l) - \gamma_x(l)\Big)^2 + \\ &+ \Big(\hat{\mathrm{E}}[h_{x1}^2]_{i+1}(l) - \mathrm{E}[h_{x1}^2]_i(l) - (2\gamma_x(l)\mathrm{E}[h_{x1}]_i(l) + \\ &-2\beta_{x1}(l)\mathrm{E}[h_{x1}h_{x2}]_i(l) - 2\beta_{x_2}(l)\mathrm{E}[h_{x1}h_{x3}]_i(l) .+ \\ &+(\sigma_{x1}(l) + \sigma_{x2}(l))^2)\Big)^2. \end{aligned} \tag{3.8.112}$$

Estimation of the DMPHM-2DF model with stationary first order moments

Let us consider the DMPHM-2DF model expressed by the moment equations (3.8.17)–(3.8.24). The estimation procedure for the DMPHM-2DF model reduces here to the minimization of square criterion

$$\begin{aligned} S = \sum_l \sum_{t \in I_l} &\Big(\hat{\mathrm{E}}[h_{x1}(t,l)] - \gamma_x(l)t - \gamma_0(l)\Big)^2 + \\ &+ \Big(\hat{\mathrm{E}}[h_{x1}^2(t,l)] - \gamma_x^2(l)t^2 - 2\gamma_x(l)\gamma_{x0}(l)t - c_{x0}(l)\Big)^2, \end{aligned} \tag{3.8.113}$$

where $\gamma_x(l), \gamma_{x0}(l)$ and $c_{x0}(l)$ are some parameters.

In the general case, criterion (3.8.113) is minimized with respect to three parameters $\gamma_x(l)$, $\gamma_{x0}(l)$ and $c_{x0}(l)$ for $l \in \mathbb{S}$. In the particular case, for $\gamma_{x0}(l) = \ln \mu_x(0,l)$ and $c_{x0}(l) = \ln^2 \mu_x(0,l)$, it is minimized with respect to parameter $\gamma_x(l)$ for $l \in \mathbb{S}$.

Iterative estimation of the DMPHM-2IF model

The estimation of the DMPHM-2IF model (3.8.101)–(3.8.107) consist of the following steps.

1^o Take some initial conditions, e.g. $\mathrm{E}[h_{x1}h_{x2}]_0(l)=0$, $\mathrm{E}[h_{x1}h_{x3}]_0(l)=0$, $\mathrm{E}[h_{x2}^2]_0(l) = 1$, $\mathrm{E}[h_{x2}h_{x3}]_0(l) = 0$, $\mathrm{E}[h_{x3}^2]_0(l) = 1$ and initial conditions $\mathrm{E}[h_{x1}]_0(l) = p_x(l)$, $\mathrm{E}[h_{x1}^2]_0(l) = p_x^2(l)$, where $p_x(l)$ is fixed.

2^o Assume initial values for $p_x(l) = \ln \mu_x(0,l)$, $\gamma_x(l)$, $\beta_{x1}(l)$ $\beta_{x2}(l)$, $q_{x1}(l)$, $q_{x2}(l)$, $\sigma_{x_1}(l)$, $\sigma_{x2}(l)$, e.g. $\gamma_x(l) = 0.1$, $\beta_{x1}(l) = 0.1$, $\beta_{x2}(l) = 0.1$, $q_{x1}(l) = 1$, $q_{x2}(l) = 1$, $\sigma_{x1}(l) = 0.01$, $\sigma_{x2}(l) = 0.01$.

3^o Calculate the values of $\mathrm{E}[h_{x1}]_i(l)$, $\mathrm{E}[h_{x1}^2]_i(l)$, $\mathrm{E}[h_{x2}^2]_i(l)$, $\mathrm{E}[h_{x3}^2]_i(l)$, $\mathrm{E}[h_{x_1}h_{x_2}]_i(l)$, $\mathrm{E}[h_{x_1}h_{x_3}]_i(l)$, $\mathrm{E}[h_{x2}h_{x3}]_i(l)$, from (3.8.101)–(3.8.107) for an i-th iteration ($i = 1, 2, ...$) and for the given values of parameters $p_x(l) = \ln \mu_x(0,l)$, $\gamma_x(l)$ $\beta_{x1}(l)$, $\beta_{x2}(l)$, $q_{x1}(l)$, $q_{x2}(l)$, $\sigma_{x_1}(l)$, $\sigma_{x2}(l)$.

4^o Determine the values of $\hat{\mathrm{E}}[h_{x1}]_i(l)$, $\hat{\mathrm{E}}[h_{x1}^2]_i(l)$, i.e. the log-central mortality rates and their squares form a sample time series.

4^o Minimize the following sum with respect to $p_x(l) = \ln \mu_x(0,l)$, $\gamma_x(l)$, $\beta_{x1}(l)$, $\beta_{x2}(l)$, $q_{x1}(l)$, $q_{x2}(l)$, $\gamma_{x1}^2(l)$, $\gamma_{x2}^2(l)$

$$
\begin{aligned}
S = \sum_l \sum_i & \Big(\hat{\mathrm{E}}[h_{x1}]_{i+1}(l) - \mathrm{E}[h_{x1}]_i(l) - \gamma_x(l)\Big)^2 + \\
& + \Big(\hat{\mathrm{E}}[h_{x1}^2]_{i+1}(l) - \mathrm{E}[h_{x1}^2]_i(l) - (2\gamma_x(l)\mathrm{E}[h_{x1}]_i(l) + \\
& - 2\beta_{x1}(l)q_{x1}(l)\mathrm{E}[h_{x1}h_{x2}]_i(l) - 2\beta_{x2}(l)q_{x2}(l)\mathrm{E}[h_{x1}h_{x3}]_i(l) + \\
& + \sigma_{x1}^2(l)q_{x1}^2(l) + \sigma_{x2}^2(l)q_{x2}^2(l))\Big)^2 ,
\end{aligned}
\tag{3.8.114}
$$

or the following sum defined for $q_{x1}(l)=q_{x2}(l)=1$

$$
\begin{aligned}
S=\sum_l\sum_i &\Big(\hat{\mathrm{E}}[h_{x1}]_{i+1}(l)-\mathrm{E}[h_{x1}]_i(l)-\gamma_x(l)\Big)^2+\\
&+\Big(\hat{\mathrm{E}}[h_{x1}^2]_{i+1}(l)-\mathrm{E}[h_{x1}^2]_i(l)-(2\gamma_x(l)\mathrm{E}[h_{x1}]_i(l)+\\
&-2\beta_{x1}(l)\mathrm{E}[h_{x1}h_{x2}]_i(l)-2\beta_{x_2}(l)\mathrm{E}[h_{x1}h_{x3}]_i(l)+\\
&+\sigma_{x1}^2(l)+\sigma_{x2}^2(l)))^2.
\end{aligned}
\tag{3.8.115}
$$

Let us notice that in the last case the unknown linear parameters in the system of moment equations (3.8.101)–(3.8.107) the unknown parameters are $p_x(l)=\ln\mu_x(0,l)$, $\gamma_x(l)$, $\beta_{x1}(l)$, $\beta_{x2}(l)$, $\sigma_{x1}^2(l)$, $\sigma_{x2}^2(l)$.

Estimation of the DMPHM-2IF model with stationary first order moments

We will consider here the DMPHM-2IF model expressed by the moment equations (3.8.41)–(3.8.48). The estimation procedure of the DMPHM-2IF model reduces to the minimization of square criterion (3.8.113) where $\gamma_x(l),\gamma_{x0}(l)$ and $c_{x0}(l)$ are some parameters.

In the general case, criterion (3.8.113) is minimized with respect to three parameters $\gamma_x(l)$, $\gamma_{x0}(l)$ and $c_{x0}(l)$ for $l\in\mathbb{S}$. For $\gamma_{x0}(l)=\ln\mu_x(0,l)$ and $c_{x0}(l)=\ln^2\mu_x(0,l)$ it can be minimized with respect to one parameter $\gamma_x(l)$ for $l\in\mathbb{S}$.

3.9. Final remarks

In this chapter new dynamic hybrid mortality models, both for realizations and for moments, are introduced. Except for the Cox–Ingersoll–Ross model, the remaining models are described by Itô's stochastic differential equations. The related discrete-time models are also presented and estimation procedures proposed.

It is worth noting that in the case of moment mortality models it is possible to estimate only two parameters $\gamma_x(l)$ and $\ln\mu_x(0,l)$ using criteria based on stationary solutions of moment equations with respect to some of the equations.

The general estimation procedures for hybrid models draw on the switching points' concept, i.e. times at which hybrid models switch from one to another state (subsystems). In Section 3.2 the switching points' identification procedure based on the self-adaptive test is proposed and illustrated on the time series of log-central mortality rates for Polish women aged 40 years.

Switchings obtained for all one-year age groups of Polish males and females with the time period 1958–2000 as well as parameters' estimation results of some hybrid models are presented in Chapter 6.

Chapter 4

Mortality model based on oriented fuzzy numbers

4.1. Introduction

Part of the discussion in Chapter 1 focuses on the theoretical foundations of the standard Lee–Carter model. The main problem in applying this model refers to the underlying assumption that the residuals are homoscedastic. The analysis of residuals provides evidence that such an assumption is mostly not fulfilled. As a result, the model may have a poor goodness-of-fit for some age groups and years. It is also known that empirical central death rates only approximate the real ones, the exact values of which are often not known (see [Rossa *et al.* 2011], Section 3.3).

All these drawbacks make it necessary to create solutions addressing the problem of heteroscedastic residuals and the approximative character of input data. One option is to assume that age-specific log-central death rates are fuzzy numbers. The approach was adopted by [Koissi, Shapiro 2006], in which they presented a fuzzy version of the Lee–Carter model (FLC) with both age-specific death rates and model's parameters being fuzzy numbers.

Since the fuzzy Lee–Carter model as modified by Koissi and Shapiro uses a fuzzy representation of the data, it allows, *inter alia*, addressing the issue of uncertainty of approximated death rates and incorporating random terms into the fuzzy structure of the model.

The parameter estimation of the Koissi–Shapiro model involves however some optimization problems, since minimization of the estimation criterion is performed on fuzzy numbers and uses a max-type operator. A modified fuzzy approach presented by Rossa, Socha and Szymański [Rossa *et al.* 2011] draws on the algebra of oriented fuzzy numbers (OFN), theoretical backgrounds of which can be found in [Kosiński *et al.* 2003, Kosiński, Prokopowicz 2004].

The OFN approach facilitates solving the optimization problem and consequently the estimation of the model's parameters.

This chapter explains basic notions relating to fuzzy numbers and oriented fuzzy numbers, as well as the underlying concepts of an Extended Fuzzy Lee–Carter model (EFLC) based on the algebra of OFN. For details on the general fuzzy set theory, the reader should refer to [Dubois, Prade 1980].

4.2. Algebra of oriented fuzzy numbers

The fuzzy set theory emerged in 1965, with the publication of Lotfi Zadeh's work [Zadeh 1965].

A classical fuzzy set is a notion that generalizes the idea of a set and allows partial membership of elements in a given set. The degree of membership is usually expressed by a function mostly denoted as μ. A 0 value of the function points to "non-membership", 1 is "full" membership, and values between 0 and 1 indicate an element's "partial" membership in the set.

Fuzzy sets, or their special case fuzzy numbers, are frequently used today as a convenient way to formally present imprecise linguistic notions, e.g. subjective notions such as cold, hot or high, low, etc.

Definition 4.1. [Zadeh 1965] A fuzzy subset A of a non-empty space $\mathcal{X}$ is a set of ordered pairs

$$A = \{\langle u, \mu_A(z)\rangle,\ u \in \mathcal{X}\}, \tag{4.2.1}$$

where $\mu_A(u) : \mathcal{X} \to [0,1]$ is a membership function assigning the degree of membership in set A to each element $u \in \mathcal{X}$.

The elements of space $\mathcal{X}$ can be arbitrarily defined objects such as persons, notions, items, or numbers.

Let us assume now that $\mathcal{X} = \mathbb{R}$, where $\mathbb{R}$ is a set of real numbers. Figure 4.1 is an example of the membership function $\mu_A(u)$ of fuzzy set A for $u \in \mathbb{R}$.

Definition 4.2. [Dubois, Prade 1980] A fuzzy subset A of real space $\mathbb{R}$ with membership function $\mu_A(u)\colon \mathbb{R} \to [0,1]$ is called a fuzzy number, if

(i) A is a normal set, i.e. $\sup_{u\in\mathbb{R}} \mu_A(u) = 1$,

(ii) A is a fuzzy-convex set, i.e.

$$\forall_{u_1,u_2\in\mathbb{R}}\forall_{\lambda\in[0,1]}\ \ \mu_A(\lambda u_1+(1-\lambda)u_2)\geq\min\{\mu_A(u_1),\mu_A(u_2)\}, \quad (4.2.2)$$

(iii) μ_A is an upper semi-continuous function, i.e. $\{u\in\mathbb{R}:\ \mu_A(u)\geq\nu\}$ is a closed set for each $\nu\in\mathbb{R}$ [Hong *et al.* 2001],

(iv) support $\mathrm{supp}A = \mathrm{cl}\{u\in\mathbb{R}:\ \ \mu_A(u) > 0\}$ of a fuzzy set A is bounded and cl is a closure operator.

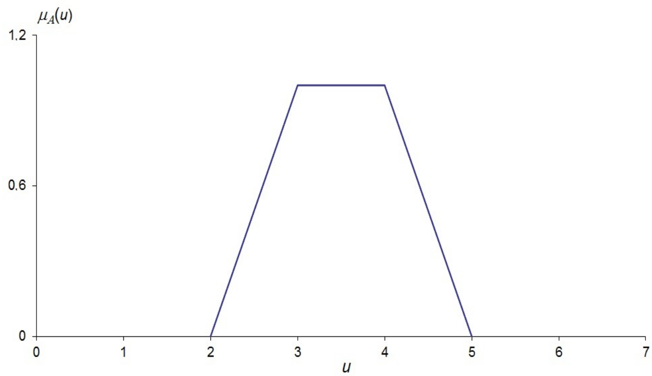

Figure 4.1. Illustrative fuzzy set $A = \{\langle u, \mu_A(z)\rangle,\ u\in\mathbb{R}\}$
Source: Developed by the authors

Definition 4.3. [Hong 2001] A fuzzy number is considered triangular if its membership function is of the following form

$$\mu_A(z) = \begin{cases} 1-\frac{|a-u|}{l_A} & \text{for} \quad a-l_A\leq u < a, \\ 1+\frac{|a-u|}{r_A} & \text{for} \quad a\leq u\leq a+r_A, \\ 0 & \text{otherwise}, \end{cases} \tag{4.2.3}$$

where $a\in\mathbb{R}$ is a central value and $l_A, r_A > 0$ are left and right spreads, respectively.

A triangular fuzzy number A (also called a triangular number) is written as

$$A = (a, l_A, r_A). \tag{4.2.4}$$

The class of popular membership functions also includes, among others, singleton, radial or ellipsoid functions.

Definition 4.4. [Hong 2001] When $l_A = r_A$ then triangular fuzzy number is called symmetric; it is denoted by

$$A = (a, s_A), \tag{4.2.5}$$

where $a\in\mathbb{R}$ is a central value and $s_A > 0$ is a spread.

Definition 4.5. [Zimmermann 2001, p. 14] The λ-cut of fuzzy number A is a set A_λ defined as

$$A_\lambda = \{u \in \mathbb{R}: \ \mu_A(u) \geq \lambda\} = [A_L(\lambda), A_R(\lambda)], \tag{4.2.6}$$

where

$$A_L(\lambda) = \inf\{u \in \mathbb{R}: \ \mu_A(u) \geq \lambda\}, \tag{4.2.7}$$

$$A_R(\lambda) = \sup\{u \in \mathbb{R}: \ \mu_A(u) \geq \lambda\}. \tag{4.2.8}$$

Figure 4.2 illustrates the λ-cut of a symmetric triangular fuzzy number (4.2.5) for $\lambda = 0.5$.

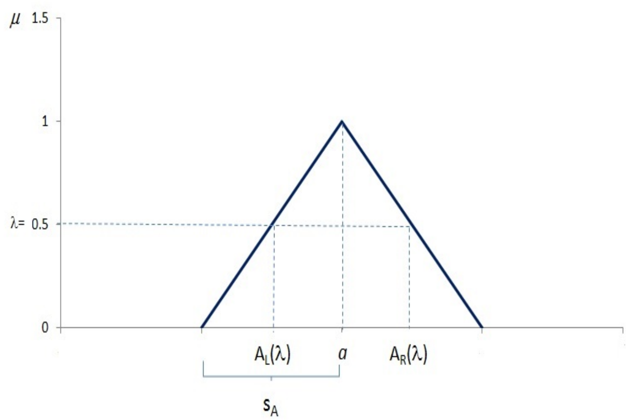

Figure 4.2. A symmetric triangular fuzzy number $A = (a, s_A)$
and its λ-cut $[A_L(\lambda), A_R(\lambda)]$ for $\lambda = 0.5$
Source: Developed by the authors

Example 4.1. If a fuzzy number is triangular and so it can be written as $A = (a, l_A, r_A)$, its λ-cut A_λ is given by

$$A_\lambda = [A_L(\lambda), A_R(\lambda)], \tag{4.2.9}$$

where

$$\begin{aligned} A_L(\lambda) &= a - l_A(1-\lambda), \\ A_R(\lambda) &= a + r_A(1-\lambda). \end{aligned} \tag{4.2.10}$$

In the special case, when A is a triangular symmetric fuzzy number, i.e. $A = (a, s_A)$

$$\begin{aligned} A_L(\lambda) &= a - s_A(1-\lambda), \\ A_R(\lambda) &= a + s_A(1-\lambda). \end{aligned} \tag{4.2.11}$$

Definition 4.6. [Koissi, Shapiro 2006] Addition and multiplication of two triangular symmetric fuzzy numbers $A = (a, s_A)$, $B = (b, s_B)$ are the following

$$A \oplus B = (a + b,\ \max(s_A, s_B)), \tag{4.2.12}$$

$$A \odot B = (ab,\ \max(s_A|b|, s_B|a|)). \tag{4.2.13}$$

Definition 4.7. [Kosiński, Prokopowicz 2004] Oriented fuzzy number $\vec{A}$ is an ordered pair

$$\vec{A} = (f, g), \tag{4.2.14}$$

where $f, g:\ [0,1] \to \mathbb{R}$ are continuous functions.

Functions f and g are the *up* part and *down* part of an oriented fuzzy number, respectively. From the continuity of both these parts it follows that the images of both functions are bounded intervals. The images are respectively called UP and DOWN.
Let us denote

$$l_A := f(0),\ 1_A^- := f(1),\ 1_A^+ := g(1),\ r_A := g(0) \tag{4.2.15}$$

and

$$\mathrm{UP} = (l_A, 1_A^-), \quad \mathrm{DOWN} = (1_A^+, r_A). \tag{4.2.16}$$

By adding a third interval

$$\mathrm{CONST} = [1_A^-, 1_A^+], \tag{4.2.17}$$

we have three subintervals in the splitting of the support of each convex fuzzy number.

In general, subintervals (4.2.16) and (4.2.17) may not satisfy the conditions

$$l_A \le 1_A^-, \quad 1_A^+ \le r_A. \tag{4.2.18}$$

However, when functions f, g are strictly monotonic and $f \le g$, then all these subintervals are proper and have the following relationships

$$l_A \le 1_A^- \le 1_A^+ \le r_A. \tag{4.2.19}$$

In this case, the sum UP $\cup$ CONST $\cup$ DOWN can represent the base of fuzzy number A in the classical sense [Kosiński, Prokopowicz 2004].

Functions f, g that are strictly monotonic on interval $[0,1]$ have inverse functions f^{-1}, g^{-1} defined on intervals UP and DOWN, respectively. Hence, a new membership function can be piecewisely defined on $\mathbb{R}$ by taking the inverse f^{-1} of function f on UP and the inverse g^{-1} of function g on DOWN.

Let us assign a constant value 1 to CONST. The relationship between the membership function μ_A of fuzzy number A and the functions f, g of oriented number $\vec{A}$ can then be written as follows

$$\mu_A(u) = \begin{cases} 1 & \text{for} \quad u \in \text{CONST}, \\ f^{-1}(u) & \text{for} \quad u \in \text{UP}, \\ g^{-1}(u) & \text{for} \quad u \in \text{DOWN}, \\ 0 & \text{for} \quad u \notin (l_A, r_A). \end{cases} \tag{4.2.20}$$

Example 4.2. A triangular fuzzy number $A = (a, l_A, r_A)$ corresponds to a oriented fuzzy number $\vec{A} = (f, g)$, where

$$f(u) = a - l_A(1-u), \quad g(u) = a + r_A(1-u), \quad u \in [0,1]. \tag{4.2.21}$$

From Example 4.1 it follows that a triangular fuzzy number $A = (a, l_A, r_A)$ has λ-cut $A_\lambda = [A_L(\lambda), A_R(\lambda)]$, where

$$A_L(\lambda) = a - l_A(1-\lambda), \quad A_R(\lambda) = a + r_A(1-\lambda), \quad \lambda \in [0,1]. \tag{4.2.22}$$

where a, l_A, r_A are known parameters.
Substituting u for λ, $f(u)$ for $A_L(u)$ and $g(u)$ for $A_R(u)$, we obtain

$$f(u) = a - l_A(1-u), \quad g(u) = a + r_A(1-u), \quad u \in [0,1]. \tag{4.2.23}$$

Hence, given that functions f, g are continuous on interval $[0,1]$, an oriented fuzzy number $\vec{A}$ is defined by an ordered pair of functions (f, g).

The implication of the above is that the triangular symmetric fuzzy number $A = (a, s_A)$ generates the oriented fuzzy number $\vec{A} = (f, g)$, where f, g are of the form

$$f(u) = a - s_A(1-u), \quad g(u) = a + s_A(1-u), \quad u \in [0,1]. \tag{4.2.24}$$

Figure 4.3 presents a symmetric triangular fuzzy number $A = (a, s_A)$ with membership function $\mu_A(u)$ of the form

$$\mu_A(u) = \begin{cases} 1 - \frac{a-u}{s_A} & \text{for} \quad u \in [a - s_A, a], \\ 1 + \frac{u-a}{s_A} & \text{for} \quad u \in (a, a + s_A], \\ 0 & \text{otherwise} \end{cases} \tag{4.2.25}$$

and a corresponding oriented fuzzy number $\vec{A} = (f, g)$, where f, g are defined in (4.2.24).

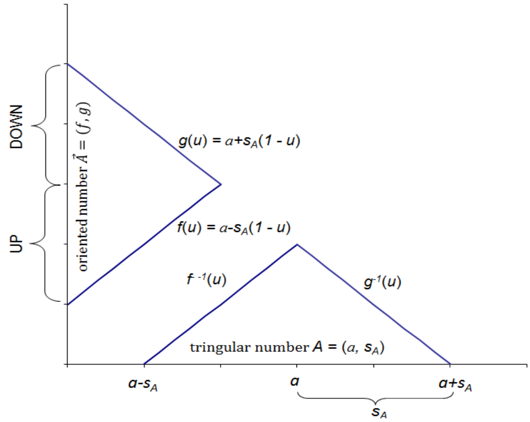

Figure 4.3. Triangular fuzzy number $A = (a, s_A)$
and oriented fuzzy number $\vec{A} = (f, g)$
Source: Developed by the authors

Definition 4.8. [Diamond 1988] The Diamond distance between oriented fuzzy numbers $\vec{A} = (f_A, g_A)$ and $\vec{B} = (f_B, g_B)$ is given by

$$D^2(\vec{A}, \vec{B}) = \int_0^1 \left[(f_A(u) - f_B(u))^2 + (g_A(u) - g_B(u))^2 \right] du, \qquad (4.2.26)$$

where f_A, f_B, g_A, g_B are integrable functions.

Definition 4.9. [Kosiński, Prokopowicz 2004] Let $\vec{A} = (f_A, g_A)$, $\vec{B} = (f_B, g_B)$, $\vec{C} = (f_C, g_C)$ be oriented fuzzy numbers. Then $\vec{C}$ is a sum of $\vec{A}$ and $\vec{B}$, what is denoted as $\vec{C} = \vec{A} \oplus \vec{B}$, if

$$f_C(u) = f_A(u) + f_B(u), \quad g_C(u) = g_A(u) + g_B(u). \qquad (4.2.27)$$

Definition 4.10. [Kosiński, Prokopowicz 2004] Let $\vec{A} = (f_A, g_A)$, $\vec{B} = (f_B, g_B)$, $\vec{C} = (f_C, g_C)$ be oriented fuzzy numbers. Then $\vec{C}$ is a product of $\vec{A}$ and $\vec{B}$, denoted as $\vec{C} = \vec{A} \otimes \vec{B}$, if

$$f_C(u) = f_A(u) f_B(u), \quad g_C(u) = g_A(u) g_B(u). \qquad (4.2.28)$$

Definition 4.11. [Kosiński, Prokopowicz 2004] $\vec{C} = (f_C, g_C)$ is a product of oriented fuzzy number $\vec{A} = (f_A, g_A)$ multiplied by scalar d, which is symbolically written as $\vec{C} = d\vec{A}$, if

$$f_C(u) = d f_A(u), \quad g_C(u) = d g_A(u). \qquad (4.2.29)$$

Definition 4.12. [Kosiński, Prokopowicz 2004] Let $\vec{A} = (f_A, g_A)$, $\vec{B} = (f_B, g_B)$, $\vec{C} = (f_C, g_C)$ be oriented fuzzy numbers. Then $\vec{C}$ is a result of dividing $\vec{A}$ by $\vec{B}$, which can be symbolically written as $\vec{C} = \vec{A} \oslash \vec{B}$, if for each argument $u \in [0,1]$ such as that $f_B(u) \neq 0$ and $g_B(u) \neq 0$, there is

$$f_C(u) = \frac{f_A(u)}{f_B(u)} \quad \text{and} \quad g_C(u) = \frac{g_A(u)}{g_B(u)}. \tag{4.2.30}$$

Example 4.3. Let $A = (a, s_A)$ and $B = (b, s_B)$ be triangular symmetric fuzzy numbers corresponding to oriented fuzzy numbers $\vec{A} = (f_A, g_A)$ and $\vec{B} = (f_B, g_B)$, where

$$\begin{aligned} &f_A(u) = a - s_A(1-u), \quad g_A(u) = a + s_A(1-u), \\ &f_B(u) = b - s_B(1-u), \quad g_B(u) = b + s_B(1-u), \quad u \in [0,1]. \end{aligned} \tag{4.2.31}$$

We can write

$$\vec{A} \oplus \vec{B} = (f_A, g_A) \oplus (f_B, g_B) = (f_A + f_B, g_A + g_B), \tag{4.2.32}$$

where

$$\begin{aligned} &f_A(u) + f_B(u) = a + b - (s_A + s_B)(1-u), \\ &g_A(u) + g_B(u) = a + b + (s_A + s_B)(1-u), \quad u \in [0,1]. \end{aligned} \tag{4.2.33}$$

Analogously, we have

$$\vec{A} \otimes \vec{B} = (f_A, g_A) \otimes (f_B, g_B) = (f_A f_B,\ g_A g_B), \tag{4.2.34}$$

where

$$\begin{aligned} &f_A(u) f_B(u) = ab - (b s_A + a s_B)(1-u) + s_A s_B (1-u)^2, \\ &g_A(u) g_B(u) = ab + (b s_A + a s_B)(1-u) + s_A s_B (1-u)^2, \quad u \in [0,1]. \end{aligned} \tag{4.2.35}$$

In turn, for a given non-zero scalar d, from Definition 4.11 we obtain

$$d\vec{A} = (d f_A,\ d g_A), \tag{4.2.36}$$

where

$$df_A(u)=d(a-s_A(1-u)),\ dg_A(u)=d(a+s_A(1-u)),\ \ u\in[0,1]. \quad (4.2.37)$$

Property 4.1. If $\vec{A}=(f_A,g_A)$ is an oriented fuzzy number, then fuzzy number $-\vec{A}$ can be expressed as follows

$$-\vec{A}=(-f_A,\ -g_A). \quad (4.2.38)$$

Let us note that $-\vec{A}$ can be taken to be the product of $\vec{A}$ and $d=-1$.

Property 4.2. Let us have two oriented fuzzy numbers Let $\vec{A}=(f_A,g_A)$, $\vec{B}=(f_B,g_B)$. The difference between $\vec{A}$ and $\vec{B}$ is written as

$$\vec{A}\ominus\vec{B}=(f_A-f_B,\ g_A-g_B). \quad (4.2.39)$$

Subtracting $\vec{B}$ from $\vec{A}$ can be assumed equivalent to adding the opposite of $\vec{B}$, i.e. adding $\vec{B}$ multiplied by scalar $d=-1$ to $\vec{A}$.

Property 4.3. By subtracting $\vec{A}$ from $\vec{A}$ we obtain

$$\vec{A}-\vec{A}=(f_A-f_A,g_A-g_A)=(0,0). \quad (4.2.40)$$

Property 4.4. If $\vec{A}\oplus\vec{C}_1=\vec{A}\oplus\vec{C}_2$, then $\vec{C}_1=\vec{C}_2$.
Indeed, let

$$\vec{A}=(f_A,g_A),\quad \vec{C}_1=(f_{C_1},g_{C_1}),\quad \vec{C}_2=(f_{C_2},g_{C_2}). \quad (4.2.41)$$

From Definition 4.9, we have

$$\vec{A}\oplus\vec{C}_1=(f_A,g_A)\oplus(f_{C_1},g_{C_1})=(f_A+f_{C_1},g_A+g_{C_1}). \quad (4.2.42)$$

The same result is obtained for $\vec{A}\oplus\vec{C}_2$, i.e.

$$\vec{A}\oplus\vec{C}_2=(f_A,g_A)\oplus(f_{C_2},g_{C_2})=(f_A+f_{C_2},g_A+g_{C_2}). \quad (4.2.43)$$

Following the assumption that $\vec{A}\oplus\vec{C}_1=\vec{A}\oplus\vec{C}_2$ we have

$$(f_A+f_{C_1},g_A+g_{C_1})=(f_A+f_{C_2},g_A+g_{C_2}) \quad (4.2.44)$$

or

$$(f_A+f_{C_1},g_A+g_{C_1})-(f_A+f_{C_2},g_A+g_{C_2})=(0,0), \quad (4.2.45)$$

i.e.

$$(f_A + f_{C_1} - f_A - f_{C_2}, g_A + g_{C_1} - g_A - g_{C_2}) = (0, 0). \tag{4.2.46}$$

This leads us to

$$f_{C_1} - f_{C_2} = 0, \quad g_{C_1} - g_{C_2} = 0 \tag{4.2.47}$$

or, equivalently,

$$f_{C_1} = f_{C_2}, \quad g_{C_1} = g_{C_2}. \tag{4.2.48}$$

From the above it follows that $\vec{C}_1 = \vec{C}_2$ which was to be proved.

Property 4.5. If $\vec{A}$ and $\vec{B}$ are oriented fuzzy numbers and $c, d \in \mathbb{R}$ are any real numbers, the following conditions are fulfilled
(i) $c(d\vec{A}) = (cd)\vec{A}$,
(ii) $d(\vec{A} \oplus \vec{B}) = d\vec{A} \oplus d\vec{B}$,
(iii) $(c + d)\vec{A} = c\vec{A} \oplus d\vec{A}$,
(iv) $1\vec{A} = \vec{A}$.

Condition (i) follows from Definition 4.11, and condition (ii) is based on Definitions 4.9 and 4.11 explaining how oriented fuzzy numbers should be added and multiplied by the scalar. Analogous reasoning applies to condition (iii). Condition (iv) follows from the property of multiplying an oriented fuzzy number by the scalar which in this case equals 1.

Let us denote by $\mathfrak{R}$ a set of oriented fuzzy numbers, with arithmetic operations of adding oriented fuzzy numbers and multiplying them by the scalar defined as above.

Property 4.6. If $\vec{A}, \vec{B}, \vec{C} \in \mathfrak{R}$ are oriented fuzzy numbers, the following conditions are fulfilled

(v) $\vec{A} \oplus \vec{B} = \vec{B} \oplus \vec{A}$ (commutative addition),
(vi) $(\vec{A} \oplus \vec{B}) \oplus \vec{C} = \vec{A} \oplus (\vec{B} \oplus \vec{C})$ (associative addition),
(vii) If $\vec{A} \oplus \vec{B} = \vec{A} \oplus \vec{C}$, then $\vec{B} = \vec{C}$ (uniqueness of addition).

Conditions (i)–(vii) are called linear space axioms. Space $\mathfrak{R}$ is a real linear space, since scalars by which oriented fuzzy numbers are multiplied are real numbers.

Let $C([0,1])$ be a set of all continuous functions defined on a bounded interval $[0,1]$. Then $\mathfrak{R} = C([0,1]) \times C([0,1])$ is a set of ordered pairs (f,g) of continuous functions, each defined on interval $[0,1]$.

Space $\mathfrak{R}$ is a linear space, because both the axiom of addition and the axiom of multiplication by the scalar are met; they are given by

$$\vec{A} \oplus \vec{B} = (f_A + f_B,\ g_A + g_B) \tag{4.2.49}$$

and

$$d\vec{A} = (df_A,\ dg_A), \tag{4.2.50}$$

where

$$\vec{A} = (f_A, g_A), \quad \vec{B} = (f_B, g_B), \quad d \in \mathbb{R}. \tag{4.2.51}$$

Let us define a norm in space $\mathfrak{R}$

$$\|(f,g)\| = \max(\sup_{u\in[0,1]} |f(u)|, \sup_{u\in[0,1]} |g(u)|). \tag{4.2.52}$$

The interval $[0,1]$ is compact and for continuous functions f,g inequalities $\sup_{u\in[0,1]} |f(u)|<\infty$, $\sup_{u\in[0,1]} |\gamma g(u)|<\infty$ hold, thus $\|(f,g)\|<\infty$. The following inequalities also hold: $\sup_{u\in[0,1]} |f(u)| > 0$ for $f(u) \neq 0$, $\sup_{u\in[0,1]} |g(u)| > 0$ for $g(u) \neq 0$, $\sup_{u\in[0,1]} |f(u)| = 0$ if $\forall_{u\in[0,1]} f(u)=0$ and $\sup_{u\in[0,1]} |g(u)| = 0$ if $\forall_{u\in[0,1]} g(u) = 0$. Hence, $\|(f,g)\| > 0$ when $(f,g) \neq (0,0)$ and $\|(f,g)\| = 0$, when $(f,g) = (0,0)$, meaning that the first axiom of the norm is met.

In general, axioms of a norm are as follows [Kolodziej 1970, p. 35]
(i) $\|x\| > 0$ for $x \neq 0$, $\|0\| = 0$,

(ii) $\|x+y\| \leq \|x\| + \|y\|$ (subadditivity),

(iii) $\|\alpha x\| = |\alpha|\ \|x\|$ (homogeneity).

To check the second axiom of the norm (4.2.52), i.e. the triangle inequality, let us assume that$\vec{A} = (f_A, g_A)$ and $\vec{B} = (f_B, g_B)$. From the definition of the norm (4.2.52), we have

$$\|\vec{A} \oplus \vec{B}\| = \max(\sup_{u\in[0,1]} |f_{A\oplus B}(u)|, \sup_{u\in[0,1]} |g_{A\oplus B}(u)|). \tag{4.2.53}$$

On the other hand,

$$f_{A\oplus B}(u) = f_A(u) + f_B(u), \tag{4.2.54}$$

consequently

$$|f_{A\oplus B}(u)| = |f_A(u) + f_B(u)| \leq \sup_{u\in[0,1]} |f_A(u)| + \sup_{u\in[0,1]} |f_B(u)|, \quad (4.2.55)$$

thus,

$$\sup_{u\in[0,1]} |f_{A\oplus B}(u)| \leq \|\vec{A}\| + \|\vec{B}\|. \quad (4.2.56)$$

Analogously, we obtain

$$\sup_{u\in[0,1]} |g_{A\oplus B}(u)| \leq \|\vec{A}\| + \|\vec{B}\|, \quad (4.2.57)$$

that is

$$\max(\sup_{u\in[0,1]} |g_{A\oplus B}(u)|, \sup_{u\in[0,1]} |g_{A\oplus B}(u)|) \leq \|\vec{A}\| + \|\vec{B}\|. \quad (4.2.58)$$

Therefore, we have

$$\|\vec{A} \oplus \vec{B}\| \leq \|\vec{A}\| + \|\vec{B}\|, \quad (4.2.59)$$

which constitutes the triangle condition.

Let us test now the third axiom of the norm, i.e. the homogeneity condition

$$\|\alpha\vec{A}\| = \max(\sup_{u\in[0,1]} |\alpha f_A(u)|, \sup_{u\in[0,1]} |\alpha g_A(u)|). \quad (4.2.60)$$

Because of the norm's properties

$$|\alpha f_A(u)| = |\alpha||f_A(u)| \leq |\alpha| \sup_{u\in[0,1]} |f_A(u)| \leq |\alpha|\|\vec{A}\| \quad (4.2.61)$$

and

$$|\alpha g_A(u)| = |\alpha||g_A(u)| \leq |\alpha| \sup_{u\in[0,1]} |g_A(u)| \leq |\alpha|\|\vec{A}\|. \quad (4.2.62)$$

Therefore

$$\|\alpha\vec{A}\| \leq |\alpha|\|\vec{A}\|. \quad (4.2.63)$$

By substituting $\alpha\frac{1}{\alpha}\vec{A}$ for $\vec{A}$ we obtain

$$\|\vec{A}\| = \|\alpha\frac{1}{\alpha}\vec{A}\| = \|\frac{1}{\alpha}(\alpha\vec{A})\| \leq \frac{1}{|\alpha|}\|\alpha\vec{A}\|. \quad (4.2.64)$$

This leads us to

$$\|\alpha\vec{A}\| \geq |\alpha|\,\|\vec{A}\|. \tag{4.2.65}$$

Taking both inequalities together, we arrive at

$$\|\alpha\vec{A}\| = |\alpha|\,\|\vec{A}\|, \tag{4.2.66}$$

which is the homogeneity axiom for the norm. Therefore, space $\mathfrak{R}$ with the norm the properties of which have been verified is a normed space.

Property 4.7. $\mathfrak{R}$ is the Banach space, because it is both a normed and complete space (i.e. each sequence of the elements in space $\mathfrak{R}$ satisfying the Cauchy condition converges to a point in that space). The proof is analogous to the proof of theorem 22.3 in [Kolodziej 1970, p. 44].

Property 4.8. A space of oriented fuzzy numbers is the Banach algebra, i.e. the Banach space with associative and continuous operation of multiplication [Żelazko 1968, p. 16] with a unit element $\vec{\mathbb{I}} = (1, 1)$, i.e. with a pair of constant functions equal 1, such that $\vec{A} \otimes \vec{\mathbb{I}} = \vec{\mathbb{I}} \otimes \vec{A} = \vec{A}$ for each $\vec{A} \in \mathfrak{R}$.

Property 4.9. Algebra $\mathfrak{R}$ is commutative, because the following equalities hold $\vec{A} \otimes \vec{B} = \vec{B} \otimes \vec{A}$ for any $\vec{A}, \vec{B} \in \mathfrak{R}$.
Indeed, we have

$$\begin{aligned} \vec{A} \otimes \vec{B} =&(f_A, g_A) \otimes (f_B, g_B) = (f_A f_B, g_A g_B) = \\ =&(f_B, g_B) \otimes (f_A, g_A) = \vec{B} \otimes \vec{A}, \end{aligned} \tag{4.2.67}$$

Property 4.10. Algebra $\mathfrak{R}$ is isomorphic with the algebra of complex numbers.

Gelfand–Mazur theorem [Alexiewicz 1969]. If the Banach space with a unit element is an algebra, then it is isometrically isomorphic with the algebra of complex numbers; more precisely, each element is written as λe, where $\lambda \in \mathbb{C}$ and e is a unit element in the space of complex numbers.

4.3. The extended Koissi–Shapiro mortality model

One of the most interesting generalizations of the Lee–Carter model referring to the algebra of fuzzy numbers is the FLC model introduced by [Koissi, Shapiro 2006]. Their version of the Lee–Carter model assumes fuzzy representation of the log-central death rates as well as model's parameters (see (1.5.48) in Chapter 1).

In 2011 Rossa, Socha and Szymański proposed a modified version of the FLC model, termed the Extended Fuzzy Lee–Carter model (EFLC model), in which mortality rates and model's parameters were represented by means of oriented fuzzy numbers (OFN) [Rossa *et al.* 2011].

The EFLC model is written by analogy to (1.5.48) as

$$\vec{Y}_{x,t} = \vec{A}_x \oplus (\vec{B}_x \otimes \vec{K}_t), \quad x = 0, 1, \ldots, X,\ t = 1, 2 \ldots, T, \tag{4.3.1}$$

where $\vec{A}_x$, $\vec{B}_x$, $\vec{K}_t$ are OFN's expressed by means of the following ordered pairs

$$\vec{A}_x = (f_{A_x}, g_{A_x}), \quad \vec{B}_x = (f_{B_x}, g_{B_x}), \quad \vec{K}_t = (f_{K_t}, g_{K_t}), \tag{4.3.2}$$

with functions f_{A_x}, g_{A_x}, f_{B_x}, g_{B_x} and f_{K_t}, g_{K_t} defined for $u \in [0, 1]$ as

$$\begin{aligned} f_{A_x}(u) &= a_x - (1-u)s_{A_x}, \quad g_{A_x}(u) = a_x + (1-u)s_{A_x}, \\ f_{B_x}(u) &= b_x - (1-u)s_{B_x}, \quad g_{B_x}(u) = b_x + (1-u)s_{B_x}, \\ f_{K_t}(u) &= k_t - (1-u)s_{K_t}, \quad g_{K_t}(u) = k_t + (1-u)s_{K_t}. \end{aligned} \tag{4.3.3}$$

For the FLC model introduced by [Koissi, Shapiro 2006], it was assumed that A_x, B_x, K_t are triangular symmetric numbers with central values a_x, b_x, k_t and spreads $s_{A_x}, s_{B_x}, s_{K_t}$, respectively (see Section 1.5.3). Thus, they are written using the notation from Definition 4.4 as

$$A_x = (a_x, s_{A_x}), \quad B_x = (b_x, s_{B_x}), \quad K_t = (k_t, s_{K_t}). \tag{4.3.4}$$

For the EFLC model (4.3.1), it is assumed that the model's parameters are still a_x, b_x, k_t and $s_{A_x}, s_{B_x}, s_{K_t}$, but they are incorporated in functions (4.3.3).

In the EFLC model it is also assumed that the log-central death rates have analogous OFN representation $\vec{Y}_{x,t} = (f_{Y_{x,t}}, g_{Y_{x,t}})$ with functions $f_{Y_{x,t}}, g_{Y_{x,t}}$ defined for $u \in [0, 1]$ as

$$\begin{aligned} f_{Y_{x,t}}(u) &= y_{x,t} - e_{x,t}(1-u), \\ g_{Y_{x,t}}(u) &= y_{x,t} + e_{x,t}(1-u), \end{aligned} \tag{4.3.5}$$

where $y_{x,t} = \ln m_{x,t}$ are (crisp) log-central death rates and $e_{x,t}$ are "fuzziness parameters" determined by means of the fuzzification procedure (see next section for more details).

After the oriented fuzzy numbers are added and multiplied according to Definitions 4.9 and 4.10, the right-hand side of (4.3.1) takes the form

$$\begin{aligned} \vec{A}_x \oplus (\vec{B}_x \otimes \vec{K}_t) &= (f_{A_x}, g_{A_x}) \oplus (f_{B_x \otimes K_t}, g_{B_x \otimes K_t}) = \\ &= (f_{A_x} + f_{B_x \otimes K_t},\ g_{A_x} + g_{B_x \otimes K_t}), \end{aligned} \tag{4.3.6}$$

where for $u \in [0,1]$ we have

$$\begin{aligned} f_{B_x \otimes K_t}(u) &= b_x k_t - (k_t s_{B_x} + b_x s_{K_t})(1-u) + s_{B_x} s_{K_t}(1-u)^2, \\ g_{B_x \otimes K_t}(u) &= b_x k_t + (k_t s_{B_x} + b_x s_{K_t})(1-u) + s_{B_x} s_{K_t}(1-u)^2 \end{aligned} \tag{4.3.7}$$

and

$$\begin{aligned} & f_{A_x}(u) + f_{B_x \otimes K_t}(u) = \\ & = a_x + b_x k_t - (s_{A_x} + k_t s_{B_x} + b_x s_{K_t})(1-u) + s_{B_x} s_{K_t}(1-u)^2, \\ & g_{A_x}(u) + f_{B_x \otimes K_t}(u) = \\ & = a_x + b_x k_t + (s_{A_x} + k_t s_{B_x} + b_x s_{K_t})(1-u) + s_{B_x} s_{K_t}(1-u)^2. \end{aligned} \tag{4.3.8}$$

Let us notice that expressions $s_{B_x} s_{K_t}(1-u)^2$ in (4.3.8) are close to 0 for small values of s_{B_x}, s_{K_t} and for $u \in [0,1]$. Given this, we can consider the following approximation

$$\begin{aligned} f_{A_x}(u) + f_{B_x \otimes K_t}(u) &\approx a_x + b_x k_t - (s_{A_x} + k_t s_{B_x} + b_x s_{K_t})(1-u), \\ g_{A_x}(u) + g_{B_x \otimes K_t}(u) &\approx a_x + b_x k_t + (s_{A_x} + k_t s_{B_x} + b_x s_{K_t})(1-u). \end{aligned} \tag{4.3.9}$$

It follows from (4.3.9) that right-hand side of the EFLC model (4.3.1), i.e. $\vec{A}_x \oplus (\vec{B}_x \otimes \vec{K}_t)$, correspond to some symmetric triangular numbers with central values $a_x + b_x k_t$ and spreads approximated by the sum $s_{A_x} + k_t s_{B_x} + b_x s_{K_t}$.

4.4. Data fuzzification with switchings

In the mortality model proposed by [Koissi, Shapiro 2006] the input data, i.e. log-central death rates, are fuzzified. Let us therefore apply the concept of fuzzification to the model with oriented fuzzy numbers [Rossa *et al.* 2011, pp. 167–174]. Since fuzzification is the basis of mortality modeling in the framework of fuzzy numbers, we shall propose a fuzzification method referring to the approach given by [Koissi, Shapiro 2006]. Essential to the discussion is the fuzzification of log-central death rates $y_{x,t} = \ln m_{x,t}$.

Using fuzzification approach by [Koissi, Shapiro 2006], each fixed value $y_{x,t}$ is transformed into a triangular, symmetric fuzzy number $Y_{x,t} = (y_{x,t}, e_{x,t})$, where $e_{x,t}$ is an unknown "fuzziness parameter" serving as a spread of fuzzy number $Y_{x,t}$.

In order to determine $e_{x,t}$, Koissi and Shapiro applied the fuzzy regression model, by introducing symmetric triangular fuzzy numbers (c_{0x}, s_{0x}) and (c_{1x}, s_{1x}) satisfying equalities

$$(y_{x,t}, e_{x,t}) = (c_{0x}, s_{0x}) \oplus (c_{1x}, s_{1x})\ t \quad \text{for each } x. \tag{4.4.1}$$

Following Definition 4.6, the above reduces to the postulate that for each x the following equalities hold

$$y_{x,t} = c_{0x} + c_{1x}t, \tag{4.4.2}$$

$$e_{x,t} = \max(s_{0x}, s_{1x}t). \tag{4.4.3}$$

What we propose here is to consider parameters $e_{x,t}$ in the framework of oriented fuzzy numbers (OFN). Accordingly, the triangular numbers $(y_{x,t}, e_{x,t}), (c_{0x}, s_{0x}), (c_{1x}, s_{1x})$ are replaced by their OFN counterparts

$$\vec{Y}_{x,t} = (f_{Y_{x,t}}, g_{Y_{x,t}}),\ \vec{C}_{0x} = (f_{C_{0x}}, g_{C_{0x}}),\ \vec{C}_{1x} = (f_{C_{1x}}, g_{C_{1x}}),$$

where $f_{Y_{x,t}}, g_{Y_{x,t}}$ are defined in (4.3.5), and functions $f_{C_{0x}}, g_{C_{0x}}, f_{C_{1x}}, g_{C_{1x}}$ have the form

$$\begin{aligned} f_{C_{0x}}(u) &= c_{0x} - s_{0x}(1-u), \quad g_{C_{0x}}(u) = c_{0x} + s_{0x}(1-u), \\ f_{C_{1x}}(u) &= c_{1x} - s_{1x}(1-u), \quad g_{C_{1x}}(u) = c_{1x} + s_{1x}(1-u). \end{aligned} \tag{4.4.4}$$

The above means that condition (4.4.1) can be replaced by the following one

$$(f_{Y_{x,t}}, g_{Y_{x,t}}) = (f_{C_{0x}}, g_{C_{0x}}) \oplus (f_{C_{1x}}, g_{C_{1x}})\; t \quad \text{for each } x. \tag{4.4.5}$$

Because Definitions 4.9 and 4.11 explain how oriented numbers should be added and multiplied by a scalar, the following equations should hold for each x and $u \in [0,1]$

$$(y_{x,t} - e_{x,t}(1-u),\ y_{x,t} + e_{x,t}(1-u)) =$$

$$= (f_{C_{0x}}(u) + t f_{C_{1x}}(u),\ g_{C_{1x}}(u) + t g_{C_{1x}}(u)) = \tag{4.4.6}$$

$$= \left(c_{0x} + c_{1x}t - (s_{0x} + s_{1x}t)(1-u),\ c_{0x} + c_{1x}t + (s_{0x} + s_{1x}t)(1-u)\right),$$

which comes down to the postulate that for each x and $t = 1, 2, \ldots, T$

$$y_{x,t} = c_{0x} + c_{1x}t, \tag{4.4.7}$$

$$e_{x,t} = s_{0x} + s_{1x}t. \tag{4.4.8}$$

According to (4.4.7), the estimates of c_{0x} and c_{1x} can be obtained using the standard least squares method (LS) which leads to the following estimation formulas

$$\hat{c}_{1x} = \frac{\overline{t \ln m_{x,t}} - \bar{t}\,\overline{\ln m_{x,t}}}{\overline{t^2} - \bar{t}^2},$$

$$\hat{c}_{0x} = \overline{\ln m_{x,t}} - \hat{c}_{1x}\bar{t}, \tag{4.4.9}$$

where $\overline{\ln m_{x,t}}$, $\overline{t \ln m_{x,t}}$, $\bar{t}$, $\overline{t^2}$ denote the respective arithmetic averages.

To find parameters s_{0x}, s_{1x} in (4.4.8), we can solve the following optimization problem. Since $e_{x,t}$ are, by assumption, non-negative numbers and the smallest value they can take is 0, we need to find such values of $\hat{s}_{0x}, \hat{s}_{1x}$, that at a given x minimize the following sum $S(s_{0x}, s_{1x}) = \sum_{t=1}^{T} e_{x,t} = T s_{0x} + s_{1x} \sum_{t=1}^{T} t$.

The optimization problem can therefore be formulated as follows

$$\text{minimize} \quad S(s_{0x}, s_{1x}) = Ts_{0x} + s_{1x}\sum_{t=1}^{T} t, \tag{4.4.10}$$

subject to the constraints

$$s_{0x}, s_{1x} \geq 0,$$

$$\hat{c}_{0x} + \hat{c}_{1x}t + (s_{0x} + s_{1x}t)\,(1-u) \geq \ln m_{x,t}, \tag{4.4.11}$$

$$\hat{c}_{0x} + \hat{c}_{1x}t - (s_{0x} + s_{1x}t)\,(1-u) \leq \ln m_{x,t}.$$

Since higher values of u result in greater spreads s_{0x}, s_{1x}, it is further assumed that $u = 0$.

Let us note that the optimization problem (4.4.10)–(4.4.11) is similar to that proposed in [Koissi, Shapiro 2006]; the difference between their approach and that introduced in this section lies in the calculation of parameters $e_{x,t}$. In our case, the values of $e_{x,t}$ are based on formula (4.4.8), i.e. $e_{x,t}$ are estimated as

$$\hat{e}_{x,t} = \hat{s}_{0x} + \hat{s}_{1x}t, \tag{4.4.12}$$

whereas in the Koissi–Shapiro approach $e_{x,t}$ are estimated from (4.4.3).

Figure 4.4 illustrates the (crisp) log-central death rates for some age groups registered in Poland for females in time period 1958–2000 and the areas of fuzziness bounded by lines

$$\begin{aligned} f_{1x}(t) &= \hat{c}_{0x} + \hat{c}_{1x}t - \hat{e}_{x,t}, \\ f_{2x}(t) &= \hat{c}_{0x} + \hat{c}_{1x}t + \hat{e}_{x,t}, \end{aligned} \tag{4.4.13}$$

where $\hat{c}_{1x}, \hat{c}_{0x}$ are given in (4.4.9) and $\hat{e}_{x,t}$ are obtained from (4.4.12) with $\hat{s}_{0x}, \hat{s}_{1x}$ solving the minimization problem (4.4.10)–(4.4.11).

Note that areas of fuzziness are slightly wider for the younger age groups, what is caused by higher variability of death rates for such ages.

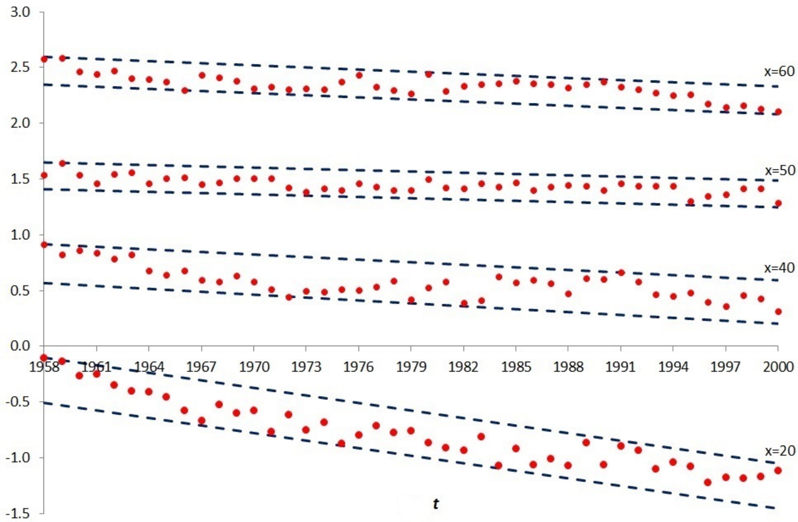

Figure 4.4. Log-central death rates and areas of fuzziness for $x = 20, 40, 50, 60$ years (females).
Source: Developed by the authors

The concept for determining switching time points introduced in the previous chapter (Section 3.2) will be used here in the fuzzification procedure of log-central death rates. To take account of switching points, the optimization problem (4.4.10)–(4.4.11) must be solved for each sub-period (mortality regime), i.e. for each time interval between two adjacent switchings.

The identification of switching points allows parameters c_{0x}, c_{1x}, s_{0x}, s_{1x} to be estimated separately for each sub-period determined by switchings. For instance, if one switching point t^* has been identified, the parameters c_{0x}, c_{1x}, $\tilde{c}_{0x}, \tilde{c}_{1x}$ of two trend lines should be estimated

$$y_{x,t} = c_{0x} + c_{1x}t, \quad t = 1, 2, \ldots, t^* - 1, \tag{4.4.14}$$

$$y_{x,t} = \tilde{c}_{0x} + \tilde{c}_{1x}t, \quad t = t^*, \ldots, T, \tag{4.4.15}$$

what leads to the estimation of two sets of parameters s_{0x}, s_{1x} and $\tilde{s}_{0x}$, $\tilde{s}_{1x}$ by solving problem (4.4.10)–(4.4.11) separately for each sub-period.

Fuzzification of the input data is a necessary step in the estimation of the mortality models presented in the next part of this book. In the approach proposed here the algebra of oriented fuzzy numbers is employed, what makes the fuzzification concept more understandable and easy. Moreover, fuzzification is performed separately for each mortality regime, therefore fuzziness parameters are flexible and better adjusted to the data.

4.5. Parameters' estimation of the EFLC model

To estimate parameters of the mortality model (4.3.1), we shall use the Diamond distances between the left and right sides of the model, i.e. between elements $\vec{Y}_{x,t}$ of the observation matrix and terms $\vec{A}_x \oplus (\vec{B}_x \otimes \vec{K}_t)$.

The task requires the minimization of the sum of Diamond distances (see Definition 4.8) written as

$$I = \sum_{x=0}^{X} \sum_{t=1}^{T} D^2(\vec{Y}_{x,t},\ \vec{A}_x \oplus (\vec{B}_x \otimes \vec{K}_t)) = \sum_{x=0}^{X} \sum_{t=1}^{T} d_{x,t}, \tag{4.5.1}$$

where

$$\begin{aligned} d_{x,t} \equiv & D^2(\vec{Y}_{x,t},\ \vec{A}_x \oplus \vec{B}_x \otimes \vec{K}_t) = \\ & = \int_0^1 \left[f_{A_x}(u) + f_{B_x \otimes K_t}(u) - f_{Y_{x,t}}(u)\right]^2 du + \\ & + \int_0^1 \left[g_{A_x}(u) + g_{B_x \otimes K_t}(u) - g_{Y_{x,t}}(u)\right]^2 du. \end{aligned} \tag{4.5.2}$$

The integrand functions are given by

$$\begin{aligned} & [f_{A_x}(u) + f_{B_x \otimes K_t}(u) - f_{Y_{x,t}}(u)]^2 = [(a_x + b_x k_t - y_{x,t}) + \\ & -(s_{A_x} + s_{B_x} k_t + s_{K_t} b_x - e_{x,t})(1-u) + s_{B_x} s_{K_t} (1-u)^2\big]^2, \\ & [g_{A_x}(u) + f_{B_x \otimes K_t}(u) - g_{Y_{x,t}}(u)]^2 = [(a_x + b_x k_t - y_{x,t}) + \\ & +(s_{A_x} + s_{B_x} k_t + s_{K_t} b_x - e_{x,t})(1-u) + s_{B_x} s_{K_t} (1-u)^2\big]^2. \end{aligned} \tag{4.5.3}$$

Let us introduce the following notations

$$\begin{aligned} U_{x,t} &= a_x + b_x k_t - y_{x,t}, \\ V_{x,t} &= s_{A_x} + s_{B_x} k_t + s_{K_t} b_x - e_{x,t}, \\ W_{x,t} &= s_{B_x} s_{K_t}. \end{aligned} \tag{4.5.4}$$

Then, we have

$$\begin{aligned}&\big[f_{A_x}(u)+f_{B_x\otimes K_t}(u)-f_{Y_{x,t}}(u)\big]^2=\big[U_{x,t}-V_{x,t}(1-u)+W_{x,t}(1-u)^2\big]^2,\\&\big[g_{A_x}(u)+g_{B_x\otimes K_t}(u)-g_{Y_{x,t}}(u)\big]^2=\big[U_{x,t}+V_{x,t}(1-u)+W_{x,t}(1-u)^2\big]^2.\end{aligned}\tag{4.5.5}$$

By adding both expressions in (4.5.5) and then denoting the sum by $\Psi_{x,t}(u)$, we obtain

$$\Psi_{x,t}(u) = 2U_{x,t}^2 + 2(2U_{x,t}W_{x,t} + V_{x,t}^2)(1-u)^2 + 2W_{x,t}^2(1-u)^4. \tag{4.5.6}$$

The integral of $\Psi_{x,t}(u)$ on interval $[0,1]$ leads to $d_{x,t}$

$$\begin{aligned}&d_{x,t} \equiv \int_0^1 \Psi_{x,t}(u)du =\\&=2U_{x,t}^2+2(2U_{x,t}W_{x,t} + V_{x,t}^2)\int_0^1(1-u)^2du + 2W_{x,t}^2\int_0^1(1-u)^4du.\end{aligned}\tag{4.5.7}$$

Since

$$\int_0^1 (1-u)^n du = \frac{1}{n+1}, \tag{4.5.8}$$

hence

$$d_{x,t} = 2U_{x,t}^2 + \frac{4}{3}U_{x,t}W_{x,t} + \frac{2}{3}V_{x,t}^2 + \frac{2}{5}W_{x,t}^2. \tag{4.5.9}$$

Given that the value of $W_{x,t}$ is close to 0, we further assume that

$$d_{x,t} \approx 2U_{x,t}^2 + \frac{2}{3}V_{x,t}^2. \tag{4.5.10}$$

Let us notice that expression (4.5.10) belongs to the minimized functional (4.5.1) and depends on coefficients a_x, b_x, k_t, s_{A_x}, s_{B_x}, s_{K_t}.

Thus, the functional F to be minimized is given by the following sum

$$F(a_x, b_x, k_t, s_{A_x}, s_{B_x}, s_{K_t}) = \sum_{t=1}^{T}\sum_{x=0}^{X} d_{x,t}. \tag{4.5.11}$$

By setting the partial derivatives of F to zero, the following system of normal equations is obtained

$$\begin{cases} \sum_{t=1}^{T}(a_x + b_x k_t - y_{x,t}) = 0, \\ \sum_{t=1}^{T}\left[(a_x+b_x k_t-y_{x,t})k_t+\frac{1}{3}(s_{B_x}k_t+s_{K_t}b_x-e_{x,t})s_{K_t}\right]=0, \\ \sum_{x=0}^{X}\left[(a_x+b_x k_t-y_{x,t})b_x+\frac{1}{3}(s_{B_x}k_t+s_{K_t}b_x-e_{x,t})s_{B_x}\right]=0, \\ \sum_{t=1}^{T}(s_{A_x} + s_{B_x}k_t + s_{K_t}b_x - e_{x,t}) = 0, \\ \sum_{t=1}^{T}(s_{A_x} + s_{B_x}k_t + s_{K_t}b_x - e_{x,t})k_t = 0, \\ \sum_{x=0}^{X}(s_{A_x} + s_{B_x}k_t + s_{K_t}b_x - e_{x,t})b_x = 0. \end{cases} \tag{4.5.12}$$

For full model identification, we impose additional restrictions on parameters b_x and k_t, the same as those we used for the Lee–Carter model, i.e.

$$\sum_{x=0}^{X} b_x = 1, \quad \sum_{t=1}^{T} k_t = 0. \tag{4.5.13}$$

We also assume that spreads $s_{A_x}, s_{B_x}, s_{K_t}$ are non-negative, i.e.

$$\forall_x \; s_{A_x}, s_{B_x} \geq 0, \quad \forall_t \; s_{K_t} \geq 0. \tag{4.5.14}$$

This set of normal equations can be solved numerically by means of an iterative procedure. In addition to numerical solution of the normal equations, there are also other minimizing algorithms, e.g. computer routines available in several mathematical packages.

4.6. Final remarks

The EFLC model presented in the chapter sets the stage for presenting other mortality models provided in the next part of the book. The estimation results of the EFLC model are contained in Chapter 6.

Chapter 5

Mortality models based on modified fuzzy numbers and complex functions

5.1. Introduction

The EFLC model and the concept of Oriented Fuzzy Numbers (OFN) can be considered as a starting point to create what we called here the algebra of Modified Fuzzy Numbers (MFN). The main difference between OFN and MFN lies in the definition of multiplication as an operation within an abstract algebra.

The modified fuzzy numbers will be used in the next section to propose the modified fuzzy Lee–Carter model MFLC. The last two sections of the chapter provide mortality models CFLC and QVLC based on the theory of complex functions.

5.2. Mortality model based on the algebra of modified fuzzy numbers

Let modified fuzzy numbers MFN be defined by analogy to the oriented fuzzy number OFN with the addition and multiplication operators $\oplus, \odot$ for MFN are given in Definition B.1 (Appendix B).

Then the Modified Fuzzy Lee–Carter model (MFLC model) is defined as follows

$$\check{Y}_{x,t} = \check{A}_x \oplus (\check{B}_x \odot \check{K}_t), \quad x = 0, 1, \ldots, X,\ t = 1, 2 \ldots, T, \tag{5.2.1}$$

where $\check{A}_x$, $\check{B}_x, \check{K}_t$ are modified fuzzy numbers represented by ordered pairs

$$\check{A}_x = (f_{A_x}, g_{A_x}), \check{B}_x = (f_{B_x}, g_{B_x}),\ \check{K}_t = (f_{K_t}, g_{K_t}), \tag{5.2.2}$$

with functions f_{A_x}, g_{A_x}, f_{B_x}, g_{B_x}, f_{K_t}, g_{K_t} defined for $u \in [0,1]$ as

$$f_{A_x}(u) = a_x - s_{A_x}(1-u), \quad g_{A_x}(u) = a_x + s_{A_x}(1-u),$$

$$f_{B_x}(u) = b_x - s_{B_x}(1-u), \quad g_{B_x}(u) = b_x + s_{B_x}(1-u), \tag{5.2.3}$$

$$f_{K_t}(u) = k_t - s_{K_t}(1-u), \quad g_{K_t}(u) = k_t + s_{K_t}(1-u).$$

Coefficients a_x, b_x, k_t and $s_{A_x}, s_{B_x}, s_{K_t}$ are the unknown parameters.

For model identification, we assume, as in the standard Lee–Carter model, that

$$\sum_{x=0}^{X} b_x = 1, \quad \sum_{t=1}^{T} k_t = 0 \tag{5.2.4}$$

and additionally

$$\forall_x \; s_{A_x}, s_{B_x} \geq 0, \quad \forall_t \; s_{K_t} \geq 0. \tag{5.2.5}$$

We will also assume that log-central death rates are represented by modified fuzzy numbers $\check{Y}_{x,t} = (f_{Y_{x,t}}, g_{Y_{x,t}})$ with

$$f_{Y_{x,t}}(u) = y_{x,t} - e_{x,t}(1-u), \quad g_{Y_{x,t}}(u) = y_{x,t} + e_{x,t}(1-u), \tag{5.2.6}$$

where $y_{x,t} = \ln m_{x,t}$ are (crisp) log-central death rates and "fuzziness parameters" $e_{x,t}$ are obtained by means of the fuzzification method.

By applying the definition of the addition and multiplication of modified fuzzy numbers (Definition B.1, Appendix B), we obtain

$$\check{B}_x \odot \check{K}_t = (f_{B_x \odot K_t}, \; g_{B_x \odot K_t}), \tag{5.2.7}$$

where

$$f_{B_x \odot K_t}(u) = b_x k_t + s_{B_x} s_{K_t}(1-u)^2,$$
$$g_{B_x \odot K_t}(u) = b_x k_t - s_{B_x} s_{K_t}(1-u)^2. \tag{5.2.8}$$

and

$$\check{A}_x \oplus (\check{B}_x \odot \check{K}_t) = (f_{A_x}, g_{A_x}) + (f_{B_x \odot K_t}, g_{B_x \odot K_t}) =$$
$$= (f_{A_x} + f_{B_x \odot K_t}, \; g_{A_x} + g_{B_x \odot K_t}), \tag{5.2.9}$$

where

$$
\begin{aligned}
f_{A_x}(u)+f_{B_x\odot K_t}(u) &= a_x+b_xk_t-s_{A_x}(1-u)+s_{B_x}s_{K_t}(1-u)^2,\\
g_{A_x}(u)+g_{B_x\odot K_t}(u) &= a_x+b_xk_t+s_{A_x}(1-u)-s_{B_x}s_{K_t}(1-u)^2.
\end{aligned}
\tag{5.2.10}
$$

The MFLC model can be then written as

$$
\check{Y}_{x,t} = \check{A}_x \oplus (\check{B}_x \odot \check{K}_t), \quad x = 0, 1, \ldots, X,\ t = 1, 2 \ldots, T, \tag{5.2.11}
$$

where

$$
\check{Y}_{x,t} = (f_{Y_{x,t}}, g_{Y_{x,t}}),\ \check{A}_x \oplus (\check{B}_x \odot \check{K}_t) = (f_{A_x\oplus B_x\odot K_t},\ g_{A_x\oplus B_x\odot K_t}) \tag{5.2.12}
$$

$$
f_{Y_{x,t}}(u) = y_{x,t} - e_{x,t}(1-u), \quad g_{Y_{x,t}}(u) = y_{x,t} + e_{x,t}(1-u), \tag{5.2.13}
$$

and

$$
\begin{aligned}
f_{A_x\oplus B_x\odot K_t}(u) &= a_x + b_xk_t - \left[s_{A_x}(1-u) - s_{B_x}s_{K_t}(1-u)^2\right],\\
g_{A_x\oplus B_x\odot K_t}(u) &= a_x + b_xk_t + \left[s_{A_x}(1-u) - s_{B_x}s_{K_t}(1-u)^2\right].
\end{aligned}
\tag{5.2.14}
$$

If $s_{A_x}, s_{B_x}, s_{K_t} \geq 0$ and $s_{A_x} - 2s_{B_x}s_{K_t} \geq 0$, expression $\check{A}_x \oplus (\check{B}_x \odot \check{K}_t)$ is a fuzzy number with a membership function close to the membership function of a triangular number with central value $a_x + b_xk_t$ and spread expressed as

$$
s_{A_x} - s_{B_x}s_{K_t}, \quad x = 0, 1 \ldots, X,\ t = 1, 2, \ldots, T. \tag{5.2.15}
$$

Example 5.1. For the sake of illustration, let us take the following parameter values of model (5.2.11): $a_x = 3,\ b_x = 0.05,\ k_t = -27,\ s_{A_x} = 0.15,\ s_{B_x} = 0.01,\ s_{K_t} = 4$.

Figure 5.1 shows a modified fuzzy number $\check{A}_x \oplus (\check{B}_x \odot \check{K}_t)$ and its corresponding symmetric fuzzy number resembling in shape a triangular number with a central value of 1.65 and a spread of 0.11.

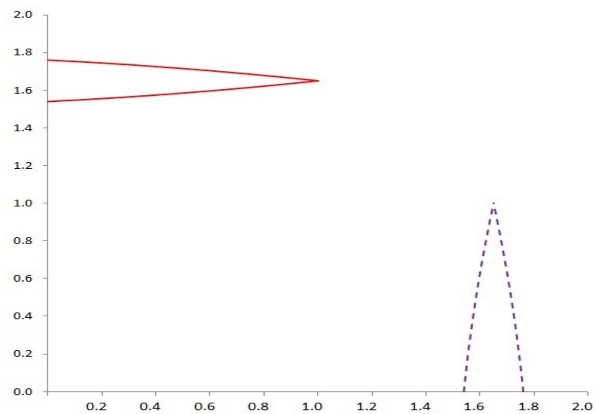

Figure 5.1. A fuzzy number $\check{A}_x \oplus (\check{B}_x \odot \check{K}_t)$ (solid line)
and the corresponding symmetric fuzzy number (dashed line)
Source: Developed by the authors

5.3. Parameters' estimation of the MFLC model

To estimate the parameters of the MFLC model (5.2.11), we shall use the Diamond distance between $\check{A}_x \oplus (\check{B}_x \odot \check{K}_t)$ and $\check{Y}_{x,t}$, i.e.

$$d_{x,t} \equiv D^2\left(\check{A}_x \oplus (\check{B}_x \odot \check{K}_t),\ \check{Y}_{x,t}\right),\ x=0,\ldots,X,\ t=1,\ldots,T. \quad (5.3.1)$$

Parameters' estimation reduces the minimization of the following criterion function

$$F(a_x, b_x, k_t, s_{A_x}, s_{B_x}, s_{K_t}) = \sum_{x=0}^{X}\sum_{t=1}^{T} d_{x,t}. \quad (5.3.2)$$

According to Definition 4.8 of the Diamond distance, (5.3.1) can be written as

$$d_{x,t} = \int_0^1 \left[f_{A_x}(u) + f_{B_x \odot K_t}(u) - f_{Y_{x,t}}(u)\right]^2 du + \quad (5.3.3)$$
$$+ \int_0^1 \left[g_{A_x}(u) + g_{B_x \odot K_t}(u) - g_{Y_{x,t}}(u)\right]^2 du.$$

To this end, we shall first transform the integrands starting with the following expressions

$$f_{A_x}(u) + f_{B_x \odot K_t}(u) - f_{Y_{x,t}}(u), \quad g_{A_x}(u) + g_{B_x \odot K_t}(u) - g_{Y_{x,t}}(u). \quad (5.3.4)$$

We have

$$f_{A_x}(u) + f_{B_x \odot K_t}(u) - f_{Y_{x,t}}(u) = \\ = a_x + k_t b_x - y_{x,t} - (s_{A_x} - e_{x,t})(1-u) + s_{B_x} s_{K_t} (1-u)^2, \tag{5.3.5}$$

where $u \in [0,1]$, $y_{x,t} = \ln m_{x,t}$.
Analogously,

$$g_{A_x}(u) + g_{B_x \odot K_t}(u) - g_{Y_{x,t}}(u) = \\ = a_x + k_t b_x - y_{x,t} + (s_{A_x} - e_{x,t})(1-u) - s_{B_x} s_{K_t} (1-u)^2. \tag{5.3.6}$$

Let us have

$$R_{x,t} = a_x + k_t b_x - y_{x,t},\ S_{x,t} = s_{A_x} - e_{x,t},\ U_{x,t} = s_{B_x} s_{K_t}. \tag{5.3.7}$$

Then, we receive

$$f_{A_x}(u) + f_{B_x \odot K_t}(u) - f_{Y_{x,t}}(u) = R_{x,t} - S_{x,t}(1-u) + U_{x,t}(1-u)^2, \\ g_{A_x}(u) + g_{B_x \odot K_t}(u) - g_{Y_{x,t}}(u) = R_{x,t} + S_{x,t}(1-u) - U_{x,t}(1-u)^2. \tag{5.3.8}$$

By squaring both sides of equation (5.3.8) and then denoting the squares of sums as $\Phi_{x,t}(u)$ and $\Psi_{x,t}(u)$, we obtain

$$\Phi_{x,t}(u) = R_{x,t}^2 - 2R_{x,t}\left[S_{x,t}(1-u) - U_{x,t}(1-u)^2\right] + \\ + \left[S_{x,t}(1-u) - U_{x,t}(1-u)^2\right]^2, \tag{5.3.9}$$

$$\Psi_{x,t}(u) = R_{x,t}^2 + 2R_{x,t}\left[S_{x,t}(1-u) - U_{x,t}(1-u)^2\right] + \\ + \left[S_{x,t}(1-u) - U_{x,t}(1-u)^2\right]^2. \tag{5.3.10}$$

The integral of $\Phi_{x,t}(u) + \Psi_{x,t}(u)$ which leads to the Diamond distance, has the following form

$$d_{x,t} = \int_0^1 \left[\Phi_{x,t}(u) + \Psi_{x,t}(u)\right] du = 2R_{x,t}^2 + 2S_{x,t}^2 \int_0^1 (1-u)^2 du + \\ = -4S_{x,t}U_{x,t} \int_0^1 (1-u)^3 du + 2U_{x,t}^2 \int_0^1 (1-u)^4 du. \tag{5.3.11}$$

We will use now a general formula

$$\int_0^1 (1-u)^n du = \int_0^1 u^n du = \frac{1}{n+1}. \qquad (5.3.12)$$

In this way, we arrive at

$$d_{x,t} = 2R_{x,t}^2 + \frac{2}{3}S_{x,t}^2 - S_{x,t}U_{x,t} + \frac{2}{5}U_{x,t}^2. \qquad (5.3.13)$$

Substituting expressions (5.3.7) for $R_{x,t}, S_{x,t}, U_{x,t}$ in (5.3.13), we have

$$\begin{aligned} d_{x,t} = {} & 2(a_x + k_t b_x - y_{x,t})^2 + \frac{2}{3}\left(s_{A_x} + e_{x,t}\right)^2 + \\ & - s_{B_x} s_{K_t}\left(s_{A_x} - e_{x,t}\right) + \frac{2}{5} s_{B_x}^2 s_{K_t}^2. \end{aligned} \qquad (5.3.14)$$

Let us notice that $d_{x,t}$ belongs to the minimized sum (5.3.2) and is a function of unknown parameters a_x, b_x, k_t, $s_{A_x}, s_{B_x}, s_{K_t}$. It is assumed, that log-central death rates $\ln m_{x,t}$ are known and fuzziness parameters $e_{x,t}$ are determined by using the data fuzzification algorithm (see the optimization problem (5.3.22)–(5.3.23) described below in this section).

Note that the sets of parameters $\{a_x, b_x, k_t\}$ and $\{s_{A_x}, s_{B_x}, s_{K_t}\}$ can be estimated separately. Let us thus first consider estimation of a_x, b_x, k_t by solving the following optimization problem

$$\text{minimize} \quad F_1(a_x, b_x, k_t) = \sum_{t=1}^{T} \sum_{x=0}^{X} (a_x + k_t b_x - y_{x,t})^2, \qquad (5.3.15)$$

with respect to b_x, k_t, given constraints (5.2.4).

To estimate b_x, k_t we can select a non-linear optimization package (e.g. one of the gradient algorithms available with the Matlab, or Excel Solver) minimizing the sum (5.3.15) given constraints (5.2.4).

To see how b_x, k_t are related each other, let us transform the following system of normal equations

$$\begin{cases} \sum_{t=1}^{T} (a_x + k_t b_x - y_{x,t}) = 0, & x = 0, 1, \ldots, X, \\ \sum_{t=1}^{T} k_t (a_x + k_t b_x - y_{x,t}) = 0, & x = 0, 1, \ldots, X, \\ \sum_{x=0}^{X} b_x (a_x + k_t b_x - y_{x,t}) = 0, & t = 1, 2, \ldots, T. \end{cases} \qquad (5.3.16)$$

We obtain

$$b_x = \frac{\sum_{t=1}^{T} y_{x,t} k_t}{\sum_{t=1}^{T} k_t^2}, \quad x = 0, 1, \ldots, X \tag{5.3.17}$$

and

$$k_t = \frac{\sum_{x=0}^{X} y_{x,t} b_x - \sum_{x=0}^{X} a_x b_x}{\sum_{x=0}^{X} b_x^2}, \quad t = 1, 2, \ldots, T. \tag{5.3.18}$$

Moreover, from the first normal equation of (5.3.16) and from (5.2.4) we have also

$$a_x = \frac{1}{T} \sum_{t=1}^{T} y_{x,t}, \quad x = 0, 1, \ldots, X. \tag{5.3.19}$$

Thus, parameters a_x represent average levels of mortality for different ages x.

In estimating the other model's parameters, $s_{A_x}, s_{B_x}, s_{K_t}$, formula (5.2.15) will be used, i.e. we assume that

$$e_{x,t} = s_{A_x} - s_{B_x} s_{K_t}, \quad x = 0, 1 \ldots, X, \ t = 1, 2, \ldots, T \tag{5.3.20}$$

and the following criterion function will be defined

$$S(s_{A_x}, s_{B_x}, s_{K_t}) = \sum_{t=1}^{T} e_{x,t} = T s_{A_x} - s_{B_x} \sum_{t=1}^{T} s_{K_t}, \tag{5.3.21}$$

According to the fuzziness assumption, $e_{x,t}$ are non-negative numbers and the smallest value they can take is 0. Estimates of parameters $s_{A_x}, s_{B_x}, s_{K_t}$ will be then calculated by solving the optimization problem

$$\text{minimize} \quad S(s_{A_x}, s_{B_x}, s_{K_t}) = T s_{A_x} - s_{B_x} \sum_{t=1}^{T} s_{K_t}, \tag{5.3.22}$$

subject to the following restrictions

$$\begin{aligned}
&\forall_x \ \forall_t \ s_{A_x}, s_{B_x}, s_{K_t} \geq 0, \ s_{A_x} - 2 s_{B_x} s_{K_t} \geq 0, \ \sum_{t=1}^{T} s_{K_t} = C, \\
&\quad a_x + b_x k_t + (s_{A_x} - s_{B_x} s_{K_t}) \geq \ln m_{x,t}, \\
&\quad a_x + b_x k_t - (s_{A_x} - s_{B_x} s_{K_t}) \leq \ln m_{x,t}, \\
&\quad s_{K_t} = \alpha t,
\end{aligned} \tag{5.3.23}$$

where a_x, b_x, k_t are replaced by their estimates and C is fixed and α is a constant parameter.

Solving the optimization problem (5.3.22)–(5.3.23) allows estimating parameters $s_{A_x}, s_{B_x}, s_{K_t}$, and makes it possible to calculate fuzziness parameters $e_{x,t}$ from (5.3.20).

5.4. Mortality model based on complex functions

In the last part of this chapter, oriented fuzzy numbers (OFN) will be represented by means of complex functions.

Let us consider a fuzzy triangular symmetric number $A = (a, s_A)$ with central value and spread a, s_A, respectively. In keeping with what [Kosiński *et al.* 2003, Kosiński, Prokopowicz 2004] proposed, the variable can be presented as an oriented fuzzy number $\vec{A} = (f_A, g_A)$, where

$$f_A(u) = a - s_A(1-u), \quad g_A(u) = a + s_A(1-u), \quad u \in [0,1]. \tag{5.4.1}$$

The OFN algebra can be easily transformed into a complex algebra using a complex form of $\vec{A} = (f_A, g_A)$, i.e.

$$A(u) = f_A(u) + ig_A(u), \quad u \in [0,1], \tag{5.4.2}$$

or a shortened form

$$A = f_A + ig_A, \tag{5.4.3}$$

where $i = \sqrt{-1}$ is an imaginary unit.

To be able to use the Gelfand–Mazur theorem (Section 4.2) that guarantees isometric isomorphism, we opted for the algebra of complex numbers $C(\mathcal{T})$ with $\mathcal{T}$ being a Hausdorff compact space, e.g. a closed interval $[0,1]$ or a Cartesian product of such intervals (Appendix B). We shall now apply a multiplication procedure appropriate for the complex numbers and use complex functions on interval $[0,1]$ as the representation of fuzzy numbers.

In the OFN algebra, the left- and right-hand side of model (4.3.1), i.e. $\vec{Y}_{x,t} = \vec{A}_x \oplus (\vec{B}_x \otimes \vec{K}_t)$, are expressed by oriented fuzzy numbers $\vec{Y}_{x,t} = (f_{Y_{x,t}}, g_{Y_{x,t}})$, $\vec{A}_x = (f_{A_x}, g_{A_x})$, $\vec{B}_x = (f_{B_x}, g_{B_x})$, $\vec{K}_t = (f_{K_t}, g_{K_t})$, (see Chapter 4) which were related to symmetric triangular fuzzy numbers with central values $y_{x,t}, a_x, b_x, k_t$ and spreads $e_{x,t}$, s_{A_x}, s_{B_x}, s_{K_t}, respectively.

Functions defining particular numbers are written as follows

$$
\begin{aligned}
f_{Y_{x,t}}(u) &= y_{x,t} - e_{x,t}(1-u), \quad g_{Y_{x,t}}(u) = y_{x,t} + e_{x,t}(1-u),\\
f_{A_x}(u) &= a_x - s_{A_x}(1-u), \quad g_{A_x}(u) = a_x + s_{A_x}(1-u),\\
f_{B_x}(u) &= b_x - s_{B_x}(1-u), \quad g_{B_x}(u) = b_x + s_{B_x}(1-u),\\
f_{K_t}(u) &= k_t - s_{K_t}(1-u), \quad g_{K_t}(u) = k_t + s_{K_t}(1-u).
\end{aligned}
\tag{5.4.4}
$$

We assume that the values of $y_{x,t} = \ln m_{x,t}$ are known and that fuzziness parameters $e_{x,t}$ were determined by using a fuzzification algorithm (i.e. the fuzzification method with switchings). The unknown model's parameters are coefficients a_x, b_x, k_t, s_{A_x}, s_{B_x}, s_{K_t}.

The discussion in this section focuses on the proposal to generalize the mortality model (4.3.1) by replacing oriented fuzzy numbers with complex functions. Therefore, we propose the Complex-Function Lee–Carter model (CFLC model) of the following form

$$Y_{x,t}(u) = A_x(u) + B_x(u)K_t(u), \quad x = 0, 1, \ldots, X,\ t = 1, \ldots, T, \tag{5.4.5}$$

where $Y_{x,t}(u), A_x(u), B_x(u), K_t(u)$ for $u \in [0,1]$ are complex functions expressed as

$$
\begin{aligned}
Y_{x,t}(u) &= f_{Y_{x,t}}(u) + i g_{Y_{x,t}}(u),\\
A_x(u) &= f_{A_x}(u) + i g_{A_x}(u),\\
B_x(u) &= f_{B_x}(u) + i g_{B_x}(u),\\
K_t(u) &= f_{K_t}(u) + i g_{K_t}(u),
\end{aligned}
\tag{5.4.6}
$$

$i = \sqrt{-1}$ is an imaginary unit and the functions on the right-hand side of (5.4.6) are defined in (5.4.4). The unknown model's parameters are coefficients a_x, b_x, k_t and $s_{A_x}, s_{B_x}, s_{K_t}$.

Product $B_x(u)K_t(u)$ on the right-hand side of (5.4.5) is obtained by multiplying complex numbers

$$
\begin{aligned}
B_x(u)K_t(u) &= \left(f_{B_x}(u) + i g_{B_x}(u)\right)\left(f_{K_t}(u) + i g_{K_t}(u)\right) =\\
&= [f_{B_x}(u) f_{K_t}(u) - g_{B_x}(u) g_{K_t}(u)] + i[f_{B_x}(u) g_{K_t}(u) + g_{B_x}(u) f_{K_t}(u)].
\end{aligned}
\tag{5.4.7}
$$

To obtain $A_x(u) + B_x(u)K_t(u)$ let us first calculate the elements in the brackets in (5.4.7). We have

$$f_{B_x}(u)f_{K_t}(u) = b_xk_t+s_{B_x}s_{K_t}(1-u)^2-(b_xs_{K_t}+k_ts_{B_x})\,(1-u),$$
$$g_{B_x}(u)g_{K_t}(u) = b_xk_t+s_{B_x}s_{K_t}(1-u)^2+(b_xs_{K_t}+k_ts_{B_x})\,(1-u). \tag{5.4.8}$$

By subtracting the respective sides, we obtain the first element of the real part of expression $B_x(u)K_t(u)$

$$f_{B_x}(u)f_{K_t}(u)-g_{B_x}(u)g_{K_t}(u)=-\left(2b_xs_{K_t} + 2k_ts_{B_x}\right)(1-u). \tag{5.4.9}$$

The first element of the real part of complex function $A_x(u)$ has the following form

$$f_{A_x}(u) = a_x - s_{A_x}(1-u) \tag{5.4.10}$$

and the formula defining the real part of complex function $A_x(u) + B_x(u)K_t(u)$ is

$$\begin{aligned} f_{A_x+B_xK_t}(u) =&f_{A_x}(u) + [f_{B_x}(u)f_{K_t}(u) - g_{B_x}(u)g_{K_t}(u)] = \\ &= a_x - \left(s_{A_x} + 2b_xs_{K_t} + 2k_ts_{B_x}\right)(1-u). \end{aligned} \tag{5.4.11}$$

A similar approach is employed to calculate the imaginary part of $A_x(u)+B_x(u)K_t(u)$. Namely, first the imaginary part of the expression on the right hand side of (5.4.7) is determined, i.e.

$$g_{B_x}(u)f_{K_t}(u) + f_{B_x}(u)g_{K_t}(u). \tag{5.4.12}$$

We have

$$g_{B_x}(u)f_{K_t}(u) = b_xk_t-s_{B_x}s_{K_t}(1-u)^2-(b_xs_{K_t}-k_ts_{B_x})\,(1-u),$$
$$f_{B_x}(u)g_{K_t}(u) = b_xk_t-s_{B_x}s_{K_t}(1-u)^2+(b_xs_{K_t}-k_ts_{B_x})\,(1-u). \tag{5.4.13}$$

Having added the respective sides, we obtain

$$g_{B_x}(u)f_{K_t}(u) + f_{B_x}(u)g_{K_t}(u) = 2b_xk_t - 2s_{B_x}s_{K_t}(1-u)^2. \tag{5.4.14}$$

Then, by adding $g_{A_x}(u)$, i.e.

$$g_{A_x}(u) = a_x + s_{A_x}(1-u), \tag{5.4.15}$$

the following imaginary part of the complex function $A_x(u){+}B_x(u)K_t(u)$ is obtained

$$g_{A_x+B_xK_t}(u) = g_{A_x}(u) + [g_{B_x}(u)f_{K_t}(u) + f_{B_x}(u)g_{K_t}(u)] = \tag{5.4.16}$$

$$= a_x + 2b_xk_t + s_{A_x}(1-u) - 2s_{B_x}s_{K_t}(1-u)^2.$$

Hence, formula (5.4.11) represents the real part of complex function $A_x(u) + B_x(u)K_t(u)$, and formula (5.4.16) its imaginary part.

5.5. Parameters' estimation of the CFLC model

Let us observe that complex functions $Y_{x,t}(u)$, $A_x(u)$, $B_x(u)$, $K_t(u)$ can be viewed as elements of the space of complex functions integrable with the square of the module.

In estimating the parameters of model (5.4.5) our interest focuses on minimizing the distance between $A_x(u) + B_x(u)K_t(u)$ and $Y_{x,t}(u)$. To achieve this, we shall use a metrics L_2 given by

$$\|(A_x{+}B_xK_t){-}Y_{x,t}\|_{L_2} = \int_0^1 |(A_x(u){+}B_x(u)K_t(u){-}Y_{x,t}(u)|^2 du, \tag{5.5.1}$$

where $|z|^2$ is the square of the module of complex function z, i.e. the sum of squares of the real and imaginary parts of z.

The sum of distances between $Y_{x,t}(u)$ and $A_x(u) + B_x(u)K_t(u)$ for $x = 0, 1, \dots, X, t = 1, 2, \dots, T$ will be taken as a criterion function, the minimization of which allows the unknown model's parameters to be estimated. The criterion function is given by

$$F(a_x, b_x, k_t, s_{A_x}, s_{B_x}, s_{K_t}) = \sum_{x=0}^{X}\sum_{t=1}^{T} \|(A_x{+}B_xK_t){-}Y_{x,t}\|_{L_2} = \tag{5.5.2}$$

$$= \sum_{x=0}^{X}\sum_{t=1}^{T}\int_0^1 |(A_x(u){+}B_x(u)K_t(u){-}Y_{x,t}(u)|^2 du.$$

The real part of expression $A_x(u) + B_x(u)K_t(u) - Y_{x,t}(u)$ is as follows

$$f_{A_x}(u){+}f_{B_x}(u)f_{K_t}(u) - f_{Y_{x,t}}(u) =$$

$$= f_{A_x}(u) + [f_{B_xK_t}(u) - g_{B_x}(u)g_{K_t}(u)] - f_{Y_{x,t}}(u) = \tag{5.5.3}$$

$$= (a_x - y_{x,t}) - (s_{A_x} + 2k_ts_{B_x} + 2b_xs_{K_t} - e_{x,t})(1-u)$$

and the imaginary part of $A_x(u) + B_x(u)K_t(u) - Y_{x,t}(u)$ is given by

$$g_{A_x}(u)+g_{B_xK_t}(u) - g_{Y_{x,t}}(u) =$$

$$=g_{A_x}(u)+[g_{B_x}(u)f_{K_t}(u) + f_{B_x}(u)g_{K_t}(u)] - g_{Y_{x,t}}(u) = \quad (5.5.4)$$

$$=(a_x - y_{x,t} + 2b_xk_t) + (s_{A_x} - e_{x,t})(1-u) - 2s_{B_x}s_{K_t}(1-u)^2.$$

To calculate the distance (5.5.1), the squares of expressions on the right-hand sides of (5.5.3) and (5.5.4) must be calculated, and next the integral of their sum. Then the criterion function (5.5.2) can be minimized with respect to unknown parameters $a_x, b_x, k_t, s_{A_x}, s_{B_x}, s_{K_t}$ by analogy to the optimization problem (5.3.15).

In the next section, we shall put forward more advanced modifications to the complex mortality model. The proposal presented in this section outlines the transition from a mortality model utilizing the algebra of oriented fuzzy numbers OFN to a model constructed in the framework of the quaternion algebra, which will be presented in the next section.

5.6. Quaternion-valued mortality model

The notion of a quaternion was introduced in 1843 by William Hamilton, an Irish mathematician, who attempted to generalize the complex algebra. The quaternion space is denoted by $\mathbb{H}$ as a tribute to the creator of quaternion theory. The basic terms and elements of the quaternion algebra are explained in Appendix B.

The Quaternion-Valued Lee–Carter model (QVLC model), will be defined by analogy to (4.3.1) or (5.4.5), i.e. as

$$\tilde{Y}_{x,t} = \tilde{A}_x + \tilde{B}_x\tilde{K}_t, \quad (5.6.1)$$

where $\tilde{Y}_{x,t}, \tilde{A}_x, \tilde{B}_x, \tilde{K}_t$ are the pairs of complex functions

$$\tilde{Y}_{x,t} = (f_{Y_{x,t}}, g_{Y_{x,t}}), \quad \tilde{A}_x = (f_{A_x}, g_{A_x}),$$
$$\tilde{B}_x = (f_{B_x}, g_{B_x}), \quad \tilde{K}_t = (f_{K_t}, g_{K_t}). \quad (5.6.2)$$

The ordered pairs of complex functions (5.6.2) are called quaternions (Definition B.6, Appendix B). Using the same symbols as in the case of the OFN algebra, functions in (5.6.2) can be written as

$$
\begin{aligned}
&f_{Y_{x,t}}(u) = y_{x,t} - i(1-u)e_{x,t}, \quad g_{Y_{x,t}}(u) = y_{x,t} + i(1-u)e_{x,t},\\
&f_{A_x}(u) = a_x - i(1-u)s_{A_x}, \quad g_{A_x}(u) = a_x + i(1-u)s_{A_x},\\
&f_{B_x}(u) = b_x - i(1-u)s_{B_x}, \quad g_{B_x}(u) = b_x + i(1-u)s_{B_x},\\
&f_{K_t}(u) = k_t - i(1-u)s_{K_t}, \quad g_{K_t}(u) = k_t + i(1-u)s_{K_t},
\end{aligned}
\tag{5.6.3}
$$

where $u \in [0,1]$ and $y_{x,t} = \ln m_{x,t}$.

Parameters $e_{x,t}$ as well as $s_{A_x}, s_{B_x}, s_{K_t}$ are determined by fuzzifying the log-central death rates. Thus, the unknown parameters of the QVLC model are only a_x, b_x, k_t. By analogy to the standard Lee–Carter model, the following restrictions are also imposed

$$
\sum_{x=0}^{X} b_x = 1, \quad \sum_{t=1}^{T} k_t = 0. \tag{5.6.4}
$$

A look at terms in (5.6.3) shows that functions $f_{Y_{x,t}}$, f_{A_x}, f_{B_x}, f_{K_t} correspond to conjugate functions $\bar{g}_{Y_{x,t}}, \bar{g}_{A_x}, \bar{g}_{B_x}, \bar{g}_{K_t}$ written as

$$
\begin{aligned}
&\bar{g}_{Y_{x,t}}(u) = y_{x,t} - i(1-u)e_{x,t},\\
&\bar{g}_{A_x}(u) = a_x - i(1-u)s_{A_x},\\
&\bar{g}_{B_x}(u) = b_x - i(1-u)s_{B_x},\\
&\bar{g}_{K_t}(u) = k_t - i(1-u)s_{K_t}.
\end{aligned}
\tag{5.6.5}
$$

Hence,

$$
\begin{aligned}
&\tilde{Y}_{x,t} = (\bar{g}_{Y_{x,t}}, g_{Y_{x,t}}), \quad \tilde{A}_x = (\bar{g}_{A_x}, g_{A_x}),\\
&\tilde{B}_x = (\bar{g}_{B_x}, g_{B_x}), \quad \tilde{K}_t = (\bar{g}_{K_t}, g_{K_t}).
\end{aligned}
\tag{5.6.6}
$$

In the matrix notation, $\tilde{A}_x(u)$ for $u \in [0,1]$ is written as

$$\tilde{A}_x(u) = \begin{bmatrix} \bar{g}_{A_x}(u) & g_{A_x}(u) \\ -\bar{g}_{A_x}(u) & g_{A_x}(u) \end{bmatrix}. \tag{5.6.7}$$

Analogously,

$$\tilde{B}_x(u) = \begin{bmatrix} \bar{g}_{B_x}(u) & g_{B_x}(u) \\ -\bar{g}_{B_x}(u) & g_{B_x}(u) \end{bmatrix} \tag{5.6.8}$$

and

$$\tilde{K}_t(u) = \begin{bmatrix} \bar{g}_{K_t}(u) & g_{K_t}(u) \\ -\bar{g}_{K_t}(u) & g_{K_t}(u) \end{bmatrix}. \tag{5.6.9}$$

The transformation of quaternion matrices $\tilde{A}_x, \tilde{B}_x, \tilde{K}_t$ leads us to

$$\begin{aligned} \tilde{A}_x(u) &= \begin{bmatrix} \bar{g}_{A_x}(u) & g_{A_x}(u) \\ -\bar{g}_{A_x}(u) & g_{A_x}(u) \end{bmatrix} = \\ &= \begin{bmatrix} a_x - is_{A_x}(1-u) & a_x + is_{A_x}(1-u) \\ -a_x + is_{A_x}(1-u) & a_x + is_{A_x}(1-u) \end{bmatrix} = \\ &= \begin{bmatrix} a_x & a_x \\ -a_x & a_x \end{bmatrix} + i(1-u) \begin{bmatrix} -s_{A_x} & s_{A_x} \\ s_{A_x} & s_{A_x} \end{bmatrix}. \end{aligned} \tag{5.6.10}$$

Analogously,

$$\begin{aligned} \tilde{B}_x(u) &= \begin{bmatrix} \bar{g}_{B_x}(u) & g_{B_x}(u) \\ -\bar{g}_{B_x}(u) & g_{B_x}(u) \end{bmatrix} = \\ &= \begin{bmatrix} b_x - is_{B_x}(1-u) & b_x + is_{B_x}(1-u) \\ -b_x + is_{B_x}(1-u) & b_x + is_{B_x}(1-u) \end{bmatrix} = \\ &= \begin{bmatrix} b_x & b_x \\ -b_x & b_x \end{bmatrix} + i(1-u) \begin{bmatrix} -s_{B_x} & s_{B_x} \\ s_{B_x} & s_{B_x} \end{bmatrix}, \end{aligned} \tag{5.6.11}$$

and

$$\tilde{K}_t(u) = \begin{bmatrix} \bar{g}_{K_t}(u) & g_{K_t}(u) \\ -\bar{g}_{K_t}(u) & g_{K_t}(u) \end{bmatrix} =$$

$$= \begin{bmatrix} k_t - is_{K_t}(1-u) & k_t + is_{K_t}(1-u) \\ -k_t + is_{K_t}(1-u) & k_t + is_{K_t}(1-u) \end{bmatrix} = \tag{5.6.12}$$

$$= \begin{bmatrix} k_t & k_t \\ -k_t & k_t \end{bmatrix} + i(1-u) \begin{bmatrix} -s_{K_t} & s_{K_t} \\ s_{K_t} & s_{K_t} \end{bmatrix}.$$

By applying the quaternion multiplication formula (Definition B.6, Appendix B) we obtain

$$\tilde{B}_x(u)\tilde{K}_t(u) =$$

$$= \begin{bmatrix} b_x & b_x \\ -b_x & b_x \end{bmatrix} \begin{bmatrix} k_t & k_t \\ -k_t & k_t \end{bmatrix} - (1-u)^2 \begin{bmatrix} -s_{B_x} & s_{B_x} \\ s_{B_x} & s_{B_x} \end{bmatrix} \begin{bmatrix} -s_{K_t} & s_{K_t} \\ s_{K_t} & s_{K_t} \end{bmatrix} + \tag{5.6.13}$$

$$+ i \left\{ (1-u) \begin{bmatrix} b_x & b_x \\ -b_x & b_x \end{bmatrix} \begin{bmatrix} -s_{K_t} & s_{K_t} \\ s_{K_t} & s_{K_t} \end{bmatrix} + (1-u) \begin{bmatrix} -s_{B_x} & s_{B_x} \\ s_{B_x} & s_{B_x} \end{bmatrix} \begin{bmatrix} k_t & k_t \\ -k_t & k_t \end{bmatrix} \right\}.$$

The multiplication of particular matrices in (5.6.13) gives the following

$$\begin{bmatrix} b_x & b_x \\ -b_x & b_x \end{bmatrix} \begin{bmatrix} k_t & k_t \\ -k_t & k_t \end{bmatrix} = b_x k_t \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = 2b_x k_t \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{5.6.14}$$

$$\begin{bmatrix} -s_{B_x} & s_{B_x} \\ s_{B_x} & s_{B_x} \end{bmatrix} \begin{bmatrix} -s_{K_t} & s_{K_t} \\ s_{K_t} & s_{K_t} \end{bmatrix} = s_{B_x} s_{K_t} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} =$$

$$= 2 s_{B_x} s_{K_t} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{5.6.15}$$

$$\begin{bmatrix} b_x & b_x \\ -b_x & b_x \end{bmatrix} \begin{bmatrix} -s_{K_t} & s_{K_t} \\ s_{K_t} & s_{K_t} \end{bmatrix} = b_x s_{K_t} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} =$$

$$= 2 b_x s_{K_t} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{5.6.16}$$

$$\begin{bmatrix} -s_{B_x} & s_{B_x} \\ s_{B_x} & s_{B_x} \end{bmatrix}\begin{bmatrix} k_t & k_t \\ -k_t & k_t \end{bmatrix} = s_{B_x} k_t \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \tag{5.6.17}$$
$$= 2 s_{B_x} k_t \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Hence, we have

$$\tilde{B}_x(u)\tilde{K}_t(u) = 2b_x k_t \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} - 2(1-u)^2 s_{B_x} s_{K_t} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \tag{5.6.18}$$
$$+ 2i(1-u)\left\{ b_x s_{K_t} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + k_t s_{B_x} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

Quaternion $\tilde{A}_x(u)$ has the form

$$\tilde{A}_x(u) = \begin{bmatrix} a_x - i s_{A_x}(1-u) & a_x + i s_{A_x}(1-u) \\ -a_x + i s_{A_x}(1-u) & a_x + i s_{A_x}(1-u) \end{bmatrix} = \tag{5.6.19}$$
$$= a_x \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} + i(1-u) s_{A_x} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}.$$

The adding of quaternion $\tilde{A}_x(u)$ to $\tilde{B}_x(u)\tilde{K}_t(u)$ leads to a different form of the right-hand side of the model (5.6.1)

$$\tilde{A}_x(u) + \tilde{B}_x(u)\tilde{K}_t(u) =$$

$$= a_x \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} + 2b_x k_t \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} - 2(1-u)^2 s_{B_x} s_{K_t} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \tag{5.6.20}$$

$$+ i(1-u)\left\{ s_{A_x} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} + 2b_x s_{K_t} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + 2k_t s_{B_x} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

By analogy, log-central death rates $y_{x,t} = \ln m_{x,t}$ can be transformed into quaternions $\tilde{Y}_{x,t}$ by adopting matrix notation

$$\tilde{Y}_{x,t}(u) = \begin{bmatrix} \bar{g}_{Y_{x,t}}(u) & g_{Y_{x,t}}(u) \\ -\bar{g}_{Y_{x,t}}(u) & g_{Y_{x,t}}(u) \end{bmatrix}, \tag{5.6.21}$$

where

$$\bar{g}_{Y_{x,t}}(u) = y_{x,t} - ie_{x,t}(1-u), \quad g_{Y_{x,t}}(u) = y_{x,t} + ie_{x,t}(1-u). \quad (5.6.22)$$

Equivalently, we can write the above as

$$\tilde{Y}_{x,t}(u) = \begin{bmatrix} y_{x,t} - ie_{x,t}(1-u) & y_{x,t} + ie_{x,t}(1-u) \\ -y_{x,t} + ie_{x,t}(1-u) & y_{x,t} + ie_{x,t}(1-u) \end{bmatrix}. \quad (5.6.23)$$

5.7. Parameters' estimation of the QVLC model

The parameters of the QVLC model (5.6.1) will be estimated using expression (5.6.20). To this end, the following norm will be introduced for the quaternion space $\mathbb{H}$

$$||F||^2_{\mathcal{L}_2} = \int_0^1 ||F(u)||^2_{\mathbb{H}} du, \quad (5.7.1)$$

where $||\cdot||^2_{\mathbb{H}}$ under the integral is the square of the norm of an element in space $\mathbb{H}$ (Definition B.10, Appendix B).

Hence, to estimate the model's parameters, the following functional will be employed

$$F(a_x, b_x, k_t, s_{A_x}, s_{B_x}, s_{K_t}) = \sum_{x=0}^{X}\sum_{t=1}^{T} ||\tilde{Y}_{x,t} - \left(\tilde{A}_x + \tilde{B}_x\tilde{K}_t\right)||^2_{\mathcal{L}_2}. \quad (5.7.2)$$

The quaternion norm in space $\mathbb{H}$ can be determined using only terms from the first row of the complex matrix representing a given quaternion. For quaternion $\tilde{A}_x(u) + \tilde{B}_x(u)\tilde{K}_t(u)$, complex functions $f_{A_x+B_xK_t}(u)$ and $g_{A_x+B_xK_t}(u)$ are as follows

$$f_{A_x+B_xK_t}(u) = a_x - 2(1-u)^2 s_{B_x}s_{K_t} - i(1-u)(s_{A_x} + 2k_t s_{B_x}),$$
$$g_{A_x+B_xK_t}(u) = a_x + 2b_x k_t + i(1-u)(s_{A_x} + 2b_x s_{K_t}). \quad (5.7.3)$$

The terms will be used to estimate the distance between the left-hand and right-hand sides of (5.6.1), i.e. between $\tilde{Y}_{x,t}$ and $\tilde{A}_x + \tilde{B}_x\tilde{K}_t$.

We already know that $\tilde{A}_x(u) + \tilde{B}_x(u)\tilde{K}_t(u)$is represented by two complex functions (5.7.3). Similarly, quaternion $\tilde{Y}_{x,t}(u)$ is defined by two complex functions (5.6.22). Therefore, $\tilde{Y}_{x,t}-(\tilde{A}_x+\tilde{B}_x\tilde{K}_t)$ can be defined by the following two functions

$$\begin{aligned}
\phi(u)&=\bar{g}_{Y_{x,t}}(u)-f_{A_x+B_xK_t}(u)=\\
&=y_{x,t}-a_x+2(1-u)^2 s_{B_x}s_{K_t}+i(1-u)(s_{A_x}-e_{x,t}+2k_t s_{B_x}),\\
\psi(u)&=g_{Y_{x,t}}(u)-g_{A_x+B_xK_t}(u)=\\
&=y_{x,t}-a_x-2b_xk_t-i(1-u)(s_{A_x}-e_{x,t}+2b_x s_{K_t}).
\end{aligned} \tag{5.7.4}$$

The formula for the squared norm of quaternion $\tilde{Y}_{x,t}-(\tilde{A}_x+\tilde{B}_x\tilde{K}_t)$ in space $\mathbb{H}$ is

$$||\tilde{Y}_{x,t}-(\tilde{A}_x+\tilde{B}_x\tilde{K}_t)||^2_{\mathbb{H}}=|\phi(u)|^2+|\psi(u)|^2. \tag{5.7.5}$$

The squared modules on the right-hand side of (5.7.5) can be transformed to

$$\begin{aligned}
|\phi(u)|^2&=\big[y_{x,t}-a_x+2(1-u)^2 s_{B_x}s_{K_t}\big]^2+\\
&+(1-u)^2(s_{A_x}-e_{x,t}+2k_t s_{B_x})^2=\\
&=(y_{x,t}-a_x)^2+4(1-u)^2 s_{B_x}s_{K_t}(y_{x,t}-a_x)+\\
&+4(1-u)^4 s^2_{B_x}s^2_{K_t}+(1-u)^2(s_{A_x}-e_{x,t}+2k_t s_{B_x})^2,\\
|\psi(u)|^2&=(y_{x,t}-a_x-2b_xk_t)^2+(1-u)^2(s_{A_x}-e_{x,t}+2b_x s_{K_t})^2.
\end{aligned} \tag{5.7.6}$$

The norm (5.7.1) of the quaternion $\tilde{Y}_{x,t}-(\tilde{A}_x+\tilde{B}_x\tilde{K}_t)$ can be then written as

$$||\tilde{Y}_{x,t}-(\tilde{A}_x+\tilde{B}_x\tilde{K}_t)||^2_{\mathcal{L}_2}=\int_0^1|\phi(u)|^2du+\int_0^1|\psi(u)|^2du, \tag{5.7.7}$$

and both integrals on the right-hand side of (5.7.7) are equal, respectively,

$$\int_0^1 |\phi(u)|^2 du = (y_{x,t} - a_x)^2 + 4 s_{B_x} s_{K_t} (y_{x,t} - a_x) \int_0^1 (1-u)^2 du +$$

$$+ 4 s_{B_x}^2 s_{K_t}^2 \int_0^1 (1-u)^4 du + (s_{A_x} - e_{x,t} + 2k_t s_{B_x})^2 \int_0^1 (1-u)^2 du = \qquad (5.7.8)$$

$$= (y_{x,t} - a_x)^2 + \frac{4}{3} s_{B_x} s_{K_t} (y_{x,t} - a_x) + \frac{4}{5} s_{B_x}^2 s_{K_t}^2 + \frac{1}{3} (s_{A_x} - e_{x,t} + 2k_t s_{B_x})^2$$

and

$$\int_0^1 |\psi(u)|^2 du = (y_{x,t} - a_x - 2b_x k_t)^2 + \int_0^1 (1-u)^2 (s_{A_x} - e_{x,t} + 2b_x s_{K_t})^2 du$$

$$= (y_{x,t} - a_x - 2b_x k_t)^2 + \frac{1}{3} (s_{A_x} - e_{x,t} + 2b_x s_{K_t})^2. \qquad (5.7.9)$$

Let us have

$$d_{x,t} \equiv ||\tilde{Y}_{x,t} - (\tilde{A}_x + \tilde{B}_x \tilde{K}_t)||_{\mathcal{L}_2}^2 = \int_0^1 |\phi(u)|^2 du + \int_0^1 |\psi(u)|^2 du =$$

$$= (y_{x,t} - a_x)^2 + \frac{4}{3} s_{B_x} s_{K_t} (y_{x,t} - a_x) + \frac{1}{3} (s_{A_x} - e_{x,t} + 2k_t s_{B_x})^2 + \qquad (5.7.10)$$

$$+ (y_{x,t} - a_x - 2b_x k_t)^2 + \frac{1}{3} (s_{A_x} - e_{x,t} + 2b_x s_{K_t})^2 + \frac{4}{5} s_{B_x}^2 s_{K_t}^2.$$

Hence, functional (5.7.2) used for estimating the model's parameters can be written as

$$F(a_x, b_x, k_t, s_{A_x}, s_{B_x}, s_{K_t}) = \sum_{x=0}^{X} \sum_{t=1}^{T} d_{x,t}. \qquad (5.7.11)$$

It is also assumed that the values of log-central death rates $y_{x,t} = \ln m_{x,t}$ are known and coefficients $e_{x,t}$ are derived by means of the switching fuzzification method described in Section 4.4 (Chapter 4).

The general concept for estimating parameters of the quaternion model requires the selection of a non-linear optimization algorithm available in several mathematical packages, to minimize (5.7.11) given restrictions (5.6.4).

To see how parameters a_x, b_x, k_t are related each other, let us derive the system of normal equations. We have

$$\begin{cases} \sum_{t=1}^{T}\left[(y_{x,t}-a_x)+\frac{2}{3}s_{B_x}s_{K_t}+(y_{x,t}-a_x-2b_xk_t)\right]=0, \\ \sum_{t=1}^{T}\left[k_t(y_{x,t}-a_x-2b_xk_t)-\frac{1}{3}s_{K_t}(s_{A_x}-e_{x,t}+2b_xs_{K_t})\right]=0, \\ \sum_{x=0}^{X}\left[b_x(y_{x,t}-a_x-2b_xk_t)-\frac{1}{3}s_{B_x}(s_{A_x}-e_{x,t}+2k_ts_{B_x})\right]=0. \end{cases} \tag{5.7.12}$$

From equations (5.7.12) we have

$$a_x=\frac{1}{T}\sum_{t=1}^{T}y_{x,t}+\frac{1}{3}s_{B_x}\frac{1}{T}\sum_{t=1}^{T}s_{K_t}=\bar{y}_x+\frac{1}{3}s_{B_x}\bar{s}_{K_t}, \tag{5.7.13}$$

$$b_x=\frac{\sum_{t=1}^{T}k_t(y_{x,t}-a_x)-\frac{1}{3}\sum_{t=1}^{T}s_{K_t}(s_{A_x}-e_{x,t})}{2\sum_{t=1}^{T}k_t^2+\frac{2}{3}\sum_{t=1}^{T}s_{K_t}^2}, \tag{5.7.14}$$

$$k_t=\frac{\sum_{x=0}^{X}b_x(y_{x,t}-a_x)-\frac{1}{3}\sum_{x=0}^{X}s_{B_x}(s_{A_x}-e_{x,t})}{2\sum_{x=0}^{X}b_x^2+\frac{2}{3}\sum_{x=0}^{X}s_{B_x}^2}. \tag{5.7.15}$$

The set of normal equations (5.7.12) can be solved numerically by means of an iterative procedure.

5.8. Final remarks

The mortality models proposed in this chapter are based on the same approach as that used to build the Extended Fuzzy Lee–Carter model (EFLC) utilizing the algebra of modified fuzzy number, complex functions or quaternions.

The parameters of the fuzzy model (MFLC) and complex models (CFLC and QVLC) are estimated by fuzzifying log-central mortality rates and optimizing some non-linear criterion functions.

The next chapter presents for illustration some real data-based estimates obtained with some of the models proposed in this book, as well as comparative analysis of prediction accuracy of the models.

Chapter 6

Models' estimation and evaluation based on the real data

6.1. Introduction

To illustrate the theoretical discussions in previous chapters presenting the proposals of new models, the following classes of mortality models will be estimated: DLCH (3.3.11), DGOBHM (3.6.16)–(3.6.19), MFLC (5.2.11)–(5.2.14) and QVLC (5.6.1)–(5.6.3).

In the estimation, real data will be used and the *ex-post* forecasting errors will be compared with the errors yielded by the SLC model (1.5.2) and, occasionally, with the errors of the DDLC (1.6.12), DGOB (1.9.26)–(1.9.27) and DMMP (1.10.18)–(1.10.19) models.

The following analysis utilizes the age-specific death rates for males and females in Poland from years 1958–2014. Data were sourced from the Human Mortality Database (www.mortality.org) and the GUS database (stat.gov.pl). The 2001–2014 death rates were only used to evaluate the models' forecasting properties (were excluded from estimations).

In the case of the MFLC and QVLC models, switching points were used to fuzzify the input data, before the estimation procedure was employed. The switchings were identified using the JL test, the theoretical foundations of which are presented in Chapter 3, Section 3.2.

For the hybrid models DLCH, DGOBH, two common switchings were adopted. They were determined from the data contained in Table 6.1 by observing the most frequent switchings. The period under consideration 1958–2000 was assumed to have two common switchings 1966 and 1991, meaning that the parameters of the hybrid models were estimated separately for three sub-periods, i.e. $I_0 = [1958, 1966), I_1 = [1966, 1991)$ and $I_2 = [1991, 2000]$. Parameters' estimates obtained for the last sub-period were next used to assess the *ex-post* prediction accuracy of the hybrid models.

6.2. Results of switching points' identification for the mortality data of Poland

Table 6.1. Switching points (years)

Males			Females		
x	m	year	x	m	year
0	33	1991	0	36	1994
4	29	1987	1	16	1974
11	12	1970	23	8	1966
15	37	1995	24	8	1966
17	30	1988	25	8	1966
20	6	1964	27	8	1966
26	33	1991	29	8	1966
27	37	1995	30	11	1969
31	32	1990	35	36	1994
33	33	1991	40	33	1991
35	33	1991	41	34	1992
37	33	1991	42	9	1967
38	33	1991	43	7	1965
38	33	1991	60	33	1991
40	34	1992	61	34	1992
41	33	1991	62	33	1991
42	33	1991	66	33	1991
43	33	1991	67	33	1991
44	33	1991	68	35	1993
45	33	1991	69	33	1991
46	33	1991	70	32	1990
47	33	1991	71	33	1991
50	33	1991	72	37	1995
51	33	1991	73	35	1993
52	33	1991	74	36	1994
53	33	1991	75	37	1995
54	33	1991	86	27	1985
55	33	1991	88	37	1995
56	33	1991	89	37	1995
57	33	1991	90	34	1992
58	34	1992			
59	33	1991			
61	33	1991			
62	33	1991			
63	33	1991			
64	33	1991			
67	33	1991			
93	26	1984			
99	9	1967			

Source: Own calculations.

Table 6.1 shows the results of the JL test obtained for the age-specific death rates for males and females noted in Poland in the calendar years 1958–2000. Switching years were selected using as the criterion statistically significant switching points at the 0.05 level. Each age group was assigned one such point.

6.3. Estimation results

6.3.1. The DLCH model

Figures 6.1–6.8 show estimates a_x, b_x, k_t, σ_x^2 of the parameters of the DLCH model, which were obtained with the age-specific mortality rates for Poland from the years 1958–2000. To make the estimation resluts easier to present, the estimates are plotted as graphs.

Curves illustrated in Figures 6.1 and 6.2 show the average log-central age-specific rates of mortality (for males and females), plotted separately for each of the mortality regimes I_0, I_1, I_2. All the curves exhibit a typical "bath tube" shape, i.e. with high values around the infant ages, followed by minimal rates at the childhood ages, higher accidental mortality at young adulthood ages and increasing mortality at adulthood and old ages with nearly constant rate of increase. The "accident hump" at adolescence stands for higher mortality rates due to accidental deaths caused by augmented risk-taking behaviour as well as increased suicide rates. Note that the observed humps, both for males and females, are more demonstrable in the last time period I_2.

The arrangement of curves in Figures 6.3 and 6.4 shows that in some age groups of males, death rates are markedly more sensitive to temporal changes in mortality than in the case of women. Especially, significant differences are observed for the time period $I_1 = [1965, 1990)$, what can be explained by the health crisis of the 1970s and 1980s in Poland.

Figures 6.5 and 6.6 indicate that in the period of analysis the trend of mortality was generally declining, with the decline being faster in the subpopulation of women, except for the period I_1, when the general mortality trend for males was rising. The estimated functions $\kappa(t, l)$ obtained with the DLCH model differ significantly from estimates of parameters κ_t obtained with the SLC model, therefore the latter are also plotted for comparison.

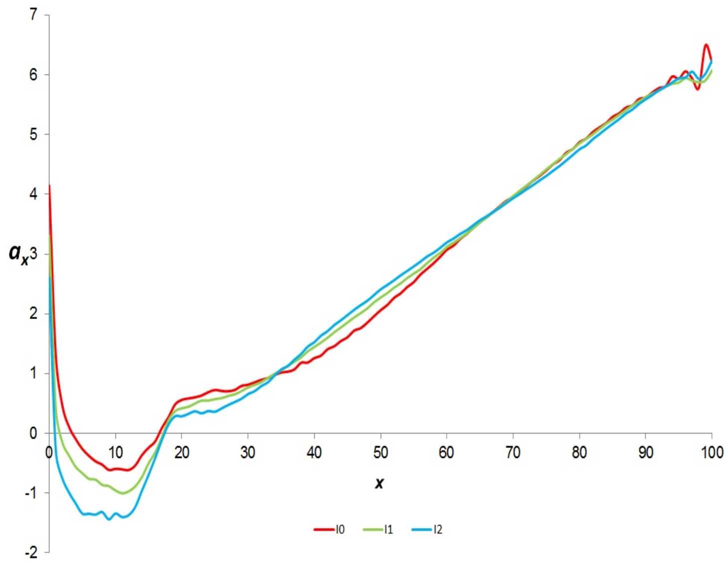

Figure 6.1. Estimates a_x, $x = 0, 1, \dots, 100$
obtained with the DLCH model for sub-periods I_0, I_1, I_2 (males)
Source: Developed by the authors

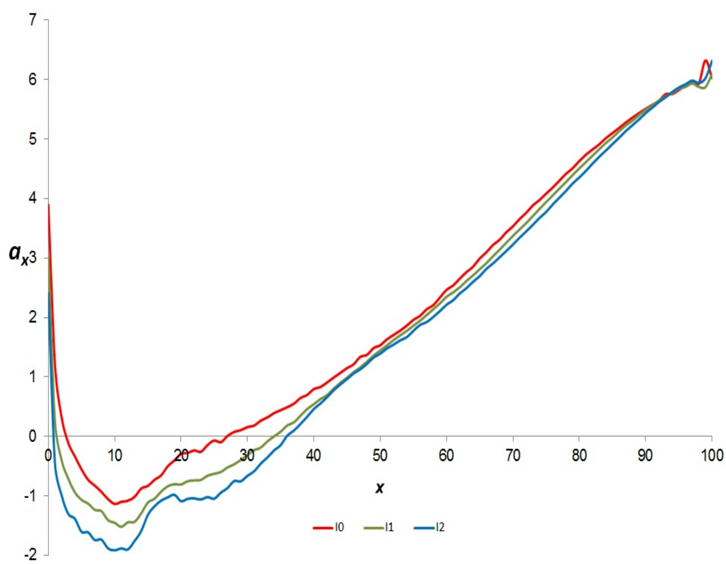

Figure 6.2. Estimates a_x, $x = 0, 1, \dots, 100$
obtained with the DLCH model for sub-periods I_0, I_1, I_2 (females)
Source: Developed by the authors

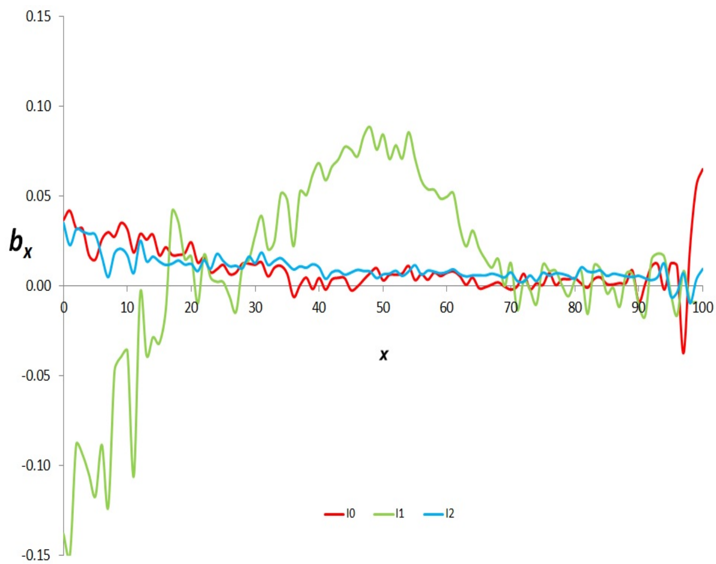

Figure 6.3. Estimates b_x, $x = 0, 1, \ldots, 100$
obtained with the DLCH model for sub-periods I_0, I_1, I_2 (males)
Source: Developed by the authors

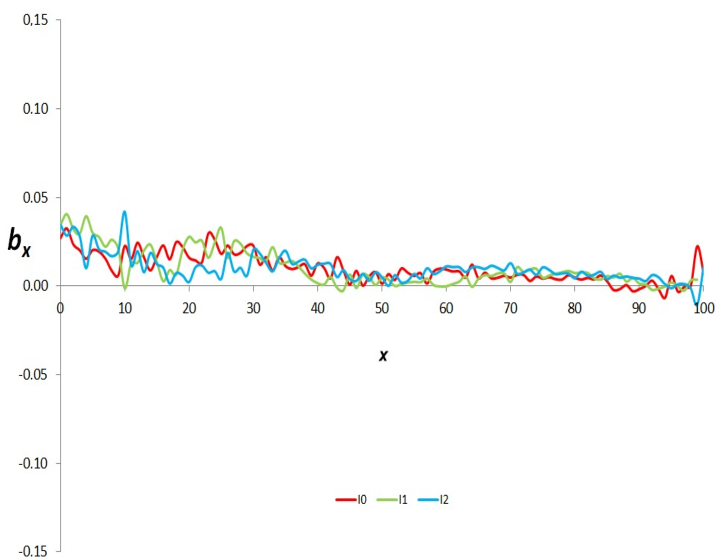

Figure 6.4. Estimates b_x, $x = 0, 1, \ldots, 100$
obtained with the DLCH model for sub-periods I_0, I_1, I_2 (females)
Source: Developed by the authors

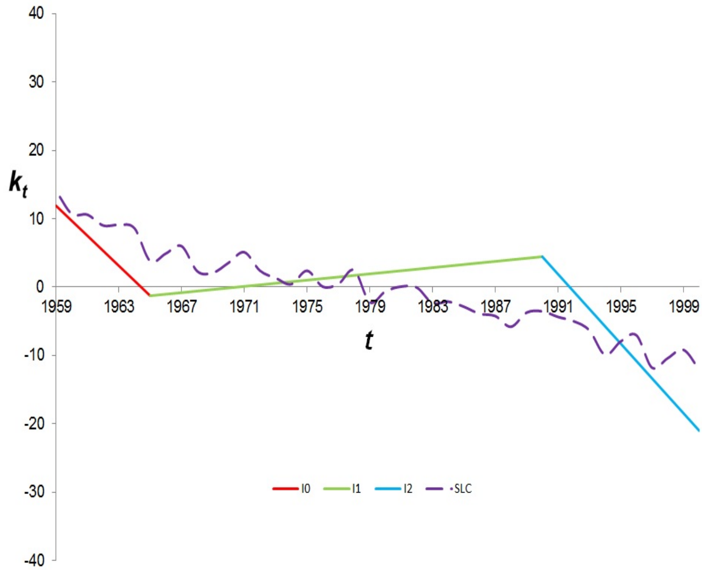

Figure 6.5. Estimated functions $\kappa(t,l)$ obtained with model DLCH for sub-periods I_0, I_1, I_2 and estimates k_t obtained with model SLC (males)
Source: Developed by the authors

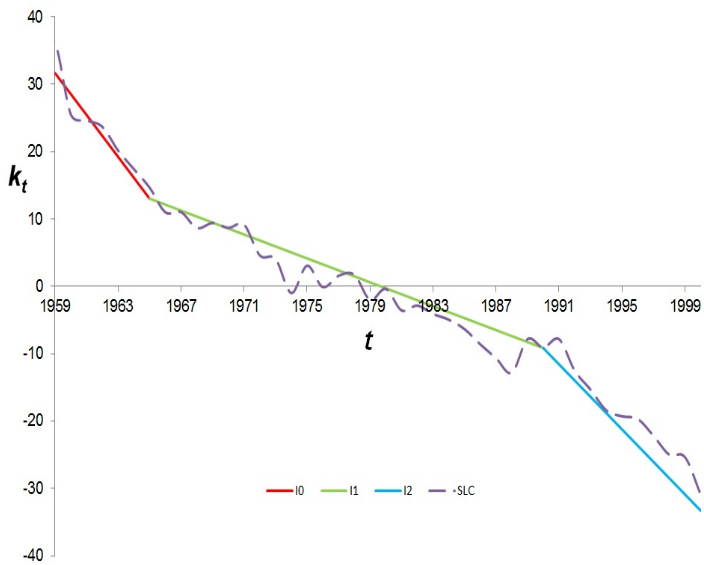

Figure 6.6. Estimated functions $\kappa(t,l)$ obtained with model DLCH for sub-periods I_0, I_1, I_2 and estimates k_t obtained with model SLC (females)
Source: Developed by the authors

Parameters σ_x^2 represent volatility of mortality rates and are illustrated in Figures 6.7 and 6.8.

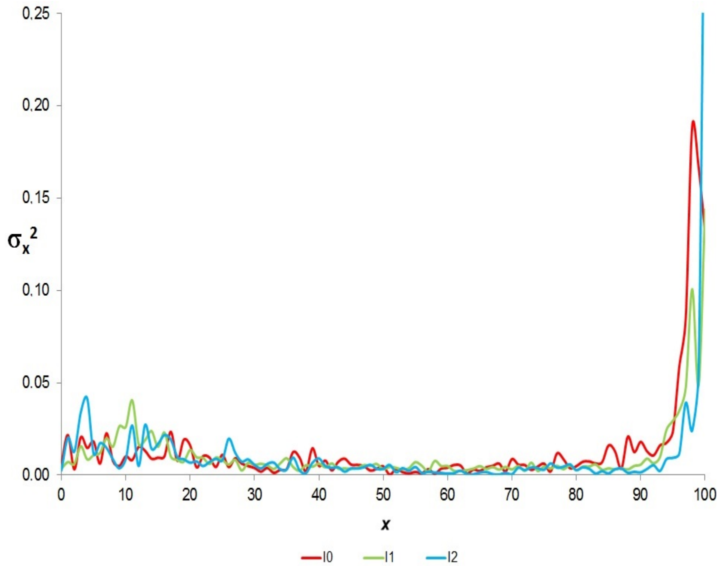

Figure 6.7. Estimates σ_x^2, $x = 0, 1, \dots, 100$
obtained with the DLCH model for sub-periods I_0, I_1, I_2 (males)
Source: Developed by the authors

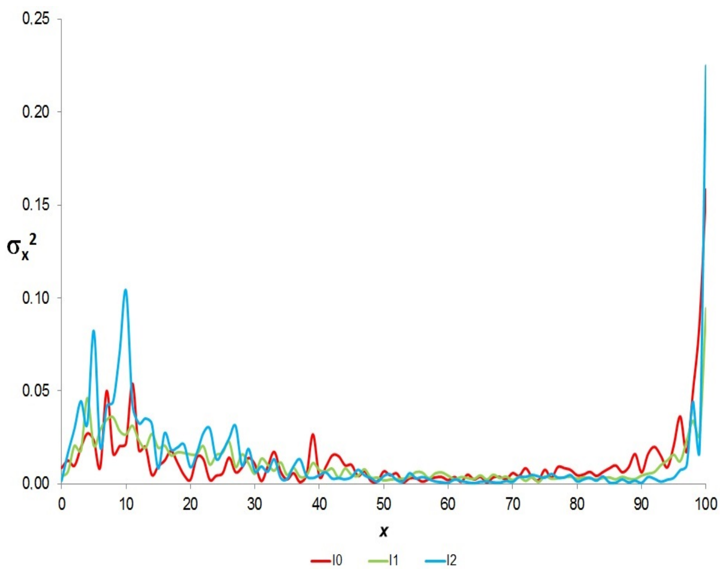

Figure 6.8. Estimates σ_x^2, $x = 0, 1, \dots, 100$
obtained with the DLCH model for sub-periods I_0, I_1, I_2 (females)
Source: Developed by the authors

The forecasting accuracy of models DLCH and SLC were compared using *ex-post* errors measured for each year in the time period 2001–2014 that was omitted from the parameters' estimation.

To estimate error sizes two types of measures were used: a mean squared error *MSE* and a mean absolute deviation *MAD*. In the case of the DLCH model they are given by the following formulas

$$MSE_t^{(\mathrm{DLCH})}=\sqrt{\frac{1}{101}\sum_{x=0}^{100}\left[\ln m_{x,t+1}(l)-(\ln m_{x,t}(l)+b_x(l)d(l))\right]^2},$$
$$MAD_t^{(\mathrm{DLCH})}=\frac{1}{101}\sum_{x=0}^{100}\left|\ln m_{x,t+1}(l)-(\ln m_{x,t}(l)+b_x(l)d(l))\right|. \quad (6.3.1)$$

In the case of the SLC model *MSE* and *MAD* take the form, respectively,

$$MSE_t^{(\mathrm{SLC})}=\sqrt{\frac{1}{101}\sum_{x=0}^{100}\left[\ln m_{x,t}-(a_x+b_xk_t)\right]^2},$$
$$MAD_t^{(\mathrm{SLC})}=\frac{1}{101}\sum_{x=0}^{100}\left|\ln m_{x,t}-(a_x+b_xk_t)\right|, \quad (6.3.2)$$

Error measures obtained with both these measures for models SLC and DLCH are shown in Tables 6.2 i 6.3.

The data in Tables 6.2 and 6.3 lead to a conclusion that the DLCH model has better forecasting properties, particularly regarding the subpopulation of males.

Columns 3 and 5 show that the mean errors (measured by means of *MSE* and *MAD*) with respect to to the predicted log-central age-specific mortality rates are mostly better to those obtained with the SLC model (columns 2 and 4).

Table 6.2. *Ex-post* comparisons of *MSE* values for models SLC and DLCH

Year	Males		Females	
	SLC	DLCH	SLC	DLCH
2001	0.197	0.096	0.098	0.110
2002	0.204	0.101	0.122	0.132
2003	0.215	0.093	0.122	0.120
2004	0.223	0.105	0.132	0.135
2005	0.230	0.119	0.146	0.186
2006	0.232	0.143	0.152	0.190
2007	0.238	0.167	0.172	0.193
2008	0.257	0.182	0.174	0.201
2009	0.281	0.170	0.191	0.245
2010	0.330	0.170	0.190	0.226
2011	0.341	0.203	0.218	0.251
2012	0.373	0.218	0.215	0.247
2013	0.406	0.226	0.246	0.286
2014	0.469	0.225	0.273	0.301

Source: Own calculations.

Table 6.3. *Ex-post* comparisons of *MAD* values for models SLC and DLCH

Year	Males		Females	
	SLC	DLCH	SLC	DLCH
2001	0.182	0.072	0.083	0.071
2002	0.185	0.076	0.107	0.087
2003	0.195	0.064	0.109	0.082
2004	0.206	0.078	0.117	0.094
2005	0.214	0.090	0.129	0.124
2006	0.214	0.105	0.130	0.130
2007	0.219	0.123	0.152	0.128
2008	0.234	0.133	0.156	0.140
2009	0.250	0.125	0.170	0.166
2010	0.302	0.126	0.167	0.168
2011	0.307	0.148	0.191	0.188
2012	0.335	0.153	0.185	0.182
2013	0.359	0.163	0.221	0.207
2014	0.430	0.168	0.245	0.221

Source: Own calculations.

6.3.2. The DGOBHM model

Figures 6.9–6.18 show the estimates of $\alpha_x(l), \ln\mu_x(0,l) = \ln_{x0}(l)$, $\beta_x(l)$, $\gamma_x(l)$ $(l = 0,1,2)$ obtained with the DGOBH moment model for males and females and for mortality regimes I_0, I_1, I_2, using the iterative estimation procedure described in Section 3.6.5.

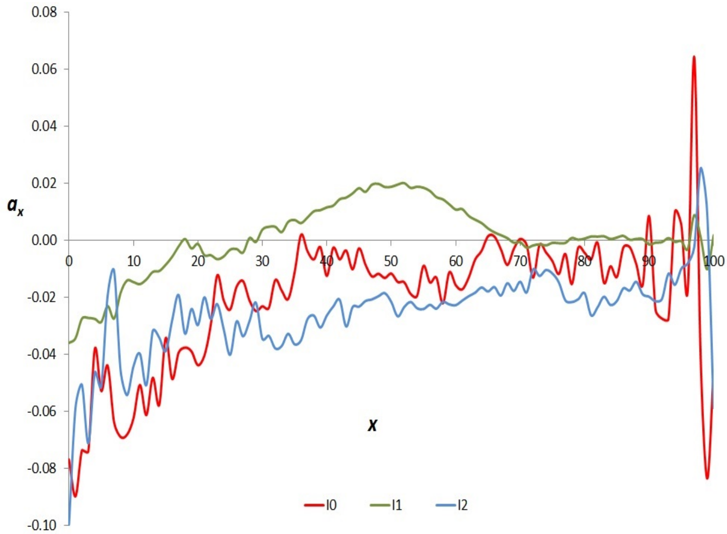

Figure 6.9. Estimates of α_x, $x = 0, 1, \ldots, 100$
obtained with the DGOBHM model for sub-periods I_0, I_1, I_2 (males)
Source: Developed by the authors

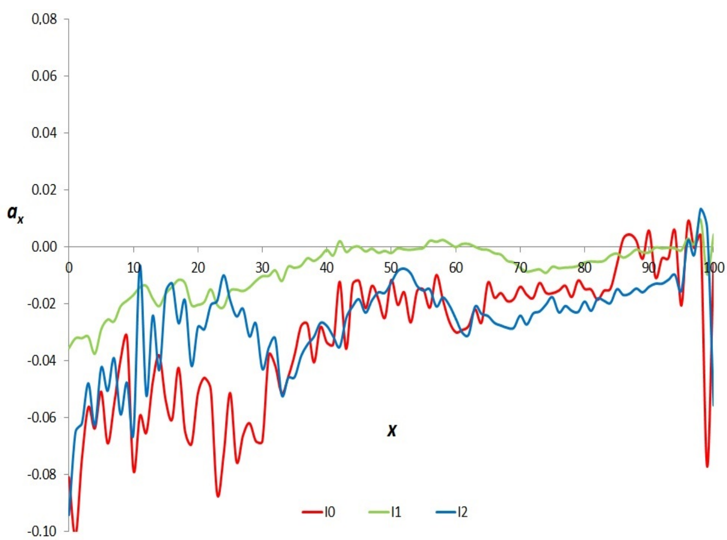

Figure 6.10. Estimates of α_x, $x = 0, 1, \ldots, 100$
obtained with the DGOBHM model for sub-periods I_0, I_1, I_2 (females)
Source: Developed by the authors

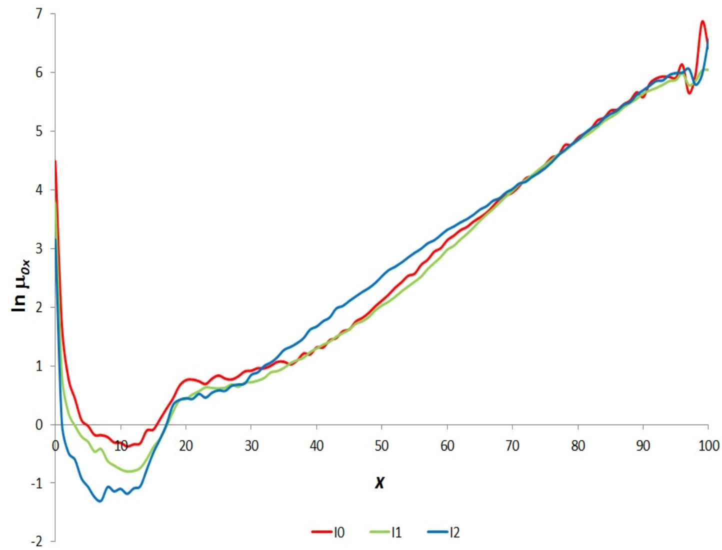

Figure 6.11. Estimates of $\ln \mu_{x0}$, $x = 0, 1, \ldots, 100$
obtained with the DGOBHM model for sub-periods I_0, I_1, I_2 (males)
Source: Developed by the authors

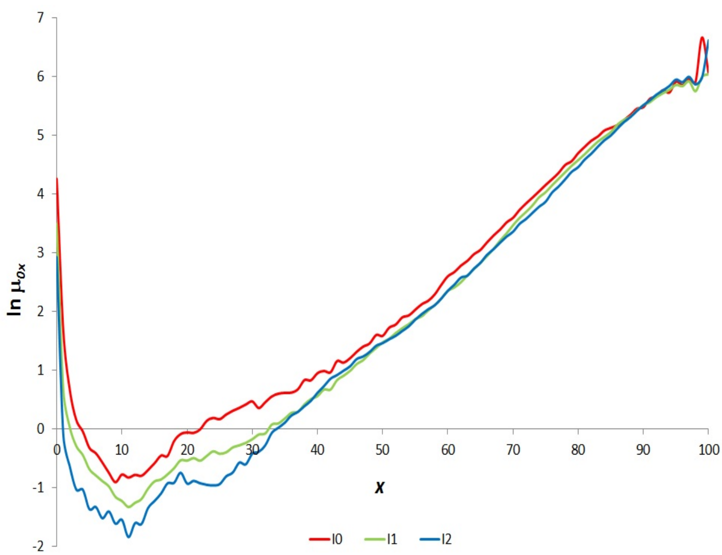

Figure 6.12. Estimates of $\ln \mu_{x0}$, $x = 0, 1, \ldots, 100$
obtained with the DGOBHM model for sub-periods I_0, I_1, I_2 (females)
Source: Developed by the authors

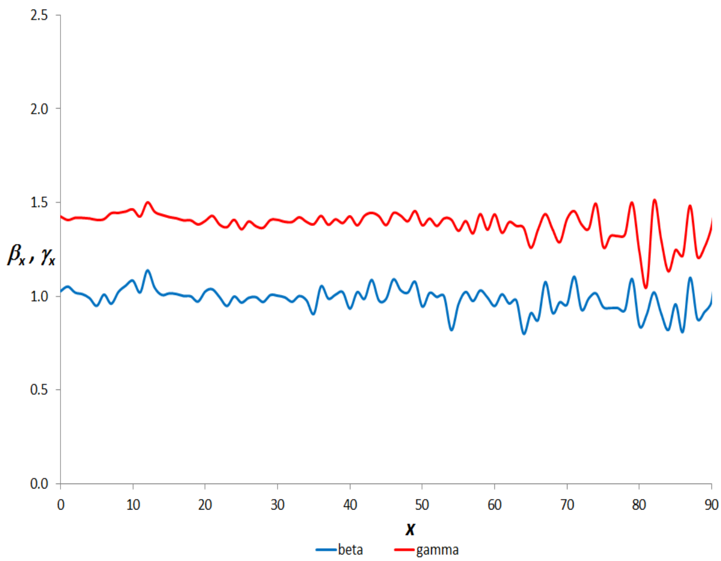

Figure 6.13. Estimates of β_x, γ_x, $x = 0, 1, \ldots, 90$
obtained with the DGOBHM model for sub-period I_0 (males)
Source: Developed by the authors

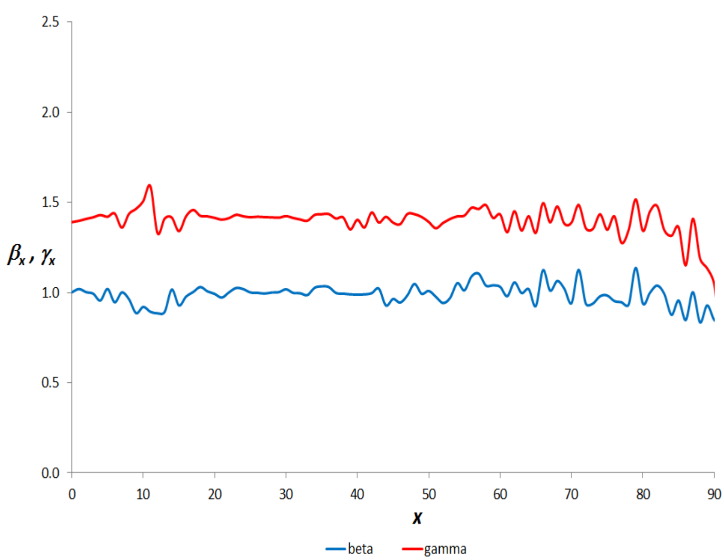

Figure 6.14. Estimates of β_x, γ_x, $x = 0, 1, \ldots, 90$
obtained with the DGOBHM model for sub-period I_0 (females)
Source: Developed by the authors

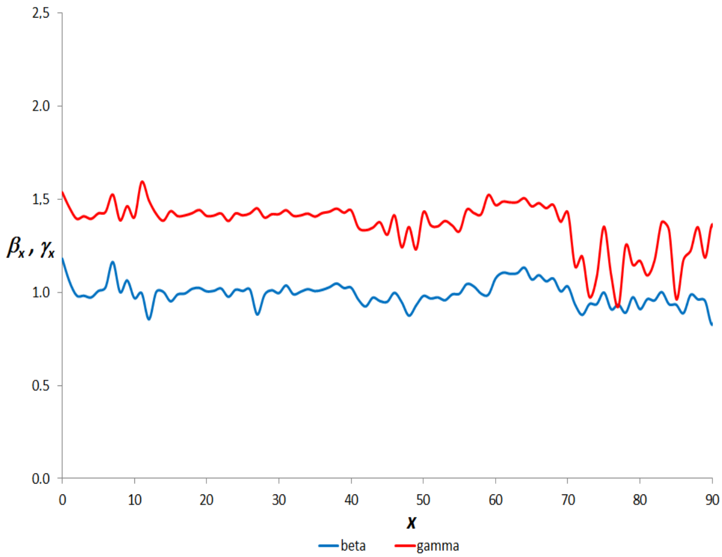

Figure 6.15. Estimates of β_x, γ_x, $x = 0, 1, \ldots, 90$
obtained with the DGOBHM model for sub-period I_1 (males)
Source: Developed by the authors

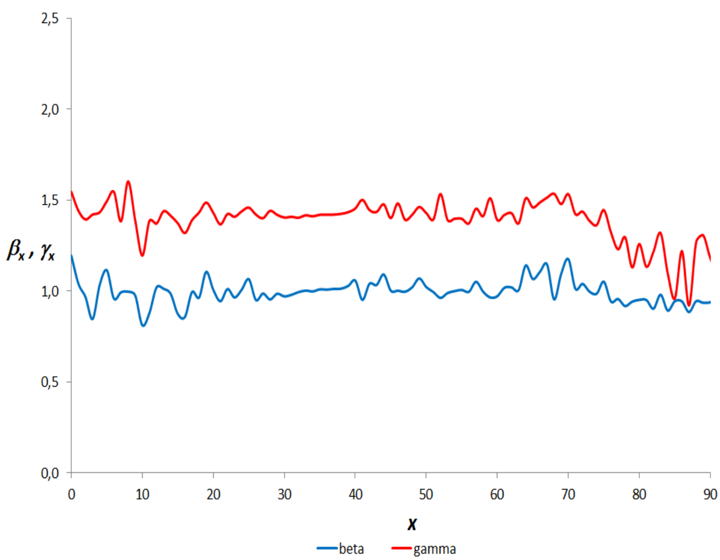

Figure 6.16. Estimates of β_x, γ_x, $x = 0, 1, \ldots, 90$
obtained with the DGOBHM model for sub-period I_1 (females)
Source: Developed by the authors

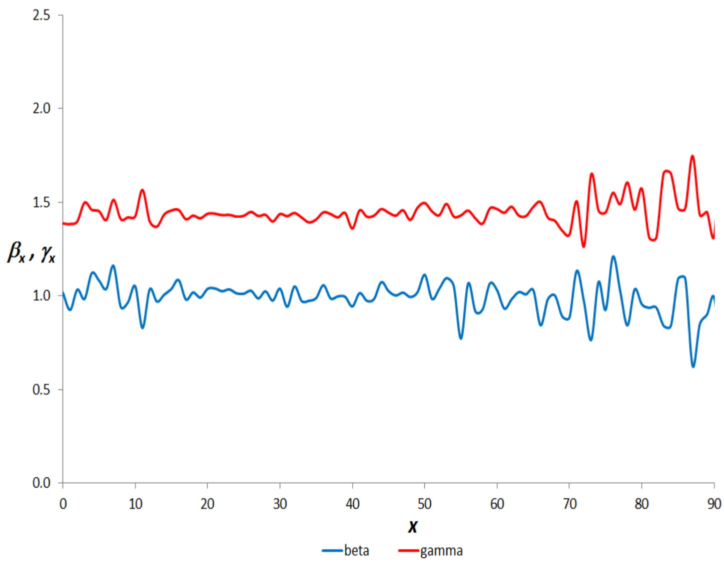

Figure 6.17. Estimates of β_x, γ_x, $x = 0, 1, \ldots, 90$
obtained with the DGOBHM model for sub-period I_2 (males)
Source: Developed by the authors

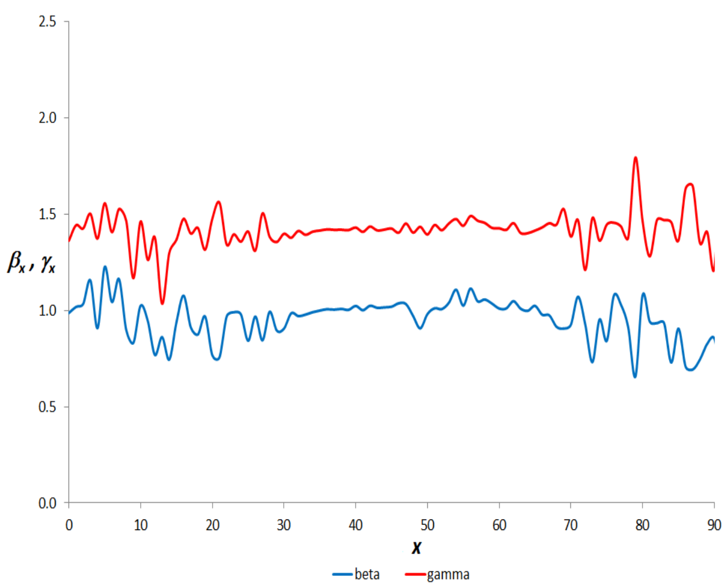

Figure 6.18. Estimates of β_x, γ_x, $x = 0, 1, \ldots, 90$
obtained with the DGOBHM model for sub-period I_2 (females)
Source: Developed by the authors

The models' parameters were estimated using the 1958–2000 data. Additionally, in the case of the DGOBHM model three sub-periods $I_0 = [1958, 1966)$, $I_1 = [1966, 1991)$, $I_2 = [1991, 2000]$ were considered and separate sets of estimates for each sub-period were obtained. Estimates for the last sub-period I_2 were next used to predict log-central mortality rates for each of the calendar years in the period 2001–2014.

To draw a comparison between prediction accuracy of the DGOBH moment model and some discrete non-hybrid models, i.e. the Giacometti–Ortobelli–Bertocchi model DGOB, the discrete modified Milevsky–Promislow model DMMP, as well as the discrete dynamic Lee–Carter model DDLC, *ex-post* forecasting errors were calculated for each of the models with respect to each year in the period 2001–2014 that was excluded from the estimation procedure.

To this end, the *MSE* and *MAD* measures for the respective hybrid and non-hybrid models were defined in a similar manner as for the DLCH and SLC models (see (6.3.1) or (6.3.2)).

From the data in Tables 6.4 and 6.5 it follows that the DGOBH moment model is more accurate in prediction than the two non-hybrid models DGOB, DMMP and seems to be comparable with the DDLC model in terms of *ex-post* forecasting errors *MSE* and *MAD*.

Table 6.4. *Ex-post* comparison of *MSE* values
for models: DGOBHM, DGOB, DMMP and DDLC

Year	Males				Females			
	DGOBHM	DGOB	DMMP	DDLC	DGOBHM	DGOB	DMMP	DDLC
2001	0.096	0.121	0.142	0.015	0.078	0.108	0.094	0.020
2002	0.090	0.138	0.119	0.100	0.099	0.139	0.103	0.109
2003	0.080	0.152	0.126	0.107	0.088	0.148	0.114	0.125
2004	0.089	0.178	0.137	0.102	0.099	0.152	0.122	0.110
2005	0.097	0.193	0.139	0.107	0.147	0.170	0.126	0.116
2006	0.115	0.194	0.135	0.111	0.147	0.173	0.132	0.154
2007	0.139	0.210	0.139	0.122	0.150	0.178	0.151	0.149
2008	0.141	0.231	0.154	0.129	0.153	0.180	0.155	0.155
2009	0.139	0.255	0.167	0.142	0.198	0.200	0.155	0.156
2010	0.155	0.306	0.206	0.147	0.175	0.227	0.197	0.183
2011	0.159	0.319	0.195	0.178	0.200	0.244	0.204	0.167
2012	0.179	0.350	0.214	0.197	0.195	0.254	0.216	0.193
2013	0.189	0.384	0.225	0.222	0.227	0.269	0.227	0.185
2014	0.201	0.451	0.289	0.247	0.239	0.288	0.243	0.215

Source: Own calculations.

Table 6.5. *Ex-post* comparison of *MAD* values
for models: DGOBHM, DGOB, DMMP and DDLC

Year	Males				Females			
	DGOBHM	DGOB	DMMP	DDLC	DGOBHM	DGOB	DMMP	DDLC
2001	0.076	0.105	0.013	0.010	0.056	0.091	0.085	0.016
2002	0.071	0.126	0.107	0.079	0.076	0.121	0.090	0.069
2003	0.057	0.140	0.140	0.088	0.068	0.131	0.099	0.087
2004	0.063	0.160	0.120	0.079	0.076	0.136	0.107	0.082
2005	0.070	0.174	0.126	0.087	0.108	0.149	0.106	0.090
2006	0.083	0.179	0.122	0.093	0.107	0.150	0.111	0.110
2007	0.098	0.193	0.124	0.103	0.106	0.153	0.127	0.110
2008	0.102	0.208	0.140	0.108	0.115	0.159	0.132	0.117
2009	0.093	0.229	0.153	0.121	0.137	0.172	0.125	0.122
2010	0.100	0.288	0.191	0.123	0.136	0.207	0.180	0.139
2011	0.109	0.277	0.181	0.157	0.156	0.218	0.170	0.140
2012	0.116	0.317	0.200	0.175	0.150	0.228	0.186	0.158
2013	0.121	0.344	0.213	0.198	0.171	0.239	0.191	0.155
2014	0.131	0.413	0.268	0.219	0.184	0.254	0.209	0.176

Source: Own calculations.

The amount of discrepancies between the theoretical and actual log-central death rates is illustrated in Figures 6.19 and 6.20.

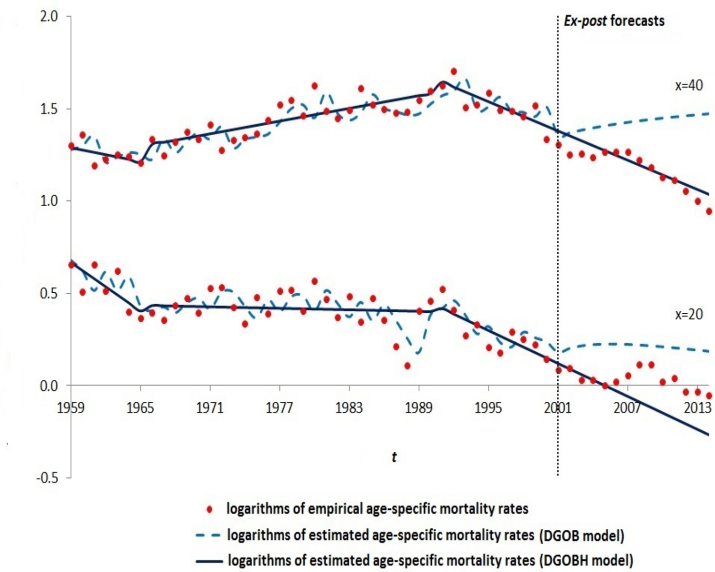

Figure 6.19. Empirical and theoretical log-central mortality rates
obtained with the DGOBH and DGOB models (males)
Source: Developed by the authors

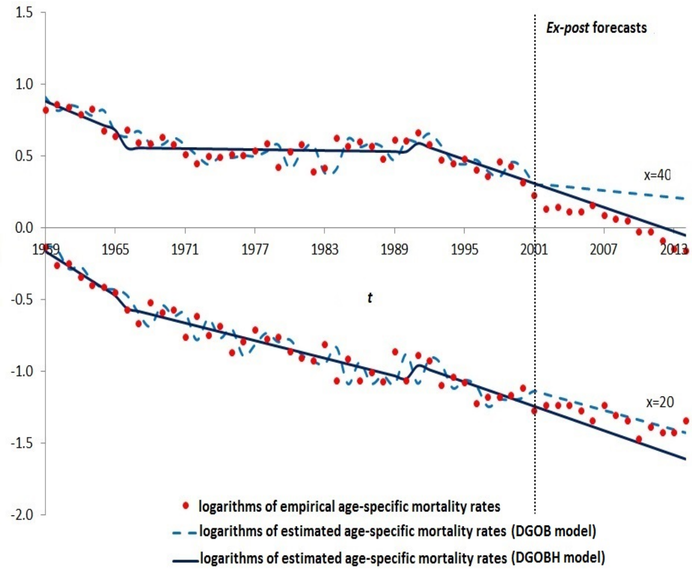

Figure 6.20. Empirical and theoretical log-central mortality rates obtained with the DGOBH and DGOB models (females)
Source: Developed by the authors

Both figures show the empirical log-central age-specific mortality rates for males and females aged $x = 20$ and $x = 40$ years in period 1958–2014 (dots) as well as the adjusted logarithms of mortality rates for the same time period and for the same age groups (dashed lines). Forecasts obtained from the DGOBHM and DGOB models for years 2001–2014 are marked by solid and dashed lines, respectively.

As can be seen, the differences between empirical values and values predicted by both models tend to grow as the forecast horizon increases. This means that rather than analyzing single trajectories, the confidence areas (or at least areas of fuzziness) should be determined for the forecasted characteristics. This functionality is available, for instance, with mortality models based on fuzzy numbers or complex functions. Some estimates obtained with such models are presented in the next sections.

6.3.3. The MFLC model

Let us consider a modified fuzzy Lee–Carter model MFLC, developed on the properties of modified fuzzy numbers (see Chapter 4). The age-specific mortality rates were fuzzified taking account of the model's switching points determining mortality regimes.

Figures 6.21–6.26 illustrate the estimates of parameters a_x, b_x, k_t, and the estimates of fuzziness parameters $s_{A_x}, s_{B_x}, s_{K_t}$ yielded by the MFLC model for men and women. Coefficients a_x, b_x, k_t have the same interpretation as in the SLC or DLCH models (see Section 1.5.3 or Section 6.3.1).

In the case under consideration, we also have estimates $s_{A_x}, s_{B_x}, s_{K_t}$, which allow the areas of fuzziness to be determined for a_x, b_x, k_t. In Figures 6.21–6.26 the areas are delimited by dashed lines.

The values of $s_{A_x} - s_{B_x} s_{K_t}$ can also be treated as the fuzziness of the forecasted logarithms of age-specific death rates generated by model, since modified fuzzy numbers $\check{A}_x$, $\check{B}_x$, $\check{K}_t$ correspond to the symmetric triangular numbers. Operations performed on these numbers according to (5.2.11) generate modified fuzzy numbers $\check{A}_x \oplus (\check{B}_x \odot \check{K}_t)$, which are equivalent to symmetric triangular fuzzy numbers resembling them in shape, with central values and spreads equal, respectively,

$$a_x + b_x k_t, \quad s_{A_x} - s_{B_x} s_{K_t}, \tag{6.3.3}$$

given $s_{A_x} - 2s_{B_x} s_{K_t} \geq 0$.

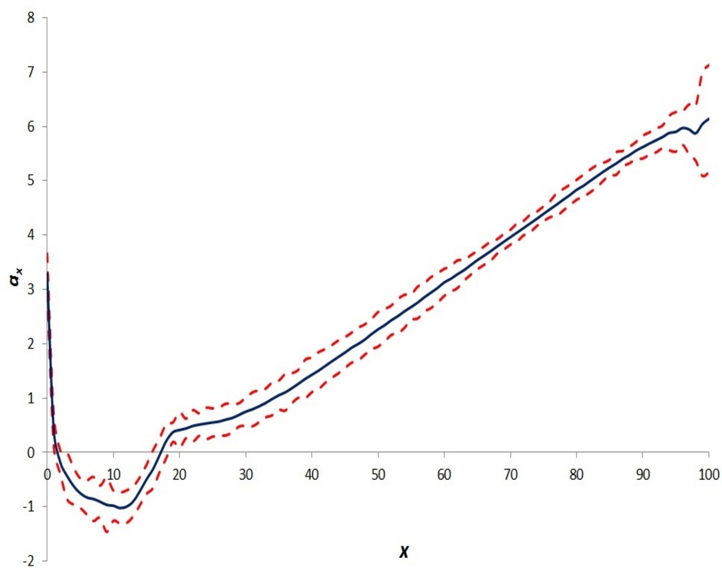

Figure 6.21. Estimates of a_x, $x = 0, 1, \ldots, 100$ and the area of fuzziness obtained with the MFLC model (males)

Source: Developed by the authors

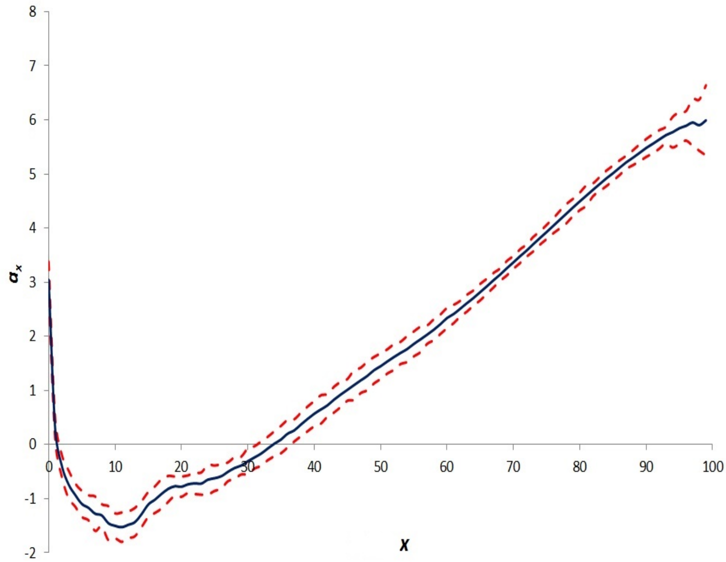

Figure 6.22. Estimates of a_x, $x = 0, 1, \ldots, 100$ and the area of fuzziness obtained with the MFLC model (females)
Source: Developed by the authors

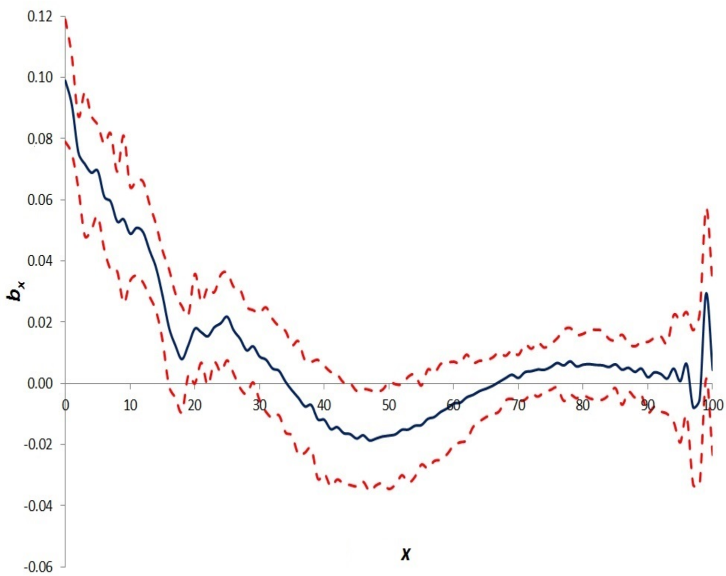

Figure 6.23. Estimates of b_x, $x = 0, 1, \ldots, 100$ and the area of fuzziness obtained with the MFLC model (males)
Source: Developed by the authors

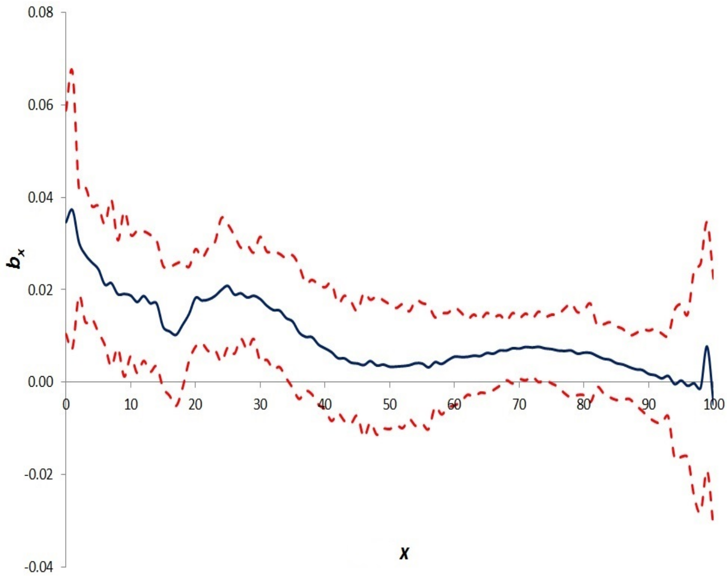

Figure 6.24. Estimates of b_x, $x = 0, 1, \ldots, 100$ and the area of fuzziness obtained with the MFLC model (females)
Source: Developed by the authors

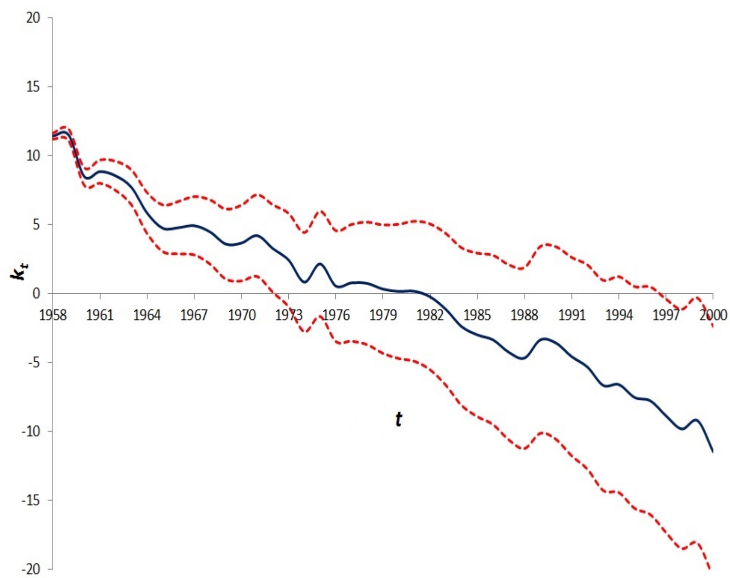

Figure 6.25. Estimates of k_t, $t = 1958, \ldots, 2000$ and the area of fuzziness obtained with the MFLC model (males)
Source: Developed by the authors

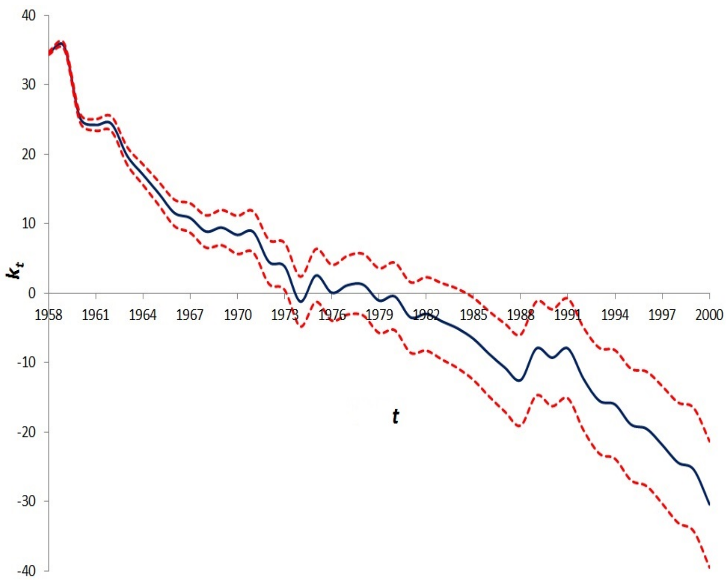

Figure 6.26. Estimates of k_t, $t = 1958, \ldots, 2000$ and the area of fuzziness obtained with the MFLC model (females)
Source: Developed by the authors

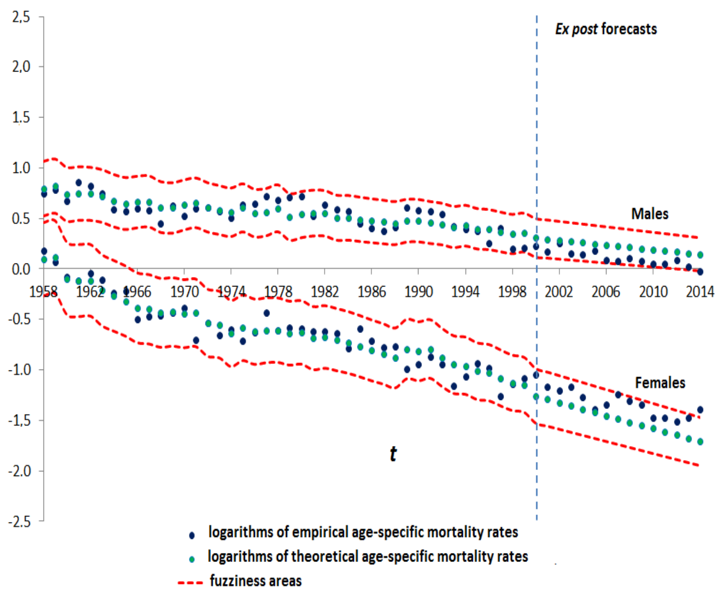

Figure 6.27. Empirical and predicted logarithms of mortality rates for the age group $x = 25$ years and the areas of fuzziness
Source: Developed by the authors

In Figure 6.27 the logarithms of empirical and predicted death rates (for males and females) are shown together with the areas of fuzziness for selected age group. In this case, as before, the estimation period is 1958–2000 and the period of *ex-post* forecast spans the years 2001–2014.

The values obtained with formula $a_x + b_x k_t$ are the central values of symmetric fuzzy numbers representing predicted fuzzy log-central mortality rates in the MFLC model. Thus, Figure 6.27 shows both the central values of predicted rates and their areas of fuzziness determined by curves arising from the following equations

$$\begin{aligned} f_{1x}(t) &= a_x + b_x k_t - (s_{A_x} - s_{B_x} s_{K_t}), \\ f_{2x}(t) &= a_x + b_x k_t + (s_{A_x} - s_{B_x} s_{K_t}). \end{aligned} \tag{6.3.4}$$

To determine k_t and s_{K_t} in (6.3.4) for the forecast period 2001–2014, a random walk model with a drift was adopted for both indicators (analogous to formula (1.5.8)).

Let us note that the areas of fuzziness in Figure 6.27 contain both the predicted values and most of empirical observations, also in the forecast period.

As values predicted by the MFLC model correspond to the triangular symmetric fuzzy numbers, their central values were used to measure errors with *MAD* and *MSE* defined similarly as in (6.3.1) or (6.3.2).

Table 6.6. *Ex-post* comparison of *MSE* values for the SLC and MFLC models

Year	Males		Females	
	SLC	MFLC	SLC	MFLC
2001	0.197	0.186	0.098	0.098
2002	0.204	0.194	0.122	0.121
2003	0.215	0.202	0.122	0.122
2004	0.223	0.209	0.132	0.132
2005	0.230	0.214	0.146	0.146
2006	0.232	0.220	0.152	0.151
2007	0.238	0.223	0.172	0.171
2008	0.257	0.240	0.174	0.173
2009	0.281	0.262	0.191	0.190
2010	0.330	0.308	0.190	0.190
2011	0.341	0.321	0.218	0.217
2012	0.373	0.351	0.215	0.215
2013	0.406	0.383	0.246	0.246
2014	0.469	0.442	0.273	0.272

Source: Own calculations.

Table 6.7. *Ex-post* comparison of *MAD* values
for the SLC and MFLC models

Year	Males		Females	
	SLC	MFLC	SLC	MFLC
2001	0.182	0.171	0.083	0.082
2002	0.185	0.175	0.107	0.107
2003	0.195	0.181	0.109	0.109
2004	0.206	0.191	0.117	0.116
2005	0.214	0.197	0.129	0.128
2006	0.214	0.203	0.130	0.129
2007	0.219	0.208	0.152	0.152
2008	0.234	0.219	0.156	0.156
2009	0.250	0.232	0.170	0.168
2010	0.302	0.281	0.167	0.166
2011	0.307	0.288	0.191	0.191
2012	0.335	0.318	0.185	0.185
2013	0.359	0.341	0.221	0.220
2014	0.430	0.410	0.245	0.245

Source: Own calculations.

It follows from the data in Tables 6.6 and 6.7 that the MFLC model generates smaller forecast errors than the SLC model. This model also allows assessing the uncertainty of the obtained estimates, i.e. the age-specific log-central mortality rates, since it can be used relatively easily to determine the areas of fuzziness.

6.3.4. The QVLC model

The estimation results for the quaternion model QVLC will be presented by plotting the estimates of a_x, b_x, k_t and $s_{A_x}, s_{B_x}, s_{K_t}$, as it was done in the previous section. The *ex-post* prediction errors obtained with *MSE* and *MAD* will be tabulated.

The algebra considered within the quaternion model was developed from the algebra of oriented fuzzy numbers by presenting oriented fuzzy numbers $\vec{A} = (f, g)$ as pairs of complex functions $\tilde{A}(u) = (f_A(u), g_A(u))$ for $u \in [0,1]$, $f_A(u) = a - is_A(1-u)$, $g_A(u) = a + is_A(1-u)$, and by adopting the definition of multiplication dedicated to quaternions (Definition B.6, Appendix B).

Since the OFN algebra satisfies the Gelfand–Mazur assumption, it is isometrically isomorphic with the algebra of complex numbers. Therefore, $s_{A_x}, s_{B_x}, s_{K_t}$ will be interpreted as the measures of fuzziness of respective model's parameters (see the MFLC model). They are used to mark the areas of fuzziness in Figures 6.28–6.33.

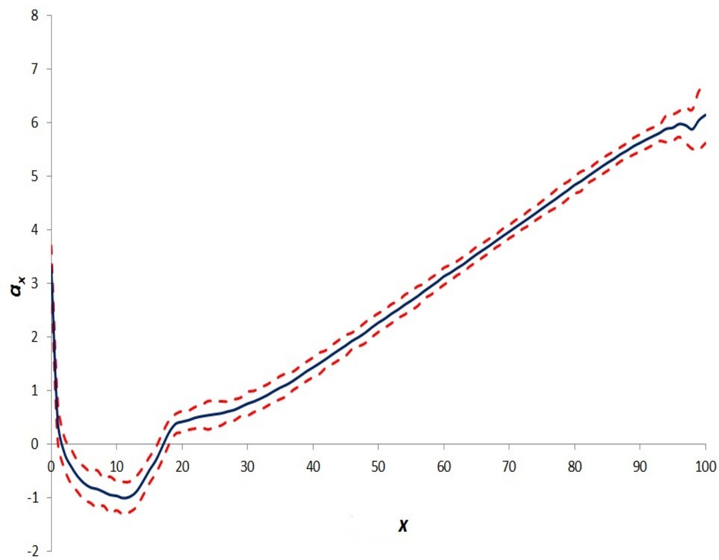

Figure 6.28. Estimates of a_x, $x = 0, 1, \ldots, 100$
and the area of fuzziness for the QVLC model (males)
Source: Developed by the authors

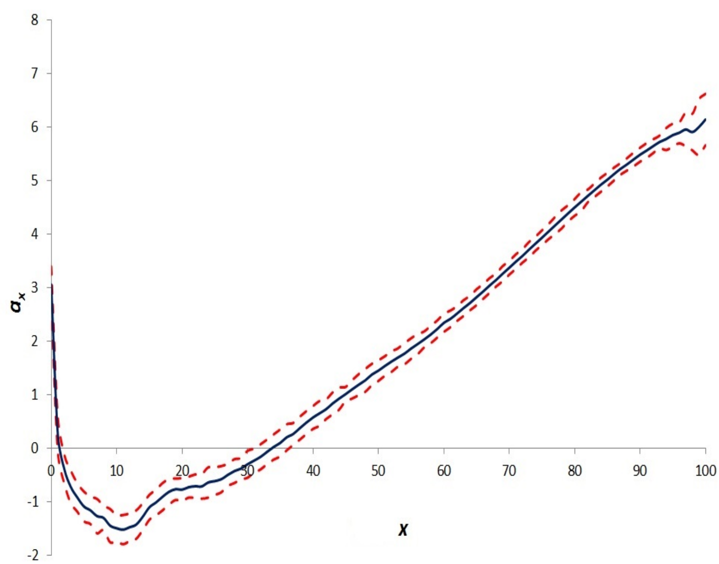

Figure 6.29. Estimates of a_x, $x = 0, 1, \ldots, 100$
and the area of fuzziness for the QVLC model (females)
Source: Developed by the authors

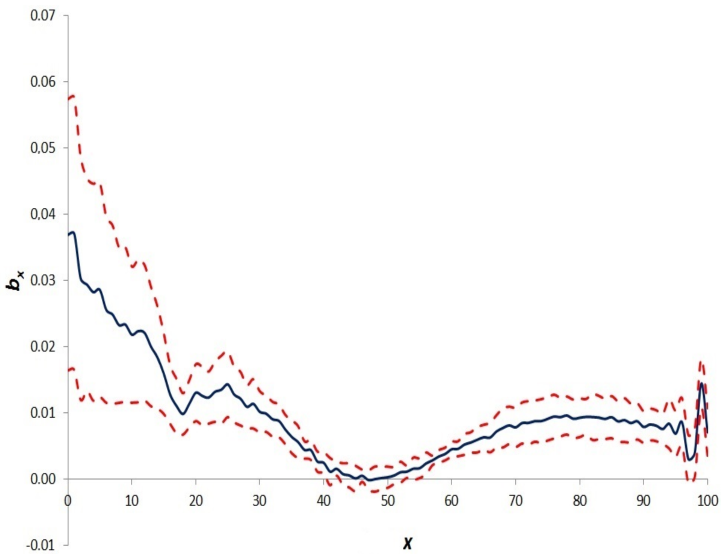

Figure 6.30. Estimates of b_x, $x = 0, 1, \ldots, 100$
and the area of fuzziness for the QVLC model (males)
Source: Developed by the authors

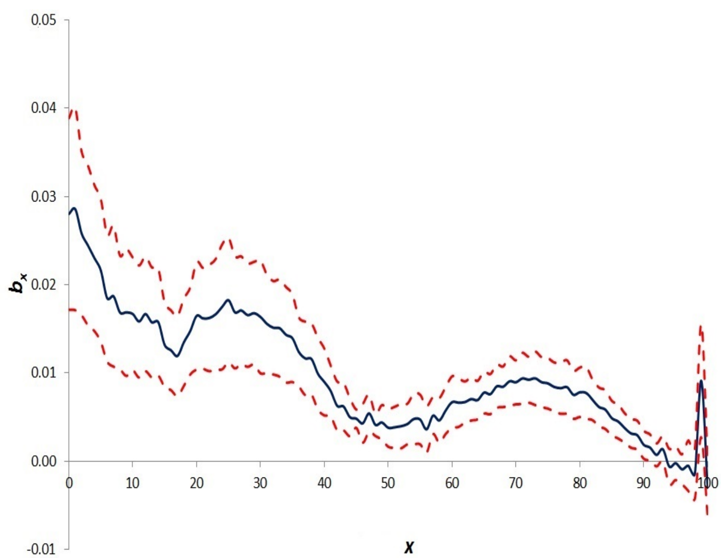

Figure 6.31. Estimates of b_x, $x = 0, 1, \ldots, 100$
and the area of fuzziness for the QVLC model (females)
Source: Developed by the authors

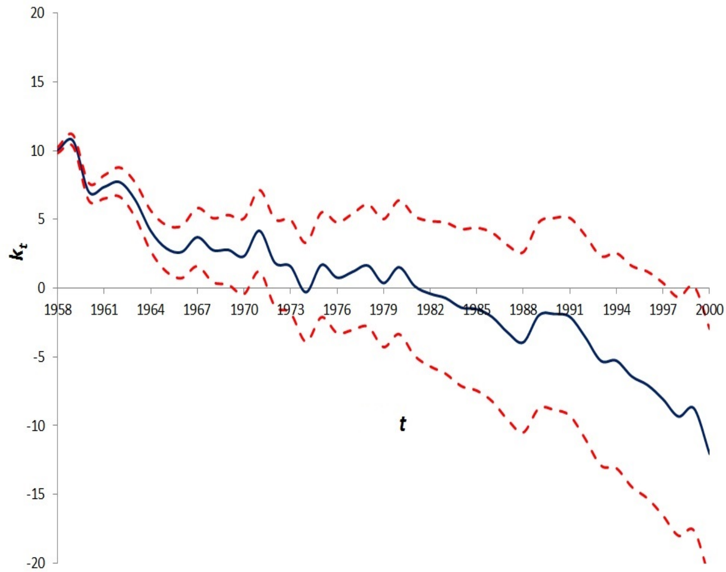

Figure 6.32. Estimates of k_t, $t = 1958, \ldots, 2000$
and the area of fuzziness for the QVLC model (males)
Source: Developed by the authors

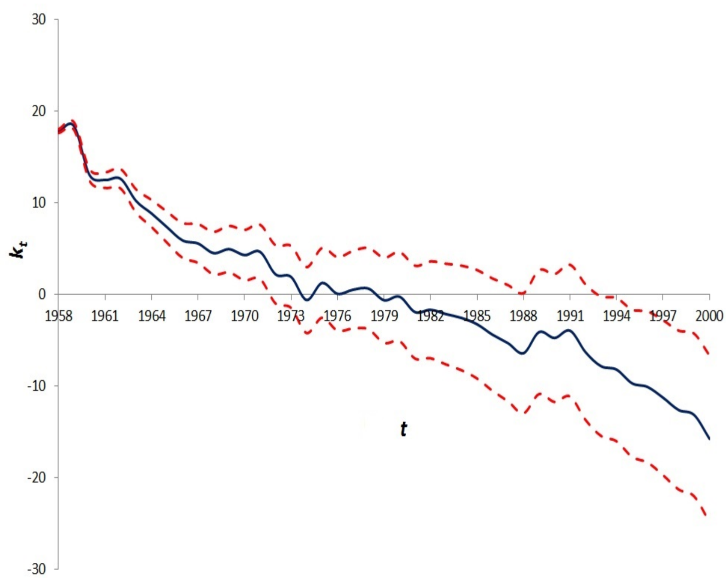

Figure 6.33. Estimates of k_t, $t = 1958, \ldots, 2000$
and the area of fuzziness for the QVLC model (females)
Source: Developed by the authors

The *MSE* and *MAD* values for the QVLC model, juxtaposed in Tables 6.8 and 6.9 with the *MSE* and *MAD* values for the SLC model, are defined by analogy to (6.3.2) or (6.3.1).

Table 6.8. *Ex-post* comparison of *MSE* values for the SLC and QVLC models

Year	Males		Females	
	SLC	QVLC	SLC	QVLC
2001	0.197	0.214	0.098	0.225
2002	0.204	0.207	0.122	0.240
2003	0.215	0.231	0.122	0.273
2004	0.223	0.250	0.132	0.289
2005	0.230	0.263	0.146	0.269
2006	0.232	0.247	0.152	0.282
2007	0.238	0.263	0.172	0.299
2008	0.257	0.273	0.174	0.314
2009	0.281	0.304	0.191	0.296
2010	0.330	0.355	0.190	0.376
2011	0.341	0.342	0.218	0.381
2012	0.373	0.368	0.215	0.415
2013	0.406	0.396	0.246	0.413
2014	0.469	0.476	0.273	0.424

Source: Own calculations.

Table 6.9. *Ex-post* comparison of *MAD* values for the SLC and QVLC models

Year	Males		Females	
	SLC	QVLC	SLC	QVLC
2001	0.182	0.159	0.083	0.199
2002	0.185	0.158	0.107	0.215
2003	0.195	0.173	0.109	0.238
2004	0.206	0.185	0.117	0.255
2005	0.214	0.192	0.129	0.247
2006	0.214	0.187	0.130	0.257
2007	0.219	0.191	0.152	0.264
2008	0.234	0.205	0.156	0.282
2009	0.250	0.228	0.170	0.269
2010	0.302	0.277	0.167	0.353
2011	0.307	0.283	0.191	0.351
2012	0.335	0.308	0.185	0.379
2013	0.359	0.333	0.221	0.377
2014	0.430	0.402	0.245	0.397

Source: Own calculations.

The data in Tables 6.8, 6.9 show that the prediction accuracy of the quaternion model is comparable with the SLC model for males and slightly worse then the SLC model for females. The quaternion model is capable of generating, like the MFLC model, areas of fuzziness of the model's parameters and consequently is useful for identifying the areas of fuzziness relating to the log-central mortality rates.

6.4. Final remarks

The recapitulation of the results of estimation and evaluation of the proposed mortality models is an opportunity to highlight their strong and weak points.

The advantage of the dynamic hybrid Lee–Carter model lies in its forecasting capabilities. The mortality forecasts it produces result in smaller or comparable prediction errors in relation to the standard Lee–Carter model. The next step of research shall be focused on confidence intervals for the predicted mortality rates.

In terms of forecasting properties, models utilizing fuzzy numbers and complex functions are similar to the standard Lee–Carter model. What makes them superior to it, however, is that they allow the areas of fuzziness of the estimated parameters to be determined, and consequently the areas of fuzziness for predicted mortality rates. Another advantage of the models is that the areas of fuzziness can be identified without employing any sophisticated methodology.

The above results encourage the authors to continue their work on developing the family of models combining the discovered capabilities of the hybrid systems and the fuzzy and complex models. The models will be analyzed more in detail in the authors' successive publications on mortality modeling.

Appendix A

Elements of the analysis of stochastic processes and stochastic equations

A.1. Basic definitions of stochastic processes

This appendix provides a review of the necessary information about stochastic processes. Information about the probability calculus has been omitted, because it is readily available in academic textbooks. The appendix has been prepared based on the material from the books [Lipcer, Sziriajew 1981, Sobczyk 1996, Socha 1993, Socha 2008].

The theory of stochastic processes is developed as the generalization of the concept of random variables. In the case of a random variable, each elementary event is assigned a number. For many real processes (physical, biological, economic, demographic etc.), this model is insufficient, as in most cases it is not a number that describes an elementary event, but rather a trajectory. This leads to the following definition of a stochastic process.

Definition A.1. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probabilistic space and $\mathbb{R}^+ = [0,\infty)$. The family $X = \{\xi(t,\omega)\}$, $t \in \mathbb{R}^+$, $\omega \in \Omega$ of random variables $\xi_t = \xi_t(\omega)$ is called a *(real) stochastic process with continuous time.* When time parameter t belongs to a set of natural numbers $\mathbb{N} = \{1, 2, ...\}$, family $X = \{\xi(t,\omega)\}$, $t \in \mathbb{N}$, $\omega \in \Omega$ is called *a random sequence* or *stochastic process with discrete time.* A stochastic process with complex outcomes is defined in the same manner.

For given $\omega \in \Omega$, a function of time $\xi(t, \cdot)$ will be called *a trajectory* or *a realization* corresponding with an elementary event ω. We shall use notation $\xi(t,\omega)$ for processes with continuous time and $\xi_t(\omega)$ for processes with discrete time, i.e. $\xi_t(\omega) = \xi(t,\omega)$, for $t \in \mathbb{N}$. Sometimes, for the sake of convenience, elementary event ω will be omitted from the notation of the processes, i.e. $\xi(t) = \xi(t,\omega)$ for $t \in \mathbb{R}^+$ or $\xi_t = \xi(t,\omega)$ for $t \in \mathbb{N}$, but this should not lead to misunderstandings. The stochastic processes will be denoted by small letters, e.g. x_1, x_2, y_1, y_2.

For given moments $t = t_1, t_2, ..., t_n$ the stochastic process $\xi(t)$ becomes a finite number of random variables $\xi(t_1), ..., \xi(t_n)$ characterized by joint probability distribution

$$F_{t_1,...,t_n}(x_1,...,x_n) \; = \; P\{\xi(t_1) < x_1, ..., \xi(t_n) < x_n\}, \qquad \text{(A.1.1)}$$

or, for continuous processes, by joint probability density

$$g(t_1, x_1, ..., t_n, x_n), \qquad \text{(A.1.2)}$$

or by characteristic function

$$\Phi(\Theta_1, t_1, ..., \Theta_n, t_n) = \mathrm{E}\left[\exp\left\{\sum_{j=1}^{n} i\Theta_j \xi(t_j)\right\}\right]. \qquad \text{(A.1.3)}$$

A natural generalization of a total characteristic function is the following characteristic functional

$$\Phi(\mathbf{\Theta}(t)) = \mathrm{E}\left[\exp\left\{i\int_{R^+} \mathbf{\Theta}(t)\xi(t)dt\right\}\right], \qquad \text{(A.1.4)}$$

where function $\mathbf{\Theta}(t)$ belongs to a class of functions for which integration on the right-hand side of (A.1.4) is well defined.

The transition from formula (A.1.4) to (A.1.3) is effected by substituting

$$\mathbf{\Theta}(t) = \sum_j \Theta_j \delta(t - t_j), \qquad \text{(A.1.5)}$$

where $\delta(t)$ is the Dirac distribution.

As in the case of other random variables, moments and cumulants for random processes are determined by differentiating the respective characteristic functional

$$\begin{aligned} \Phi(\mathbf{\Theta}(t)) = \Phi(\mathbf{\Theta}(t)) = 1 + i\sum_{j=1}^{n} \Theta_j(t)\mathrm{E}[x_j(t)] + \\ + \frac{i^2}{2!}\sum_{j=1}^{n}\sum_{k=1}^{n} \Theta_j(t)\Theta_k(t)\mathrm{E}[x_j(t)x_k(t)] + \ldots \end{aligned} \qquad \text{(A.1.6)}$$

For common density $g(t_1, x_1, ..., t_n, x_n)$, the higher order mixed moments are of the form

$$\mathrm{E}[x_1^{p_1}(t_1)\dots x_n^{p_n}(t_n)] = \\ = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} [x_1^{p_1}(t_1)\dots x_n^{p_n}(t_n)] g(x_1, t_1, ..., x_n, t_n) dx_1 ... dx_n. \tag{A.1.7}$$

Based on the various definitions of convergence of random variables, several definitions of the continuity of stochastic process can be formulated.

Definition A.2. Stochastic process $\xi(t)$, $t \in \mathbb{R}^+$ is called *continuous almost everywhere*, if

$$\mathrm{P}\{\omega : \lim_{t\to s} \xi(t,\omega) = \xi(s,\omega) = 0\} = 1. \tag{A.1.8}$$

Definition A.3. Stochastic process $\xi(t)$, $t \in \mathbb{R}^+$, is called *continuous in probability*, if

$$\forall \varepsilon > 0 \quad \lim_{t\to s} \mathrm{P}\{|\xi(t,\omega) - \xi(s,\omega)| > \varepsilon\} = 0. \tag{A.1.9}$$

A.1.1. Second-order processes

Because of its applications, of special importance is the class of second-order processes with complex values, i.e. with restricted second moments

$$\mathrm{E}[|x(t,\omega)|^2] < \infty, \quad t \in \mathbb{R}^+. \tag{A.1.10}$$

The values characterizing the second-order processes are the auto-correlation function and the auto-covariance function, which are respectively defined as

$$R_{xx}(t_1, t_2) = \mathrm{E}[x(t_1)\,\overline{x(t_2)}] \tag{A.1.11}$$

and

$$K_{xx}(t_1, t_2) = \mathrm{E}[(x(t_1) - \mathrm{E}[x(t_1)])\,\overline{(x(t_2) - \mathrm{E}[x(t_2)])}], \tag{A.1.12}$$

Expressions (A.1.11) and (A.1.12) are sometimes succinctly called the correlation and covariance functions and are written as $R_x(t_1, t_2)$ and $K_x(t_1, t_2)$ or $R(t_1, t_2)$ and $K(t_1, t_2)$; the upper dash represents a complex conjugate.

For $t_1 = t_2 = t$,

$$K_{xx}(t,t) = \mathrm{E}[(x(t) - \mathrm{E}[x(t)])^2] = \sigma_x^2(t), \tag{A.1.13}$$

where $\sigma_x(t)$ is a process standard deviation $x(t)$.

For two dissimilar processes $x(t)$, $y(t)$, *the cross-correlation functions* and *the cross-covariance functions* are respectively introduced

$$R_{xy}(t_1,t_2) = \mathrm{E}[x(t_1)\,\overline{y(t_2)}], \tag{A.1.14}$$

$$K_{xy}(t_1,t_2) = \mathrm{E}[(x(t_1) - \mathrm{E}[x(t_1)])\,\overline{(y(t_2) - \mathrm{E}[y(t_2)])}]. \tag{A.1.15}$$

For a vector process $x(t)$ with complex values, *the matrix correlation functions* and *the matrix covariance functions* are defined as follows

$$\mathbf{R_{xx}}(t_1,t_2) = \mathrm{E}[\mathbf{x}(t_1)\,\mathbf{x}^*(t_2)], \tag{A.1.16}$$

$$\mathbf{K_{xx}}(t_1,t_2) = \mathrm{E}[(\mathbf{x}(t_1) - \mathrm{E}[\mathbf{x}(t_1)])\,(\mathbf{x}(t_2) - \mathrm{E}[\mathbf{x}(t_2)])^*], \tag{A.1.17}$$

where the asterisk denotes feedback and transposition.

The matrix cross-correlation function and *the matrix cross-covariance function* are defined similarly

$$\mathbf{R_{xy}}(t_1,t_2) = \mathrm{E}[\mathbf{x}(t_1)\,\mathbf{y}^*(t_2)], \tag{A.1.18}$$

$$\mathbf{K_{xy}}(t_1,t_2) = \mathrm{E}[(\mathbf{x}(t_1) - \mathrm{E}[\mathbf{x}(t_1)])\,(\mathbf{y}(t_2) - \mathrm{E}[\mathbf{y}(t_2)])^*]. \tag{A.1.19}$$

For the second-order processes continuity is defined in the mean square sense.

Definition A.4. A stochastic second-order process $x(t)$, $t \in \mathbb{R}^+$ is called *continuous in the mean square sense* at point t, if

$$\begin{aligned} &\underset{\Delta t\to 0}{l.i.m}(x(s+\Delta t,\omega) - (x(s,\omega)) = \\ &\qquad = \lim_{\Delta t\to 0} \mathrm{E}[|x(s+\Delta t,\omega) - x(s,\omega)|^2] = 0, \end{aligned} \tag{A.1.20}$$

where $l.i.m$ denotes a limit in the mean square sense.

In examining continuity in the mean square sense, the following theorem can be of assistance.

Theorem A.1. A necessary and sufficient condition for continuity in the mean square sense of process $x(t)$ is the existence of an autocorrelation function $R_x(t_1,t_2)$ continuous over set $\{(t_1,t_2) :\ t_1 = t_2\}$.

For p-order processes, continuity is defined as below.

Definition A.5. A p-order stochastic process $x(t)$, $t \in \mathbb{R}^+$ is called *continuous at points in the sense of the p-th moment*, $0 < p < \infty$, if

$$\lim_{t \to s} \mathrm{E}[|x(t,\omega) - x(s,\omega)|^p] = 0. \tag{A.1.21}$$

In the special case, i.e. for $p = 2$, it is called *continuous in the mean square sense.*

A.1.2. Stationary processes

Stationary processes are a broad class of stochastic processes where probabilistic processes depend not on the present value of variable t, but on the difference $t - s$.

Definition A.6. Stochastic process $x(t)$, $t \in \mathbb{R}^+$ is called *weakly stationary* or *stationary in the broad sense*, if for any $\Delta \in \mathbb{R}$ and any $t, s \in \mathbb{R}^+$ the following relationships take place

$$\begin{aligned} &\mathrm{E}[|x(t)|^2] < \infty, \\ &\mathrm{E}[x(t)] = \mathrm{E}[x(t+\Delta,\omega)], \\ &\mathrm{E}[x(t+\Delta)\,\overline{x(s+\Delta)}] = \mathrm{E}[x(t)\,\overline{x(s)}], \end{aligned} \tag{A.1.22}$$

i.e. if the first and second moments do not change with the shift of variable t. For the sake of simplification, the word "weakly" will be mostly omitted, which should not lead to misunderstandings.

An immediate conclusion from this definition is that the mean value and variance are constant in time and that the correlation and covariance functions only depend on the difference $t_2 - t_1$, i.e.

$$\mathrm{E}[x(t)] = m_x = \mathrm{const}, \tag{A.1.23}$$

$$\mathrm{E}\left[(x(t) - \mathrm{E}[x(t)])^2\right] = \sigma_x^2 = \mathrm{const}, \tag{A.1.24}$$

$$R_x(t_1,t_2) = R_x(t_2 - t_1) = R_x(\tau), \tag{A.1.25}$$

$$K_x(t_1,t_2) = K_x(t_2 - t_1) = K_x(\tau), \tag{A.1.26}$$

where $t_1 = t$, $t_2 = t + \tau$.

A.1.3. Gaussian processes

Gaussian processes (normal) constitute a very important class of stochastic processes because of their wide applications. The literature provides several definitions of a Gaussian process. We shall present one of them.

Definition A.7. A vector stochastic process $\mathbf{x}(t)$, $\mathbf{x} \in \mathbb{R}^r$, $t \in \mathbb{R}^+$ is called *Gaussian (or normal)*, if for any natural $n \in \mathbb{N}$ and any subset $\{t_1, ..., t_n\}, t_i \in \mathbb{R}^+$, $n \geq 1$ vector random variables $\mathbf{x}(t_1), ..., \mathbf{x}(t_n)$ have joint Gaussian distribution, i.e. their characteristic function for any real vectors $\mathbf{\Theta}_1, ..., \mathbf{\Theta}_n$ is

$$\Phi(\mathbf{\Theta}_1, t_1, ..., \mathbf{\Theta}_n, t_n) = \mathrm{E}\left[\exp\left\{\sum_{j=1}^{n} i\, \mathbf{\Theta}_j^T \mathbf{x}(t_j)\right\}\right] = \tag{A.1.27}$$

$$= \exp\left\{\sum_{j=1}^{n} i\mathbf{\Theta}_j^T \mathbf{m}(t_j) - \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} i\, \mathbf{\Theta}_j^T \mathbf{K}(t_j, t_k)\mathbf{\Theta}_k\right\},$$

where $\mathbf{m}(t)$ and $\mathbf{K}(t_1, t_2)$ are, respectively, a vector of mean values (a mean) and a covariance matrix of a vector process $\mathbf{x}(t)$, $t \in \mathbb{R}^+$, $\mathbf{\Theta} = [\mathbf{\Theta}_1^T, ..., \mathbf{\Theta}_n^T]^T$, $\mathbf{m} = [\mathbf{m}_1^T, ..., \mathbf{m}_n^T]^T$, $\mathbf{K} = [\mathbf{K}(t_1, t_2)]$.

If covariance matrix $\mathbf{K}(t_i, t_j)$, $i, j = 1, ..., n$ is non-singular, the joint probability density of vector variables $\mathbf{x}(t_1)$,...,$\mathbf{x}(t_n)$ is of the form

$$g_G(\mathbf{x}_1, t_1, ..., \mathbf{x}_n, t_n) = $$
$$= [(2\pi)^{n^2}|\mathbf{K}|]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{u} - \mathbf{m}_\mathbf{x})^T \mathbf{K}^{-1}(\mathbf{u} - \mathbf{m}_\mathbf{x})\right\}, \tag{A.1.28}$$

where $\mathbf{u} = [\mathbf{x}_1^T, ..., \mathbf{x}_n^T]^T$, $\mathbf{m}_\mathbf{x} = [\mathbf{m}_1^T, ..., \mathbf{m}_n^T]^T$, and $|\mathbf{K}|$ is the determinant of an $n^2 \times n^2$ block covariance matrix written as $\mathbf{K} = [\mathbf{K}(t_i, t_j)]$, $i, j = 1, ..., n$.

In the special case, when the elements of the matrix are one dimensional, i.e. $[\mathbf{K}(t_i, t_j)] = K(t_i, t_j)$, the covariance matrix $\mathbf{K}$ is of the form

$$\mathbf{K} = \begin{bmatrix} K(t_1, t_1) & K(t_1, t_2) & \dots & K(t_1, t_n) \\ \vdots & \vdots & \ddots & \vdots \\ K(t_n, t_1) & K(t_n, t_2) & \dots & K(t_n, t_n) \end{bmatrix}. \tag{A.1.29}$$

A.1.4. Markov processes

Let us discuss now a broad class of stochastic processes in which "the future" does not depend on "the past" if "the present" is known. We will start with presenting a general definition of such a process.

Definition A.8. A vector stochastic r-dimensional process $\boldsymbol{\xi}(t)$, $t \in \mathbb{R}^+$, is called *a Markov process*, if for $n \in \mathbb{N}$ and any values of parameter $t_m \in \mathbb{R}^+$, $m = 1, ..., n$, where $t_0 < t_1 < ... < t_n$ and for any real vectors $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbb{R}^r$, the following relationship takes place

$$P\{\boldsymbol{\xi}(t_n) < \mathbf{x}_n | \, \boldsymbol{\xi}(t_{n-1}) = \mathbf{x}_{n-1}, ..., \boldsymbol{\xi}(t_1) = \mathbf{x}_1\} = \\ = P\{\boldsymbol{\xi}(t_n) < \mathbf{x}_n | \, \boldsymbol{\xi}(t_{n-1}) = \mathbf{x}_{n-1}\}, \tag{A.1.30}$$

i.e. the conditional distribution of $\boldsymbol{\xi}(t_n)$, for the given values of $\boldsymbol{\xi}(t_0)$, $\boldsymbol{\xi}(t_1), ..., \boldsymbol{\xi}(t_{n-1})$, only depends on the process value at the previous moment, meaning that it does not depend on all values that process $\boldsymbol{\xi}(t)$ took until t_{n-1}, i.e. it only depends on $\boldsymbol{\xi}(t_{n-1})$.

Let us introduce notations

$$\mathcal{P}(s, \mathbf{x}; t, \mathbf{B}) = \mathrm{P}\{\boldsymbol{\xi}(t) \in \mathbf{B} | \, \boldsymbol{\xi}(s) = \mathbf{x}\}, \quad s \le t, \tag{A.1.31}$$

$$F(s, \mathbf{x}; t, \mathbf{y}) = \mathrm{P}\{\boldsymbol{\xi}(t) < \mathbf{y} | \, \boldsymbol{\xi}(s) = \mathbf{x}\}, \tag{A.1.32}$$

where $\mathbf{B} \in \mathcal{B}^r$, $\mathcal{B}^r$ is the σ-algebra of Borel sets in $\mathbb{R}^r$.

Functions $\mathcal{P}(s, \mathbf{x}; t, \mathbf{B})$ and $F(s, \mathbf{x}; t, \mathbf{y})$ are called *a transition probability function*, or briefly *a transition function* related to Markov process $\boldsymbol{\xi}(t)$.

Markov processes are analyzed using homogeneity properties.

Definition A.9. Markov process $\boldsymbol{\xi}(t), t \in \mathbb{R}^+$ is called *homogeneous* (in terms of time), if for any $s, t \in \mathbb{R}^+$, $s < t$, the transition function only depends on the difference between time arguments $t - s = \tau$, i.e.

$$\mathcal{P}(s, \mathbf{x}; t, \mathbf{B}) = \mathcal{P}(\mathbf{x}, \tau, \mathbf{B}), \tag{A.1.33}$$

$$F(s, \mathbf{x}; t, \mathbf{y}) = F(\mathbf{x}, \tau, \mathbf{y}). \tag{A.1.34}$$

An important class of the Markov processes is continuous-time processes and continuous state space processes.

Definition A.10. A vector Markov process $\boldsymbol{\xi}(t), t \in \mathbb{R}^+$ with values in $\mathbb{R}^r$ is *an r-dimensional diffusion process* if its transition function $F(s, \mathbf{x}; t, \mathbf{y})$ for each $t \in \mathbb{R}^+$ and each $\varepsilon > 0$ satisfies the following conditions
(i)

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|\mathbf{y}-\mathbf{x}| \geq \varepsilon} d_{\mathbf{y}} F(t, \mathbf{x}; t+\Delta t, \mathbf{y}) = 0, \tag{A.1.35}$$

(ii) there is some vector function $\mathbf{A}(\mathbf{x}, t)$, for which

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|\mathbf{y}-\mathbf{x}| < \varepsilon} (\mathbf{y}-\mathbf{x})\, d_{\mathbf{y}} F(t, \mathbf{x}; t+\Delta t, \mathbf{y}) = \mathbf{A}(\mathbf{x}, t), \tag{A.1.36}$$

(iii) here is some vector function $\boldsymbol{\sigma}(\mathbf{x}, t)$, for which

$$\begin{aligned} \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|\mathbf{y}-\mathbf{x}| < \varepsilon} (\mathbf{y}-\mathbf{x})(\mathbf{y}-\mathbf{x})^T d_{\mathbf{y}} F(t, \mathbf{x}; t+\Delta t, \mathbf{y}) = \\ = \boldsymbol{\sigma}(\mathbf{x}, t) \boldsymbol{\sigma}^T(\mathbf{x}, t) = \mathbf{B}(\mathbf{x}, t) > 0, \end{aligned} \tag{A.1.37}$$

where $|\cdot|$ is the Euclidean norm in $\mathbb{R}^r$ and convergence under conditions given by (A.1.36), (A.1.37) is monotonous because of $\mathbf{x}$, and $d_{\mathbf{y}} F(t, \mathbf{x}; t+\Delta t, \mathbf{y})$ is a differential of function F with respect to $\mathbf{y}$.

Functions $\mathbf{A}(\mathbf{x}, t)$ and $\mathbf{B}(\mathbf{x}, t)$ are called *a drift vector* and *a diffusion matrix*, respectively.

When $\mathbf{A}(\mathbf{x}, t)$ and $\mathbf{B}(\mathbf{x}, t)$ are known, a transition function density can be determined for a diffusion process.

A.1.5. Processes with independent increments

A particularly importance case of Markov processes are processes with independent increments.

Definition A.11. A stochastic process $\xi(t), t \in \mathbb{R}^+$, is called *a process with independent increments*, if for any $t_i \in \mathbb{R}^+$, such as $t_0 < t_1 < \ldots < t_n$, random variables representing the increments of process $\xi(t)$, i.e. $\xi(t_0)$, $\xi(t_1) - \xi(t_0), \ldots, \xi(t_n) - \xi(t_{n-1})$ are independent.

Definition A.12. A stochastic process $\xi(t), t \geq 0, \xi(0) = 0$ defined on the probabilistic space $(\Omega, \mathcal{F}, \mathrm{P})$ is called *a process with increments* *independent of the past*, if for any $t,s\in\mathbb{R}^+$, $0\le s\le t<\infty$, random variables representing the increments of process $\xi(t)$, i.e. $\xi(t)-\xi(s)$ are independent of $\mathcal{F}$.

Definition A.13. A stochastic process $\xi(t)$, $t\in\mathbb{R}^+$ with independent increments is called *a process with stationary independent increments*, if its increments $\xi(t_1)-\xi(t_0)$,..., $\xi(t_n)-\xi(t_{n-1})$ only depend on differences $t_1-t_0,...,t_n-t_{n-1}$, respectively.

Among processes with independent increments the Wiener, Poisson and Lévy processes are particularly important. Let us discuss the Wiener process more in detail.

Definition A.14. A stochastic process $\xi(t,\omega)$, $t\in\mathbb{R}^+$ defined on probabilistic space $(\Omega,\mathcal{F},\mathrm{P})$ is called *a Wiener process* or *a Brownian motion*, if:

(i) $\mathrm{P}\{\xi(0,\omega)=0\}=1$,

(ii) $\xi(t,\omega)$ is a process with stationary increments independent of the past,

(iii) increments $\xi(t,\omega)-\xi(s,\omega)$ have Gaussian distribution for which

$$\mathrm{E}[\xi(t,\omega)-\xi(s,\omega)]=0, \tag{A.1.38}$$

$$\mathrm{E}[(\xi(t,\omega)-\xi(s,\omega))^2]=\sigma^2|t-s|, \quad \sigma^2=\mathrm{const}>0, \tag{A.1.39}$$

(iv) for almost all $\omega\in\Omega$ realizations $\xi(t,\omega)$ are continuous with respect to $t\in\mathbb{R}^+$.

Some authors used properties (i)–(iii) to define a Wiener process, arguing that a Wiener process $\xi(t,\omega)$ defined in this way has a modification the realizations of which are continuous almost everywhere.

For $\sigma^2=1$ process $\xi(t,\omega)$ is called *a standard Wiener process.* The existence of such a process results from the proposition presented by Lipcer and Shiryaev [Lipcer, Sziriajew 1981].

Let $\eta_1,\eta_2,...$ be a sequence of Gaussian random variables with the mean values equal zero and unit variances and let $\phi_1(t),\phi_2(t),...$, $t\in[0,T]$ be any complete and orthogonal sequence in $L^2[0,T]$. Then the following theorem takes place.

Theorem A.2. For each $t\in[0,T]$, the following series

$$\xi(t,\omega)=\sum_{j=1}^{\infty}\eta_j(\omega)\int_0^t\phi_j(s)ds, \tag{A.1.40}$$

is convergent almost everywhere and defines a Wiener process over interval $[0, T]$.

From the definition of a standard Wiener process the following properties of the process can be derived

$$\mathrm{E}[\xi(t)] = 0, \tag{A.1.41}$$

$$K(s,t) = \mathrm{E}[\xi(s)\xi(t)] = \min(s,t), \tag{A.1.42}$$

$$P\{\xi(t) \leq x\} = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{x} \exp\left\{-\frac{y^2}{2t}\right\} dy, \tag{A.1.43}$$

$$\mathrm{E}[|\xi(t)|] = \sqrt{\frac{2t}{\pi}}, \tag{A.1.44}$$

$$\begin{aligned} \mathrm{E}[(\xi(t+\Delta t)-\xi(t))^{2p}] &= \frac{1}{\sqrt{2\pi\Delta t}} \int_{-\infty}^{+\infty} z^{2p} \exp\left\{-\frac{z^2}{2\Delta t}\right\} dz = \\ &= (2n-1)!!\,(\Delta t)^p. \end{aligned} \tag{A.1.45}$$

It follows that almost all realizations of the Wiener process are continuous. Moreover, the Wiener process can be demonstrated to have an important property given by the following theorem.

Theorem A.3. Even though almost all realizations of the Wiener process are continuous, they are not differentiable for all $t \geq 0$ and on each finite interval they have infinite oscillations.

Definition A.15. A stochastic process $\boldsymbol{\xi}(t)$ is called *an r-dimensional Wiener process* $\boldsymbol{\xi}(t) = [\xi_1(t), \ldots, \xi_r(t)]^T$, if each of its components $\xi_i(t)$, $i = 1, \ldots, r$ is a scalar Wiener process and, additionally, all $\xi_i(t)$ are mutually independent processes.

A.1.6. White noise

The fundamental mathematical tool for analyzing stochastic dynamic systems (comparable with imaginary unit $i = \sqrt{-1}$ in electrical engineering) is an abstract stochastic process called white noise.

The literature, particularly the technical literature, offers several definitions of white noise. The definition given below has been derived from [Sobczyk 1996].

Let $D(T)$ be a space of test functions, i.e. all infinitely differentiable functions $\phi : T \to R^1$ vanishing identically outside a finite closed interval. The topology for this space is the same as in the regular Schwartz distribution spaces, meaning that $D(T)$ is a topological vector space. Let H be a Hilbert space of all P-equivalent random variables defined on $(\Omega, \mathcal{F}, \mathrm{P})$ with a finite second moment.

Definition A.16. Continuous linear projection $\Phi : D(T) \to H$ is called *a generalized stochastic process on set* T. The value of generalized stochastic process Φ in ϕ is denoted as $\{\phi,\ \Phi\}$ or $\Phi(\phi)$.

The advantage of a generalized stochastic process is that it always has a derivative, which is also a generalized stochastic process. The definition is the following.

Definition A.17. Derivative $\dot{\Phi}$ of generalized process Φ with respect to t (a generalized derivative in space $D(T)$) is given by the following relationship

$$\{\phi, \dot{\Phi}\} = \{\frac{d\phi}{dt},\ \Phi\} \text{ for all } \phi \in D(T). \tag{A.1.46}$$

By applying the definition of a generalized derivative of the Wiener processes, the compound Poisson process and α-stable Lévy process, new generalized stochastic processes can be obtained.

Gaussian white noise

Definition A.18. A generalized derivative of the Wiener process $w(t)$, $t \in [0,\infty)$ denoted as $\eta_w(t) = \dot{w}(t) = \frac{dw(t)}{dt}$, i.e.

$$\{\phi,\ \eta\} = \{\phi,\ \dot{\xi}\} = -\{\frac{d\phi}{dt},\ \xi\} \quad \text{for all } \phi \in D(T), \tag{A.1.47}$$

is called *Gaussian white noise.*

Equality (A.1.47) can be also written as

$$d\xi(t) = \eta(t)dt. \tag{A.1.48}$$

Without going in details, it is possible to demonstrate that for each $\phi \in D(T)$ integral of $\{\phi, \eta\}$ is a Gaussian random variable and that for each finite number of functions $\phi_1, ..., \phi_n \in D(T)$ random variables $\{\phi_i, \eta\}$, $1 \leq i \leq n$ have a joint Gaussian distribution. Moreover, the expected value of $\eta(t)$ is

$$\mathrm{E}[\eta(t)] = 0 \tag{A.1.49}$$

and the covariance function is given by the Dirac distribution

$$K_{\eta\eta}(t_1, t_2) = c\delta(t_2 - t_1) = c\delta(\tau), \quad c = \text{const}, \quad c > 0, \tag{A.1.50}$$

From the last equality it directly follows that the variance of Gaussian white noise is infinite $K_{\eta\eta}(t,t) = \delta(0) = \infty$ and the power spectral density function of a process is a constant function equal to c, i.e.

$$S_{\xi\xi}(\lambda) = c. \tag{A.1.51}$$

Property (A.1.50) confirms the "physical infeasibility" of such a process, whereas property (A.1.51) explains the origin of the term "white noise", which has been coined by analogy to "white light" made up of electromagnetic waves (colours) of all frequencies.

A.2. Differential and integral calculus of stochastic processes

A.2.1. Integrating and differentiating in the mean square sense

While discussing the second-order processes we presented a definition of the continuity of a stochastic process in the mean square sense. Differentiability and integrability in the mean square sense are defined likewise.

Definition A.19 (differentiating in the mean square sense). *A derivative in the mean square sense* of stochastic process $x(t)$ is defined by

$$\frac{d}{dt}x(t) = \dot{x}(t) = \underset{\Delta t \to 0}{l.i.m.} \frac{x(t + \Delta t) - x(t)}{\Delta t}. \tag{A.2.1}$$

Theorem A.4. A necessary and sufficient condition of differentiability (the existence of a derivative) in the mean square sense of process $x(t)$ is the existence a second derivative of autocorrelation function $\frac{\partial^2 R_x(t_1,t_2)}{\partial t_1 \partial t_2}$ limited and continuous on set $\{(t_1, t_2): \ t_1 = t_2\}$.

Definition A.20 (integrating in the mean square sense). Let $f(t)$ be a complex function on set $[a, b]$ and let $\{T_n\}$ be a sequence of divisions of interval $[a, b]$, i.e.

$$T_n = \{a = t_0^{(n)} < t_1^{(n)} < ... < t_n^{(n)} = b\},$$
$$\lim_{n \to +\infty} \max_{1 \le i \le n} (t_i^{(n)} - t_{i-1}^{(n)}) = 0. \tag{A.2.2}$$

We shall define *integral in the mean square sense* on interval $[a, b]$ as a limit of the Riemann sums

$$\int_a^b f(t)x(t)dt = \underset{n \to +\infty}{l.i.m} \sum_{i=0}^{n-1} f(t'_{in})x(t'_{in})(t_{i+1}^{(n)} - t_i^{(n)}), \tag{A.2.3}$$

where t'_{in} is any sequence satisfying the following inequalities

$$t_i^{(n)} \le t'_{in} \le t_{i+1}^{(n)}. \tag{A.2.4}$$

Theorem A.5. An integral in the mean square sense $\int_a^b f(t)x(t)dt$ exists if and only if there is an ordinary finite double Riemann integral

$$\int_a^b \int_a^b f(t_1)\overline{f(t_2)}R(t_1, t_2)dt_1 dt_2. \tag{A.2.5}$$

The averaging procedure is commutative with respect to differentiation and integration in the mean square sense, i.e.

$$\frac{d\mathrm{E}[x(t)]}{dt} = \mathrm{E}\left[\frac{dx(t)}{dt}\right], \tag{A.2.6}$$

$$\mathrm{E}\left[\int_a^b f(t)x(t)dt\right] = \int_a^b f(t)\mathrm{E}[x(t)]dt. \tag{A.2.7}$$

The above property holds true for any linear operation L, i.e. if L_t is a linear operator transforming second-order process $x(t)$ into second-order process $y(t)$, i.e.

$$y(t) = L_t[x(t)], \tag{A.2.8}$$

then

$$\mathrm{E}[y(t)] = L_t[\mathrm{E}[x(t)]]. \tag{A.2.9}$$

Moreover, the autocorrelation function of process $y(t)$ is given by the following relationship

$$R_{yy}(t_1, t_2) = L_{t_1} L_{t_2} R_{xx}(t_1, t_2). \tag{A.2.10}$$

In the special case of $L_t = L$

$$R_{yy}(t_1, t_2) = L^2 R_{xx}(t_1, t_2). \tag{A.2.11}$$

A.2.2. Stochastic integrals with respect to diffusion processes

In the analysis of stochastic processes, both the construction of integrals with respect to stochastic processes and the differentiation rules are different from those defined for the deterministic functions. Let us refer here only to the basic definitions and theorems, starting with the historically earliest stochastic integrals with respect to diffusion processes called the Itô and Stratonovich integrals.

Definition A.21. *The Itô stochastic integral* of non-anticipating function ($\mathcal{F}_t$-measurable) $f(x(t), t)$ on interval $[0, T]$ with respect to some diffusion process $x(t)$ is a mean-square limit of the Riemann sum, i.e.

$$\int_0^T f((x(t), t)dx(t) = \underset{\Delta t \to 0}{l.i.m} \sum_{i=0}^{N} f(x(t_i), t_i)[x(t_{i+1}) - x(t_i)], \tag{A.2.12}$$

where

$$\mathrm{E}\left[\int_0^T f^2((x(t), t)dt\right] < \infty, 0 = t_1 < \ldots < t_{N+1} = T, \Delta t = \max_i (t_{i+1} - t_i)$$

and the limit does not depend on the way how the points t_i are selected.

Unlike a mean-square stochastic integral given by (A.2.3), the value of the stochastic integral of non-anticipating function $f(x(t), t)$ on interval $[0, T]$ with respect to diffusion process $x(t)$ depends on the selection of intermediate points t'_i, for which the values of functions $f(x(t), t)$ in the Riemann sums are determined, meaning that interval $[t_i,\ t_{i+1}]$ is treated as some convex set and that any value t'_i from this interval can be presented as a combination of convex points t_i and t_{i+1}, i.e. as $t'_i = \beta t_i + (1 - \beta) t_{i+1}$, and the value of function $f(x(t), t)$ in this interval as $f(\beta x(t_i) + (1-\beta)x(t_{i+1}),\ \beta t_i + (1-\beta)t_{i+1})$, where β is a real parameter $0 \le \beta \le 1$. Then, the formula defining the stochastic integral of non-anticipating function $f(x(t),t)$ ($\mathcal{F}_t$-measurable) on interval $[0,T]$ with respect to diffusion process $x(t)$ is defined as the Itô stochastic integral, in which equality (A.2.12) is replaced by

$$\int_0^T f(x(t),t)d_\beta x(t) = \\ \underset{\Delta t \to 0}{l.i.m} \sum_{i=0}^{N} f(\beta x(t_i) + (1-\beta)x(t_{i+1}), \beta t_i + (1-\beta)t_{i+1})[x(t_{i+1}) - x(t_i)]. \tag{A.2.13}$$

This general definition of a stochastic integral takes account of two special cases $\beta = 1$ and $\beta = \frac{1}{2}$, which are called, respectively, the Itô stochastic integral and the Stratonovich stochastic integral. The mutual relationship between these integrals is described by the following theorem.

Theorem A.6. If $x(t), t \in [0,T]$ is a diffusion process $f(x(t),t)$ is a non-linear non-anticipating function on interval $[0,T]$ that has continuous derivatives because of both arguments and

$$\mathrm{E}\left[\int_0^T f^2(x(t),t)dt\right] < \infty, \tag{A.2.14}$$

then the equality occurs

$$\int_0^T f(x(t),t)d_\beta x(t) = \int_0^T f(x(t),t)dx(t) + \\ +(1-\beta)\int_0^T \frac{\partial f}{\partial x}(x(t),t)B(x(t),t)dt, \tag{A.2.15}$$

where $B(x(t),t)$ is a diffusion coefficient defined by relationship (A.1.37), $0 \le \beta \le 1$.

The definitions of the stochastic integral and Theorem A.6 can be extended to vector processes.

Definition A.22. Let $\mathbf{x}(t)$ be an r-dimensional diffusion process for $t \in [0,T]$, where drift vector $\mathbf{A}(\mathbf{x},t)$, diffusion matrix $\mathbf{B}(\mathbf{x},t)$ and the first derivatives $\partial\mathbf{B}(\mathbf{x},t)/\partial x_j$, $j=1,\dots,r$ are continuous with respect to both arguments. Let $\mathbf{f}(\mathbf{x},t)$ be a non-linear, non-anticipating function of values in $\mathbb{R}^r$, continuous with respect to $\mathbf{x}$ and satisfying the following conditions for $t\in[0,T]$

(i) there exist partial derivatives $\frac{\partial\mathbf{f}(\mathbf{x},t)}{\partial x_j}$ $j=1,\dots,r$,

(ii) $\int_0^T \mathrm{E}[|\mathbf{f}^T(\mathbf{x}(s),s)\mathbf{A}(\mathbf{x}(s),s)|]ds<\infty$,

(iii) $\int_0^T \mathrm{E}[|\mathbf{f}^T(\mathbf{x}(s),s)\mathbf{B}(\mathbf{x}(s),s)\mathbf{f}(\mathbf{x}(s),s)|]ds<\infty$,

then the stochastic vector integral is defined by formula

$$\int_0^T \mathbf{f}^T(\mathbf{x}(s),s)d_\beta\mathbf{x}(t)= \\ \underset{\Delta t\to 0}{l.i.m}\sum_{i=0}^{N-1}\mathbf{f}^T(\beta\mathbf{x}(t_i)+(1-\beta)\mathbf{x}(t_{i+1}),\beta t_i+(1-\beta)t_{i+1})[\mathbf{x}(t_{i+1})-\mathbf{x}(t_i)], \tag{A.2.16}$$

where $\Delta t=\max[t_{i+1}-t_i]$, $0=t_0<...<t_N=T$.

As in the case of the scalar function, the Itô and Stratonovich stochastic integrals are defined for $\beta=1$ (the Itô vector integral) and for $\beta=\frac{1}{2}$ (the Stratonovich vector integral).

Stratonovich has demonstrated that the mutual relationship between the Itô and Stratonovich vector stochastic integrals written, respectively, as

$$I_I=\int_0^T \mathbf{f}^T(\mathbf{x}(s),s)d_1\mathbf{x}(t),\quad I_S=\int_0^T \mathbf{f}^T(\mathbf{x}(s),s)d_{\frac{1}{2}}\mathbf{x}(t), \tag{A.2.17}$$

is the following

$$I_S=I_I+\frac{1}{2}\sum_{j=1}^{r}\sum_{k=1}^{r}\int_0^T\frac{\partial f_j}{\partial x_k}(\mathbf{x}(t),t)b_{jk}(\mathbf{x}(t),t)dt, \tag{A.2.18}$$

where $b_{jk}(\mathbf{x}(t),t)$ are the elements of diffusion process matrix $\mathbf{x}(t)$.

A.2.3. Itô's formula for diffusion processes

Let us present now the formula for differentiating the vector function of a diffusion process.

Let $\mathbf{x}(t)$ be an n-dimensional stochastic process given for $t\in[0,T]$, i.e. $\mathbf{x}(t)=[x_1(t),...,x_n(t)]^T$ with a stochastic differential

$$d\mathbf{x}(t)=\mathbf{a}(t,\omega)dt+\boldsymbol{\sigma}(t,\omega)d\boldsymbol{w}(t), \tag{A.2.19}$$

where $\boldsymbol{\xi}(t)$ is r-dimensional Wiener process for $t \in [0, T]$.

Vector $\mathbf{a}(t,\omega) = [a_1(t,\omega), \ldots, a_n(t,\omega)]^T$ as well as matrix $\boldsymbol{\sigma}(t,\omega) = [\sigma_{ij}(t,\omega)]$, $i = 1, \ldots, n$, $j = 1, \ldots, r$ consist of non-linear, non-anticipating functions satisfying the following conditions

$$P\left\{\int_0^T |a_i(t,\omega)|dt < \infty\right\} = 1, \quad i = 1, ..., n,$$

$$P\left\{\int_0^T |\sigma_{ij}^2(t,\omega)|dt < \infty\right\} = 1, \quad i = 1, ..., n, \quad j = 1, ..., r. \tag{A.2.20}$$

The exact differential of some non-linear function of many variables $f(\mathbf{x}(t), t)$ is then determined according to the following theorem.

Theorem A.7. Let $f(t, y_1, ..., y_n,)$ be continuous and have continuous derivatives $\partial f/\partial t$, $\partial f/\partial y_i$, $\partial^2 f/\partial y_i \partial y_j$, $i, j = 1, ..., n$. Stochastic process $f(t, x_1(t), ..., x_n(t))$ has a stochastic differential of the form

$$df(t, x_1(t), \ldots, x_n(t)) =$$

$$= \left[\frac{\partial f}{\partial t}(t, x_1(t), \ldots, x_n(t)) + \sum_{i=1}^{n} \frac{\partial f}{\partial y_i}(t, x_1(t), \ldots, x_n(t)) + \right.$$

$$\left. + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 f}{\partial y_i \partial y_j}(t, x_1(t), \ldots, x_n(t))\sum_{k=1}^{r}\sigma_{ik}(t,\omega)\sigma_{jk}(t,\omega)\right] dt + \tag{A.2.21}$$

$$+ \sum_{i=1}^{n}\sum_{k=1}^{r}\frac{\partial f}{\partial y_i}(t, x_1(t), ..., x_n(t))\sigma_{ik}(t,\omega)dw_k.$$

In the special case when $r = n$ and function $f(\mathbf{x}(t), t)$ is a quadratic form, i.e.

$$f(\mathbf{x}(t), t) = \mathbf{x}^T(t)\mathbf{H}(t)\mathbf{x}(t), \tag{A.2.22}$$

where $\mathbf{H}(t)$ is a deterministic matrix, $\mathbf{x}(t)$ is a diffusion process with differential given by

$$d\mathbf{x}(t) = \mathbf{a}(t)dt + \boldsymbol{\sigma}(t)d\boldsymbol{\xi}(t). \tag{A.2.23}$$

Thus,

$$d(\mathbf{x}^T(t)\mathbf{H}(t)\mathbf{x}(t)) = \Big[\mathbf{x}^T(t)\mathbf{H}(t)\mathbf{a}(t) + \mathbf{a}^T(t)\mathbf{H}(t)\mathbf{x}(t) +$$

$$+\mathbf{x}^T(t)\frac{d\mathbf{H}(t)}{dt}\mathbf{x}(t) + tr(\boldsymbol{\sigma}(t)\boldsymbol{\sigma}^T(t)\mathbf{H}(t))\Big]\,dt + \qquad \text{(A.2.24)}$$

$$+\,[\mathbf{x}^T(t)\mathbf{H}(t)\boldsymbol{\sigma}(t) + \boldsymbol{\sigma}^T(t)\mathbf{H}(t)\mathbf{x}(t)]d\boldsymbol{w}(t),$$

where $tr(\mathbf{A})$ is a trace of matrix $\mathbf{A}$, i.e. if $\mathbf{A}=[a_{ij}]$, $i,j=1,\ldots,n$; then

$$tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}. \qquad \text{(A.2.25)}$$

Assuming that $\mathbf{H}(t)$ is a unitary matrix independent of t, i.e. $\mathbf{H}(t)=\mathbf{I}$, then

$$d(|\mathbf{x}(t)|^2) = \left[\mathbf{x}^T\mathbf{a}(t) + \mathbf{a}^T(t)\mathbf{x}(t) + \sum_{i=1}^{r}\sigma_{ii}^2(t)\right]dt + \qquad \text{(A.2.26)}$$

$$+\,[\mathbf{x}^T\boldsymbol{\sigma}(t) + \boldsymbol{\sigma}^T(t)\mathbf{x}(t)]d\boldsymbol{w}(t).$$

A.2.4. The Itô and Stratonovich stochastic differential equations for diffusion processes

Let us consider the Itô differential stochastic vector equation

$$d\mathbf{x}(t) = \mathbf{F}(t,\mathbf{x})dt + \sum_{k=1}^{m}\mathbf{G}_k(t,\mathbf{x})dw_k(t), \qquad \mathbf{x}(t_0) = \mathbf{x}_0, \qquad \text{(A.2.27)}$$

where $\mathbf{F}, \mathbf{G}_k : [0,T]\times\mathbb{R}^n \to \mathbb{R}^n$ are non-linear deterministic vector functions $\mathbf{F}=[F_1,\ldots,F_n]^T$, $\mathbf{G}_k=[\sigma_k^i]$, $i=1,\ldots,n$, $k=1,\ldots,m$, w_k are the standard independent Wiener processes measurable with respect to a non-decreasing family of σ-algebras $\Im_t$, $t\in[0,T]$.

Let as denote by $(\mathcal{C}_T, \mathcal{B}_T)$ a measurable space of functions $\bar{\mathbf{x}} = (\mathbf{x}(t), t\in[0,T])$ continuous on $[0,T]$ with σ-algebra $\mathcal{B}_T = \sigma(\mathbf{x} : \mathbf{x}(s), s \le T)$. Let us similarly denote $\mathcal{B}_t = \sigma(\mathbf{x} : \mathbf{x}(s), s \le t)$. Let $F_i(t,\mathbf{x})$ and $\sigma_k^i(t,\mathbf{x})$ be non-anticipating functionals, i.e. $\mathcal{B}_t$-measurable for all $t\in[0,T]$.

Definition A.23. Stochastic process $\mathbf{x}(t)$, $t \in [0, T]$, continuous with probability 1, is called *a robust solution* or *a solution of a stochastic differential process* (A.2.27) with $\mathcal{F}_0$-measurable initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$, if for all $t \in [0, T]$ vector random variables $\mathbf{x}(t)$ are $\mathcal{F}_t$-measurable and

$$P\left\{\int_0^T |\mathbf{F}(t, \mathbf{x})| dt < \infty\right\} = 1, \tag{A.2.28}$$

$$P\left\{\sum_{k=1}^{m} \int_0^T |\mathbf{G}_k(t, \mathbf{x})|^2 dt < \infty\right\} = 1, \tag{A.2.29}$$

($|.|$ is a Euclidean norm) and with probability 1 for $t \in [0, T]$ we have

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t \mathbf{F}(s, \mathbf{x}) ds + \int_0^t \sum_{k=1}^{m} \mathbf{G}_k(s, \mathbf{x}) dw_k(s). \tag{A.2.30}$$

Definition A.24. A stochastic differential equation (A.2.27) has *a unique strong solution*, if for any of its robust solutions $\mathbf{x}(t)$, $\tilde{\mathbf{x}}(t)$, $t \in [0, T]$ the following relationship occurs

$$P\left\{\sup_{t \in [0,T]} |\mathbf{x}(t) - \tilde{\mathbf{x}}(t)| > 0\right\} = 0. \tag{A.2.31}$$

Let us present the simplest theorem on the existence and uniqueness of solutions.

Theorem A.8. Let the coordinates of vectors $\mathbf{F}(\mathbf{x}, t)$ and $\mathbf{G}_k(\mathbf{x}, t)$, $k = 1, ..., m$ be non-anticipating functionals $\bar{\mathbf{x}} \in \mathcal{C}_T$, $t \in [0, 1]$ satisfying the Lipschitz condition

$$\begin{aligned} &|F_i(t, \bar{\mathbf{x}}) - F_i(t, \bar{\mathbf{y}})|^2 + |\sigma_k^i(t, \bar{\mathbf{x}}) - \sigma_k^i(t, \bar{\mathbf{y}})|^2 \leq \\ &\leq L_1 \int_0^t |\mathbf{x}(s) - \mathbf{y}(s)|^2 dK(s) + L_2 |\mathbf{x}(t) - \mathbf{y}(t)|^2 \end{aligned} \tag{A.2.32}$$

and a growth condition

$$F_i^2(t, \bar{\mathbf{x}}) + (\sigma_k^i)^2(t, \bar{\mathbf{x}}) \leq L_1 \int_0^t (1 + |\mathbf{x}(s)|^2) dK(s) + L_2 (1 + (\mathbf{x}(t))^2, \tag{A.2.33}$$

for all $i = 1, \ldots, n$, $k = 1, \ldots, m$, where $L_1 > 0$ and $L_2 > 0$ are constants, $K(s)$ is a non-decreasing right continuous function, $0 \leq K(s) \leq 1$ and

$\bar{\mathbf{x}}, \bar{\mathbf{y}} \in \mathcal{C}_T$. Let initial condition $\mathbf{x}_0 = \mathbf{x}_0(\omega)$ be $\mathcal{F}_0$-measurable random vector variable, such as

$$\mathrm{P}\{\sum_{i=1}^{n} |x_{0i}| < \infty\} = 1. \tag{A.2.34}$$

In this case, equation (A.2.27) has unique strong solution $\mathbf{x}(t)$, measurable with respect to $\mathcal{F}_t$, $t \in [0, 1]$.

If we assume that there is a function $V(t, \mathbf{x})$ that has continuous and limited derivatives of first order with respect to t and of second order with respect to the coordinates of vector $\mathbf{x}$ for $t \in [0, T]$ and $\mathbf{x} \in \mathbb{R}^n$, which is denoted by $V \in \mathbf{C}_2$ then, from Theorem A.7 it follows that

$$\begin{aligned} &V(t, \mathbf{x}(t)) - V(s, \mathbf{x}(s)) = \\ &= \int_s^t \mathcal{L}(V(u, \mathbf{x}(u))du + \int_s^t \sum_{k=1}^{M} \frac{\partial V}{\partial \mathbf{x}} \mathbf{G}_k(u, \mathbf{x}(u))dw_k(u), \end{aligned} \tag{A.2.35}$$

where the operator $\mathcal{L}(.)$ is defined depending on the definition of the stochastic integral on the right-hand side of equation (A.2.30). This means that if the stochastic integral is an Itô integral, then stochastic equation (A.2.27) is called *the Itô stochastic equation*, but if the stochastic integral is a Stratonovich integral, then equation (A.2.27) is called *the Stratonovich stochastic equation.* For simplicity, the equations will be called *the Itô equation* and *the Stochanovich equation*, respectively. The same correspondence of terms is maintained both for differentials $d_I w_k$, $d_S w_k$ and for operators $\mathcal{L}_I$, $\mathcal{L}_S$.

Because of equations (A.2.21), (A.2.30) and (A.2.35) it follows that the Itô and Stratonovich operators are defined as follows

$$\mathcal{L}_I(.) = \frac{\partial}{\partial t} + \mathbf{F}^T(t, \mathbf{x}) \frac{\partial}{\partial \mathbf{x}} + \frac{1}{2} \sum_{k=1}^{m} \left\langle \frac{\partial}{\partial \mathbf{x}},\ \mathbf{G}_k(t, \mathbf{x}) \right\rangle^2, \tag{A.2.36}$$

$$\begin{aligned} \mathcal{L}_S(.) = &\frac{\partial}{\partial t} + \left(\mathbf{F}(t, \mathbf{x}) \frac{\partial}{\partial \mathbf{x}} + \frac{1}{2} \sum_{k=1}^{m} \frac{\partial \mathbf{G}_k(t, \mathbf{x})}{\partial \mathbf{x}} \mathbf{G}_k(t, \mathbf{x}) \right)^T \frac{\partial}{\partial \mathbf{x}} + \\ &+ \frac{1}{2} \sum_{k=1}^{m} \left\langle \frac{\partial}{\partial \mathbf{x}},\ \mathbf{G}_k(t, \mathbf{x}) \right\rangle^2. \end{aligned} \tag{A.2.37}$$

where $\langle \mathbf{x}, \mathbf{y} \rangle$ denotes a scalar product of vectors $\mathbf{x}$ and $\mathbf{y}$.

Stochastic differential equation (A.2.27) is equivalent in the "Stratonovich sense" to the following Itô equation

$$d\mathbf{x} = \left[\mathbf{F}(t,\mathbf{x}) + \frac{1}{2} \sum_{k=1}^{m} \frac{\partial \mathbf{G}_k(t,\mathbf{x})}{\partial \mathbf{x}} \mathbf{G}_k(t,\mathbf{x}) \right] dt + \sum_{k=1}^{m} \mathbf{G}_k(t,\mathbf{x}) dw_k(t). \quad \text{(A.2.38)}$$

This means that all results obtained for the Itô equations can be used to analyse the Stratonovich equations. It needs to be stressed that when functions $\mathbf{G}_k$ are not dependent on vector $\mathbf{x}$, then Itô and Stratonovich equations are identical.

Because in this book the Itô equations are used the most frequently, for simplicity we shall use the following notation $\mathcal{L}_I(.) = \mathcal{L}(.)$.

Given that linear systems are particularly important for modeling dynamic systems, let us present a theorem on the existence of strong solutions of a linear vector stochastic differential equation.

Theorem A.9. Let the elements of vector function

$$\mathbf{A}_0(t) = [a_0^1(t), ..., a_0^n(t)]^T \quad \text{(A.2.39)}$$

and matrix

$$\mathbf{A} = [a_{ij}], \;\; \mathbf{G}_k = [\sigma_{k0}^i], \;\; i,j = 1,...,n, \;\; k = 1,...,m \quad \text{(A.2.40)}$$

be measurable (deterministic) functions of variable $t \in [0,1]$ satisfying the following conditions

$$\int_0^1 |a_0^i(t)|dt < \infty, \; \int_0^1 |a_{ij}(t)|dt < \infty, \; \int_0^1 |(\sigma_{k0}^i)^2(t)|dt < \infty. \quad \text{(A.2.41)}$$

Then the vector stochastic differential equation

$$d\mathbf{x} = [\mathbf{A}_0(t) + \mathbf{A}(t)\mathbf{x}(t)]dt + \sum_{k=1}^{M} \mathbf{G}_{k0}(t)\, dw_k(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0, \quad \text{(A.2.42)}$$

where $\xi_k(t)$ are independent Wiener processes measurable with respect to $\mathcal{F}_t$, $t \in [0,1]$, has an unique strong solution given by an integral formula

$$\mathbf{x}(t) = \mathbf{\Phi}(t,0) \left[\mathbf{x}(t_0) + \int_0^t \mathbf{\Phi}^{-1}(s,0)\mathbf{A}_0(s)ds + \right. \\ \left. + \int_0^t \sum_{k=1}^{m} \mathbf{\Phi}^{-1}(s,0)\mathbf{G}_{k0}(s)dw_k(s) \right], \quad \text{(A.2.43)}$$

where $\mathbf{\Phi}(t,0)$ is $n \times n$ fundamental matrix

$$\mathbf{\Phi}(t,0) = \mathbf{I} + \int_0^t \mathbf{A}(s)\mathbf{\Phi}(s,0)ds, \tag{A.2.44}$$

and $\mathbf{I}$ is $n \times n$ identity matrix.

Solution (A.2.43) can be extended for any $t \in [0,T]$, assuming that the elements of vector $\mathbf{A}_0$ and matrices $\mathbf{A}, \mathbf{G}_k$ are measurable limited functions for $t \in [0,T]$.

Physical interpretation of the Stratonovich equation

Differential equations with stochastic parameters are frequently used for modeling real processes, e.g. physical, chemical, biological, technical or economic. An example of such an equation is the scalar Langevin equation

$$\frac{dx(t)}{dt} = F(t,x(t)) + G(t,x(t))\dot{w}(t), \tag{A.2.45}$$

where $F(t,x)$ and $G(t,x)$ are non-linear scalar functions, and $\dot{w}(t)$ is a Gaussian white noise.

Because white noise is an abstract notion and in the real world only coloured noise or non-stationary noise occurs, the following problem is encountered. Let us consider a family of differential equations (A.2.45), where process $\dot{w}(t)$ has been replaced by a sequence of stationary wideband Gaussian processes (i.e. processes for which power spectral density has a "wide range") $\{\eta^n(t)\}$, $n = 1, 2, ...$ and let us assume that sequence $\{\eta^n(t)\}$ converges in some sense to the Gaussian white noise (for instance, as n grows the band length of the power spectral density of process $\{\eta^n(t)\}$ also grows). Let us assume that for each n process $\eta^n(t)$ has regular realizations. Then, the appropriate sequence $\{x_n(t)\}$ is a solution of this family of differential equations

$$\frac{dx_n(t)}{dt} = F(t,x_n(t)) + G(t,x_n(t))\eta^n(t). \tag{A.2.46}$$

Let us assume that sequence $\{x_n(t)\}$ converges to a process $\tilde{x}(t)$. Two questions arise then: what is the nature of process $\tilde{x}(t)$ and how does it relate to process $x(t)$, solving the corresponding Itô stochastic differential equation

$$dx(t) = F(t,x(t))dt + G(t,x(t))d_I w(t). \tag{A.2.47}$$

Wong and Zakai have demonstrated that the sequence of solutions $\{x_n(t)\}$ converges to the solution of the corresponding Stratonovich equation

$$dx(t) = F(t, x(t))dt + G(t, x(t))d_S w(t) \tag{A.2.48}$$

or the equivalent Itô equation

$$dx(t) = \left[F(t, x(t)) + \frac{1}{2}\frac{\partial G}{\partial x}(t, x(t))G(t, x(t))\right] dt + \\ + G(t, x(t))dw(t). \tag{A.2.49}$$

This important finding has been generalized for the multidimensional case by [Papanicolau, Kohler 1974].

A.3. Moment equations for linear stochastic dynamic systems

The determination of the moments of solutions (system responses) in linear dynamic stochastic systems is one of the problems faced in the stochastic analysis. In this section, we shall discuss the basic methods used to solve linear vector stochastic differential equations and the methods for determining their moments.

A.3.1. Linear systems with additive excitation

Let us consider a linear vector stochastic differential equation with additive excitation

$$d\mathbf{x}(t) = [\mathbf{A}_0(t) + \mathbf{A}(t)\mathbf{x}(t)]dt + \sum_{k=1}^{m} \mathbf{G}_{k0}(t)dw_k(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0, \tag{A.3.1}$$

where

$\mathbf{x}(t) = [x_1(t), ..., x_n(t)]^T$,

$\mathbf{A}_0(t) = [a_0^1(t), ..., a_0^n(t)]^T$,

$\mathbf{A}(t) = [a_{ij}(t)]$,

$\mathbf{G}_{k0}(t) = [\sigma_{k0}^1(t), ..., \sigma_{k0}^n(t)]^T$ are n-dimensional vectors,

$w_k(t)$ are independent standard Wiener processes, $i, j = 1,...,n$, $k = 1, ..., m$,

$\mathbf{x}_0 = [x_{01}, ..., x_{0n}]^T$ is an initial condition, i.e. a vector random variable independent of $w_k(t)$, $k = 1, ..., m$,

a_0^i, a_{ij} and σ_{k0}^i are limited, measurable deterministic functions of $t \in \mathbb{R}^+$.

The solution (strong) is given by the relationship

$$\mathbf{x}(t) = \mathbf{\Psi}(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{\Psi}(t, s)\mathbf{A}_0(s)ds + \\ + \int_{t_0}^{t} \mathbf{\Psi}(t, s)\sum_{k=1}^{m} \mathbf{G}_{k0}(s)dw_k(s), \tag{A.3.2}$$

where $\mathbf{\Psi}(t, t_0)$ is the $n \times n$ fundamental matrix of homogeneous equation

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}(t)\mathbf{x}(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0. \tag{A.3.3}$$

In particular, when $\mathbf{A}$ is a constant matrix, then

$$\mathbf{\Psi}(t, t_0) = \mathbf{\Psi}(t - t_0) = \exp\{\mathbf{A}(t - t_0)\} = \sum_{l=0}^{\infty} \frac{1}{l!}\mathbf{A}^l(t - t_0)^l \tag{A.3.4}$$

and relationship (A.3.2) is reduced to

$$\mathbf{x}(t) = \exp\{\mathbf{A}(t - t_0)\}\mathbf{x}_0 + \int_{t_0}^{t} \exp\{\mathbf{A}(t - s)\}\mathbf{A}_0(s)ds + \\ + \int_{t_0}^{t} \exp\{\mathbf{A}(t - s)\} \sum_{k=1}^{m} \mathbf{G}_{k0}(s)d\boldsymbol{\xi}(s). \tag{A.3.5}$$

Using the Itô formula and an averaging procedure, we obtain equations for the mean values and second-order moments

$$\frac{d\mathbf{m}(t)}{dt} = \mathbf{A}_0(t) + \mathbf{A}(t)\mathbf{m}(t), \quad \mathbf{m}(t_0) = \mathbf{m}_0, \tag{A.3.6}$$

$$\frac{d\mathbf{\Gamma}(t)}{dt} = \mathbf{m}(t)\mathbf{A}_0^T(t) + \mathbf{A}_0(t)\mathbf{m}^T(t) + \mathbf{\Gamma}(t)\mathbf{A}^T(t) + \mathbf{A}(t)\mathbf{\Gamma}(t), \tag{A.3.7}$$

$$+\sum_{k=1}^{m}\mathbf{G}_{k0}(t)\mathbf{G}_{k0}^T(t) \quad \mathbf{\Gamma}(t_0)=\mathbf{\Gamma_0}, \tag{A.3.8}$$

where

$$\begin{aligned}\mathbf{m}(t) &= \mathrm{E}[\mathbf{x}(t)], \quad \mathbf{\Gamma}(t) = \mathrm{E}[\mathbf{x}(t)\mathbf{x}^T(t)],\\ \mathbf{m}_0 &= \mathrm{E}[\mathbf{x}(t_0)], \quad \mathbf{\Gamma}_0 = \mathrm{E}[\mathbf{x}(t_0)\mathbf{x}^T(t_0)].\end{aligned} \tag{A.3.9}$$

A.3.2. Linear systems with additive and parametric excitation

Let us start by solving two simple examples of scalar linear stochastic differential equations.

Homogeneous case. Let us consider the Itô scalar linear homogeneous stochastic equation

$$dx(t) = a(t)x(t)dt + \sigma(t)x(t)dw(t), \quad x(t_0) = x_0, \tag{A.3.10}$$

where $t \in [t_0, \infty)$, $a(t)$ and $\sigma(t)$ are non-linear functions of variable t and initial condition x_0 is a random variable independent of standard Wiener process $w(t)$.

Using the Itô formula, we can demonstrate that the solution of equation (A.3.10) is a stochastic process

$$x(t) = \psi(t, t_0)x_0, \tag{A.3.11}$$

where

$$\psi(t, t_0) = \exp\left\{\int_{t_0}^{t}\left[a(s) - \frac{\sigma^2(s)}{2}\right]ds + \int_{t_0}^{t}\sigma(s)dw(s)\right\}, \tag{A.3.12}$$

the p-th moment of which is of the form

$$\mathrm{E}[x^p(t)] = \mathrm{E}[x_0^p]\exp\left\{p\int_{t_0}^{t}\left[a(s) - \frac{\sigma^2(s)}{2}\right]ds + \frac{p^2}{2}\int_{t_0}^{t}\sigma^2(s)ds\right\}. \tag{A.3.13}$$

Heterogeneous case. Let us consider the Itô linear scalar heterogeneous stochastic equation

$$dx(t) = [a(t)x(t) + b(t)]dt + [\sigma(t)x(t) + q(t)]dw(t), \; x(t_0) = x_0, \tag{A.3.14}$$

where $t \in [t_0, \infty)$, $b(t)$ and $q(t)$ are non-linear functions of time and all other notations are the same as in (A.3.10).

Let us introduce a new variable

$$z(t) = x(t)\psi^{-1}(t, t_0), \tag{A.3.15}$$

where $\psi(t, t_0)$ is given by relationship (A.3.12). Using the Itô formula, we obtain the equation for process $z(t)$

$$dz(t) = \{[b(t) - q(t)\sigma(t)]dt + q(t)d\xi(t)\}\psi^{-1}(t, t_0). \tag{A.3.16}$$

By integrating equation (A.3.16) and introducing a transform inverse to (A.3.15), we arrive at the following solution

$$\begin{aligned} x(t) = \psi(t, t_0)\{x(t_0) + \int_{t_0}^{t} \psi^{-1}(s, t_0)[b(s) - q(s)\sigma(s)]ds + \\ + \int_{t_0}^{t} \psi^{-1}(s, t_0)q(s)dw(s)\}. \end{aligned} \tag{A.3.17}$$

Unlike the homogeneous case, it is more convenient to find a differential equation for the p-th moment rather than to average the p-th power of (A.3.17) ($p > 0$). Therefore, by applying, again, the Itô formula to function x^p for $p > 0$, using equation (A.3.14) and then averaging the obtained result, we arrive at

$$\begin{aligned} \frac{d\mathrm{E}[x^p(t)]}{dt} = \mathrm{E}[x^p(t)]\left[pa(t) + \frac{p(p-1)}{2}\sigma^2(t)\right] + \\ + \mathrm{E}[x^{p-1}(t)][pb(t) + p(p-1)q(t)\sigma(t)] + \\ + \mathrm{E}[x^{p-2}(t)]\frac{p(p-1)}{2}q^2(t), \quad \mathrm{E}[x^p(t_0)] = \mathrm{E}[x_0^p]. \end{aligned} \tag{A.3.18}$$

Unfortunately, in the general case of a linear vector stochastic differential equation with parametric excitation (the noise coefficients are dependent on the state vector) it is not possible to find an analytical solution. This possibility is only available in the case of a linear vector stochastic differential equation with additive excitation (the noise coefficients are independent of the state vector). For the vector case, we shall consider equations of the first- and second-order moments.

Let us consider the Itô linear vector stochastic equation with additive and parametric excitation

$$d\mathbf{x}(t) = [\mathbf{A}_0(t) + \mathbf{A}(t)\mathbf{x}(t)]dt+$$

$$+\sum_{k=1}^{m}[\mathbf{G}_{k0}(t) + \mathbf{G}_k(t)\mathbf{x}(t)]dw_k(t), \tag{A.3.19}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0,$$

where

$\mathbf{x}(t) = [x_1(t), ..., x_n(t)]^T$,

$\mathbf{A}_0(t) = [a_0^1(t), ..., a_0^n(t)]^T$,

$\mathbf{G}_{k0}(t) = [\sigma_{k0}^1(t), ..., \sigma_{k0}^n(t)]^T$ are n-dimensional vectors,

$\mathbf{A}(t) = [a_{ij}(t)]$, $i, j = 1, ..., n$,

$\mathbf{G}_k(t) = [\sigma_{kj}^i(t)]$ are $n \times n$ matrices,

$w_k(t)$, $k = 1, ..., m$ are independent standard Wiener processes,

$\mathbf{x}_0 = [x_{01}, ..., x_{0n}]^T$ is an initial condition,

a_0^i, a_{ij}, σ_{k0}^i are limited measurable deterministic functions of variable $t \in \mathbb{R}^+$.

For simplicity, let us assume that $\mathbf{x}_0 = [x_{01}, ..., x_{0n}]^T$ is a vector random variable independent of $w_k(t)$, $k = 1, 2, \ldots, m$. Using the Itô formula and an averaging procedure, we obtain equations for mean values and second-order moments

$$\frac{d\mathbf{m}(t)}{dt} = \mathbf{A}_0(t) + \mathbf{A}(t)\mathbf{m}(t), \quad \mathbf{m}(t_0) = \mathbf{m}_0, \tag{A.3.20}$$

$$\frac{d\mathbf{\Gamma}(t)}{dt} = \mathbf{m}(t)\mathbf{A}_0^T(t) + \mathbf{A}_0(t)\mathbf{m}^T(t) + \mathbf{\Gamma}(t)\mathbf{A}^T(t)+$$

$$+\mathbf{A}(t)\mathbf{\Gamma}(t) + \sum_{k=1}^{m}[\mathbf{G}_{k0}(t)\mathbf{G}_{k0}^T(t) + \mathbf{G}_k(t)\mathbf{m}(t)\mathbf{G}_{k0}(t)+$$

$$+ \mathbf{G}_{k0}(t)\mathbf{m}^T(t)\mathbf{G}_k^T(t) + \mathbf{G}_k(t)\mathbf{\Gamma}(t)\mathbf{G}_k^T(t)], \tag{A.3.21}$$

$$\mathbf{\Gamma}(t_0) = \mathbf{\Gamma_0},$$

where

$$\mathbf{m}(t) = \mathrm{E}[\mathbf{x}(t)], \quad \mathbf{\Gamma}(t) = \mathrm{E}[\mathbf{x}(t)\mathbf{x}^T(t)],$$
$$\mathbf{m}_0 = \mathrm{E}[\mathbf{x}(t_0)], \quad \mathbf{\Gamma}_0 = \mathrm{E}[\mathbf{x}(t_0)\mathbf{x}^T(t_0)]. \tag{A.3.22}$$

Equations (A.3.20) and (A.3.21) for the coordinates are of the form

$$\frac{dm_i(t)}{dt} = a_0^i(t) + \sum_{j=1}^{n} a_{ij}(t)m_j(t), \quad m_i(t_0) = m_{i0} \tag{A.3.23}$$

and

$$\frac{d\Gamma_{ij}(t)}{dt} = a_0^i(t)m_j(t) + a_0^j(t)m_i(t) +$$
$$+\sum_{l=1}^{n}[a_{il}(t)\Gamma_{lj}(t) + a_{jl}(t)\Gamma_{li}(t)] +$$
$$+\sum_{k=1}^{m}\sum_{\alpha=1}^{n}[\sigma_{k\alpha}^i(t)\sigma_{k0}^j(t)m_\alpha(t) + \sigma_{k\alpha}^j(t)\sigma_{k0}^i(t)m_\alpha(t)] + \tag{A.3.24}$$
$$+\sum_{k=1}^{m}\sigma_{k0}^i(t)\sigma_{k0}^j(t) + \sum_{k=1}^{m}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\sigma_{k\alpha}^i(t)\sigma_{k\alpha}^j(t),$$
$$\Gamma_{\alpha\beta}(t), \;\; \Gamma_{ij}(t_0) = \Gamma_{ij0}, \;\; i,j = 1,\ldots,n,$$

where

$$m_i(t) = \mathrm{E}[x_i(t)], \Gamma_{ij}(t) = \mathrm{E}[x_i(t)x_j(t)],$$
$$m_{i0} = \mathrm{E}[x_i(t_0)], \Gamma_{ij0} = \mathrm{E}[x_i(t_0)x_j(t_0)]. \tag{A.3.25}$$

Let us notice that the obtained moment equations are closed, i.e. there are no moments on the right-hand side that would be of higher order than the moments on the left-hand side and the second-order moments only depend on variable t.

A.4. Methods of discretization of stochastic differential equations

Given that analytical solutions are only known for few stochastic solutions, it is necessary to work out methods approximating equation solutions, especially the numerical methods. The main idea underlying these methods consists in replacing the stochastic differential equation with its discrete representation, which is some difference equation. In the special case, the Itô scalar stochastic equation

$$dx(t) = F(x(t),t)dt + \sum_{k=1}^{m} G_k(x(t),t)dw_k(t), \quad x(t_0) = x_0 \quad \text{(A.4.1)}$$

is replaced by an appropriate discrete difference equation

$$x_{i+1} = x_i + F_i \Delta t_i + \sum_{k=1}^{m} G_{k_i} \Delta w_{k_i}, \quad i = 0, 1, ... \quad \text{(A.4.2)}$$

In the case of the Stratonovich scalar stochastic differential equation

$$dx(t) = F(x(t),t)dt + \sum_{k=1}^{m} G_k(x(t),t)dw_k^*(t), \quad x(t_0) = x_0, \quad \text{(A.4.3)}$$

its discrete-time representation is written as

$$x_{i+1} = x_i + \left[F_i + \frac{1}{2} \sum_{k=1}^{m} \left(\frac{\partial G_k}{\partial x} \right)_i G_{k_i} \right] \Delta t_i + \sum_{k=1}^{m} G_{k_i} \Delta w_{k_i}, \quad \text{(A.4.4)}$$

where $i = 0, 1, ...,$ $dw_k(t)$ and $dw_k^*(t)$, $k = 1, 2, .., m$ are stochastic differentials of standard Wiener processes in the Itô and Stratonovich sense, respectively.

$$x_i = x(t_i), \quad F_i = F(x(t_i), t_i), \quad G_i = G(x(t_i), t_i),$$

$$\left(\frac{\partial G_k}{\partial x} \right)_i = \frac{\partial G_k(x(t),t)}{\partial x}_{|x=x_i}, \quad \Delta t_i = t_{i+1} - t_i, \quad \text{(A.4.5)}$$

$$t_0 < t_1 < ... < t_N = T, \quad \Delta w_{k_i} = w(t_{i+1}) - w(t_i), \quad i = 0, 1, ...$$

In estimating the parameters of mortality models represented by multidimensional stochastic differential equations containing one Wiener process it is convenient to use only the two simplest methods of discretization, i.e. the Milstein method

$$x_{i+1}^j = x_i^j + F_i^j \Delta t + G_i^j \Delta w + \frac{1}{2}\left(\sum_{l=1}^{n} G_i^l \frac{\partial G_i^j}{\partial x^l}\right)\left\{(\Delta w)^2 - \Delta\right\}, \quad \text{(A.4.6)}$$

or the Euler method

$$x_{i+1}^j = x_i^j + F_i^j \Delta t + G_i^j \Delta w, \quad \text{(A.4.7)}$$

where $j = 1, ...n$, $i = 0, 1, ...$, $\mathbf{x} = [x_1, ..., x_n]^T$ is a state vector and the coordinates of the drift and diffusion vectors $\mathbf{F} = [F^1, ..., F^n]^T$, $\mathbf{G} = [G^1, ..., G^n]^T$ are non-linear functions of the state vector.

Appendix B

Elements of the algebra of modified fuzzy and complex numbers

B.1. Modified fuzzy numbers

The algebra of modified fuzzy numbers (MFN) is constructed along the same lines as the algebra of oriented fuzzy numbers (OFN) proposed by [Kosiński *et al.* 2003, Kosiński, Prokopowicz 2004].

The representation of a fuzzy number is a pair of continuous functions (f_A, g_A). The difference between MFN and OFN lies in the differently defined multiplication of their elements. Whereas in the algebra of OFN predecessors and successors are multiplied by each other

$$\vec{A} \otimes \vec{B} = (f_A f_B,\ g_A g_B), \tag{B.1.1}$$

for $\vec{A} = (f_A, g_A)$, $\vec{B} = (f_B, g_B)$, in the algebra of MFN elements $\check{A} = (f_A, g_A)$ and $\check{B} = (f_B, g_B)$ are multiplied as follows

$$\check{A} \odot \check{B} = \left(\frac{1}{2}(f_A f_B + g_A g_B),\ \frac{1}{2}(f_A g_B + g_A f_B) \right). \tag{B.1.2}$$

Definiction B.1. The term "a modified fuzzy number $\check{A}$" will apply to each ordered pair of continuous functions

$$\check{A} = (f, g), \tag{B.1.3}$$

where $f, g:\ [0,1] \to \mathbb{R}$ satisfy axioms (i)–(iii), defining the equality, sum and product of modified fuzzy numbers:

(i) Modified fuzzy numbers $\check{A} = (f_A, g_A)$ and $\check{B} = (f_B, g_B)$ are equal if and only if $f_A(u) = f_B(u)$ and $g_A(u) = g_B(u)$ for each $u \in [0,1]$.

(ii) A sum of modified fuzzy numbers $\check{A} = (f_A, g_A)$ and $\check{B} = (f_B, g_B)$ is a modified fuzzy number $\check{A} \oplus \check{B}$ of the form

$$\check{A} \oplus \check{B} = (f_A, g_A) \oplus (f_B, g_B) = (f_A + f_B,\ g_A + g_B). \tag{B.1.4}$$

(iii) A product of modified fuzzy numbers $\breve{A} = (f_A, g_A)$ and $\breve{B} = (f_B, g_B)$ is a modified fuzzy number $\breve{A} \odot \tilde{B}$ of the form

$$\breve{A} \odot \breve{B} = (f_A, g_A) \odot (f_B, g_B) = \\ = \left(\frac{1}{2}(f_A f_B + g_A g_B),\ \frac{1}{2}(f_A g_B + g_A f_B)\right). \tag{B.1.5}$$

Definition B.2. $\breve{C} = (f_C, g_C)$ is obtained by multiplying a modified fuzzy number $\breve{A} = (f_A, g_A)$ by scalar d, which is symbolically written as $\breve{C} = d\tilde{A}$, if

$$f_C(u) = df_A(u), \quad g_C(u) = dg_A(u). \tag{B.1.6}$$

In the algebra of MFN, Definition B.1 describes the product of modified fuzzy numbers. In the algebra of CFN (*Complex-Valued Fuzzy Numbers*), the product is defined as

$$A \odot B = (f_A, g_A) \odot (f_B, g_B) = (f_A f_B - g_A g_B,\ f_A g_B + g_A f_B). \tag{B.1.7}$$

Let us now consider the algebra of GFN (*Generalized Fuzzy Numbers*) that generalizes the multiplication of elements by introducing numerical coefficients $\alpha, \beta, \gamma, \delta \in \mathbb{R}$. In the algebra of GFN, multiplication is defined as

$$(f_A, g_A) \odot (f_B, g_B) = (\alpha f_A f_B + \beta g_A g_B,\ \gamma f_A g_B + \delta g_A f_B). \tag{B.1.8}$$

In the special case, when

$$\alpha = \beta = \gamma = \delta = \frac{1}{2}, \tag{B.1.9}$$

(B.1.5) is obtained and

$$\alpha = \gamma = \delta = 1, \quad \beta = -1 \tag{B.1.10}$$

leads to (B.1.7).

Example B.1. Let $\breve{A} = (f_A, g_A)$ and $\breve{B} = (f_B, g_B)$, where

$$f_A(u) = a - s_A(1-u), \quad g_A(u) = a + s_A(1-u), \\ f_B(u) = b - s_B(1-u), \quad g_B(u) = b + s_B(1-u),\ u \in [0,1]. \tag{B.1.11}$$

From

$$(f_A, g_A) \oplus (f_B, g_B) = (f_A(u) + f_B(u), g_A(u) + g_B(u)), \tag{B.1.12}$$

we have

$$\begin{aligned} f_A(u) + f_B(u) &= a + b - (s_A + s_B)(1-u), \\ g_A(u) + g_B(u) &= a + b + (s_A + s_B)(1-u),\ u \in [0,1]. \end{aligned} \tag{B.1.13}$$

The result of algebraic multiplication for $\check{A}, \check{B}$ being MFN's is written as

$$(\check{A} \odot \check{B})(u) = (ab + s_A s_B(1-u)^2,\ ab - s_A s_B(1-u)^2), \tag{B.1.14}$$

and for A, B being CFN's we have

$$(A \odot B)(u) = (-2(as_B + bs_A)(1-u),\ 2ab - 2s_A s_B(1-u)^2). \tag{B.1.15}$$

It is essential that for each algebra the null element and the unity element be identified.

In the algebra of oriented fuzzy numbers, the null element is denoted as $\vec{\mathbb{O}} = (0,0)$, where $0(u) = 0$ for each $u \in [0,1]$ and the unity element is written as $\vec{\mathbb{I}} = (1,1)$, so that for each $\vec{A}$ there is

$$\vec{A} \otimes \vec{\mathbb{I}} = \vec{\mathbb{I}} \otimes \vec{A}. \tag{B.1.16}$$

Let us first determine the general form of the null element $(f_{\mathbb{O}},\ g_{\mathbb{O}})$ in the algebra of GFN's. From the definition of the null element it follows that there should be

$$(f_A, f_B) \oplus (f_{\mathbb{O}},\ g_{\mathbb{O}}) = (f_A, f_B). \tag{B.1.17}$$

Hence, we have

$$(f_A,\ g_A) \oplus (f_{\mathbb{O}},\ g_{\mathbb{O}}) = (f_A + f_{\mathbb{O}},\ g_A + g_{\mathbb{O}}). \tag{B.1.18}$$

(B.1.17) implies equality

$$(f_A + f_{\mathbb{O}},\ g_A + g_{\mathbb{O}}) = (f_A,\ g_A), \tag{B.1.19}$$

so

$$f_A(u)+f_{\mathbb{O}}(u) = f_A(u), \quad g_A(u)+g_{\mathbb{O}}(u) = g_A(u), \quad u \in [0,1]. \tag{B.1.20}$$

For $u \in [0,1]$ we have $f_{\mathbb{O}}(u) = 0$, $g_{\mathbb{O}}(u) = 0$, i.e. the null element is $(0,\ 0)$.

The same procedure is applied to find the general form of the unity element $(f_{\mathbb{I}}, g_{\mathbb{I}})$ in the algebra of GFN's, where multiplication is performed according to formula (B.1.8). Thus, the following condition should be satisfied

$$(f_A, f_B) \odot (f_{\mathbb{I}},\ g_{\mathbb{I}}) = (f_{\mathbb{I}},\ g_{\mathbb{I}}) \odot (f_A, f_B) = (f_A, f_B). \tag{B.1.21}$$

We have

$$(f_A, f_B) \odot (f_{\mathbb{I}},\ g_{\mathbb{I}}) = (\alpha f_A f_{\mathbb{I}} + \beta g_A g_{\mathbb{I}},\ \gamma f_A g_{\mathbb{I}} + \delta g_A f_{\mathbb{I}}). \tag{B.1.22}$$

(B.1.21) implies that the following equality exists

$$(\alpha f_A f_{\mathbb{I}} + \beta g_A g_{\mathbb{I}},\ \gamma f_A g_{\mathbb{I}} + \delta g_A f_{\mathbb{I}}) = (f_A,\ g_A), \tag{B.1.23}$$

i.e. for each $u \in [0,1]$ we have

$$\alpha f_A(u) f_{\mathbb{I}}(u) + \beta g_A(u) g_{\mathbb{I}}(u) = f_A(u) \tag{B.1.24}$$

and

$$\gamma f_A(u) g_{\mathbb{I}}(u) + \delta g_A(u) f_{\mathbb{I}}(u) = g_A(u). \tag{B.1.25}$$

The above equalities can be written as a matrix

$$\begin{bmatrix} \alpha f_A(u) & \beta g_A(u) \\ \delta g_A(u) & \gamma f_A(u) \end{bmatrix} \begin{bmatrix} f_{\mathbb{I}}(u) \\ g_{\mathbb{I}}(u) \end{bmatrix} = \begin{bmatrix} f_A(u) \\ g_A(u) \end{bmatrix}. \tag{B.1.26}$$

Let us denote the left hand side of matrix (B.1.26) as $\mathbb{G}(u)$, i.e.

$$\mathbb{G}(u) = \begin{bmatrix} \alpha f_A(u) & \beta g_A(u) \\ \delta g_A(u) & \gamma f_A(u) \end{bmatrix}. \tag{B.1.27}$$

Let us now consider the inverse of the matrix $\mathbb{G}(u)$ for $u \in [0,1]$. The condition of matrix inversion is defined as

$$\det\left[\mathbb{G}(u)\right] \neq 0. \tag{B.1.28}$$

We have

$$\det\left[\mathbb{G}(u)\right] = \alpha\gamma f_A^2(u) - \beta\delta g_A^2(u). \tag{B.1.29}$$

Hence, matrix $\mathbb{G}(u)$ is invertible for $\alpha\gamma f_A^2(u) \neq \beta\delta g_A^2(u)$.

From (B.1.10) we obtain for the algebra of MFN's $\alpha\gamma = \frac{1}{4}$, $\beta\delta = \frac{1}{4}$, so (B.1.29) should be written as

$$\det\left[\mathbb{G}(u)\right] = \frac{1}{4}(f_A^2(u) - g_A^2(u)). \tag{B.1.30}$$

In the case of the algebra of CFN's, we have $\alpha\gamma = 1$, $\beta\delta = -1$ and (B.1.29) is written as

$$\det\left[\mathbb{G}(u)\right] = f_A^2(u) + g_A^2(u). \tag{B.1.31}$$

Let us find matrix $\mathbb{D}(u)$ of the cofactors of matrix (B.1.27). We have

$$\mathbb{D}(u) = \begin{bmatrix} \gamma f_A(u) & -\delta g_A(u) \\ -\beta g_A(u) & \alpha f_A(u) \end{bmatrix}, \tag{B.1.32}$$

the transposition of which results in a matrix $\mathbb{D}^T(u)$

$$\mathbb{D}^T(u) = \begin{bmatrix} \gamma f_A(u) & \beta g_A(u) \\ -\delta g_A(u) & \alpha f_A(u) \end{bmatrix}. \tag{B.1.33}$$

Thereby, the inverse matrix $\mathbb{G}^{-1}(u)$ has the following form

$$\mathbb{G}^{-1}(u) = \frac{1}{\det\left[\mathbb{G}(u)\right]} \begin{bmatrix} \gamma f_A(u) & -\beta g_A(u) \\ -\delta g_A(u) & \alpha f_A(u) \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{\gamma f_A(u)}{\alpha\gamma f_A^2(u) - \beta\delta g_A^2(u)} & -\frac{\beta g_A(u)}{\alpha\gamma f_A^2(u) - \beta\delta g_A^2(u)} \\ -\frac{\delta g_A(u)}{\alpha\gamma f_A^2(u) - \beta\delta g_A^2(u)} & \frac{\alpha f_A(u)}{\alpha\gamma f_A^2(u) - \beta\delta g_A^2(u)} \end{bmatrix}. \tag{B.1.34}$$

The matrix equation (B.1.26) can now be written as

$$\mathbb{G}(u) \begin{bmatrix} f_{\mathbb{I}}(u) \\ g_{\mathbb{I}}(u) \end{bmatrix} = \begin{bmatrix} f_A(u) \\ g_A(u) \end{bmatrix}, \tag{B.1.35}$$

and after multiplying (B.1.35) from the left side by $\mathbb{G}^{-1}(u)$ we arrive at

$$\begin{bmatrix} f_{\mathbb{I}}(u) \\ g_{\mathbb{I}}(u) \end{bmatrix} = \mathbb{G}^{-1}(u) \begin{bmatrix} f_A(u) \\ g_A(u) \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{\gamma f_A(u)}{\alpha\gamma f_A^2(u) - \beta\delta g_A^2(u)} & -\frac{\beta g_A(u)}{\alpha\gamma f_A^2(u) - \beta\delta g_A^2(u)} \\ -\frac{\delta g_A(u)}{\alpha\gamma f_A^2(u) - \beta\delta g_A^2(u)} & \frac{\alpha f_A(u)}{\alpha\gamma f_A^2(u) - \beta\delta g_A^2(u)} \end{bmatrix} \begin{bmatrix} f_A(u) \\ g_A(u) \end{bmatrix} = \quad \text{(B.1.36)}$$

$$= \begin{bmatrix} \frac{\gamma f_A^2(u)}{\alpha\gamma f_A^2(u) - \beta\delta g_A^2(u)} - \frac{\beta g_A^2(u)}{\alpha\gamma f_A^2(u) - \beta\delta g_A^2(u)} \\ -\frac{\delta f_A(u) g_A(u)}{\alpha\gamma f_A^2(u) - \beta\delta g_A^2(u)} + \frac{\alpha f_A(u) g_A(u)}{\alpha\gamma f_A^2(u) - \beta\delta g_A^2(u)} \end{bmatrix}.$$

Let us notice that in the algebra of GFN's the right-hand side reduces to

$$\begin{bmatrix} \frac{f_A^2(u)}{f_A^2(u) + g_A^2(u)} + \frac{g_A^2(u)}{f_A^2(u) + g_A^2(u)} \\ -\frac{f_A(u) g_A(u)}{f_A^2(u) + g_A^2(u)} + \frac{f_A(u) g_A(u)}{f_A^2(u) + g_A^2(u)} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \text{(B.1.37)}$$

whereas for the algebra of MFN's we have

$$\begin{bmatrix} \frac{\frac{1}{2} f_A^2(u)}{\frac{1}{4}(f_A^2(u) - g_A^2(u))} - \frac{\frac{1}{2} g_A^2(u)}{\frac{1}{4}(f_A^2(u) - g_A^2(u))} \\ -\frac{\frac{1}{2} f_A(u) g_A(u)}{\frac{1}{4}(f_A^2(u) - g_A^2(u))} + \frac{\frac{1}{2} f_A(u) g_A(u)}{\frac{1}{4}(f_A^2(u) - g_A^2(u))} \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}. \quad \text{(B.1.38)}$$

Let us verify, whether elements $\mathbb{I} = (1, 0)$ and $\check{\mathbb{I}} = (2, 0)$ are unity in the algebra of CFN's and MFN's, respectively.

For CFN, we have

$$A \odot \mathbb{I} = (f_A,\ g_A) \odot (1,\ 0) =$$
$$= (f_A \cdot 1 - g_A \cdot 0,\ g_A \cdot 1 + f_A \cdot 0) = (f_A,\ g_A) = A \quad \text{(B.1.39)}$$

and

$$\mathbb{I} \odot A = (1 \cdot f_A - 0 \cdot g_A,\ 1 \cdot g_A + 0 \cdot f_A) = (f_A,\ g_A) = A, \quad \text{(B.1.40)}$$

so element $\mathbb{I} = (1, 0)$ is unity in the algebra of CFN's.

By running similar calculations for the MFN numbers, we obtain

$$\check{A} \odot \check{\mathbb{I}} = (f_A,\ g_A) \odot (2,\ 0) = = (\frac{1}{2}(f_A \cdot 2 + g_A \cdot 0),\ \frac{1}{2}(g_A \cdot 2 + f_A \cdot 0)) = (f_A,\ g_A) = \check{A} \tag{B.1.41}$$

and

$$\check{\mathbb{I}} \odot \check{A} = (2,\ 0) \odot (f_A,\ g_A) = = (\frac{1}{2}(2 \cdot f_A + 0 \cdot g_A),\ \frac{1}{2}(2 \cdot g_A + 0 \cdot f_A)) = (f_A,\ g_A) = \check{A}, \tag{B.1.42}$$

which indicates that element $\check{\mathbb{I}} = (2, 0)$ is unity in the algebra of MFN's.

Since the algebras of OFN's and CFN's have essentially different unity elements, inverse elements to each non-zero elements are different too.

In the OFN algebra, an inverse element to a given element $\vec{A}$, i.e. an element for which

$$\vec{A} \otimes \vec{A}^{-1} = \vec{A}^{-1} \otimes \vec{A} = \vec{\mathbb{I}}, \tag{B.1.43}$$

is given by the formula

$$\vec{A}^{-1} = \left(\frac{1}{f}, \frac{1}{g}\right). \tag{B.1.44}$$

This means that if any value of $f(u)$, $g(u)$ for some $u \in [0, 1]$ is zero, an inverse element to $\vec{A}$ does not to exist.

In the case of the algebra of CFN's, the null and unity elements are written as $\mathbb{O} = (0, 0)$ and $\mathbb{I} = (1, 0)$, respectively, where $0(u) = 0$, $\mathbb{I}(u) = 1$ for each $u \in [0, 1]$. Moreover, for any $A \neq (0, 0)$ we have

$$A^{-1} = \left(\frac{f}{f^2 + g^2}, -\frac{g}{f^2 + g^2}\right), \tag{B.1.45}$$

so

$$A \odot A^{-1} = \left(\frac{f^2}{f^2+g^2} + \frac{g^2}{f^2+g^2}, -\frac{fg}{f^2+g^2} + \frac{fg}{f^2+g^2}\right) = (1, 0) = \mathbb{I}. \tag{B.1.46}$$

As we can see, in the algebra of CFN's an inverse element does not exist only for $A = (0, 0)$. This fact is meaningful because Gelfand and Mazur used it as the main assumption of their theorem stating that each Banach complex algebra with the unit element, in which every non-zero element is inversible, is isometrically isomorphic with the algebra of complex numbers. Accordingly, there exists an isomorphic projection of these algebras onto each other and the distances between given elements are equal, which essentially fulfills the notion of "isometry". The last fact is important for selecting appropriate metrics enabling the parameters of a new mortality model to be determined.

B.2. Complex numbers and complex functions

The content of this section is based on [Sierpiński 1968, Chapter VI] and two monographs [Sakai 1971] and [Żelazko 1968].

Let us use $\mathcal{A}$ as a symbol representing any linear algebra for field of complex numbers $\mathbb{C}$.

When a new operation termed involution (hereafter denoted by *) is introduced into algebra $\mathcal{A}$ the algebra is called the Banach *–algebra. If that algebra's norm fulfills condition $\|A^*A\| = \|A\|^2$, the algebra is called the C^*–algebra. More Banach algebras of this kind can be created, but only some of them are interesting to us. Of special importance is the $C(\mathcal{T})$–algebra. Other specific cases of Banach algebras will be denoted by $\mathcal{A}_i$ $(i = 1, 2, ...)$.

B.2.1. The Banach C^*–algebra

Let $\mathcal{A}$ be a linear algebra on the field of complex numbers $\mathbb{C}$. This means that each element $A \in \mathcal{A}$ is assigned a real number $\|A\|$ called the norm of element A, which meets the following conditions

(1) $||A|| \geq 0$,

(2) $||A|| = 0$, if and only if $A = 0$,

(3) $||A + B|| \leq ||A|| + ||B||$,

(4) $||\alpha A|| = |\alpha| \cdot ||A||, \ \alpha \in \mathbb{C}$,

(5) Space $\mathcal{A}$ is a complete space in the norm $|| \cdot ||$.

Definition B.3. Projection $A \to A^*$ of algebra $\mathcal{A}$ onto itself is called involution, if

(i) $(A^*)^* = A$,
(ii) $(A + B)^* = A^* + B^*$,
(iii) $(AB)^* = B^*A^*$,
(iv) $(\alpha A)^* = \alpha A^*$, $\alpha \in \mathbb{C}$.

The Banach algebra with involution is called the Banach *-algebra.

Definition B.4. The Banach algebra $\mathcal{A}$ is termed the C^*–algebra, if it fulfills the following condition

$$||A^*A|| = ||A||^2 \quad \text{for each } A \in \mathcal{A}. \tag{B.2.1}$$

B.2.2. The Banach $C(\mathcal{T})$–algebra

Let $\mathcal{T}$ be a compact Hausdorff space. An example of this space can be interval $[0, 1]$ on the real line or the Cartesian product of n intervals $[0, 1]$. Let $C(\mathcal{T})$ denote the algebra of continuous and complex functions in $\mathcal{T}$.

Definition B.5. A norm in the Banach $C(\mathcal{T})$ space is defined as

$$||A|| = \max_{\tau \in \mathcal{T}} |A(\tau)| \tag{B.2.2}$$

and involution * is written as

$$A^*(\tau) = \bar{A}(\tau) \quad \text{for each } \tau \in \mathcal{T}. \tag{B.2.3}$$

Since

$$(A^*A)(\tau) = A^*(\tau)A(\tau) = \bar{A}(\tau)A(\tau) = |A(\tau)|^2 \tag{B.2.4}$$

and

$$|A^*A(\tau)| = |A(\tau)|^2, \tag{B.2.5}$$

therefore $C(\mathcal{T})$ is the C^*–algebra.

Let us consider a space of all pairs $\vec{A} = (f, g)$ of oriented fuzzy numbers. Each oriented fuzzy number $\vec{A} = (f, g)$ can be treated as a complex function and by using the following representation

$$A = f + ig \tag{B.2.6}$$

or

$$A(u) = f(u) + ig(u), \quad u \in [0, 1], \tag{B.2.7}$$

where $i = \sqrt{-1}$ is an imaginary unit.

The space $\mathcal{A}_1$ is a Banach space, if the norm of element $A \in \mathcal{A}_1$ is defined as

$$||A|| = \max_{u \in [0,1]} |A(u)|, \tag{B.2.8}$$

where $|A(u)|$ is a module of the complex number $A(u) = f(u) + ig(u)$, i.e.

$$|A(u)|^2 = f^2(u) + g^2(u), \quad u \in [0,1]. \tag{B.2.9}$$

Example B.2. Let $\vec{A}$ be an oriented fuzzy number corresponding to a symmetric triangular fuzzy number with central value a and spread s. An oriented fuzzy number $\vec{A}$ is written as

$$\vec{A} = (f, g), \tag{B.2.10}$$

where

$$\begin{aligned} f(u) &= a - s(1-u), \\ g(u)&a + s(1-u). \end{aligned} \tag{B.2.11}$$

Squaring of both equalities leads to

$$\begin{aligned} f^2(u) &= a^2 + (1-u)^2 s^2 - 2a(1-u)s, \quad u \in [0,1], \\ g^2(u) &= a^2 + (1-u)^2 s^2 + 2a(1-u)s, \quad u \in [0,1]. \end{aligned} \tag{B.2.12}$$

Thus, we have

$$f^2(u) + g^2(u) = 2a^2 + 2(1-u)^2 s^2. \tag{B.2.13}$$

Let us denote

$$F(u) = |A(u)|^2 = 2a^2 + 2(1-u)^2 s^2. \tag{B.2.14}$$

To determine the norm of element A we need to find the maximum of the function

$$F(u) = |A(u)|^2. \tag{B.2.15}$$

In this case, $F(u)$ is of the form

$$F(u) = 2a^2 + 2(1-u)^2 s^2, \quad u \in [0,1]. \tag{B.2.16}$$

Assume that $a = 1.7494$ and $s = 0.0411$. Hence, for $u \in [0, 1]$ we obtain

$$f(u) = 1.7494 - 0.0411(1 - u),$$
$$g(u) = 1.7494 + 0.0411(1 - u). \tag{B.2.17}$$

The complex functions $(f(u), g(u))$ and $F(u)$ are plotted in Figures B.1 and B.2, respectively.

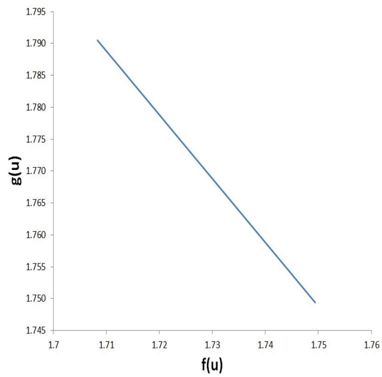

Figure B.1. Complex function $(f(u), g(u))$ for $a = 1.7494$, $s = 0.0411$
Source: Developed by the authors

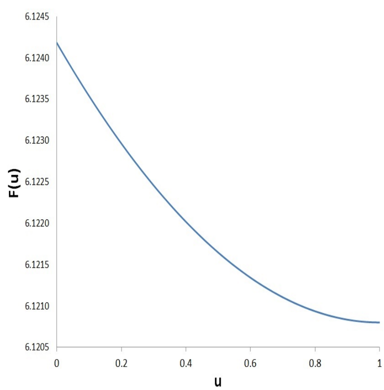

Figure B.2. Function $F(u)$ for $a = 1.7494$, $s = 0.0411$
Source: Developed by the authors

Example B.3. Let $\vec{A}$ be an oriented fuzzy number corresponding to a symmetric Gaussian fuzzy number with parameters a and s. The membership function $\mu_A(x)$ has the form

$$\mu_A(x) = e^{-\left(\frac{x-a}{s}\right)^2}. \tag{B.2.18}$$

By taking logarithms on both sides, we obtain

$$\ln \mu_A(x) = -\left(\frac{x-a}{s}\right)^2. \tag{B.2.19}$$

Hence,

$$\frac{x-a}{s} = \pm\sqrt{-\ln \mu_A(x)}. \tag{B.2.20}$$

Let us have $\mu_A(x) = u$ and

$$\begin{aligned} x(u) &= f(u) \quad \text{for } x < a, \\ x(u) &= g(u) \quad \text{for } x \geq a. \end{aligned} \tag{B.2.21}$$

Then

$$f(u) = a \pm s\sqrt{-\ln u}. \tag{B.2.22}$$

The definition of function $f(u)$ leads to condition $f(u) < a$, so

$$f(u) = a - s\sqrt{-\ln u} \quad \text{for } f(u) < a. \tag{B.2.23}$$

Analogously, for function g we have

$$g(u) = a + s\sqrt{-\ln u} \quad \text{for } g(u) \geq a. \tag{B.2.24}$$

By squaring both equalities and using the addition method, we arrive at

$$f^2(u) + g^2(u) = 2a^2 - 2s^2 \ln u. \tag{B.2.25}$$

Let us denote

$$F(u) = |A(u)|^2 = 2a^2 - 2s^2 \ln u. \tag{B.2.26}$$

To determine the norm of element A, the maximum of function $F(u)$ should be found. Let $a = 1.7494$, $s = 0.0411$. Figures B.3 and B.4 illustrate the complex function $(f(u), g(u))$ and function $F(u)$, respectively.

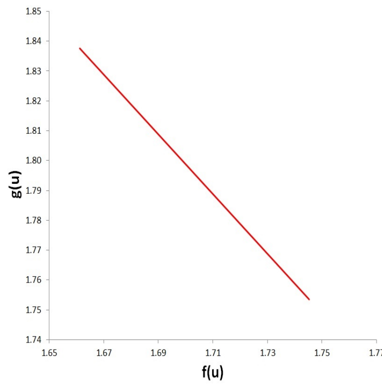

Figure B.3. Complex function $(f(u), g(u))$ for $a = 1.7494$, $s = 0.0411$
Source: Developed by the authors

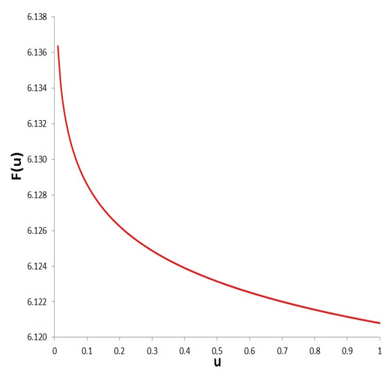

Figure B.4. Function $F(u)$ for $a = 1.7494$, $s = 0.0411$
Source: Developed by the authors

Example B.4. Let $\vec{A}$ be an oriented fuzzy number corresponding to a non-symmetric Gaussian fuzzy number. The membership function $\mu_A(x)$ is written as

$$\mu_A(x) = \begin{cases} e^{-\left(\frac{x-a}{s_L}\right)^2} & \text{for } x < a, \\ e^{-\left(\frac{x-a}{s_R}\right)^2} & \text{for } x \geq a. \end{cases} \tag{B.2.27}$$

Let us have

$$\mu_A(x) = u, \tag{B.2.28}$$

$$\begin{aligned} x(u) &= f(u) \quad \text{for } x < a, \\ x(u) &= g(u) \quad \text{for } x \geq a. \end{aligned} \tag{B.2.29}$$

Proceeding in the same way as in examples B.1 i B.2, we obtain

$$f(u) = a - s_L\sqrt{-\ln u} \quad \text{for } f(u) < a, \qquad \text{(B.2.30)}$$
$$g(u) = a + s_R\sqrt{-\ln u} \quad \text{for } f(u) \geq a$$

and

$$F(u) = 2a^2 + 2(s_R - s_L)\sqrt{-\ln u} - 2(s_R^2 + s_L^2)\ln u. \qquad \text{(B.2.31)}$$

Figures B.5 and B.6 show complex function $(f(u), g(u))$ and function $F(u)$ plotted for $a = 1.7494$, $s_L = 0.05$, $s_R = 0.03$.

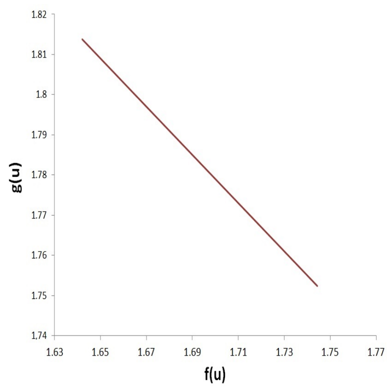

Figure B.5. Complex function $(f(u), g(u))$ for $a = 1.7494, s_L = 0.05, s_R = 0.03$
Source: Developed by the authors

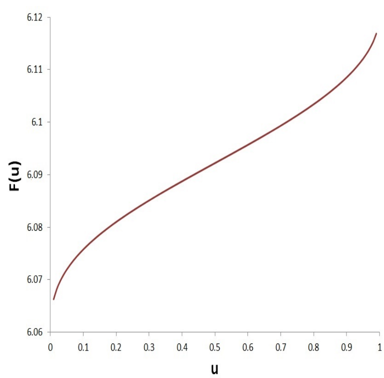

Figure B.6. Function $F(u)$ for $a = 1.7494, s_L = 0.05, s_R = 0.03$
Source: Developed by the authors

B.2.3. The quaternion space

Definition B.6. A quaternion is an ordered pair of complex numbers $\tilde{A} = (\phi, \psi)$ that fulfills the following axioms

(i) $(\phi, \psi) = (f, g) \;\Leftrightarrow\; \phi = f,\; \psi = g,$

(ii) $(\phi, \psi) + (f, g) = (\phi + f, \psi + g),$

(iii) $(\phi, \psi)(f, g) = (\phi f - \psi \bar{g}, \phi g + \psi \bar{f}),$

(iv) $\alpha(\phi, \psi) = (\alpha\phi, \alpha\psi),\; \alpha \in \mathbb{R}.$

Let $\mathbb{H}$ be a space of quaternions and let us assume that the space has a base of two elements 1 and j. Each quaternion $\tilde{A} \in \mathbb{H}$ can therefore be explicitly written as

$$\tilde{A} = \phi + j\psi. \tag{B.2.32}$$

A quaternion can also be presented as an algebraic sum, a complex matrix and a real matrix.

There are three types of operations that are defined in the quaternion space $\mathbb{H}$: adding, multiplying quaternions by complex numbers and multiplying quaternions.

From axiom (iv) it follows that for $\alpha \in \mathbb{R}$ and $\tilde{A} = \phi + j\psi$ there is

$$\alpha(\phi, \psi) = (\alpha\phi, \alpha\psi) = \alpha\phi + j\alpha\psi. \tag{B.2.33}$$

For $\phi, \psi \in \mathbb{C}$ of the form

$$\phi = a + ib,\;\; \psi = c + id, \tag{B.2.34}$$

an element $\tilde{A} \in \mathbb{H}$ can be expressed as

$$\tilde{A} = \phi + j\psi = (a + ib) + j(c + id) = a + ib + jc + ijd. \tag{B.2.35}$$

Let $k = ij$, then we obtain

$$\tilde{A} = a + ib + jc + kd. \tag{B.2.36}$$

Consequently, a pair of two complex numbers (ϕ, ψ) can be explicitly written as an algebraic sum, i.e.

$$\tilde{A} = a + ib + jc + kd. \tag{B.2.37}$$

Note that $\mathbb{H} = \mathbb{R}^4$ means that every quaternion can be explicitly determined by four numbers a, b, c, d (hence its name).

The multiplication of the elements of base i, j, k in the space $\mathbb{H}$ can be presented as

$$i^2 = j^2 = k^2 = ijk = -1 \tag{B.2.38}$$

$$ij = k,\ \ ji = k,\ \ jk = i,\ \ kj = i,\ \ ki = j,\ \ ik = j. \tag{B.2.39}$$

According to axiom (iii), the formula for multiplying two quaternions $\tilde{A} = (\phi, \psi)$ and $\tilde{B} = (f, g)$ is the following

$$\tilde{A}\tilde{B} = (\phi, \psi)(f, g) = (\phi f - \psi \bar{g},\ \phi g + \psi \bar{f}). \tag{B.2.40}$$

Let $\tilde{A} = (i, 0)$ and $\tilde{B} = (0, 1)$. Then

$$\tilde{A}\tilde{B} = (i, 0)(0, 1) = (i \cdot 0 - 0 \cdot \bar{1},\ i \cdot 1 + 0 \cdot \bar{0}) = (0, i) \tag{B.2.41}$$

$$\tilde{B}\tilde{A} = (0, 1)(i, 0) = (0 \cdot i - 1 \cdot \bar{0},\ 0 \cdot 0 + 1 \cdot \bar{i}) = (0, -i), \tag{B.2.42}$$

i.e. $\tilde{A}\tilde{B} \neq \tilde{B}\tilde{A}$, meaning that the multiplication of quaternions is not commutative.

Definition B.7. Let us define a conjugate quaternion for any given $\tilde{A} \in \mathbb{H}$

$$\tilde{A}^* = \phi - j\psi. \tag{B.2.43}$$

It follows from the above that the conjugation of quaternions is involution in $\mathbb{H}$, so it satisfies the following conditions

(i) $(\tilde{A}^*)^* = \tilde{A}$, for each $\tilde{A} \in \mathbb{H}$,

(ii) $(\tilde{A}\tilde{B})^* = \tilde{B}^*\tilde{A}^*$, for $\tilde{A}, \tilde{B} \in \mathbb{H}$,

(iii) $(\tilde{A} + \tilde{B})^* = \tilde{A}^* + \tilde{B}^*$,

(iv) for $\lambda \in \mathbb{R}$ and $\tilde{A} \in \mathbb{H}$, $(\lambda\tilde{A})^* = \lambda\tilde{A}^*$,

(v) for $\lambda \in \mathbb{C}$ and $\tilde{A} \in \mathbb{H}$, $(\lambda\tilde{A})^* = \tilde{A}^*\bar{\lambda}$.

The quaternion $\tilde{A} = (\phi, \psi)$, $\phi, \psi \in \mathbb{C}$ can be presented not only as an algebraic sum, but also as a complex matrix

$$\tilde{A} = \begin{bmatrix} \phi & \psi \\ -\bar{\psi} & \bar{\phi} \end{bmatrix}, \quad \phi, \psi \in \mathbb{C}. \tag{B.2.44}$$

That there is isomorphism between quaternions represented by a complex matrix and an algebraic sum can be demonstrated by introducing the following base in the space $\mathbb{R}^4$

$$\mathbf{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ \ \mathbf{i} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix},\ \ \mathbf{j} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},\ \ \mathbf{k} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}. \tag{B.2.45}$$

Starting out from the algebraic version of quaternion $\tilde{A}$ and utilizing the properties of operations on matrices and complex numbers, we successively obtain

$$\tilde{A} = a \cdot \mathbf{1} + b \cdot \mathbf{i} + c \cdot \mathbf{j} + d \cdot \mathbf{k} =$$

$$= a \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} + c \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} = \tag{B.2.46}$$

$$= \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} + \begin{bmatrix} bi & 0 \\ 0 & -bi \end{bmatrix} + \begin{bmatrix} 0 & c \\ -c & 0 \end{bmatrix} + \begin{bmatrix} 0 & di \\ di & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} a + bi & 0 \\ 0 & a - bi \end{bmatrix} + \begin{bmatrix} 0 & c + di \\ -c + di & 0 \end{bmatrix} = \tag{B.2.47}$$

$$= \begin{bmatrix} \phi & 0 \\ 0 & \bar{\phi} \end{bmatrix} + \begin{bmatrix} 0 & \psi \\ -\bar{\psi} & 0 \end{bmatrix} = \begin{bmatrix} \phi & \psi \\ -\bar{\psi} & \bar{\phi} \end{bmatrix}.$$

Conjugation for quaternions represented by the matrices is written as

$$\overline{\begin{bmatrix} \phi & \psi \\ -\bar{\psi} & \bar{\phi} \end{bmatrix}} = \begin{bmatrix} \bar{\phi} & -\psi \\ \bar{\psi} & \phi \end{bmatrix} \tag{B.2.48}$$

and for quaternions represented by an algebraic sum as

$$\bar{\tilde{A}} = \overline{a \cdot \mathbf{1} + b \cdot \mathbf{i} + c \cdot \mathbf{j} + d \cdot \mathbf{k}} = a \cdot \mathbf{1} - b \cdot \mathbf{i} - c \cdot \mathbf{j} - d \cdot \mathbf{k}, \tag{B.2.49}$$

whereas for a pair of complex numbers as

$$(\overline{\phi, \psi}) = (\bar{\phi}, -\psi). \tag{B.2.50}$$

Conjugation is also denoted by an asterisk; then $\bar{\tilde{A}}$ is replaced by $\tilde{A}^*$. A quaternion represented by a real matrix can also be written as

$$\begin{bmatrix} a & b & -d & -c \\ -b & a & -c & d \\ d & c & a & b \\ c & -d & -b & a \end{bmatrix}, \quad a, b, c, d \in \mathbb{R}. \tag{B.2.51}$$

Let us consider a general form of a complex matrix representing a quaternion $\tilde{A} \in \mathbb{H}$

$$\tilde{A} = \begin{bmatrix} \phi & \psi \\ -\bar{\psi} & \bar{\phi} \end{bmatrix}. \tag{B.2.52}$$

Definition B.8. A fuzzy quaternion is the quaternion

$$\tilde{A}(u) = (\phi(u), \bar{\phi}(u)), \quad u \in [0,1], \tag{B.2.53}$$

corresponding to a fuzzy number A.

The complex function $\phi(u)$ has two parts: the real part is represented by the central value of the fuzzy number and the imaginary part is represented by its membership function.

Definition B.9. A triangular, symmetric fuzzy quaternion is a quaternion $\tilde{A}$ corresponding to a triangular, symmetric fuzzy number $A = (a, s_A)$, where a is the central value and s_A is the spread of the fuzzy number (see e.g. Figure 4.2). The complex matrix of a triangular, symmetric fuzzy quaternion has the following form

$$\tilde{A} = \begin{bmatrix} a - is_A(1-u) & a + is_A(1-u) \\ -a + is_A(1-u) & a + is_A(1-u) \end{bmatrix}. \tag{B.2.54}$$

Definition B.10. The square root of the product of quaternion $\tilde{A}$ and its conjugation $\tilde{A}^*$ is called the norm of element $\tilde{A}$. It is denoted by $||\tilde{A}||$ and written as

$$||\tilde{A}||_{\mathbb{H}} = \sqrt{\tilde{A}\tilde{A}^*} = \sqrt{|\phi|^2 + |\psi|^2}. \tag{B.2.55}$$

Quaternion function $\tilde{A}(\cdot): \ [0,1] \to \mathbb{C}$ can be written as

$$\tilde{A}(u) = \begin{bmatrix} \phi(u) & \psi(u) \\ -\bar{\psi}(u) & \bar{\phi}(u) \end{bmatrix}, \quad \phi(u), \psi(u) \in \mathbb{C}, \ u \in [0,1]. \tag{B.2.56}$$

Based on the definition of the determinant of complex 2×2 square matrix, we can write

$$\det \tilde{A}(u) = \det \begin{bmatrix} \phi(u) & \psi(u) \\ -\bar{\psi}(u) & \bar{\phi}(u) \end{bmatrix} = \tag{B.2.57}$$

$$= \phi(u)\bar{\phi}(u) + \psi(u)\bar{\psi}(u) = |\phi(u)|^2 + |\psi(u)|^2 = \|\tilde{A}\|^2_{\mathbb{H}},$$

for the algebraic sum

$$\det \tilde{A} = \det(a \cdot \mathbf{1} + b \cdot \mathbf{i} + c \cdot \mathbf{j} + d \cdot \mathbf{k}) = a^2 + b^2 + c^2 + d^2, \tag{B.2.58}$$

and for a pair of complex numbers

$$\det \tilde{A}(u) = \det(\phi(u), \psi(u)) = |\phi(u)|^2 + |\psi(u)|^2. \tag{B.2.59}$$

The matrix of algebraic cofactors of matrix $\tilde{A}(u)$ has the form

$$D(u) = \begin{bmatrix} \bar{\phi}(u) & \bar{\psi}(u) \\ -\psi(u) & \phi(u) \end{bmatrix}, \tag{B.2.60}$$

and transposition turns $D(u)$ into

$$D^T(u) = \begin{bmatrix} \bar{\phi}(u) & -\psi(u) \\ \bar{\psi}(u) & \phi(u) \end{bmatrix}. \tag{B.2.61}$$

Since

$$\det \tilde{A}(u) = \|\tilde{A}(u)\|^2_{\mathbb{H}}, \tag{B.2.62}$$

the inverse matrix $\tilde{A}^{-1}(u)$ has the form

$$\tilde{A}^{-1}(u) = \frac{1}{\|\tilde{A}(u)\|^2_{\mathbb{H}}} \begin{bmatrix} \bar{\phi}(u) & -\psi(u) \\ \bar{\psi}(u) & \phi(u) \end{bmatrix} = \tag{B.2.63}$$

$$= \frac{1}{\|\tilde{A}(u)\|^2_{\mathbb{H}}} \overline{\begin{bmatrix} \phi(u) & \psi(u) \\ -\bar{\psi}(u) & \bar{\phi}(u) \end{bmatrix}} = \frac{1}{\|\tilde{A}(u)\|^2_{\mathbb{H}}} \overline{\tilde{A}(u)}.$$

With the notion of the norm and conjugation of a quaternion, it is possible to define an inverse element to a non-zero quaternion $\tilde{A}$ of the form

$$\tilde{A}^{-1}(u) = \frac{1}{||\tilde{A}||^2_{\mathbb{H}}} \overline{\tilde{A}(u)}, \quad u \in [0,1] \tag{B.2.64}$$

or

$$\tilde{A}^{-1}(u) = \frac{1}{||\tilde{A}||^2_{\mathbb{H}}} \tilde{A}^*(u), \quad u \in [0,1]. \tag{B.2.65}$$

Element $\tilde{A}$ can have only one inverse element. If two inverse elements $\tilde{B}_1$ and $\tilde{B}_2$ existed, we would have

$$\tilde{B}_1\tilde{A} = \tilde{A}\tilde{B}_1 = \mathbb{I} \tag{B.2.66}$$

and

$$\tilde{B}_2\tilde{A} = \tilde{A}\tilde{B}_2 = \mathbb{I}. \tag{B.2.67}$$

Taking product $\tilde{B}_2\tilde{B}_1\tilde{A}$, we have

$$\tilde{B}_2\tilde{B}_1\tilde{A} = \tilde{B}_2\left(\tilde{B}_1\tilde{A}\right) = \tilde{B}_2\mathbb{I} = \tilde{B}_2. \tag{B.2.68}$$

On the other hand, the same product $\tilde{B}_2\tilde{B}_1\tilde{A}$ can be written as

$$\tilde{B}_2\tilde{B}_1\tilde{A} = \tilde{B}_2\tilde{A}\tilde{B}_1 = (\tilde{B}_2\tilde{A})\tilde{B}_1 = \mathbb{I}\tilde{B}_1 = \tilde{B}_1. \tag{B.2.69}$$

From the above it follows that since there must be $\tilde{B}_2 = \tilde{B}_1$, the quaternion space $\mathbb{H}$ has the property that each non-zero quaternion $\tilde{A}$ has only one inverse element, which is also called an invertible or regular quaternion. The statement is the founding assumption of the Gelfand–Mazur theorem.

Because of the Gelfand–Mazur theorem, the Banach algebra in which each non-zero element is invertible is isometrically isomorphic with the Banach algebra of complex numbers. Hence, the quaternion algebra is such an algebra.

Let us demonstrate now that if quaternions $\tilde{A}$, $\tilde{B}$ are invertible, the product $\tilde{A}\tilde{B}$ is invertible too and

$$(\tilde{A}\tilde{B})^{-1} = \tilde{B}^{-1}\tilde{A}^{-1}. \tag{B.2.70}$$

Let $\tilde{A}(u)$ and $\tilde{B}(u)$, $u \in [0,1]$ be non-zero quaternions. Let us have

$$\tilde{C}^{-1}(u) = \left[\tilde{A}(u)\tilde{B}(u)\right]^{-1} = \frac{1}{\|\tilde{A}(u)\tilde{B}(u)\|^2}\left[\tilde{A}(u)\tilde{B}(u)\right]^*. \tag{B.2.71}$$

Since the quaternion algebra $\mathbb{H}$ is the C^*–algebra, we obtain

$$(\tilde{A}(u)\tilde{B}(u))^* = \tilde{B}^*(u)\tilde{A}^*(u). \tag{B.2.72}$$

From the definitions of a quaternion norm (B.2.55) and (B.2.72) it follows that

$$\begin{aligned}
&\|\tilde{A}(u)\tilde{B}(u)\|^2 = \Big[\tilde{A}(u)\tilde{B}(u)\Big]\Big[\tilde{A}(u)\tilde{B}(u)\Big]^* = \\
&= \tilde{A}(u)\tilde{B}(u)\tilde{B}^*(u)\tilde{A}^*(u) = \tilde{A}(u)\|\tilde{B}(u)\|^2\tilde{A}^*(u) = \\
&= \tilde{A}(u)\tilde{A}^*(u)\|\tilde{B}(u)\|^2 = \|\tilde{A}(u)\|^2\|\tilde{B}(u)\|^2.
\end{aligned} \tag{B.2.73}$$

Thus, we receive

$$\|\tilde{A}(u)\tilde{B}(u)\|^2 = \|\tilde{A}(u)\|^2\|\tilde{B}(u)\|^2. \tag{B.2.74}$$

Therefore, based on (B.2.72) and (B.2.74) we can rewrite (B.2.71) as

$$\begin{aligned}
&\tilde{C}^{-1}(u) = \frac{1}{\|\tilde{A}(u)\tilde{B}(u)\|^2}\Big[\tilde{A}(u)\tilde{B}(u)\Big]^* = \\
&= \frac{1}{\|\tilde{A}(u)\|^2\|\tilde{B}(u)\|^2}\tilde{B}^*(u)\tilde{A}^*(u) = \\
&= \frac{1}{\|\tilde{B}(u)\|^2}\tilde{B}^*(u)\frac{1}{\|\tilde{A}(u)\|^2}\tilde{A}^*(u) = \tilde{B}^{-1}(u)\tilde{A}^{-1}(u),
\end{aligned} \tag{B.2.75}$$

i.e.

$$\Big[\tilde{A}(u)\tilde{B}(u)\Big]^{-1} = \tilde{B}^{-1}(u)\tilde{A}^{-1}(u), \tag{B.2.76}$$

which was to be proved.

Definition B.11. The unitary space $U(\mathbb{H})$ is a vector space over the quaternion space $\mathbb{H}$ containing formula

$$(\tilde{A}, \tilde{B}) \to \langle \tilde{A}, \tilde{B} \rangle \in \mathbb{C}, \quad \tilde{A}, \tilde{B} \in \mathbb{H}, \tag{B.2.77}$$

fulfilling the conditions

1^o. $\langle \tilde{A}, \tilde{B} \rangle = \overline{\langle \tilde{B}, \tilde{A} \rangle}$,
2^o. $\langle \tilde{A} + \tilde{B}, \tilde{C} \rangle = \langle \tilde{A}, \tilde{C} \rangle + \langle \tilde{B}, \tilde{C} \rangle$,
3^o. $\langle a\tilde{A}, \tilde{B} \rangle = a\langle \tilde{A}, \tilde{B} \rangle, \quad a \in \mathbb{C}$,
4^o. $\langle \tilde{A}, \tilde{A} \rangle > 0$ for $\tilde{A} \neq \mathbb{O}$ and if $\langle \tilde{A}, \tilde{A} \rangle = 0 \;\Rightarrow\; \tilde{A} = \mathbb{O}$.

The complex number $\langle \tilde{A}, \tilde{B} \rangle$ is then called a scalar product of elements $\tilde{A}, \tilde{B} \in \mathbb{H}$ and conditions $1^o - 4^o$ are called the axioms of the scalar product ([Kolodziej 1970], p. 62).

Definition B.10 establishes the notion of a norm in space $\mathbb{H}$. According to [Mlak 1970], the norm expresses the scalar product by means of the following formula

$$\langle \tilde{A}, \tilde{B} \rangle = \frac{1}{4}\Big\{\|\tilde{A}+\tilde{B}\|^2 + i\|\tilde{A}+i\tilde{B}\|^2 - \|\tilde{A}-\tilde{B}\|^2 - i\|\tilde{A}-i\tilde{B}\|^2\Big\}, \quad \text{(B.2.78)}$$

where $\tilde{A}, \tilde{B} \in \mathbb{H}$ and $\mathbb{H}$ is the quaternion space.

To better understand how the norm of the elements of space $\mathbb{H}$ and the scalar product induced by the norm of space $\mathbb{H}$ are related to each other, we shall test the formula (B.2.78).

Using the scalar product axioms, we successively obtain

$$\begin{aligned} \|\tilde{A} + \tilde{B}\|^2 &= \langle \tilde{A} + \tilde{B}, \tilde{A} + \tilde{B} \rangle = \langle \tilde{A}, \tilde{A} + \tilde{B} \rangle + \langle \tilde{B}, \tilde{A} + \tilde{B} \rangle = \\ &= \langle \tilde{A}, \tilde{A} \rangle + \langle \tilde{A}, \tilde{B} \rangle + \langle \tilde{B}, \tilde{A} \rangle + \langle \tilde{B}, \tilde{B} \rangle = \qquad \text{(B.2.79)} \\ &= \|\tilde{A}\|^2 + \langle \tilde{A}, \tilde{B} \rangle + \langle \tilde{B}, \tilde{A} \rangle + \|\tilde{B}\|^2. \end{aligned}$$

Analogously, we have

$$\begin{aligned} \|\tilde{A} - \tilde{B}\|^2 &= \langle \tilde{A} - \tilde{B}, \tilde{A} - \tilde{B} \rangle = \langle \tilde{A}, \tilde{A} - \tilde{B} \rangle - \langle \tilde{B}, \tilde{A} - \tilde{B} \rangle = \\ &= \langle \tilde{A}, \tilde{A} \rangle - \langle \tilde{A}, \tilde{B} \rangle - \langle \tilde{B}, \tilde{A} \rangle + \langle \tilde{B}, \tilde{B} \rangle = \qquad \text{(B.2.80)} \\ &= \|\tilde{A}\|^2 - \langle \tilde{A}, \tilde{B} \rangle - \langle \tilde{B}, \tilde{A} \rangle + \|\tilde{B}\|^2. \end{aligned}$$

By subtracting (B.2.80) from (B.2.79), we get

$$\|\tilde{A}+\tilde{B}\|^2 - \|\tilde{A} - \tilde{B}\|^2 = 2\langle \tilde{A}, \tilde{B} \rangle + 2\langle \tilde{B}, \tilde{A} \rangle. \qquad \text{(B.2.81)}$$

The same procedure is applied to the other terms of (B.2.78). We have

$$
\begin{aligned}
\|\tilde{A}+i\tilde{B}\|^2 &= \langle\tilde{A}+i\tilde{B},\tilde{A}+i\tilde{B}\rangle = \langle\tilde{A},\tilde{A}+i\tilde{B}\rangle+i\langle\tilde{B},\tilde{A}+i\tilde{B}\rangle = \\
&= \langle\tilde{A},\tilde{A}\rangle + \langle\tilde{A},i\tilde{B}\rangle + i\langle\tilde{B},\tilde{A}\rangle + i\langle\tilde{B},i\tilde{B}\rangle = \\
&= \|\tilde{A}\|^2 + \bar{i}\langle\tilde{A},\tilde{B}\rangle + i\langle\tilde{B},\tilde{A}\rangle + i\bar{i}\|\tilde{B}\|^2.
\end{aligned}
\tag{B.2.82}
$$

By multiplying $\|\tilde{A}+i\tilde{B}\|$ by i, we obtain

$$
\begin{aligned}
i\|\tilde{A}+i\tilde{B}\|^2 &= i\|\tilde{A}\|^2 - ii\langle\tilde{A},\tilde{B}\rangle + ii\langle\tilde{B},\tilde{A}\rangle + i\|\tilde{B}\|^2 = \\
&= i\|\tilde{A}\|^2 + \langle\tilde{A},\tilde{B}\rangle - \langle\tilde{B},\tilde{A}\rangle + i\|\tilde{B}\|^2.
\end{aligned}
\tag{B.2.83}
$$

After substituting $-i$ for i, there is

$$
-i\|\tilde{A}-i\tilde{B}\|^2 = -i\|\tilde{A}\|^2+\langle\tilde{A},\tilde{B}\rangle-\langle\tilde{B},\tilde{A}\rangle-i\|\tilde{B}\|^2. \tag{B.2.84}
$$

From the sum of (B.2.83) and (B.2.84) we arrive at

$$
\begin{aligned}
i\|\tilde{A}+i\tilde{B}\|^2-i\|\tilde{A}-i\tilde{B}\|^2 &= i\|\tilde{A}\|^2+\langle\tilde{A},\tilde{B}\rangle-\langle\tilde{B},\tilde{A}\rangle+i\|\tilde{B}\|^2 + \\
&\quad -i\|\tilde{A}\|^2 + \langle\tilde{A},\tilde{B}\rangle - \langle\tilde{B},\tilde{A}\rangle - i\|\tilde{B}\|^2 = \\
&= 2\langle\tilde{A},\tilde{B}\rangle - 2\langle\tilde{B},\tilde{A}\rangle.
\end{aligned}
\tag{B.2.85}
$$

Finally, the sum of (B.2.81) and (B.2.85) leads to

$$
\begin{aligned}
\|\tilde{A}+\tilde{B}\|^2-\|\tilde{A}-\tilde{B}\|^2 &+ i\|\tilde{A}+i\tilde{B}\|^2-i\|\tilde{A}-i\tilde{B}\|^2 = \\
&= 2\langle\tilde{A},\tilde{B}\rangle + 2\langle\tilde{B},\tilde{A}\rangle + 2\langle\tilde{A},\tilde{B}\rangle - 2\langle\tilde{B},\tilde{A}\rangle = \\
&= 4\langle\tilde{A},\tilde{B}\rangle.
\end{aligned}
\tag{B.2.86}
$$

In this way, the formula for a scalar product in the unitary space is derived.

Index

Bibliography

Aalen O. O., (1978), *Nonparametric Inference for a Family of Counting Processes*, Annals of Statistics, 6, 701–726.

Akushevich I., Akushevich L., Manton K., Yashin A., (2003), *Stochastic process model of mortality and aging: Application to longitudinal data*, Nonlinear Phenomena in Complex Systems, 6, 515–523.

Alexiewicz A., (1969), *Analiza funkcjonalna*, PWN, Warszawa.

Alho J. M., Spencer B. D., (2005), *Statistical Demography and Forecasting*, Springer.

Antoch J., Huškova M., Janic A., Ledwina T., (2008), *Data driven rank test for the change point problem*, Metrika, Vol. 68 (1), 1–15.

Arnold B. C., (1983), *Pareto Distributions*, International Cooperative Publishing House, Fairland, Maryland.

Arnold B. C., Press S. J., (1989), *Bayesian Estimation and Prediction for Pareto Data*, Journal of the American Statistical Association, 84, 1079–1084.

Bain L. J., (1974), *Analysis for the Linear Failure-Rate Life-Testing Distribution*, Technometrics, 4, 551–559.

Balicki A., (2006), *Analiza przeżycia i tablice wymieralności*, PWE, Warszawa.

Bargiela A., Pedrycz W., Nakashima T., (2007), *Multiple regression with fuzzy data*, Fuzzy Sets and Systems, 158, 2169–2188.

Bayraktar E., Milevsky M. A., Promislow S. D., Young V. R., (2009),*Valuation of mortality risk via the instantaneous Sharpe ratio: Applications to life annuities*, Journal of Economics Dynamics and Control, 33, 676–691.

Benjamin B., Pollard J., (1993), *The analysis of mortality and other actuarial statistics*, The Institute of Actuaries, Oxford.

Bennett S., (1983), *Log-Logistic Regression models for Survival Data*, Applied Statistics, 32, 165–171.

Biffis E., (2005), *Affine processes for dynamic mortality and actuarial valuations*, Insurance: Mathematics and Economics, 37, 443–468.

Biffis E., Denuit M., (2006), *Lee–Carter goes risk-neutral. An application to the Italian annuity market*, Giornalle dell'Instituton Italiano degli Attuari, LXIX, 1–21.

Biffis E., Denuit M., Devolder P., (2010), *Stochastic mortality under measure changes*, Scandinavian Actuarial Journal, 4, 284–311.

Blom H.A.P., Lygeros J., (eds.) (2006), *Stochastic Hybrid Systems. Theory and Safety Critical Applications*, Springer, Berlin.

Booth H., (2006), *Demographic forecasting: 1980 to 2005 in review*, International Journal of Forecasting, 22, 547–581.

Boukas E. K., (2005), *Stochastic Hybrid Systems: Analysis and Design*, Birkhauser, Boston.

Bravo J. M., (2009), *Modelling mortality using multiple stochastic latent factors*, Proceedings of 7th International Workshop on Pension, Insurance and Saving, Paris, May 28–29, 1–15.

Bravo J. M., Braumann C. A., (2007), *The value of a random life: modelling survival probabilities in a stochastic environment*, Bulletin of the International Statistical Institute, LXII.

Brazauskas V., Serfling R., (2000), *Robust and Efficient Estimation of the Tail Index of a Single Parameter Pareto Distribution*, North American Actuarial Journal, 4, 12–27.

Brazauskas V., Serfling R., (2001), *Small Sample Performance of Robust Estimators of Tail Parameters For Pareto and Exponential models*, Journal of Statistical Computation and Simulation, 70, 1–19.

Brouhns N., Denuit M., Vermunt J. K., (2002), *A Poisson log-bilinear regression approach to the construction of projected lifetables*, Insurance: Mathematics and Economics, 31, 373–393.

Cairns A. J. G., Blake D., Dowd K., (2006), *A Two-Factor for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration*, The Journal of Risk and Insurance, Vol. 73(4), 687–718.

Cairns A. J. G., Blake D., Dowd K., (2008), *Modelling and management of mortality risk: a review*, Scandinavian Actuarial Journal, 2–3, 79–113.

Cairns A. J. G., Blake D., Dowd K., Coughlan G. D., Epstein D., (2011), *Mortality density forecasts: An analysis of six stochastic mortality models*, Insurance: Mathematics and Economics, 48, 355–367.

Cairns A. J. G., Blake D., Dowd K., Coughlan G., Epstein D., Khallaf-Allah M., (2008), *The plausibility of mortality density forecasts: an analysis of six stochastic mortality models*, Pensions Institute,Discussion Paper PI-0801.

Cairns A. J. G., Blake D., Dowd K., Coughlan G. D., Epstein D., Ong A., Balevich I., (2009), *A quantitative comparison of stochastic mortality models using data from England and Wales and the United States*, North American Actuarial Journal, 13, 1–35.

Chen Z., (2000), *A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function*, Statistics & Probability Letters, 49(2), 155–161.

Chiang C. L., (1968), *Introduction to Stochastic Processes in Biostatistics*, John Wiley, New York.

Coelho E., Magalhaes M. G., Bravo J. M., (2010), *Mortality projections in Portugal*, Proceeding of the Conference of European Statisticians, Working Session on Demographic Projections, Lisbon, Portugal, EUROSTAT (Series Forecasting demographic components: mortality, April 28–30, 1–11.

Cox D. R., (1972), *Regression models and Life-Tables*, Journal of the Royal Statistical Society, Ser. B, 34, 187–220.

Cox D. R., (1975), *Partial Likelihood*, Biometrika, 62, 269–276.

Cox J. C., Ingersoll J. E., Ross S. A., (1985a), *A theory of the term structure of interest rates*, Econometrica, 53, 385–407.

Cox J. C., Ingersoll J. E., Ross S. A., (1985b), *An intertemporal general equilibrium model of asset prices*, Econometrica, 53, 363–384.

Currie I. D., (2006), *Smoothing and forecasting mortality rates with p-splines*, Talk given at the Institute of Actuaries in 2006, available from http://www.ma.hw.ac.uk/ iain/research/talks.html

Dahl M., (2004), *Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts*, Insurance: Mathematics and Economics, 35, 113–136.

De Moivre A., *Annuities upon Lives*, Printed by W. P. and fold by Francis Fayram, London.

Debon A., Montes F., Puig F., (2008), *Modelling and forecasting mortality in Spain*, European Journal of Operational Research, 189, 624–637.

Denuit M., (2007), *Life Annuities with Stochastic Survival Probabilities: A Review*, Working Paper. Institut de Statistique & Institut des Sciences Actuarielles Universite Catholique de Louvain, Belgium.

Diamond P., (1988), *Fuzzy least-squares*, Information Sciences, 46(3), 141–157.

DiCiccio T., (1987), *Approximate Inference for the Generalized Gamma Distribution*, Technometrics, 29, 33–40.

Dubois D., Prade H., (1980), *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, London, Toronto, Sydney, San Francisco.

Evlanov L., Konstantinov V., (1976), *System with Random Parameters*, Nauka, Moskwa (In Russian).

Feigl P., Zelen M., (1965), *Estimation of the Exponential Survival Probabilities with Concomitant Information*, Biometrics, 21, 826–838.

Frątczak E., (1997), *Analiza historii zdarzeń – elementy teorii, wybrane przykłady zastosowań z wykorzystaniem pakietu TDA*, Szkoła Główna Handlowa, Warszawa.

Fung M. C., Peters G. W., Shevchenko P. V., (2015),*A State-Space Estimation of the Lee–Carter Mortality Model and Implications for Annuity Pricing*, Annals of Actuarial Science, vol. 11(2), 343–389.

Giacometti R., Ortobelli S., Bertocchi M., (2011), *A stochastic model for mortality rate on Italian Data*, Journal of Optimization Theory and Applications, 149, 216–228.

Girosi F., King G., (2006), *Demographic Forecasting*, Cambridge University Press, Cambridge.

Glasser M., (1967), *Exponential Survival with Covariance*, Journal of the American Statistical Association, 62, 501–568.

Gompertz B., (1825), *On the nature of the function expressive of the law of human mortality and on a new mode of determining life contingencies*, Philosophical Transactions of the Royal Society of London, Ser. A, CXV, 513–583.

Gupta R. C., Kannan N., Raychaudhuri A., (1997), *Analysis of lognormal survival data*, Mathematical Biosciences, 139(2), 103–115.

Graunt J. (1662), *Natural and political observations mentioned in a following index, and made upon the bills of mortality*, see London: John Martyn and James Allestry, 1973.

Haberman S., Renshaw A., (2008), *Mortality, longevity and experiments with the Lee–Carter model*, Lifetime Data Analysis, 14, 286–315.

Haberman S., Renshaw A., (2009), *On age-period-cohort parametric mortality rate projections*, Insurance: Mathematics and Economics, 45, 255–270.

Haberman S., Renshaw A., (2011), *A comparative study of parametric mortality projection models*, Insurance: Mathematics and Economics, 48, 3–55.

Hainaut D., (2012), *Multi dimensions Lee–Carter model with switching mortality processes*, Insurance: Mathematics and Economics, 47, 409–418.

Hainaut D., Devolder P., (2007), *Management of a pension fund under mortality and financial risks*, Insurance: Mathematics and economics, 41(1), 134–155.

Hainaut D., Devolder P., (2008), *Mortality modelling with Levy processes*, Insurance: Mathematics and Economics, 42, 409–418.

Halley E., (1693), *An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslaw with an attempt to ascertain the price of annuities upon lives*, Phil. Trans. Roy. Soc. London, 17, 596–610.

Hansen L. P. (1982), *Large sample properties of Generalized Method of Moments Estimators*, Econometrica, 50, 1029–1054.

Harper F. S., (1936), *An actuarial study of infant mortality*, Scandinavian Actuarial Journal, vol. 3–4, 234–270.

Harter H. L., (1967), *Maximum Likelihood Estimation of the Parameters of a Four-Parameter Generalized Gamma Population from Complete and Censored Samples*, Technometrics, 9, 159–165.

Hatzopoulos P., Haberman S., (2011), *A dynamic parametrization modeling for the age-period-cohort mortality*, Insurance: Mathematics and Economics, 49, 155–174.

Heligman L., Pollard J. H., (1980), *The age pattern of mortality*, Journal of the Institute of Actuaries, 107, 49–80.

Hespanha J.P., Tiwari A., (2006), *Hybrid Systems: Computations and Control*, Springer, Berlin.

Hjorth U., (1980), *A Reliability Distribution with Increasing, Decreasing, Constant and Bathtub-Shaped Failure Rates*, Technometrics, 22, 99–107.

Hong H. D., (2001), *Shape preserving multiplications of fuzzy numbers*, Fuzzy Sets and Systems, 123, 81–84.

Hong H. D., Song J. K., Do H. Y., (2001), *Fuzzy least-squares regression analysis using shape preserving operations*, Information Sciences, 138, 185–193.

Huffer F. W., McKeague I. W., (1991), *Weighted Least Squares Estimation for Aalen's Additive Risk Model*, Journal of the American Statistical Association, 86, 114–129.

Ibrahim R., (1985), *Parametric Random Vibration*, Research Studies Press, Letchworth, United Kingdom.

Itkis Y., (1983), *Dynamic switching of type I/ type II structures in tracking servosystems*, IEEE Transactions on Automatic Control, 28, 531–534.

Janic-Wróblewska A., Ledwina T., (2000), *Data driven rank test for two-sample problem*, Scandinavian Journal of Statistics, 27, 281–297.

Janssen J., Skiadas C. H., (1995), *Dynamic modelling of life table data*, Applied Stochastic models and Data Analysis, 11, 35–49.

Kannisto V., (1994), *Development of oldest-old mortality, 1950-1990: evidence from 28 developed countries*, Odense University Press, Odense, Denmark.

Kaplan E. L., Meier P., (1958), *Nonparametric Estimation from Incomplete Observations*, Journal of the American Statistical Association, 53, 457–481.

Kazakov I. E., Artemiev B. M., (1980), *Optimization of Dynamic Systems with Random Structure*, Nauka, Moscow (w j. rosyjskim).

Keilman N., (1990), *Uncertainty in population forecasting: issues, backgrounds, analyses, recommendations*, Swets & Zeitlinger, Amsterdam.

Keilman N. (ed.), (2005), *Perspectives on Mortality Forecasting. Probabilistic models*, Försäkringskassan, Swedish Social Insurance Agency.

Kędelski M., Paradysz J., (2006), *Demografia*, Wyd. AE, Poznań.

Kodlin D., (1967), *A New Response Time Distribution*, Biometrics, 2, 227–239.

Koissi M. C., Shapiro A. F., (2006), *Fuzzy formulation of the Lee–Carter model for mortality forecasting*, Insurance: Mathematics and Economics, 39, 287–309.

Kołodziej W., (1970), *Wybrane rozdziały analizy matematycznej*, Biblioteka Matematyczna, t. 36, PWN, Warszawa.

Kosiński W., Prokopowicz P., (2004), *Algebra liczb rozmytych*, Matematyka Stosowana, 46, 37–63.

Kosiński W., Prokopowicz P., Ślęzak D., (2003), *Ordered fuzzy numbers*, Bulletin of the Polish Academy of Sciences – Mathematics, 51, 327–338.

Krane S. A., (1963), *Analysis of Survival Data by Regression Techniques*, Technometrics, 5, 161–174.

Ladde G. S., Wu L., (2009), *Development of modified Geometric Brownian Motion model by using stock price data and basic statistics*, Nonlinear Analysis, 71, 1203–1208.

Lambert J. H., (1776), *Dottrina degli azzardi*, Gaeta and Fontana, Milan.

Lee R. D., Carter L., (1992), *Modeling and forecasting the time series of U.S. mortality*, Journal of the American Statistical Association, 87, 659–671.

Lee R. D., Miller T., (2001), *Evaluating the performance of the Lee–Carter method for forecasting mortality*, Demography, 38, 537–549.

Liberzon D., (2003), *Switchings in Systems and Control*, Birkhauser, Boston, Basel, Berlin.

Lin D. Y., (1991), *Goodness-of-Fit Analysis for the Cox Regression Model Based on a Class of Parameter Estimators*, Journal of the American Statistical Association, 86, 725–728.

Lin H., Antsaklis P. J., (2009), *Stability and stabilizability of switched linear systems: a survey of recent results*, IEEE Transactions on Automatic Control, 54, 308–322.

Lipcer R., Sziriajew A., (1981), *Statystyka procesów stochastycznych*, PWN, Warszawa.

Loparo K. A., Aslanis J. T., Hajek O., (1987), *Analysis of switching linear systems in the plane, part 1: local behavior of trajectories and local cycle geometry*, Journal of Optimization Theory and Applications, 52, 365–394.

Luciano E., Spreeuw J., Vigna E., (2008), *Modelling stochastic mortality for dependents lives*, Insurance: Mathematics and Economics, 43, 234–244.

Luciano E., Vigna E., (2008), M*ortality risk via affine stochastic intensities: calibration and empirical relevance*, Belgian Actuarial Biulletin, vol. 8(1), 5–16.

Makeham W. M., (1867), *On the Law of Mortality and the Construction of Annuity Tables*, J. Inst. Actuaries and Assur. Mag., 8, 301–310.

Malik H. J., (1970), *Estimation of the Parameters of the Pareto Distribution*, Metrika, 15, 126–132.

Mao X., Yuan C., (2006), *Stochastic Differential Equations with Markovian Switching*, Imperial College Press. London.

Milevsky M. A., Promislow S. D., (2001), *Mortality derivatives and the option annuities*, Insurance: Mathematics and Economics, 29, 299–318.

Mlak W., (1970), *Wstęp do teorii przestrzeni Hilberta*, PWN, Warszawa.

Nelson W., (1969), *Hazard Plotting for Incomplete Failure Data*, Journal of Quality Technology, 1, 27–52.

Nielsen B., Nielsen J. P., (2010), *Identification and Forecasting in the Lee–Carter Model*, Economics Series Working Papers No 2010-W07, University of Oxford, Department of Economics, available on-line: www.nuffield.ox.ac.uk

Okólski M. (ed.), (1990), *Determinanty umieralności w świetle teorii i badań empirycznych*, Wyd. SGPiS, Warszawa.

Okólski M., (2003), *Kryzys zdrowotny w Polsce*, Polityka Społeczna, nr 1.

Oksendal B., (2003), *Stochastic Differential Equations: An Introduction with applications*, 6th ed., Springer Verlag, Heidelberg–New York.

Opperman L., (1870), *Insurance record 1870*, 42–43, 46–47.

Ostasiewicz S., (2011), *Aproksymacja czasu trwania życia w populacjach niejednorodnych*, Zeszyty Naukowe WSOWL, 4 (162), 342–358.

Papanicolau G., Kohler W.,(1974), *Asymptotic theory of mixing stochastic ordinary differential equation*, Commun. Pure Applied Math., 27, 641–668.

Perks W. (1932), *On some experiments in the graduation of mortality statistics*, Journal of the Institute of Actuaries, 63, 12–40.

Pettitt A. N., (1984), *Proportional Odds Model for Survival Data and Estimates Using Ranks*, Applied Statistics, 33, 169–175.

Pitacco E., (2004), *Survival models in a dynamic context: a survey*, Insurance: Mathematics and Economics, 35, 279–298.

Plat R., (2009), *On stochastic mortality modeling*, Insurance: Mathematics and Economics, 45, 393–404.

Polovko A. M., (1968), *Fundamentals of Reliability Theory*, Academic Press, New York.

Prentice R. L., (1974), *A Log-Gamma Model and Its Maximum Likelihood Estimation*, Biometrika, 61, 539–544.

Preston S. H., Heuveline P., Guillot M., (2001), *Demography. Measuring and Modeling Population Processes*, Blackwell Publishing Ltd., Malden–Oxford–Carlton.

Proschan F., (1963), *Theoretical Explanation of Observed Decreasing Failure Rate*, Technometrics, 5, 375–385.

Pugachev V., Sinitsyn I., (1987), *Stochastic Differential Systems: Analysis and Filtering*, John Willey and Sons, Chichester.

Quandt R. E., (1966), *Old and New Methods of Estimation and the Pareto Distribution*, Metrika, 10, 55–82.

Renshaw A., Haberman S., (2003a), *Lee–Carter mortality forecasting with age-specific enhancement*, Insurance: Mathematics and Economics, 33, 255–272.

Renshaw A., Haberman S., (2003b), *Lee–Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections*, Applied Statistics, 52, 119–137.

Renshaw A., Haberman S., (2003c), *On the forecasting of mortality reduction factors*, Insurance: Mathematics and Economics, 32, 379–401.

Renshaw A., Haberman S., (2006), *A cohort-based extension to the Lee–Carter model for mortality reduction factor*, Insurance: Mathematics and Economics, 38, 556–570.

Renshaw A., Haberman S. (2008), *On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee–Carter modelling*, Insurance: Mathematics and Economics, 42(2), 797–816.

Rossa A., Socha L. Szymański A., (2011), *Analiza i modelowanie umieralności w ujęciu dynamicznym*, Wyd. UŁ, Łódź.

Rossa A., Socha L., (2013), *Proposition of Hybrid Stochastic Lee–Carter Mortality Model*, Advances in Methodology and Statistics, 10(1), 1–16.

Rosset E., (1979), *Trwanie życia ludzkiego*, PWN, Warszawa.

Russo V., Giacometti R., Ortobelli S., Rachev S., Fabozzi F. J., (2011), *Calibrating affine stochastic mortality models using term assurance premiums*, Insurance: Mathematics and Economics, 49, 53–60.

Sakai S., (1971), *C*-algebras and W*-algebras*, Springer Verlag, Berlin–Heidelberg–New York.

Schrager D., (2006), *Affine stochastic mortality*, Insurance: Mathematics and Economics, 38, 81–97.

Shi P., Li F., (2015), *A survey on Markovian jump systems: modeling and design*, Int. J. Cont. Autom. Syst., 13, 1–16.

Sierpiński W., (1968), *Arytmetyka teoretyczna*, PWN, Warszawa.

Sobczyk K., (1996), *Stochastyczne równania różniczkowe*, WNT, Warszawa.

Socha L., (1993), *Równania momentów w stochastycznych układach dynamicznych*, PWN, Warszawa.

Socha L., (2008), *Linearization Methods for Stochastic Dynamic Systems*, Springer, Berlin.

Stacy E. W., (1962), *A Generalization of the Gamma Distribution*, The Annals of Mathematical Statistics, 33, 1187–1192.

Stacy E. W., Mihram G. A., (1965), *Parameter Estimation for a Generalized Gamma Distribution*, Technometrics, 7, 349–358.

Steffensen, J. (1930), *Om sandsynligheden for at afkommet uddor*, Matematisk tidsskrift. B, 19–23.

Steward G. W. (1993), *On the Early History of the Singular Value Decomposition*, SIAM Review, 35(4), 551–566.

Stratonovich R., (1960), *A new form of representation of stochastic integral and equations*, SIAM Journal on Control and Optimization, 4, 362–371.

Szymański A., Rossa A. (2014), *Fuzzy mortality model based on Banach algebra*, International Journal of Intelligent Technologies and Applied Statistics, 7(3), 241–265.

Tabeau E., Berg Jehs A., Heathcote Ch. (eds.), (2001), *Forecasting Mortality in Developed Countries. Insights from Statistical, Demographic and Epidemiological Perspective*, Kluwer Academic Publishers, London.

Thatcher A. R., Kannisto V., Vaupel J. W., (1998), *The force of mortality at ages 80 to 120*, Monographs on Population Aging, vol. 5, Aging Research Center, Centre for Health and Social Policy, Odense University.

Teel A., Subbaraman A., Sferlazza A., (2014), *Stability analysis for stochastic hybrid systems: A survey*, Automatica, 50, 2435–2456.

Thiele T. N., Sprague T., (1871), *On a mathematical formula to express the rate of mortality throughout the whole of life tested by a series of observations made use of by the danish life insurance company of 1871*, Journal of the Institute of Actuaries, 16(5), 313–329.

Wang Ch.-W., Huang H.-Ch, Liu I-Ch., (2011), *A Quantitative Comparison of the Lee–Carter Model under Different Types of Non-Gaussian Innovations*, The Geneva Papers on Risk and Insurance - Issues and Practice, vol. 36(4), 675–696.

Weibull W., (1939), *Statistical Theory of the Strength of Materials*, Ingenioor Vetenskps Akademiens Handlingar, 151, 1–45.

Wilmoth J. R., (1993), *Computational methods for fitting and extrapolating the Lee–Carter model of mortality change.* Technical report, University of California, in: http://www.demog.berkeley.edu/ jrw/Papers/LCtech.pdf

Wilmoth J. R., Horiuchi S., (1999), *Rectangularization revised: variability of age at death within human populations*, Demography, 36(4), 475–495.

Winkelbauer A., (2014), *Moments and Absolute Moments of the Normal Distribution*, Working Paper, Cornell University Library, available on-line: https://arxiv.org/abs/1209.4340

Wittstein J., Bumsted D. (1883), *The mathematical law of mortality*, Journal of the Institute of Actuaries and Assurance Magazine, 24(3), 153–173.

Wong E., (1971), *Stochastic Processes in Information Theory and Dynamical Systems*, McGraw–Hill, New York.

Wu S. J., (2003), *Estimation for the Two-Parameter Pareto Distribution under Progressive Censoring with Uniform Removals*, Journal of Statistical Computation and Simulation, 73, 125–134.

Wunsch G., Mouchart M., Duchene J. (eds.), 2002, *The Life Table: Modelling Survival and Death*, Springer-Science+Business Media B.V.

Van Berkum F., Antonio K., Vellekoop M., (2013), *Structural changes in mortality rates with an application to Dutch and Belgian data*, Workin Paper, KU Leuven.

Van der Maen W. J., (1943), *Het berekenen van sterftekansen*, Verz.-Arch., 24, 281–300.

Vasiček O., (1977), *An equilibrium characterization of the term structure*, Journal of Financial Economics, 5, 177–188.

Vaupel J. W., Zhang Z., van Raalte A. A., *Life expectancy and disparity: an international comparison of life table data*, in: BMJ Open 2011;1:e000128, doi:10.1136/bmjopen-2011-000128.

Yashin A. I., Arbeev K. G., Akushevich I., Kulminski A., Akushevich L., Ukraintseva S. V., (2207), *Stochastic model for analysis of longitudinal data on aging and mortality*, Mathematical Biosciences, vol. 208(2), 538–551.

Yin G., Zhang Q., Yang H., Yin K., (2002), *A class of hybrid market models: simulation, identification and estimation*, Proceedings of the American Control Conference, Anchorage, May 8–10, 2571–2576.

Yin G., Zhang Q., Yang H., Yin K., (2003), *Constrained stochastic estimation algorithms for a class of hybrid stock market models*, Journal of Optimization Theory and Applications, 118, 157–182.

Yin G. G., Zhu C., (2010), *Hybrid Switching Diffusions: Properties and Applications*, Springer Science + Bussiness Media, New York.

Zadeh L., (1965), *Fuzzy Sets*, Information and Control, 8, 338–353.

Zimmermann H.-J., (1996), *Fuzzy Set Theory – and its Applications*, fourth edition, Springer Science + Bussiness Media, New York.

Zippin C., Armitage P., (1966), *Use of Concomitant Variables and Incomplete Survival Information in the Estimation of an Exponential Survival Parameter*, Biometrics, 22, 665–672.

Żelazko W., (1968), *Algebry Banacha*, Biblioteka Matematyczna, t. 32, PWN, Warszawa.