Stochastic Mortality Modelling with Cointegrated Vector Autoregressive Processes and Characterizations of Logistic-type Hazard Rate Distributions

Von der Fakultät für Mathematik und Physik
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades

Doktor der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

von

Dipl. -Phys. Thomas Salfeld

2019

Referent: Prof. Dr. S. Weber
Korreferent: Prof. Dr. M. Scherer

Tag der Promotion: 15. Februar 2018

Stochastic Mortality Modelling with Cointegrated Vector Autoregressive Processes and Characterizations of Logistic-type Hazard Rate Distributions

Dipl.-Phys. Thomas Salfeld
Leibniz Universität Hannover

Abstract

This thesis is devoted to stochastic mortality modelling. The first part considers the popular family of GAPC models and identifies several conceptual difficulties of most well-established models. The GAPC models are embedded in the framework of generalized linear models. However, the vast majority of the literature only considers the canonical link function and by that omits an important modelling factor. In our study, we also incorporate a non-canonical link function and demonstrate its advantages on the fitting performance. While the first part focuses on the static component of the modelling approach, where the main objective is to identify the influencing factors that drive the mortality structure, the second part is devoted to the dynamical part of the modelling approach. For the proposed model we identify appropriate multivariate stochastic processes for the dynamics of the involved stochastic factors. We study cointegration relations between the individual components and compare the forecasting performance with the common GAPC approach. The last part of this thesis can be considered independently of the previous content. There, we provide an extensive characterization of the lifetime distribution which is induced by logistic-type hazard rates of the proposed Kannisto model. Furthermore, we reveal multiple connections to other well-known lifetime distributions.

Keywords: stochastic mortality modelling, cointegrated vector autoregressive processes, logistic hazard rate functions

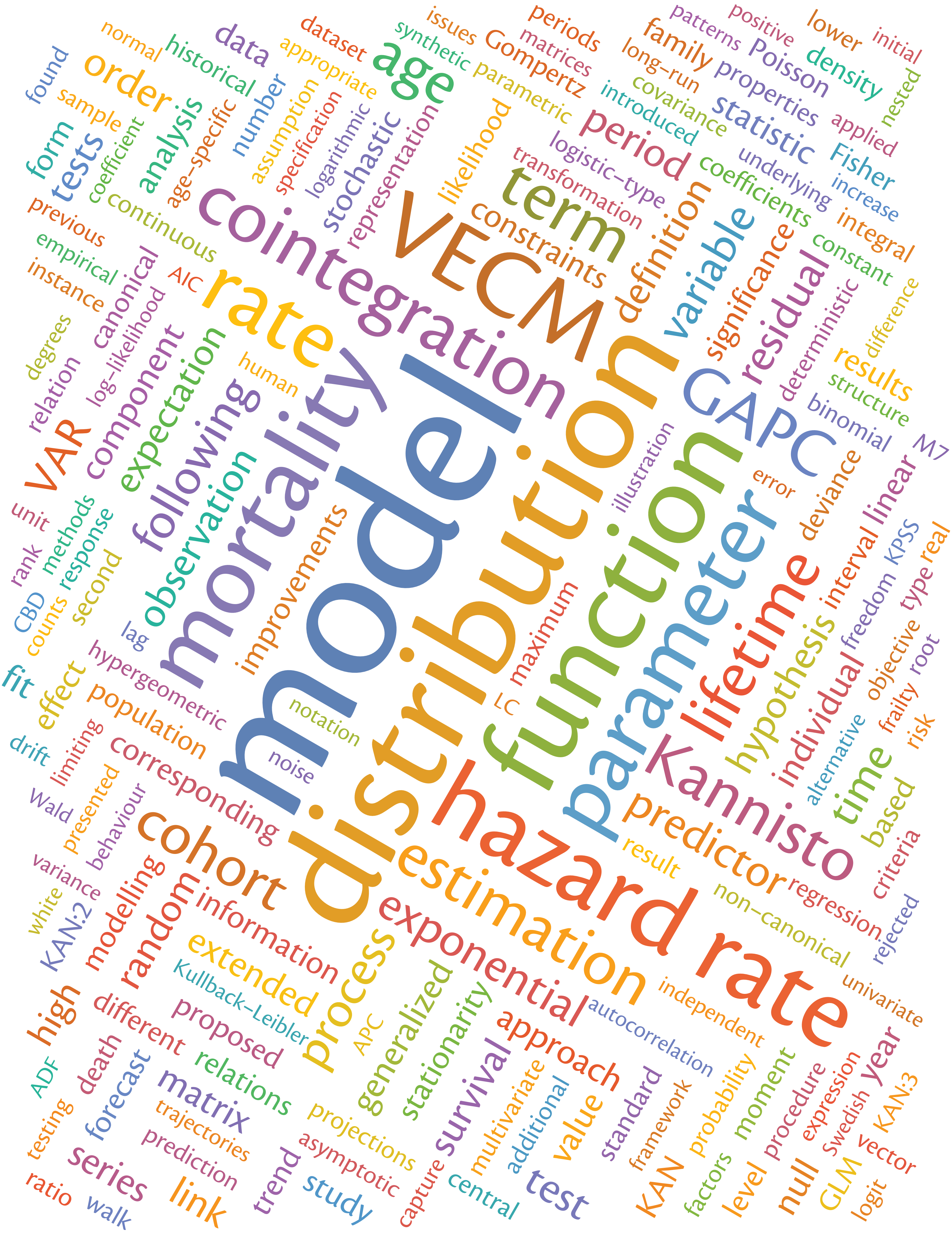
Stochastic Mortality Modelling with Cointegrated Vector Autoregressive Processes and Characterizations of Logistic-type Hazard Rate Distributions

Dipl.-Phys. Thomas Salfeld
Leibniz Universität Hannover

Kurzfassung

Die vorliegende Dissertation behandelt stochastische Mortalitätsmodelle. Im ersten Teil werden zunächst die Familie der GAPC-Modelle vorgestellt und anschließend einige konzeptionelle Probleme der in der Literatur und Praxis etablierten Modelle herausgestellt. Die GAPC-Modelle sind im Rahmen von generalisierten linearen Modellen formuliert. Der Großteil der Literatur vernachlässigt jedoch einen wichtigen Freiheitsgrad der Modellierung, indem nur kanonische Link-Funktionen betrachtet werden. Unsere Analyse schließt eine nicht kanonische Link-Funktion ein, welche in vielen Fällen zu einer verbesserten Güte der Regression führt. Während sich der erste Teil der Arbeit mit der statischen Komponente des Modellierungsansatzes beschäftigt, in dem der Fokus darauf liegt, die Haupteinflussfaktoren für die Struktur der Mortalität zu identifizieren, beschäftigt sich der zweite Teil der Arbeit mit deren dynamischen Entwicklungen. Für das von uns vorgeschlagene Modell analysieren wir, welche multivariaten Prozesse sich eignen, um die Charakteristiken der Dynamik abzubilden. Dazu wird eine Kointegrationsanalyse durchgeführt, um eventuelle langfristige Beziehungen zwischen den einzelnen Faktoren aufzudecken. Der abschließende Teil der Arbeit kann unabhängig von dem Vorhergehenden betrachtet werden. In diesem dritten Teil wird eine umfangreiche Charakterisierung der Lebenszeit-Verteilung durchgeführt, die durch logistische Hazard-Raten des vorgeschlagenen Kannisto Models impliziert wird. Darüber hinaus werden mehrere Verbindungen zu weiteren bekannten Verteilungen hergestellt.

Schlagwörter: stochastische Mortalitätsmodelle, kointegrierte vektorautoregressive Prozesse, logistische Hazard-Raten-Modelle


model
distribution
function
hazard rate
mortality
cointegration
VECM
rate
parameter
lifetime
Kannisto
estimation
exponential
GAPC
term
definition
variable
residual
significance
predictor
hypothesis
individual
time
cohort
corresponding
following
observation
expectation
component
improvements
VAR
age
period
statistic
family
Gompertz
Poisson
properties
coefficients
underlying
Fisher
integral
constant
deterministic
results
structure
binomial
interval
linear
deviance
freedom
KPSS
type
real
objective
frailty
root
risk
based
criteria
alternative
regression
non-canonical
rejected
univariate
independent
autocorrelation
approach
survival
stationarity
generalized
process
extended
information
random
modelling
proposed
relations
different
matrix
forecast
death
high
series
link
trend
study
asymptotic
projections
trajectories
prediction
capture
central
test
value
standard
additional
multivariate
framework
KAN
probability
moment
factors
level
procedure
null
expression
Swedish
year
KAN:3
vector
GLM
logit
transformation
logistic-type
constraints
likelihood
representation
stochastic
logarithmic
specification
assumption
number
age-specific
analysis
coefficient
tests
form
found
sample
order
normal
historical
data
appropriate
dataset
synthetic
issues
parametric
matrices
periods
long-run
covariance
introduced
patterns
positive
lower
initial
density
nested
applied
increase
difference
M7
error
illustration
maximum
LC
notation
noise
human
AIC
previous
continuous
empirical
instance
canonical
log-likelihood
degrees
relation
unit
rank
methods
response
second
CBD
counts
lag
hypergeometric
fit
effect
population
drift
limiting
Wald
presented
behaviour
variance
white
KAN:2
Kullback-Leibler
APC
ADF
testing
ratio
walk
result

Contents

Abbreviations

Notation	Description
AR(p)	autoregressive process of order p
AC	autocorrelation
ACF	autocorrelation function
ADF	augmented Dickey-Fuller (test)
AIC	Akaike information criterion
ARCH	autoregressive conditional heteroscedasticity
ARIMA(p,d,q)	autoregressive integrated moving-average of order (p,d,q)
ARMA(p,q)	autoregressive moving-average of order (p,q)
BIC	Bayesian information criterion
D.F.	degrees of freedom
DGP	data generation process
FPE	final prediction error (criterion)
GAPC	Generalized Age-Period-Cohort (stochastic mortality models)
GLM	generalized linear model
HMD	Human Mortality Database
HQ	Hannan-Quinn (criterion)
IRLS	iteratively reweighted least squares algorithm
KPSS	Kwiatkowski-Phillips-Schmidt-Shin (test)
LJB	Lomnicki-Jarque-Bera (test)
LM	Lagrange multiplier (test)
LR	likelihood ratio (test)
MLE	maximum likelihood estimation
MSE	mean squared error (matrix)
NR	Newton-Raphson method
OLS	ordinary least squares
P.I.	prediction interval
RW	random walk
SC	Schwarz criterion
s.e.	standard error

Notation	Description
VAR(p)	vector autoregressive process of order p
VECM(p)	vector error correction model of order p
WLS	weighted least squares

Symbols

Notation	Description
L	lag operator
Δ	differencing operator
$I(d)$	integrated of order d
i.i.d..	independently and identically distributed
$\xrightarrow{\mathbb{P}}$	converges in probability to
$\xrightarrow{a.s.}$	converges almost surely to
$\xrightarrow{\mathcal{D}}$	converges in distribution to
$\mathbb{1}(\cdot)$	indicator function
plim	probability limit
$\mathbb{E}$	expectation
$\mathbb{P}$	probability
$\mathbb{V}$	variance
Cov	covariance matrix
Corr	correlation matrix
H_0	null hypothesis
H_1	alternative hypothesis
p-value	tail probability of a statistic
${}_pF_q$	generalized hypergeometric function
$\Gamma(z)$	gamma function
$(a)_n$	Pochhammer symbol
$B_z(a,b)$	incomplete beta function
$\psi(z)$	polygamma function

Vector/Matrix Operations and Related Symbols

Notation	Description
$\mathbb{1}_m$	$(m \times m)$ identity matrix
M'	transpose of M
M^{-1}	inverse of M
$M_\perp$	orthogonal complement of M
$M^{1/2}$	square root of M
M^k	k-th power of M
$\otimes$	Kronecker product
$\det(M), \|M\|$	determinant of M
$\mathrm{rk}(M)$	rank of M
$\mathrm{tr}(M)$	trace of M
vech	half-vectorization operator (column stacking operator of the lower triangular part of symmetric matrices)
$\boldsymbol{\alpha}$	loading matrix of VECM
$\boldsymbol{\beta}$	cointegration matrix
$\boldsymbol{\Pi}$	$:= \boldsymbol{\alpha}\boldsymbol{\beta}'$
$\boldsymbol{A}_i$	coefficient matrix of VAR
$\boldsymbol{D}_i$	long-run form coefficient matrix of VECM
$\boldsymbol{\Gamma}_i$	short-run form coefficient matrix of VECM (transitory form)
u_t, ε_t	multidimensional white noise process
y_t	multidimensional stochastic process
$\bar{y}$	$:= \frac{1}{T}\sum_{t=1}^{T} y_t$ sample mean (vector)
$y_t(h)$	h-step forecast of y_{t+h} at origin t
Σ_u	$:= \mathbb{E}[u_t u_t'] = \mathrm{Cov}[u_t]$, white noise covariance matrix
Σ_y	$:= \mathbb{E}\left[(y_t - \mu)(y_t - \mu)'\right] = \mathrm{Cov}(y_t)$, covariance matrix of a stationary process y_t
$\Omega(h)$	correction term for MSE matrix of h-step forecast
$\Sigma_y(h)$	MSE or forecast error covariance matrix of h-step forecast of y_t

Distributions and Test Statistics

Notation	Description
$\mathcal{N}(\mu,\Sigma)$	multivariate normal distribution with mean μ and covariance matrix Σ
$\chi^2_m, \chi^2(m)$	chi-squared distribution with $m \in \mathbb{R}_+$ degrees of freedom
$\mathcal{P}(\lambda), \mathrm{Poi}(\lambda)$	Poisson distribution with parameter $\lambda \in \mathcal{R}_+$
$\mathcal{B}(n,p), \mathrm{Bin}(n,p)$	binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$
$F(m,n)$	Fisher distribution with (m,n) degrees of freedom
$t_m, t(m)$	Student's t-distribution with m degrees of freedom
$\mathcal{K}(\alpha,\beta)$	Kannisto distribution with parameters α and β
$\mathcal{G}(\lambda,\xi)$	Gompertz distribution with parameters λ and ξ
$\mathcal{E}(\gamma,\theta,\beta)$	extended exponential distribution with parameters γ, θ and β
$f_{\mathcal{D}}$	probability density function (PDF) of distribution $\mathcal{D}$
$F_{\mathcal{D}}$	cumulative distribution function (CDF) of distribution $\mathcal{D}$
$S_{\mathcal{D}}$	survival function of distribution $\mathcal{D}$
$h_{\mathcal{D}}$	hazard rate of distribution $\mathcal{D}$
$H_{\mathcal{D}}$	cumulative hazard rate of distribution $\mathcal{D}$
$v_{\mathcal{D}}$	mean residual life function of distribution $\mathcal{D}$
λ_{LM}	Lagrange multiplier statistic
λ_{LR}	likelihood ratio statistic
Q_h, Q_h^*	portmanteau statistic, modified portmanteau statistic

Actuarial Notation

Notation	Description
l_x	number of living individuals at age x
l_0	starting point for l_x: the number of individuals alive at age 0 (radix)
d_x	number of individuals who die between age x and age $x + 1$; $d_x = l_x - l_{x+1}$
q_x	probability of death between age x and age $x + 1$, i.e., $q_x = {}^{d_x}/_{l_x}$
p_x	probability that an individual with age x will survive to age $x + 1$, i.e., $p_x = {}^{l_{x+1}}/_{l_x}$
${}_nq_x$	probability of death between age x and age $x + n$
${}_np_x$	probability that a person with age x will survive to age $x + n$, i.e., ${}_np_x = {}^{l_{x+n}}/_{l_x}$
e_x	$= \sum_{t=1}^{\infty} {}_tp_x$; expectation of life for a person alive at age x
$m(t,x)$	central death rate

Thesis Structure

In the first part of this dissertation, we study stochastic mortality models in the framework of Generalized Age-Period-Cohort (GAPC) models. These models are represented in terms of generalized linear models and decompose the mortality across the dimensions age, period, and cohort. We give a review on well-established models and provide a comparative case study to highlight the strengths and weaknesses of various model predictors. While some of the GAPC models provide a good fitting accuracy to historical data, almost all of them share the same conceptional issues, which are mainly implied by imposed constraints on their parameters to ensure identifiability or some structural properties. We identify these issues and offer a detailed discussion of their implications. In the second case study, we investigate how a non-canonical link function impacts the quality-of-fit. This particular degree of freedom is mostly ignored in the literature. By drawing conclusions from both case studies, we propose a model which does not suffer from the common identifiability issues and employs a non-canonical link. We denote this model as the Kannisto model since it implies logistic-type growth of mortality rates, originally studied by the Finnish demographer Kannisto. We compare the fitting performance to the well-established models and demonstrate the advantages of the Kannisto model.

In the second part of the thesis, we focus on multivariate dynamics of the system of Kannisto variables. The objective is to identify an appropriate discrete time stochastic process which is capable of capturing the characteristics of the underlying stochastic factors that determine the mortality structure. We will investigate the presence of cointegration relations between the individual components. Furthermore, we demonstrate that, through the ability of capturing common trends, vector error correction models (VECMs) are better suited to model the dynamics of those factors compared to the standard choice of a random walk with drift.

The third chapter, on the field of survival analysis, is largely independent of the previous content. There, we provide an extensive characterization of the Kannisto lifetime distribution and its generalization, the so-called extended exponential distribution, which are both specified by logistic-type hazard rate functions as proposed by the Kannisto model of Section 1.9. We study the connections of these distributions to other well-known life and non-life distributions and prove further characterizing properties.

The individual chapters of this thesis are self-contained and organized as follows.

Chapter 1 | GAPC Models

Throughout Sections 1.1 to 1.3, we give a brief introduction to some basic concepts of mortality modelling and provide a review of the historical evolution of human mortality. In Section 1.4, we formally introduce the family of GAPC mortality models, where we first review their building blocks and subsequently present several well-established models. In Section 1.5, we discuss the Newton-Raphson and Fisher Scoring algorithms for parameter estimation of GAPC models and in Section 1.6, we provide a brief introduction to common statistical tests for model assessment and validation. Then, we provide two quantitative case studies with the focus on elderly populations in Section 1.7. First, we compare various predictors and then analyze the influence of a non-canonical link on the fitting accuracy. In Section 1.8 we identify several common issues of GAPC models which emerge due to the imposed parameter constraints. Based on the conclusions of the case studies, we propose a model family in Section 1.9.1, which still belongs to the GAPC family, but does not share the common problematic properties. Subsequently, a goodness-of-fit analysis is provided for the proposed Kannisto models, and we employ the standard GAPC approach for forecasting. An important observation which provides the connection to the second part of the thesis is that the estimated parameter trajectories suggest that there might be common stochastic trends between the factors of the Kannisto models. Conclusions on this chapter are given in Section 1.10.

Chapter 2 | Cointegration Analysis for the Kannisto Model

In order to identify an appropriate multivariate time series model for the dynamics of the Kannisto factors, we first provide an overview of some concepts of discrete multivariate stochastic processes in Section 2.2. In Section 2.3, we offer a detailed discussion on the specification procedure for VAR/VECM time series models. This includes methods for lag order selection, unit root and stationarity tests, parameter estimation under rank restrictions, cointegration tests, and residual tests for model diagnostics. In Section 2.4, we employ the VECM specification procedure on the Kannisto models and prove the existence of cointegration relations between their components. This implies that a VECM process, which is capable of capturing these relations, is an appropriate modelling approach. In Section 2.5 VECM driven projections of the Kannisto models are conducted and compared to the standard GAPC approach. Conclusions on this chapter are given in Section 2.6.

Chapter 3 | Characterization of the Kannisto and the Extended Exponential Distribution

This chapter can be considered independently of the previous content. The connection to the previous content is that the proposed Kannisto model implies a logistic-type hazard rate. Equivalently to density functions, hazard rate functions are representatives of distributions. However, the resulting continuous lifetime distribution, which is based on the logistic-type hazard rate, remained widely uncharacterized in the literature. In Section 3.2.1, we first provide an introduction to central concepts of survival analysis by considering different representatives of lifetime distributions and reviewing their properties. In Section 3.3, we reveal several connections of the Kannisto distribution and its generalization, the so-called extended exponential distribution, to other well-known distributions. In Section 3.5, we provide an extensive characterization of the Kannisto and the extended exponential distribution by deriving analytic expressions for the mean residual life function, moment generating function, central moments, Fisher information matrix, and Kullback-Leibler divergence. Section 3.6 gives a conclusion.

Contributions of the Thesis

The following provides a summary of the individual contributions made in this thesis. In Chapter 1, we investigate the GAPC family of stochastic mortality models. This popular and well-established class of models decomposes mortality across the dimensions age, period, and cohort, and has been widely studied by, e.g., Alai and Sherris (2014), Aro and Pennanen (2011), Berkum, Antonio and Vellekoop (2014), Börger, Fleischer and Kuksin (2014), Cairns, Blake and Dowd (2006), Cairns, Blake, Dowd et al. (2009), Haberman and Renshaw (2009), Lee and Carter (1992), Lovász (2011), O'Hare and Y. Li (2012), Plat (2009) and Renshaw and Haberman (2003, 2006). Recent articles, see, e.g., Currie (2016), Hunt and Blake (2014) and Villegas, Kaishev and Millossovich (2015) showed that the GAPC models can be embedded in the framework of generalized linear models, as introduced by the seminal paper of McCullagh and Nelder (1989). This conceptional generalization extends the classical modelling approaches with a feature which has been widely ignored in the literature. Our contribution in Section 1.7.2, shows that a non-canonical link function can improve the quality-of-fit for a variety of predictors. A similar conclusion has also been obtained by Currie (2016). An improved fitting accuracy is demonstrated, in particular, for populations aged above 60. The fact that a non-canonical logit link (with Poisson distributed response variable) performs better than the canonical logarithmic link implies that for high ages the historical age-related mortalities obey a logistic-type growth rather than an exponential.

Another main contribution of the first chapter is the analysis of several conceptional issues of GAPC models. The vast majority of GAPC models, especially those with a higher fitting performance, have an underdetermined predictor function. Thus, further parameter constraints are required to ensure parameter identification. This implies that the estimated parameters do not solely depend on the underlying data, but also on arbitrarily imposed constraints. Hence, particular patterns in the paths of the parameters may only occur due to those constraints. Related studies on the impact of the constraints can be found in, e.g., Hunt and Villegas (2015). Apart from parameter interpretability, the lack of identifiability also implies further issues. In particular, we demonstrate that the cohort term fails to serve its intended purpose to capture or reveal cohort effects, although it increases the quality-of-fit. Another point of criticism is that for forecasting purposes, the cohort term is usually assumed to be independent of periodic terms, which appears to be highly questionable, see also Currie (2012). Another contribution of this chapter is the significance analysis of individual parameters, which, to our knowledge, is considered for the first time for GAPC models. The surprising result of the Wald-type tests shows that one of the best performing models on our reference dataset has only about 6% (26 of 428) individually significant parameters.

Based on the conclusions of the conducted case studies we propose a mortality model with identifiable parameters and a non-canonical Poisson link. The so-called Kannisto model family implies a logistic-type growth of the age-related mortality rates and contains only parametric age-modulation terms in order to reflect the regular mortality structure of elderly populations. We provide a comparative analysis and highlight several advantages of the model.

In the second part of the thesis, we switch our focus to the dynamics of the stochastic factors that drive the mortality structure. In Chapter 2, we treat the period-related parameters of GAPC models as components of multivariate time series. The objective is to identify appropriate processes which are able to capture essential features of these time series. In the literature, the standard modelling approach for the period factors has been a random walk with drift, see, e.g., Cairns, Blake and Dowd (2006), Cairns, Blake, Dowd et al. (2009), Dowd, Cairns, Blake et al. (2010a,b) and Haberman and Renshaw (2011). This approach is motivated by the fact that the corresponding time series are non-stationary, and thus, a random walk is one of the most elementary potential candidates. However, the observation of the obtained time series from the proposed Kannisto model indicates long-run dependencies between the components. An appropriate framework to capture these dependencies is given by cointegrated processes as proposed by Engle and Granger (1987) and Granger (1981). Cointegration methods in the context of mortality modelling have been applied by orthogonal approaches in, e.g., Gaille and Sherris (2012), Lazar and Denuit (2009) and Salhi and Loisel (2011). The contribution of Chapter 2 is to provide statistical evidence for the existence of cointegration relations by using the tests proposed by Johansen (1988, 1995) and Johansen and Juselius (1990). In our analysis, we find cointegration relations for all proposed Kannisto predictors. Consequently, the individual periodic terms follow long-run equilibrium relations that cannot be represented by a random walk process, which is only capable to capture dependencies as instantaneous correlations. We provide an analysis of the forecasting performance, comparing the standard random walk approach with cointegrated VECM processes, after proper specification and validation procedures. The results show that by using VECM processes, we obtain forecasts which are more consistent with previous developments in terms of central forecasts and prediction intervals. We show that the framework of cointegration can be successfully applied to stochastic mortality modelling by using more sophisticated time series processes for the periodic terms. This result is not limited to the Kannisto predictor and can be applied to other predictors with multiple periodic terms, see, e.g., Gaille and Sherris (2011).

The contribution of the third part of the thesis is an extensive characterization of the distribution which is induced by a logistic-type hazard rate function, as proposed in the first part. Logistic-type hazard rate functions have been originally studied by Kannisto (1992) and Thatcher, Kannisto and Vaupel (1998). However, the primary objective of these studies was to find a parametric function which minimizes the Euclidean distance to empirical age-related mortalities. The authors did not consider the properties of the lifetime distribution which is induced by the corresponding logistic-type hazard function. Therefore, despite the practical importance of logistic-type hazard rates for mortality modelling of elderly populations, the implied distribution remained widely uncharacterized. A few specific contributions can be found in Marshall and Olkin (2007) and Missov (2013). In Chapter 3, we study the relations of the Kannisto distribution and its three-parameter generalization, the so-called extended exponential distribution, as proposed by Marshall and Olkin (2007). We show several connections to other well-known life and non-life distributions. Furthermore, we derive analytic expressions for the mean residual life function, moment generating function, central moments, Fisher information matrix, and Kullback-Leibler divergence. Moreover, we prove that the extended exponential and the Kannisto distribution belong to the minimum domain of attraction of the Weibull distribution, and to the maximum domain of attraction of the Gumbel distribution. The obtained contributions provide deeper insights to parametric hazard rate models for higher ages, which hopefully will find beneficial applications in actuarial science or life insurance industry.

Chapter 1 | GAPC Models

Contents

1.1 | Motivation

The development of human mortality rates has shown continuing improvements over the past centuries. The life expectancy, as well as the maximum lifespan, have strongly increased. As mentioned in Wilmoth (1997), demographic studies do not indicate certain biologic imposed bounds of the lifetime. The life expectancy at birth has experienced substantial gains and has almost tripled over the human history. These improvements are mainly based on general enhancements of living standards and medical developments. The continuous increase of the life expectancy can be attributed to significant improvements of infant and child mortalities at the beginning of the 20th century as well as to the reduction of mortality rates of the elderly population due to improved medical diagnostics methods, treatments of cardiovascular diseases and cancer (see Wilmoth (2000) for more details). As Oeppen and Vaupel (2002) point out, the female life expectancy at birth in the record-holding country has increased for 160 years at a steady pace of almost 3 months per year.

Extrapolative methods for projections of mortality rates have been used by actuaries for centuries. Traditionally, deterministic modelling approaches have been used by life insurers for the calculation of premiums and reserves. The risk of a deviation between forecasted mortality rates and the eventually realized mortality rates has been assumed to be diversified over time and individuals. As historical data of the past century shows, mortality rates improved quite unpredictable under classic models, so that the mortality risk remained undiversified. Traditional deterministic approaches appeared to be inadequate since the projections made by these models showed a substantial underestimation of the life expectancy trend. The world's oldest insurance company, the Equitable Life, which was established in 1762, declared bankruptcy in 2000, as a result of overestimated mortality rates and falling interest rates, see Roberts (2012).

Mortality modelling has a long history, see, e.g., Gompertz (1825) and Moivre (1725). However, significant developments of mathematical methods were achieved only recently, see Booth (2006) and Booth and Tickle (2008) for detailed reviews of the methodological developments of demographic forecasting since 1980. Reviews on different approaches to mortality modelling and forecasting methods can be found in Pitacco, Denuit, Haberman and Olivieri (2009), Pollard, Benjamin and Soliman (1987), Tabeau, Berg Jeths and Heathcote (2001), Tuljapurkar and Boe (1998) and Wong-Fupuy and Haberman (2004).

Historical lifetime data shows clearly that mortality rates decrease over time. However, the reduction differs for different ages and there are also variations among various cohorts. In the past two decades, enormous efforts have been made to explore stochastic models with the objective to describe the dynamics of human mortalities and to develop pricing tools for classical and modern mortality-linked securities. The increasing amount of academic attention was substantially triggered by the life insurance industry and their new challenges in risk management and internal models brought by Solvency 2.

Research with the focus on the impact of mortality decrease on mortality linked securities such as annuities and life insurance can be found among others in Ballotta and Haberman (2003), Gatzert and Wesker (2014) and Olivieri (2001). For valuations of mortality-contingent claims, see, e.g., Ballotta and Haberman (2006) and Milevsky and Promislow (2001). Publications on stochastic forward mortality models, inspired by modelling approaches of interest rates in finance, can be found in, e.g., Bauer, Benth and Kiesel (2012), Biffis (2005), Biffis and Millossovich (2006), Dahl (2004), Dahl and Møller (2006), Norberg (2010) and Tappe and Weber (2014).

Before we provide an introduction to basic concepts for mortality models, we first bring clarity in different types of involved mortality-related risks. We follow the conventions of Cairns, Blake and Dowd (2006) to distinguish between the following risks.

- The term *mortality risk* covers all forms of uncertainty about the future mortality rates, including increases or decreases of mortality rates.
- The term *longevity risk* encompasses the uncertainty in the long-term trend of mortality rates and indicates that future survival rates turn out to be higher than anticipated.
- The term *short-term* or *catastrophic mortality risk* covers the risk of catastrophic events leading to significantly higher mortality rates. These include, in particular, influenza pandemics such as the Spanish flu in 1918 or the recent swine flu pandemic and also all kinds of natural catastrophes such as the earthquake at the west coast of Sumatra and the resulting tsunami in December 2004.

The remarkable evolution of mortality improvements beyond previously anticipated limits has shown the need for sophisticated mortality models. The main advantage of the development of stochastic mortality modelling is that they produce forecasts in terms of probability distributions rather than deterministic point forecasts and thus allowing quantification of the forecast uncertainty.

Thus, the focus of this thesis is to provide additional methods for the quantification of mortality and longevity risks. Both risks are not only generally important for life insurers, but they are also used for the determination of the *Solvency Capital Requirement* (SCR) as requested by Solvency 2.

The following content of this chapter is organized as follows. In Sections 1.2 and 1.3, we give a brief introduction to elementary concepts of mortality modelling and review the historical evolution of human mortality. In Section 1.4, we formally introduce the family of GAPC mortality models, where we first introduce their building blocks and subsequently present several well-established models. In Section 1.5, we discuss the Newton-Raphson and Fisher Scoring algorithms for parameter estimation of GAPC models in the framework of generalized linear models, and in Section 1.6, we provide a brief introduction to common statistical tests for model assessment and validation. The comprehensive surveys in Sections 1.4 to 1.6 serve as preparations, and for the most part, do not contain any original research. After the theoretical foundations of stochastic mortality models, parameter estimations for GLMs, and statistical tests for model comparison, we provide two quantitative case studies with the focus on elderly populations in Section 1.7. In the first case study, we compare various predictors, and subsequently, analyze the influence of a non-canonical link function on the fitting accuracy. In Section 1.8 we identify several common issues of GAPC models which emerge due to the imposed parameter constraints. Based on the conclusions of the case studies, we propose a model family, the so-called Kannisto model, in Section 1.9.1. Subsequently, a goodness-of-fit analysis for the proposed model is provided. In Section 1.9.4 we consider the standard GAPC approach for forecasting which will be used as a reference for a more sophisticated time series approach that will be introduced in Chapter 2. Conclusions and an outlook are provided in Section 1.10.

1.2 | Basic Concepts and Source of Data

To identify patterns and trends in human mortality a precise and reliable data source over an extended period is required. For comparative studies, the data should also be available for several groups, distinguished by country and gender. For most developed countries mortality data is collected by official authorities or population registers. To facilitate research on human mortality the Human Mortality Database (HMD) was initiated aiming to provide continuous collections and aggregations of mortality data. HMD is a joint project of the Department of Demography at the University of California, Berkeley, USA, and the Max Planck Institute for Demographic Research in Rostock, Germany. Detailed population and mortality data is freely available for researchers at www.mortality.org. Currently, the HMD database offers life tables for 38 countries. These tables represent an empirical record of mortality-related quantities with data grouped by countries, gender, periods or cohorts. The data covers the ages from 0 to 110 for each country with yearly increments in age and time. The most extensive collection of life data, which starts from the year 1751, is provided for Sweden. This particular dataset will serve as our reference mortality dataset for the illustration of the historical changes in human mortality, as will be provided in Section 1.3, and also for the case studies of Generalized Age-Period-Cohort models in Section 1.7. We took that particular set of data, firstly because it covers the longest available period and on the other hand, we focused on the female population, since they were not strongly involved in military conflicts. In general, mortality data of male populations show substantial distortions in times of wars. Thus, to highlight long-term effects, male data is mostly excluded from the demonstrations. Further illustrations of mortality improvements, that clearly show the dramatic impact of epidemic diseases and wars on human life over the last century, can be found in Appendix A.2.

Initially, we provide some notes on how HMD life tables are structured. The concept of life tables belongs to the most established tools in the empirical survival analysis. One of the first documented developments of a life table goes back to the well-known astronomer Edmund Halley, see Halley (1693). Estimation methods for hazard rates, survival functions, and other lifetime representatives (see Section 3.2.1) are designed for situations, where a large sample size is available, but the exact times of the events are unknown, see, e.g., Rinne (2014). That means that generally, life tables do not contain information on the level of individuals. Mortality rates are rather aggregated population-wise and grouped by ages and periods. Thus, in a discrete setting, a person of age x has an exact age in the interval $[x,x+1)$. Similarly, an event that occurs in year t, takes place at some time in the interval $[t,t+1)$. For each calendar year t and age x the quantities included in life tables are:

- $l_{t,x}$, the number of individuals aged x during the year t. Note that in life tables $l_{t,0}$ is usually normalized to a particular value (e.g., 10^5), the so-called radix.
- $d_{t,x}$, the number of deaths occurred during the year t at age x.
- $p_{t,x}$, the conditional probability at the calendar year t of surviving to age $x+1$ given the survival to age x.
- $q_{t,x}$, the conditional probability at time t for an individual of age x not surviving up to age $x+1$. Note that $q_{t,x} = 1 - p_{t,x}$.
- $e_{t,x}$, the expectation of remaining life at age x in calendar year t.

Detailed information on the methods used to create HMD life tables can be found in Wilmoth, Andreev, Jdanov et al. (2007).

1.2.1 | Real Cohorts vs. Synthetic Cohorts

In the following, we will distinguish between the concepts of *real cohorts* and *synthetic cohorts* and also introduce a slightly more general meaning of the term cohort.

In the usual sense, a cohort T is a group of individuals with the same year of birth, namely T, or in other words, a group of individuals aged 0 in the period T. We will use the tuple $(T,0)^*$ to denote the *real birth cohorts*. A group of people of the same birth cohort will turn the same age in future periods in case of survival. The subgroup of individuals of the birth cohort $(T,0)^*$, that survive y years, such that in year $T+y$ their age is y, will be denoted as $(T+y,y)^*$ and called a *generalized real cohort* or simply *real cohort*. For instance, the $(1950,0)^*$ and the $(2015,65)^*$ real cohorts are both groups of people with the birth year 1950, but the later cohort contains only the subgroup of individuals who survived up to age 65 in 2015. This conceptual distinction of real cohorts will be used later.

Next, we introduce an artificial cohort type, the so-called *synthetic cohort*. Synthetic cohorts are groups of people aged x at a reference period T and which, at every further age throughout their life, experience the age-specific death/survival rates of that period T (cf., Mokyr, 2003). Synthetic cohorts will be denoted by (T,x).

Table 1.1: An illustration of mortality rates referring to real and synthetic cohorts. The columns of the table contain age-specific mortality rates for fixed periods and the rows contain mortality rates for fixed ages at different periods, respectively. The essential difference between a real and a synthetic cohort is the assumption which age-specific mortality rates a group of individuals will experience during their lives. Mortality rates of real cohorts are arranged on diagonals, see the orange marked entries of the real cohort $(2006{,}60)^*$. Mortality rates of synthetic cohorts are arranged on verticals, see the blue marked entries of the synthetic cohort $(2006{,}60)$.

$\cdots$	$q_{2006,110}$	$q_{2007,110}$	$q_{2008,110}$	$q_{2009,110}$	$q_{2010,110}$	$q_{2011,110}$	$q_{2012,110}$	$q_{2013,110}$	$q_{2014,110}$	?
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\cdots$	$q_{2006,75}$	$q_{2007,75}$	$q_{2008,75}$	$q_{2009,75}$	$q_{2010,75}$	$q_{2011,75}$	$q_{2012,75}$	$q_{2013,75}$	$q_{2014,75}$	?
$\cdots$	$q_{2006,74}$	$q_{2007,74}$	$q_{2008,74}$	$q_{2009,74}$	$q_{2010,74}$	$q_{2011,74}$	$q_{2012,74}$	$q_{2013,74}$	$q_{2014,74}$	?
$\cdots$	$q_{2006,73}$	$q_{2007,73}$	$q_{2008,73}$	$q_{2009,73}$	$q_{2010,73}$	$q_{2011,73}$	$q_{2012,73}$	$q_{2013,73}$	$q_{2014,73}$	?
$\cdots$	$q_{2006,72}$	$q_{2007,72}$	$q_{2008,72}$	$q_{2009,72}$	$q_{2010,72}$	$q_{2011,72}$	$q_{2012,72}$	$q_{2013,72}$	$q_{2014,72}$	?
$\cdots$	$q_{2006,71}$	$q_{2007,71}$	$q_{2008,71}$	$q_{2009,71}$	$q_{2010,71}$	$q_{2011,71}$	$q_{2012,71}$	$q_{2013,71}$	$q_{2014,71}$	?
$\cdots$	$q_{2006,70}$	$q_{2007,70}$	$q_{2008,70}$	$q_{2009,70}$	$q_{2010,70}$	$q_{2011,70}$	$q_{2012,70}$	$q_{2013,70}$	$q_{2014,70}$	?
$\cdots$	$q_{2006,69}$	$q_{2007,69}$	$q_{2008,69}$	$q_{2009,69}$	$q_{2010,69}$	$q_{2011,69}$	$q_{2012,69}$	$q_{2013,69}$	$q_{2014,69}$	?
$\cdots$	$q_{2006,68}$	$q_{2007,68}$	$q_{2008,68}$	$q_{2009,68}$	$q_{2010,68}$	$q_{2011,68}$	$q_{2012,68}$	$q_{2013,68}$	$q_{2014,68}$	?
$\cdots$	$q_{2006,67}$	$q_{2007,67}$	$q_{2008,67}$	$q_{2009,67}$	$q_{2010,67}$	$q_{2011,67}$	$q_{2012,67}$	$q_{2013,67}$	$q_{2014,67}$	?
$\cdots$	$q_{2006,66}$	$q_{2007,66}$	$q_{2008,66}$	$q_{2009,66}$	$q_{2010,66}$	$q_{2011,66}$	$q_{2012,66}$	$q_{2013,66}$	$q_{2014,66}$	?
$\cdots$	$q_{2006,65}$	$q_{2007,65}$	$q_{2008,65}$	$q_{2009,65}$	$q_{2010,65}$	$q_{2011,65}$	$q_{2012,65}$	$q_{2013,65}$	$q_{2014,65}$	?
$\cdots$	$q_{2006,64}$	$q_{2007,64}$	$q_{2008,64}$	$q_{2009,64}$	$q_{2010,64}$	$q_{2011,64}$	$q_{2012,64}$	$q_{2013,64}$	$q_{2014,64}$	?
$\cdots$	$q_{2006,63}$	$q_{2007,63}$	$q_{2008,63}$	$q_{2009,63}$	$q_{2010,63}$	$q_{2011,63}$	$q_{2012,63}$	$q_{2013,63}$	$q_{2014,63}$	?
$\cdots$	$q_{2006,62}$	$q_{2007,62}$	$q_{2008,62}$	$q_{2009,62}$	$q_{2010,62}$	$q_{2011,62}$	$q_{2012,62}$	$q_{2013,62}$	$q_{2014,62}$	?
$\cdots$	$q_{2006,61}$	$q_{2007,61}$	$q_{2008,61}$	$q_{2009,61}$	$q_{2010,61}$	$q_{2011,61}$	$q_{2012,61}$	$q_{2013,61}$	$q_{2014,61}$	?
$\cdots$	$q_{2006,60}$	$q_{2007,60}$	$q_{2008,60}$	$q_{2009,60}$	$q_{2010,60}$	$q_{2011,60}$	$q_{2012,60}$	$q_{2013,60}$	$q_{2014,60}$	?

To make the above definitions more precise, we consider an array of mortality rates $q_{t,x}$ as illustrated in Table 1.1. The columns of the table contain age-specific mortality rates for fixed periods, and the rows contain the mortality rates for fixed ages at different periods. Mortality rates can be grouped according to different arrangements, cf., e.g., Liu (2008).

- A *diagonal* arrangement

$$\{q_{t,x}, q_{t+1,x+1}, \ldots, q_{t+n,x+n}\},$$

corresponds to a sequence of mortality rates of individuals with the same year of birth $t - x$. For instance, the orange marked entries $\{q_{2006,60}, \ldots, q_{2014,68}\}$ of Table 1.1 correspond to a real cohort born in $t - x = 1946$. We can alternatively define a real cohort $(T, x)^*$ as a group of individuals, which experience the diagonal-arranged age-specific mortality rates $q_{T,x}, q_{T+1,x+1},$

$q_{T+2,x+2}, \ldots, q_{T+c,x+c}$, for some integer c such that $x + c$ is the highest available age of the life table.

- A *vertical* arrangement,

$$\{q_{t,x}, q_{t,x+1}, \ldots, q_{t,x+n}\},$$

corresponds to a sequence of age-specific mortality rates at some fixed period t. The column entries contain mortality rates corresponding to distinct real cohorts, see the blue marked entries of Table 1.1. By the above definition, a synthetic cohort (T, x) is a fictive group of individuals with age x in T, which experience the vertical-arranged age-specific mortality rates of the period T, i.e., $q_{T,x}, q_{T,x+1}, q_{T,x+2}, \ldots, q_{T,x+c}$, for some integer c, such that $x + c$ is the highest available age of the life table. Synthetic cohorts are auxiliary constructions and are often used when the actual mortality rates needed for statements on real cohort are not available. The most popular example is the life expectancy at birth. In order to provide the life expectancy of, say a newborn in 2015, one would need future mortality rates $q_{2015,0}, q_{2016,1}, \ldots, q_{2015+c,c}$. These are obviously not available at present time. To avoid the problem of missing data, one can make the assumption, that a newborn will experience the same age-specific mortality rate in year $2015 + y$ like individuals aged y in 2015. This is equivalently to the assumption that the newborn belongs to the synthetic cohort $(2015,0)$.

- A *horizontal* arrangement

$$\{q_{t,x}, q_{t+1,x}, \ldots, q_{t+n,x}\}$$

corresponds to a time series of mortality rates referring to a given age x.

As a concluding remark, note that a synthetic cohort is an auxiliary construction and due to mortality changes over time, a real cohort $(T,y)^*$ will not experience the same mortality rates as those of the synthetic cohort (T,y). In the further course of the thesis, we will estimate parametric hazard rates for fixed periods. The advantage of considering mortality rates of synthetic cohorts over real cohorts is that observations exist on the entire age range for any given reference time. For real cohorts on the other side, only a few observations might be available. The non-linear structure of mortality rates makes the estimation on shorter ranges less reliable. Thus, projections of life tables are often based on modelling mortalities of synthetic cohorts rather than of real cohorts. However, projections based on synthetic cohorts can be used to obtain forecasts for real cohorts, see, e.g., Section 2.5.5.

1.3 | Historical Evolution of Human Mortality

The objective of this section is to present an empirical study on the evolution of human mortalities that have experienced significant changes over the past century.

We will demonstrate several aspects of historical mortality improvements based on some lifetime characteristics, as the life expectancy, and lifetime representatives, as survival functions and the hazard rate functions (see Section 3.2.1). The survival functions and life expectancies will be illustrated for both, real and synthetic cohorts. For the following presentation, we use our reference HMD of the Swedish female population. For a comparative study of the Swedish male population, see, e.g., Liu (2008).

Figure 1.1 illustrates the death counts $d_{t,x}$ for the full available age range $x \in \{0, \ldots, 109\}$ and the periods $t \in \{1860, 1910, 1960, 2010\}$. Note that the total numbers are normalized to 10^5 and that for the periods 1860 and 1910 the infant mortalities are out of plot range. The data show that in 1860 nearly 20% of the children did not survive until the age of 5, whereas in 1910 this number decreased to almost 10%, while present rates are below 0.3%. Qualitatively, one can observe that the modes of the curves move towards higher ages, while the dispersion around the modes decreases. Wilmoth (2000) terms this behaviour as the *compression of mortality* because not only the level of longevity has increased but also the certainty about the timing of death.

Figure 1.2 illustrates empirical survival functions, i.e., the probability to survive up to a particular age, for several real birth cohorts (dashed lines) and synthetic cohorts (solid lines). Note, there would be no difference between survival functions of real cohorts $(T,0)^*$ and those of synthetic cohorts $(T,0)$ if the age-specific mortality rates would remain unchanged. Since the age-specific mortality rates have generally shown a decreasing trend, the values of the survival functions for recent cohorts turn out to be higher. In the perspective of survival functions, the effect of mortality compression is often referred to as the *rectangularization* of the survival function. By comparing the survival curves of the real $(1910,0)^*$-cohort (orange dashed) and the synthetic $(1910,0)$-cohort (orange), we can make another important observation, which is, that mortalities at different ages changed at different extents. Another observation, which demonstrates the magnitude of the overall improvements, can be made by comparing quantiles of ancient and recent cohorts. For instance, the empirical 0.2 quantile of the $(1860,0)$-cohort is, due to high infant mortalities, only the age of 5, whereas the corresponding quantile of $(2010,0)$-cohort is the age of 76.

In Figures 1.3 and 1.4 historical hazard rates, alternatively called the force of mortality curves, are illustrated. Hazard rates will be covered in detail in Section 3.2. For now, they should be understood as the instantaneous risk of dying associated with a particular age. Qualitatively, hazard rates show the same pattern as the mortality curves. In actuarial science, the hazard rates belong to the mainly preferred representations of the lifetime, since they illustrate the age-based risk profile experienced by individuals.

Figure 1.3 illustrates, on a logarithmic scale, the smoothed mortality rates of the Swedish female population for the reference periods $t \in \{1860, 1910, 1960, 2010\}$. Each of the presented periods shows a relatively high infant mortality risk followed by an initial decrease. After reaching the lowest values around the age of 10, the values eventually increase. Note that newborns in 1910 subjected the same age-specific mortality risk, as individuals around the age of 70 at the same period. Exactly one century later, in 2010, the infant hazard rates were reduced by approximately a factor of 30. However, it is still as high as of individuals aged around 50 in 2010. Moreover, the illustration shows that for each period the hazard rates are roughly linear for ages above 60, which means that on a linear scale the risk is exponentially increasing in age. However, the slopes of the hazard rates tend to decrease for ages above 80, which indicates a sub-exponential age-specific growth. Studies on mortalities at very high ages are available in, e.g., Himes, Preston and Condran (1994), Kannisto (1992) and Thatcher, Kannisto and Vaupel (1998). Figure 1.3 also shows that the general decay in mortalities

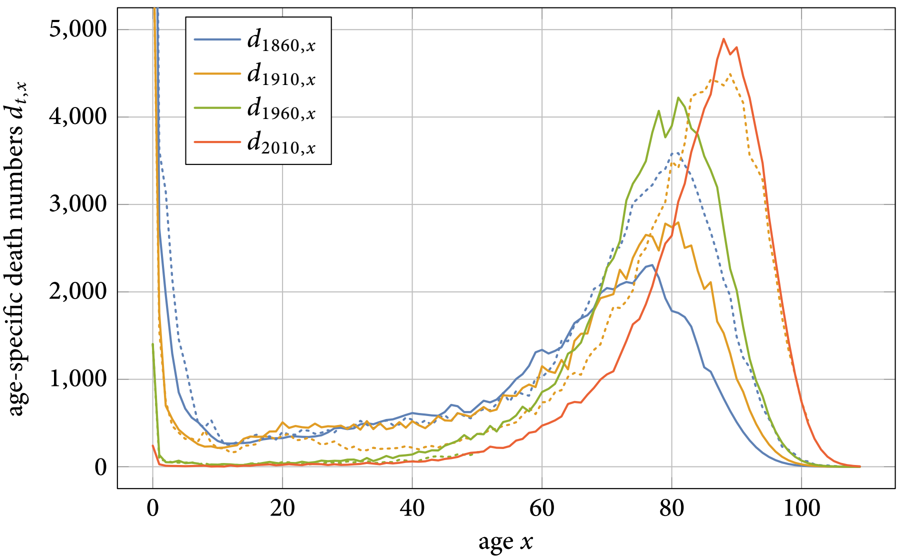


Figure 1.1: Empirical number of deaths $d_{x,t}$ at age $x \in \{0,\dots,109\}$ for the periods $t \in \{1860, 1910, 1960, 2010\}$ of Swedish female population. Note, the total numbers are normalized to 10^5. The dotted curves represent the age-specific death numbers of real birth cohorts, e.g., 1860 (blue), 1910 (orange) and 1960 (green). For the 1860 and 1910 birth cohorts, there is a clear shift of the mode to higher ages compared to periodic data.

is affected by catastrophic events as in 1918. The outbreak of an influenza pandemic, known as the Spanish flu in 1918, is considered as the deadliest natural disaster in human history, and is responsible for more than 50 million deaths, see Taubenberger and Morens (2006). Figure 1.3 shows that in contrast to other influenzas, primary the young and healthy part of the population were affected. Notice, in particular, the spread of the 1910 curve (orange) and the 1918 curve (blue) between the ages 10 up to 40. The spread between the hazard rate in 1910 and that of the time of the largest pandemic in human history also stresses out how remarkable the mortality improvements have been in the course of the past 100 years.

Figure 1.4 shows the hazard rates for all periods since 1850. Here, one can observe the continuous changes of the age-specific instantaneous mortality risk, experienced by the Swedish female population. Figure 1.5 shows the post-age 60 hazard rates on a linear scale. Note that the rates appear highly regular. We will focus on this particular age range in the further course of the thesis and provide a comparative analysis between models with logistic-type hazard rates and exponential increasing hazard rates.

In the actuarial literature, there are essentially two different approaches to quantify mortality improvements. The first is, to choose a reference period, say for example the period 1910, and consider the ratio of the hazard rates of that particular period and a second period of interest, say the period 2010. This approach allows a quantification of mortality changes for widely separated periods by considering the hazard ratios.

Through the thesis, we will use an alternative definition of mortality improvements. The rigorous definition will be provided in Section 2.5.4. Descriptive spoken, the alternative approach characterizes

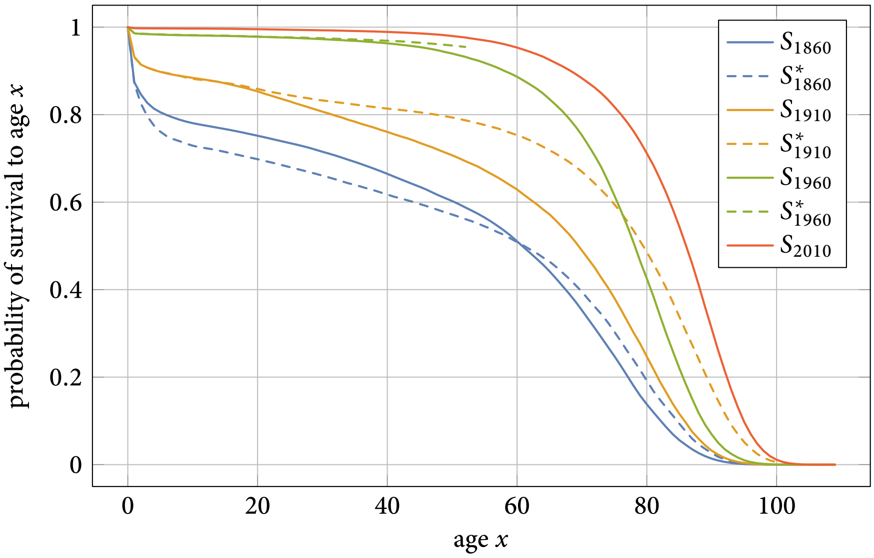


Figure 1.2: Survival functions of Swedish female population. The dashed curves S^*_T represent survival functions of the birth cohorts 1860, 1910 and 1960 and the solid lines the survival functions S_T, i.e., the survival probabilities up to age x, of the synthetic cohorts from the reference periods $T \in \{1860,1910,1960,2010\}$.

mortality improvements $j_t(x)$ as the infinitesimal relative changes of the hazard rate in time. For discrete data, as provided by life tables, this approach corresponds to the following expression:

$$j_t(x) = -\frac{(h_t(x) - h_{t-1}(x))}{h_{t-1}(x)}.$$

The minus sign is a convention to express that for decreasing mortality rates the improvements considered to be positive, and for increasing mortality rates the improvements are negative. The historical annual mortality improvements of the Swedish female or male population are shown in Figures 1.7 and 1.8. The figures illustrate the mortality improvements, in the sense of relative changes of the hazard rates, for the periods 1900-2010 and ages 25-100. Note, these representations are obtained by applying two-dimensional B-spline smoothing methods to raw data. For additional details on B-spline smoothing with focus on mortality modelling see, Currie, Durban and Eilers (2004). B-spline smoothing yields data with reduced fluctuations and reveals local periodic patterns as well as cohort effects. Without smoothing, these patterns would be covered by high fluctuations especially at younger ages.

Figures 1.7 and 1.8 can be read as follows. Blue areas imply that mortality is deteriorating. Green regions represent almost unchanged mortality rates. Yellow parts imply smaller rates of improvement, orange and red state for stronger rates of improvement. For example, the influenza pandemic of 1918 and the corresponding affected ages are clearly noticeable by those representation. Another example for a negative development can be observed for the male population at ages around 40 in 1965 of Figure 1.8. The local increase of the mortality can be attributed to a rise in cardiovascular mortality

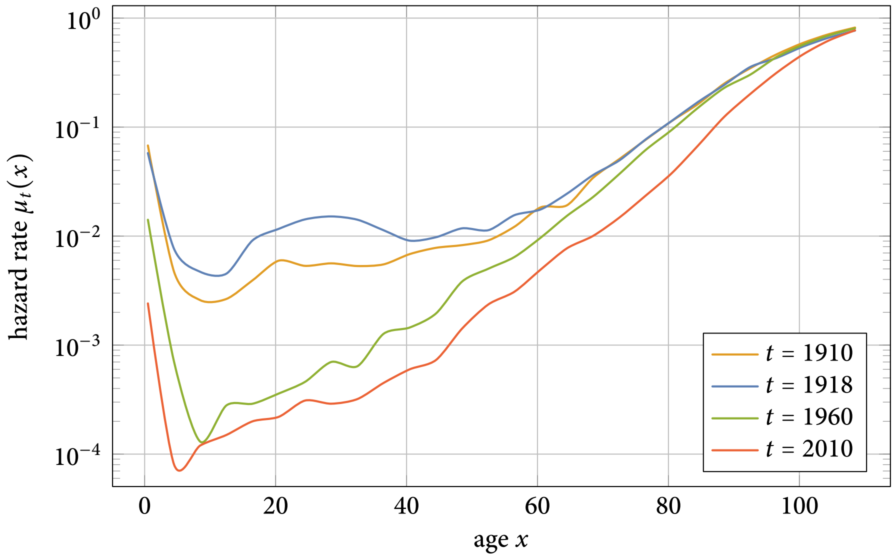


Figure 1.3: Smoothed hazard rates (force of mortality) of Sweden's female population for the reference periods $t \in \{1910, 1918, 1960, 2010\}$.

among industrial workers, cf., Diderichsen and Hallqvist (1997). High mortality improvements can be registered for both genders up to the age of 60 immediately after the ending of the Second World War. That fact can be primarily attributed to the discovery of penicillin and further antibiotics and vaccines.

For more examples of local deviations of the positive trend, see Figures 1.9 and 1.10 for the mortality improvements of the UK-Wales population, where one can clearly observe the impacts of both World Wars on the male population. Noticeable as well, is the increased mortality for the male population around 1990 for ages about 30. This increase in mortality occurred mainly due to the sexually transmitted human immunodeficiency virus infection (HIV).

Cohort effects can also be observed more clearly for the UK-Wales population compared to the Swedish population. For instance, notice the diagonal-arranged patterns of annual mortality improvements above 3%. These improvements can be attributed to birth cohorts centred around the year 1935. For more insights and explanations why these cohorts experienced higher improvements in mortality than other generations, see Willets (2004).

Further illustrations of mortality improvements of other countries can be found in Appendix A.2, where similar trends of mortality improvements and local effects can be detected. For an outlining country-specific behaviour, see the mortality improvements of Russia on page 246. Russia has experienced a negative development of the life expectancy due to social disruptions and instability resulted by the collapse of the Soviet Union.

Finally, we demonstrate in Figures 1.11 and 1.12 the time evolution of the life expectancy for synthetic and real Swedish female cohorts. Figure 1.11 illustrates the life expectancy at age x, denoted by $e_{T,x}$. The following characteristics are worth mentioning: compared to the life expectancies of higher ages, the life expectancy at birth made the largest progress, from values around 45 in 1850's to

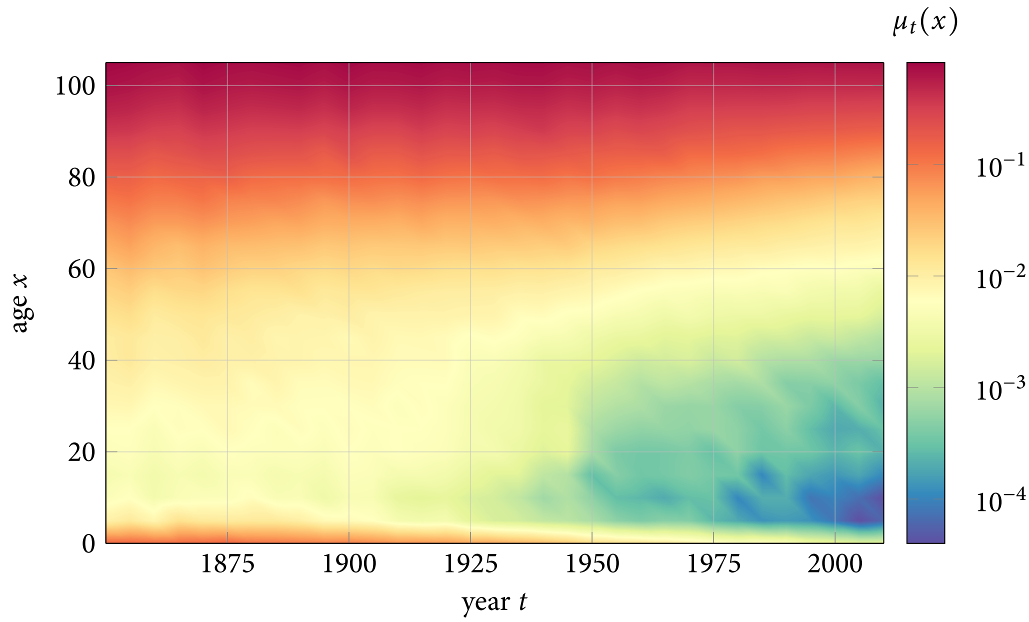


Figure 1.4: Hazard rates of Swedish female population on a logarithmic scale.

almost the age of 84 for the most recent periodic data. The down peaks at 1918, resulted due to the flu pandemic, are noticeable for the life expectancies at birth and at age 20. Note also, the life expectancy at birth in 2011 is 83.9 years, whereas at age 60 the expectancy is with 85.7 years only slightly larger. The progress of the curves emphases the fact, the mortality improvements had a higher impact on younger ages.

The characteristics of the life expectancy for real cohorts are illustrated in Figure 1.12. Note that the life expectancy of real cohorts is only fully observable after all individuals of that cohort have passed away. Since we do not include projected mortality values in our presentation, the paths terminate at the beginning of the 20th century. Projections of the life expectancy of real cohorts can be found in Figure 2.19.

In summary, we have presented the empirical changes in the human mortality based on several lifetime representatives. Chapter 3 will contain a deeper discussion of those representatives including their connections. The next section will review several stochastic mortality models which have been successfully applied on empirical data.

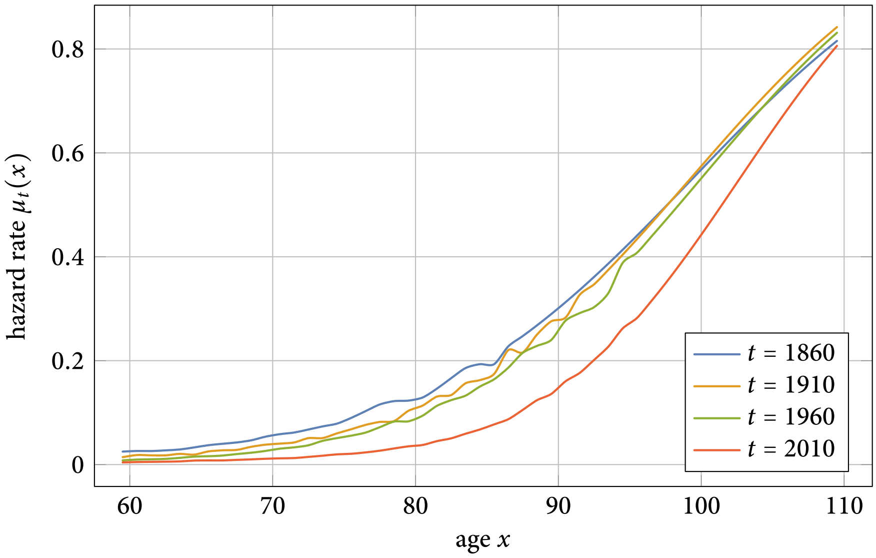


Figure 1.5: Post-age-60 hazard rates (force of mortality) of Sweden's female population for the reference periods $t \in \{1860, 1910, 1960, 2010\}$ on a linear scale. Note that data show a highly regular structure. We will use that fact to propose a parametric logistic-type hazard rate model.

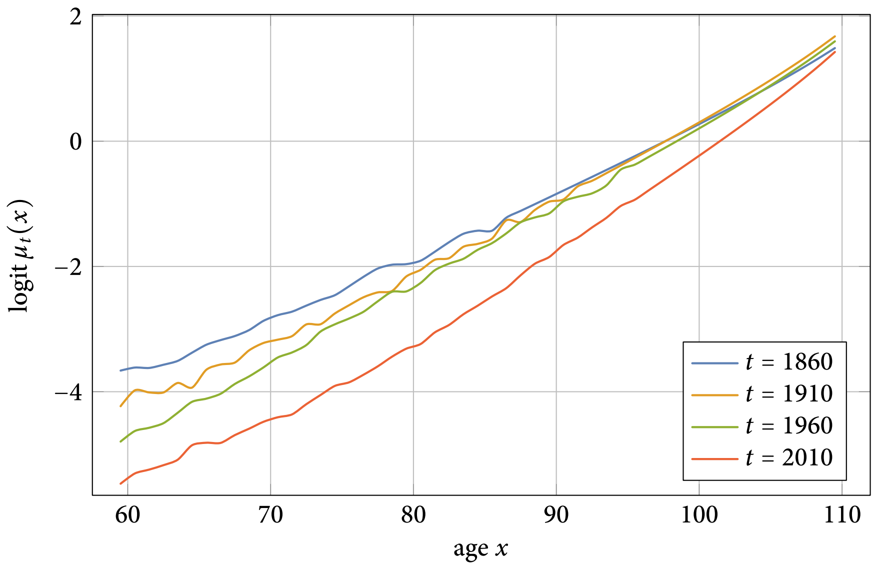


Figure 1.6: Logit transformed post-age-60 hazard rates of Sweden's female population for the reference periods $t \in \{1860, 1910, 1960, 2010\}$. Notice that data show a highly regular linear structure. Thus, we will use that fact to propose a parametric logistic-type hazard rate model, which linearizes under the logit transformation.

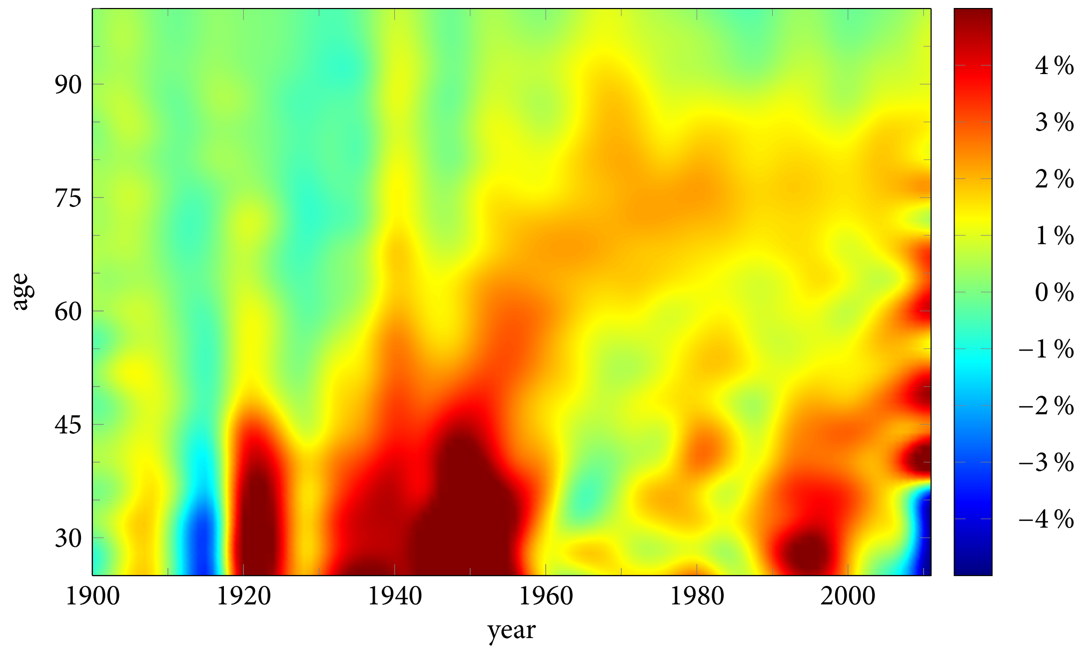


Figure 1.7: Annual mortality improvements of the Swedish female population.

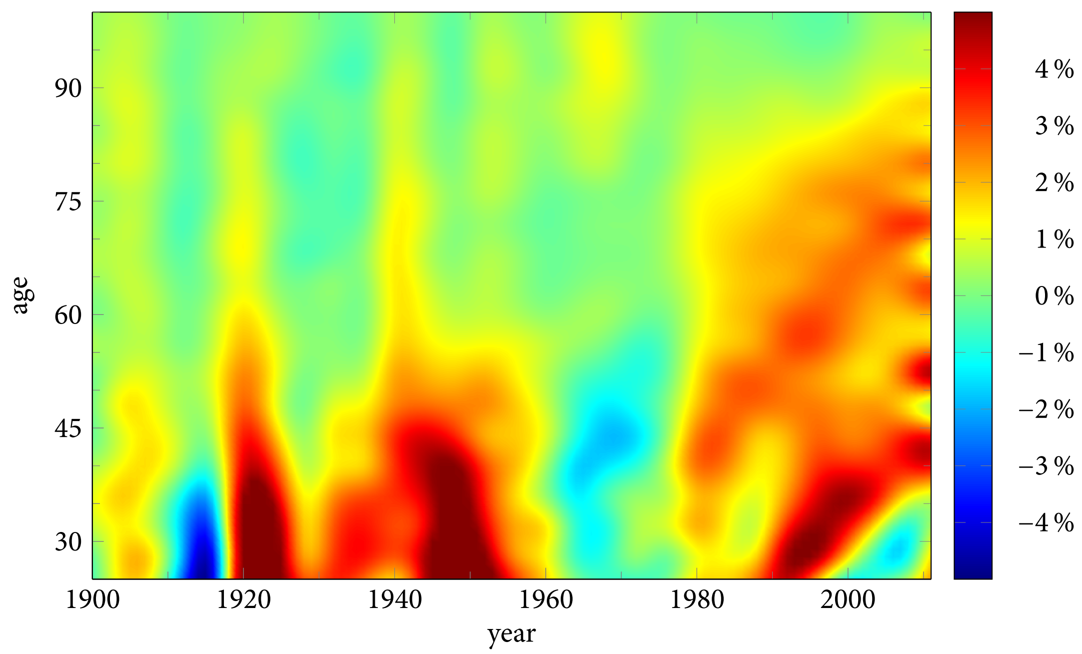


Figure 1.8: Annual mortality improvements of the Swedish male population.

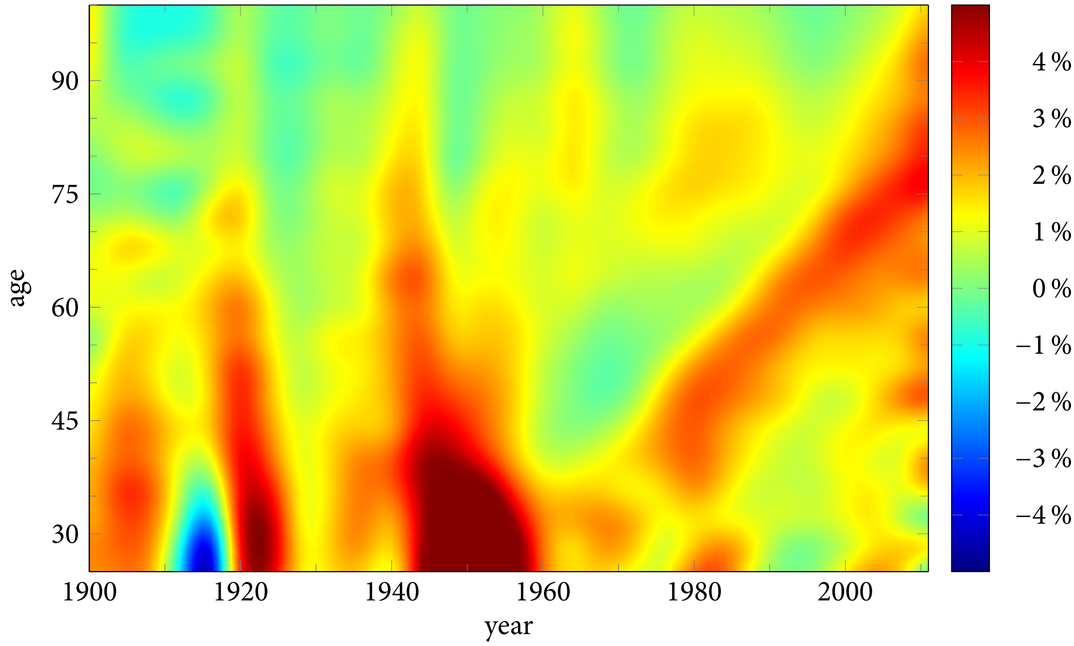


Figure 1.9: Annual mortality improvements of the UK-Wales female population.

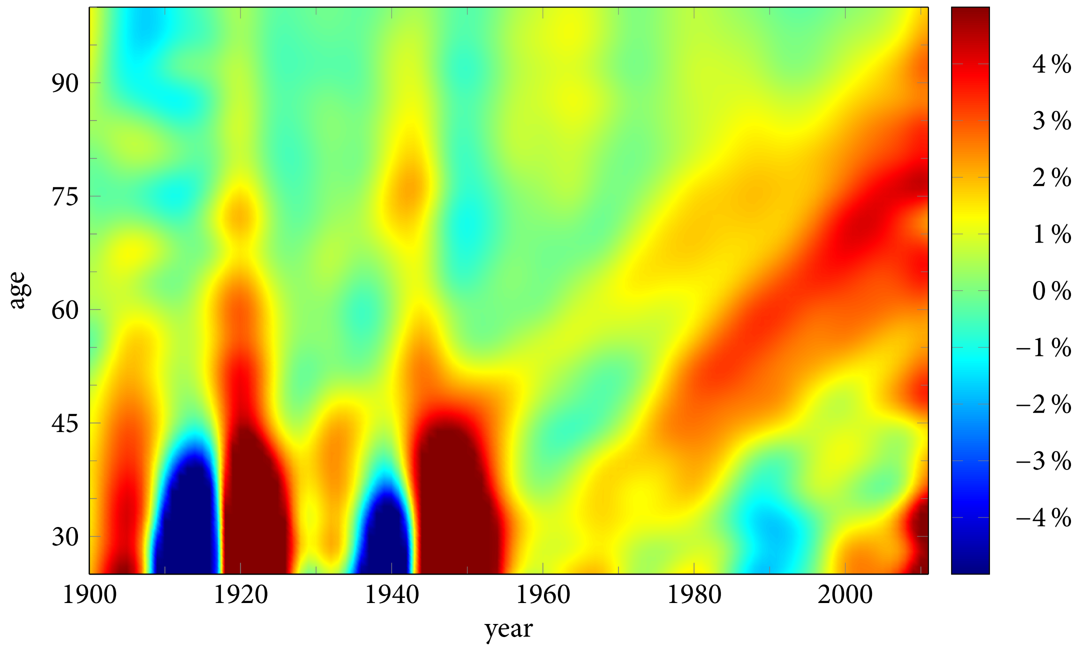


Figure 1.10: Annual mortality improvements of the UK-Wales male population.

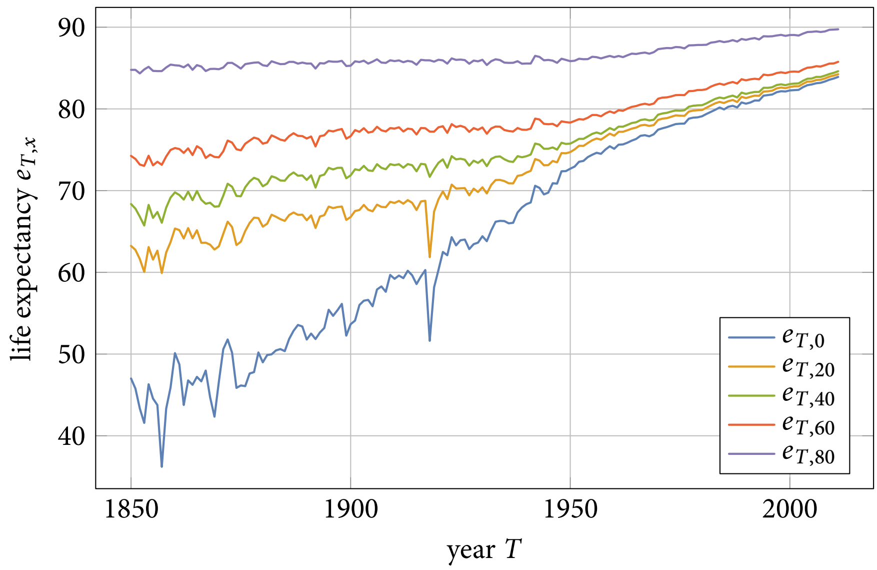


Figure 1.11: Life expectancy at birth, age 20, 40, 60, and 80 of Swedish females (synthetic cohorts) from the periods 1850-2011.

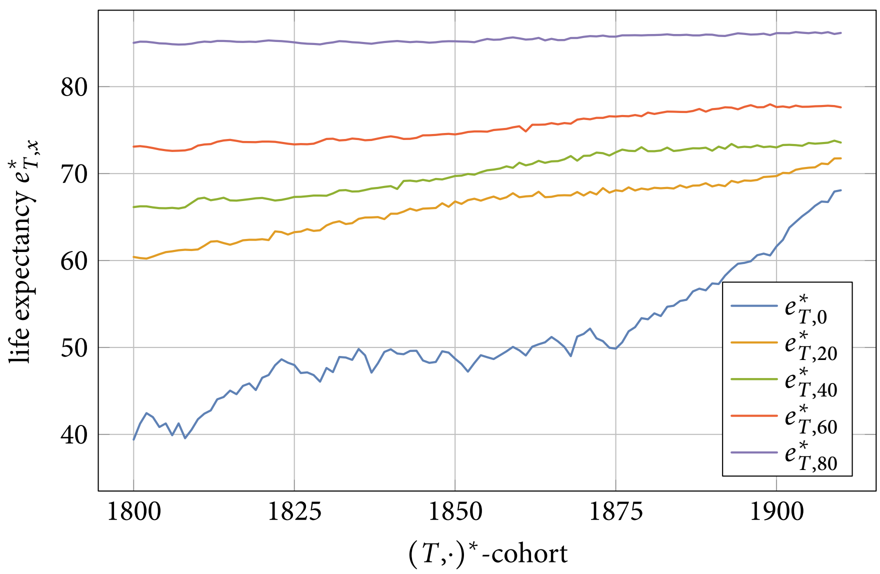


Figure 1.12: Real cohort life expectancy $e^*_{T,x}$ at age $x \in \{0, 20, 40, 60, 80\}$ for Swedish female $(T,x)^*$-cohorts.

1.4 | Stochastic Mortality Models

The development of human mortality, as presented in the previous section, has triggered a lot of academic attention. Thus, a vast of mortality models have been developed over the last 25 years. In this section, an overview of well-established and recently proposed mortality models is given.

1.4.1 | Model Quality Criteria

Before presenting some modelling approaches, we first provide a list of criteria which can be used to evaluate and compare different models. The following collection of criteria has been proposed and thoroughly discussed by Cairns, Blake and Dowd (2008, p. 87). Their requirements for a good model are:

- Mortality rates should be positive.
- The model should be consistent with historical data.
- Long-term dynamics under the model should be biologically reasonable.
- Parameter estimates should be robust relative to the period of data and range of ages employed.
- Model forecasts should be robust relative to the period of data and range of ages employed.
- Forecast levels of uncertainty and central trajectories should be plausible and consistent with historical trends and variability in mortality data.
- The model should be straightforward to implement using analytical methods or fast numerical algorithms.
- The model should be relatively parsimonious.
- It should be possible to use the model to generate sample paths and calculate prediction intervals.
- The structure of the model should make it possible to incorporate parameter uncertainty in simulations.
- At least for some countries, the model should incorporate a stochastic cohort effect.
- The model should have a non-trivial correlation structure, i.e., the mortality improvements should not be perfectly correlated for all ages.

Most of the provided criteria are self-explanatory and reasonable. For some criteria, a subjective interpretation remains, such as for a biologically reasonable long-term behaviour. Other points have contrary objectives and require trade-offs, such as the criteria for model parsimoniousness and accuracy. Plat (2009) points out that the existing models meet most of the listed criteria, but none of the models fulfils all of them.

1.4.2 | Review of Stochastic Mortality Models

In the following, we introduce the notation and common components of stochastic mortality models. Stochastic mortality models are either focusing on central death rates $m_{t,x}$ or initial death rates $q_{t,x}$ as measures of mortality.

Let the random variable $D_{t,x}$ denote the number of deaths of a population group at age x during the calendar year t. Moreover, let $d_{t,x}$ denote the observed death count, $E^c_{t,x}$ the central exposed to risk, i.e., the average population size aged x in calendar year t, and $E^0_{t,x}$ the initial number exposed to risk. In this notation, the one-year initial mortality rate $q_{t,x}$ is defined as the ratio $q_{t,x} = D_{t,x}/E^0_{t,x}$ and can be estimated by $\hat{q}_{t,x} = d_{t,x}/E^0_{t,x}$. Alternatively, the central mortality rate $m_{t,x}$ is defined as $m_{t,x} = D_{t,x}/E^c_{t,x}$, with the empirical estimate $\hat{m}_{t,x} = d_{t,x}/E^c_{t,x}$. If the underlying population data only contains the initial exposures, one usually uses an approximation of the central exposures as the mean of the initial exposures of two consecutive years t and $t+1$, i.e.,

$$E^c_{t,x} \approx \frac{E^0_{t,x} + E^0_{t+1,x}}{2} = \frac{E^0_{t,x} + E^0_{t,x} - d_{x,t}}{2} = E^0_{t,x} - \frac{1}{2} d_{t,x}. \tag{1.1}$$

Consistently to Brouhns, Denuit and Vermunt (2002), we assume for the following that the number of deaths $D_{t,x}$ is Poisson distributed

$$D_{t,x} \sim \mathrm{Poi}\left(m_{t,x} E^c_{t,x}\right). \tag{1.2}$$

with mean $m_{t,x}E^c_{t,x}$. Despite the fact that human hazard rates follow complicated structures, they vary slowly during a period of one year. A common simplification is therefore that individuals with the same age x at year t experience the same piecewise hazard rates $\mu_{t,x}$. By that assumption, the central mortalities provide decent approximations of the hazard rates $\mu_{t,x}$. Moreover, we have the following relation between the two mortality measures

$$q_{t,x} = 1 - e^{-m_{t,x}}.$$

The random variable of the death counts $D_{t,x}$ in terms of the initial mortality rate $q_{t,x}$ follows a binomial distribution, i.e.,

$$D_{t,x} \sim \mathrm{Bin}\left(q_{t,x}, E^0_{t,x}\right), \tag{1.3}$$

see, Currie (2016). The maximum likelihood estimates of $m_{t,x}$ and $q_{t,x}$, derived from eqs. (1.2) and (1.3), are given by (Rinne, 2014, Section 5.1)

$$\hat{m}_{t,x} = \frac{d_{t,x}}{E^c_{t,x}},$$
$$\hat{q}_{t,x} = \frac{d_{t,x}}{E^0_{t,x}}.$$

Therefore, the central or initial death rates are sometimes directly defined as the maximum likelihood estimators of the hazard rate, which is assumed to be constant on small intervals of time and age. Models based on eqs. (1.2) and (1.3) are therefore called Poisson or binomial regression models, respectively. The choice of the mortality measure is primarily affected by the available data. The Poisson model requires central exposures to risk $E^c_{t,x}$, which are more common (e.g., in Human Mortality Database) than the initial exposures to risk $E^0_{t,x}$.

1.4.3 | Generalized Age-Period-Cohort Models

The vast majority of stochastic mortality models decompose mortality rates or other mortality measures across the dimensions of age, period, and cohort. Recent contributions, for instance, Hunt and Blake (2014), Currie (2016) and Villegas, Kaishev and Millossovich (2015) showed that many proposed stochastic mortality models are covered by the family of the Generalized Age-Period-Cohort (GAPC) mortality models which can be expressed in the framework of generalized linear and non-linear models (see, e.g., McCullagh and Nelder, 1989)). The framework of the following section is known and largely based on the contributions of Currie (2016), Hunt and Blake (2014) and Villegas, Kaishev and Millossovich (2015). We start with an introduction to the GAPC model class and its building blocks and subsequently review some popular models which have been proposed in the literature.

Following the definition of Villegas, Kaishev and Millossovich (2015), a GAPC model is composed of four components:

(a) The *random component* encompasses a distribution assumption on the death count $D_{t,x}$. A common assumption is that the number of deaths $D_{t,x}$ follow a Poisson distribution or a binomial distribution, i.e.,

$$D_{t,x} \sim \text{Poi}(\mu_{t,x} E^c_{t,x}) \tag{1.4}$$

or

$$D_{t,x} \sim \text{Bin}(q_{t,x}, E^0_{t,x}) \tag{1.5}$$

with expectations $\mathbb{E}[D_{t,x}/E^c_{t,x}] = \mu_{t,x}$ and $\mathbb{E}[D_{t,x}/E^0_{t,x}] = q_{t,x}$, respectively. More precisely, one assumes that conditionally on $\mu_{t,x}$, or similarly on $q_{t,x}$, the random variables $D_{t,x}$ are independent and follow Poisson or binomial distributions, respectively. Note, in more general settings, as for GLMs, for some response variable Y and predictor variable X, the conditional expectation $\mathbb{E}[Y \mid X]$, is a member of the exponential family distribution. This conditional factor structure of the GAPC models will be discussed in more detail in Section 1.5.

(b) The *systematic component* captures the effects of the dimensions age x, calendar year t, and cohort $c = t - x$ through a linear or bilinear *predictor function* $\eta_{t,x}$ of the form

$$\eta_{t,x} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} + \beta_x^{(0)} \gamma_{t-x}.$$

The predictor contains the following components:

- The first term α_x is called the static age function, aiming to capture the general shape of the corresponding mortality measure.
- The second part contains a set of $N \geq 0$ bilinear age/periodic terms $\beta_x^{(i)} \kappa_t^{(i)}$, for $i = 1, \ldots, N$, where the period functions $\kappa_t^{(i)}$ determine the mortality change through time, and the age functions $\beta_x^{(i)}$ serve as modulations of periodic trends across the ages.

- The last term γ_{t-x} is the cohort term aiming to capture mortality effects based on the year of birth. These effects can include age-specific modifications through the term $\beta_x^{(0)}$.

As we will see later by reviewing the proposed models of recent decades, the age-specific modulating term $\beta_x^{(i)}$, $i = 0,1,\ldots,N$, can either be a specific analytic function, i.e., $\beta_x^{(i)} \equiv f^{(i)}(x)$ or non-parametric without any pre-specified structure. A key assumption of the GAPC family is that the periodic terms κ_t and the cohort terms γ_{t-x} are all model factors for each period t and cohort $t-x$, rather than smooth functions of time or cohort (cf., e.g., Villegas, Kaishev and Millossovich, 2015). This assumption enables stochastic projections of mortality rates by applying time series methods for the estimated coefficients of those factors. This approach leads to probabilistic rather than to deterministic forecasts. It is important to point out that B-Spline models do not belong to GAPC family since they impose a polynomial functional form for the periodic terms.

(c) The *link function* g provides a connection between the random component and the systematic predictor. The link is a monotone and differentiable function that describes how the mean of the regression objective depends on the linear predictor, i.e.,

$$g\left(\mathbb{E}\left[\frac{D_{t,x}}{E_{t,x}}\right]\right) = \eta_{t,x}.$$

For the Poisson model, the canonical choice for the link function is the logarithmic function and for the binomial model the logit function, respectively. One of the requirements for an appropriate link function is that the transformed data are approximately linear (or multiplicative bilinear) since they are passed to a linear regressor. In Section 1.7, we will demonstrate that for high ages the logit link function leads to be fit than the canonical Poisson link, for the vast majority of stochastic mortality models.

(d) A set of *parameter constraints* to ensure model identification. An important characteristic of the most proposed stochastic mortality models is that their parameters are only identifiable up to transformations, i.e., the model parameters

$$\theta := \left(\alpha_x, \beta_x^{(1)}, \ldots, \beta_x^{(N)}, \kappa_t^{(1)}, \ldots, \kappa_t^{(N)}, \beta_x^{(0)}, \gamma_{t-x}\right)$$

can be transformed by a map ν to equivalent parameters

$$\nu(\theta) = \tilde{\theta} = \left(\tilde{\alpha}_x, \tilde{\beta}_x^{(1)}, \ldots, \tilde{\beta}_x^{(N)}, \tilde{\kappa}_t^{(1)}, \ldots, \tilde{\kappa}_t^{(N)}, \tilde{\beta}_x^{(0)}, \tilde{\gamma}_{t-x}\right),$$

such that the predictor $\eta_{t,x}$ remains unchanged. The implications and drawbacks of required parameter constraints will be discussed later.

Many proposed mortality models can be expressed within the GAPC modelling framework. Some models assume Poisson distributed death counts with a log link function, for instance, the model proposed by Lee and Carter (1992) and the extensions proposed by Renshaw and Haberman (2003, 2006). Other models assume binomial distributed death counts and a logit link to target the logit transformed initial death rates as the regression objective, see, e.g., Cairns, Blake and Dowd (2006) and the extensions in Cairns, Blake, Dowd et al. (2009), and also Aro and Pennanen (2011). Further models of the GAPC family can be found in Alai and Sherris (2014), Berkum, Antonio and Vellekoop (2014), Börger, Fleischer and Kuksin (2014), Haberman and Renshaw (2009), Lovász (2011), O'Hare and Y. Li (2012) and Plat (2009).

Some examples of proposed mortality models, which do not belong to the GAPC family, can be found in Currie, Durban and Eilers (2004), Renshaw, Haberman and Hatzopoulos (1996) and Sithole, Haberman and Verrall (2000). As Hunt and Blake (2014) point out, the models assume the periodic functions $\kappa_t^{(i)}$ to be cubic B/P-splines or Legendre polynomials, respectively. Projections based on these models are obtained by extrapolation of deterministic functions. Thus, the application of these models is restricted to short-term forecasts. On the other hand, the P-splines approach of Currie, Durban and Eilers (2004) proved to be very useful for smoothing and data regulation to identify general trends of mortality change, as well as, reveal cohort effects, which are difficult to detect with crude mortality rates. An application of P-spline smoothing is shown in Figures 1.7 to 1.10.

In the following, we describe some established GAPC models which have received considerable attention from the scientific community. The review on proposed models is based on the survey articles of Hunt and Blake (2014) and Villegas, Kaishev and Millossovich (2015). The second paper introduces an **R** package called **StMoMo**[1], which provides an excellent tool for parameter estimation of GAPC models, assessing their goodness-of-fit, and performing projections using ARIMA models.

The Lee-Carter Model

The Lee-Carter (LC) model was one of the first stochastic mortality models and still remains widely used. The LC model in the original form, as proposed by Lee and Carter (1992), does not fit into the GAPC family, since it misses a random component. Thus, we use the implementation of Brouhns, Denuit and Vermunt (2002), which assumes Poisson distributed death counts, the log link function to target the force of mortality $\mu_{t,x}$, and the original predictor function with a static age function α_x, one non-parametric age-periodic term $\beta_x^{(1)}\kappa_t^{(1)}$, and no cohort effect. More precisely, the predictor is given by

$$\eta_{t,x} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)}. \tag{1.6}$$

The LC model belongs to the simplest GAPC models. As we will see later, it forms a basis for further model extensions. One key property of that model, and therefore also for all generalizations, is that the parameters are not identifiable without additional constraints, i.e., the model response variable remains unchanged under the parameter transformation ν, with

$$\nu : \left(\alpha_x, \beta_x^{(1)}, \kappa_t^{(1)}\right) \mapsto \left(\alpha_x + c_1\beta_x^{(1)}, \frac{1}{c_2}\beta_x^{(1)}, c_2(\kappa_t^{(1)} - c_1)\right), \tag{1.7}$$

1 The acronym stands for **St**ochastic **Mo**rtality **Mo**delling. The source code is available at https://github.com/amvillegas/StMoMo.

for arbitrary real constants c_1 and $c_2 \neq 0$. To overcome the identification issue, Lee and Carter suggest to impose the following two parameter constraints

$$\sum_x \beta_x^{(1)} = 1, \qquad \sum_t \kappa_t^{(1)} = 0. \tag{1.8}$$

These constraints can be achieved by

$$c_1 = \frac{1}{n}\sum_t \kappa_t^{(1)}, \qquad c_2 = \sum_x \beta_x^{(1)}$$

in the transformation of eq. (1.7). LC model forecasts are attained by using ARIMA processes for the time index $\kappa_t^{(1)}$. A common choice proposed in the literature, is to use a random walk with drift, i.e.,

$$\kappa_t^{(1)} = \delta + \kappa_{t-1}^{(1)} + \xi_t, \qquad \xi_t \sim \mathcal{N}(0,\sigma_\kappa),$$

with δ being the drift parameter and ξ_t a Gaussian white noise process with variance σ_κ. The strength of the LC model and its extensions is their flexibility to capture age-specific mortality pattern over large ranges of ages. However, one of the disadvantages of the LC model is that it only allows a trivial correlation structure of the projected mortality rates, i.e., the changes in the mortality are perfectly correlated across all ages (cf., Cairns, Blake and Dowd, 2006).

The Cairns-Black-Dowd Model

To overcome the issue of a single age-period factor with trivial correlated projected mortality rates, Cairns, Blake and Dowd (2006) proposed a model with two age-periodic terms and parametric age-modulations $\beta_x^{(1)} = 1$ and $\beta_x^{(2)} = x - \bar{x}$, where $\bar{x}$ denotes the average age in the underlying data. The model does not include an age-specific function α_x, nor a cohort function γ_{t-x}. The predictor function of the Cairns-Black-Dowd (CBD) model is given by

$$\eta_{t,x} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}. \tag{1.9}$$

The absence of a static age function and the assumption of parametric age-modulations, results in a more parsimonious model compared to the LC model. The CBD assumption of approximately linear log-transformed mortality rates on the entire age range restricts the application of the model to high ages only. The key characteristics of the CBD model, and the unique property across other models covered here, is that the model does not require any parameter constraints to be identifiable. The original approach of Cairns, Blake and Dowd (2006) used ordinary least squares parameter estimation targeting the logit transformed initial death rates $q_{t,x}$. Haberman and Renshaw (2011) used the original predictor function and the assumption of binomial distributed death counts with a logit link function, to adapt the model to the framework of generalized linear models. In order to obtain forecasts of mortality rates, Cairns, Blake and Dowd (2006) employed a two-dimensional random walk for the periodic terms $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. Empirical studies have shown that the obtained time series $(\kappa_t^{(1)},\kappa_t^{(2)})$ are not stationary. Using a two-dimensional random walk leads to diverging prediction intervals of the forecasts. In Chapter 2, we will propose an approach, which uses cointegration relations between the components of time series to obtain smaller prediction intervals.

The Renshaw and Haberman Model

Renshaw and Haberman (2006) proposed an extension of the Lee-Carter model by an incorporation of an additional cohort term to the LC predictor in eq. (1.6). The predictor of the (generalized) Renshaw-Haberman (RH) model satisfies the equation

$$\eta_{t,x} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(0)}\gamma_{t-x}. \tag{1.10}$$

The predictor is invariant under the following transformation

$$v : \left(\alpha_x, \beta_x^{(1)}, \kappa_t^{(1)}, \beta_x^{(0)}, \gamma_{t-x}\right) \mapsto \left(\alpha_x + c_1\beta_x^{(1)} + c_2\beta_x^{(1)}, \frac{1}{c_3}\beta_x^{(1)}, c_3(\kappa_t^{(1)} - c_1), \frac{1}{c_4}\beta_x^{(0)}, c_4(\gamma_{t-x} - c_2)\right), \tag{1.11}$$

with real constants c_1, c_2, $c_3 \neq 0$ and $c_4 \neq 0$. The suggested parameter constraints by Cairns, Blake, Dowd et al. (2009) in order to ensure model identification are

$$\sum_x \beta_x^{(1)} = 1, \qquad \sum_t \kappa_t^{(1)} = 0, \qquad \sum_x \beta_x^{(0)} = 1, \qquad \sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c = 0.$$

These can be imposed by the following choice of parameters

$$c_1 = \frac{1}{n}\sum_t \kappa_t^{(1)}, \quad c_2 = \frac{1}{n+k-1}\sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c, \quad c_3 = \sum_x \beta_x^{(1)}, \quad c_4 = \sum_x \beta_x^{(0)}$$

in the transformation of eq. (1.11), where n denotes the number of periods and k the number of ages covered in the mortality data.

The RH model has been criticized for its slow convergence and the lack of robustness due to high sensitivity to the choice of the initial parameters, see, e.g., Cairns, Blake, Dowd et al. (2011, 2009) and Hunt and Villegas (2015). Since the RH model generalizes the LC model, Currie (2016) suggested using the estimated parameters of the LC model as starting values to overcome convergence issues which have been also encountered by, e.g., Macdonald, Gallop, Miller et al. (2007). Renshaw and Haberman (2006) also considered nested models obtained by restricting the age modulating terms $\beta_x^{(i)}$, $i = 0{,}1$ to constants. Of particular interest, is the submodel

$$\eta_{t,x} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \gamma_{t-x}, \tag{1.12}$$

which resolves some stability problems due to the simplification $\beta_x^{(0)} = 1$ in eq. (1.10), see Hunt and Villegas (2015) for a deeper discussion on robustness and convergence issues of RH models. The restricted model with the predictor given in eq. (1.12) will be referred to as the RH model in the further course of the thesis. Mortality projections of the RH model are obtained by assuming that the age-specific effects α_x and $\beta_x^{(1)}$ remain constant over time, whereas $\kappa_t^{(1)}$ and γ_{t-x} are modelled by univariate ARIMA processes, under the assumption of independence between periodic and cohort effects. The independence assumption is common in the literature of GAPC models. However, it is questionable and has been criticized by, e.g., Currie (2012).

Extensions of the predictor function in eq. (1.10) have been proposed in Berkum, Antonio and Vellekoop (2014) by postulating a predictor of the form

$$\eta_{t,x} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \gamma_{t-x}. \tag{1.13}$$

Further extensions to multiple cohort terms with predictors of the type

$$\eta_{t,x} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)}\kappa_t^{(i)} + \sum_{j=N+1}^{M} \beta_x^{(j)}\gamma_{t-x}^{(j)}.$$

have been considered by Hatzopoulos and Haberman (2011). According to Hunt and Blake (2014), there was no evidence found in practice for the demand of multiple cohort terms.

The Classic Age-Period-Cohort Model

Another nested model of the generalized RH model, with predictor given in eq. (1.10), is the model obtained by restricting both age-specific effects to $\beta_x^{(0)} = \beta_x^{(1)} = 1$. This leads to the predictor

$$\eta_{t,x} = \alpha_x + \kappa_t^{(1)} + \gamma_{t-x}. \tag{1.14}$$

As Hunt and Blake (2014) and Villegas, Kaishev and Millossovich (2015) point out, despite the fact that the model of the APC type has been traditionally used by, e.g., Clayton and Schifflers (1987) and Hobcraft, Menken and Preston (1985), it was Currie (2006) who introduced the particular model type to the actuarial science. One can show that the response variable $\eta_{t,x}$ is invariant under the parameter transformations

$$\nu_\phi : \left(\alpha_x, \kappa_t^{(1)}, \gamma_{t-x}\right) \mapsto \left(\alpha_x + \phi_1 - \phi_2 x, \kappa_t^{(1)} + \phi_2 t, \gamma_{t-x} - \phi_1 - \phi_2(t-x)\right)$$

and

$$\nu_c : \left(\alpha_x, \kappa_t^{(1)}, \gamma_{t-x}\right) \mapsto \left(\alpha_x + c_1, \kappa_t^{(1)} - c_1, \gamma_{t-x}\right), \tag{1.15}$$

with real constants c_1, ϕ_1 and ϕ_2. One possible set of parameter constraints to ensure identification is

$$\sum_t \kappa_t^{(1)} = 0, \qquad \sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c = 0, \qquad \sum_{c=t_1-x_k}^{t_n-x_1} c\gamma_c = 0, \tag{1.16}$$

where the last two constraints are employed to obtain a cohort effect with zero mean and no linear trend. To impose the two constraints on the cohort term Haberman and Renshaw (2011) suggest to use the transformation ν_ϕ with constants ϕ_1, ϕ_2 obtained by the regression

$$\gamma_{t-x} = \phi_1 + \phi_2(t-x) + \varepsilon_{t-x},$$

with i.i.d. $\varepsilon_{t-x} \sim \mathcal{N}(0,\sigma^2)$. Subsequently, the first constraint in eq. (1.16) can be achieved by using the transformation of eq. (1.15) with

$$c_1 = \frac{1}{n}\sum_t \kappa_t^{(1)}.$$

Comparing the parameter estimation of the RH model and the APC model, which only differ by the functional restriction of the age-specific term of the latter model to $\beta_x^{(1)} \equiv 1$, we observe a significant difference of the running time. The convergence time of the parameter estimation for the RH model is 2 order of magnitudes larger than the parameter estimation time of the APC model. These convergence difficulties combined with robustness issues make the RH model less appealing for practical use.

M7 Model

The M7 model is an extended CBD model with additional quadratic age effect and a cohort term proposed by Cairns, Blake, Dowd et al. (2009). The predictor of the model is given by

$$\eta_{t,x} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \left((x-\bar{x})^2 - \tilde{\sigma}_x^2\right)\kappa_t^{(3)} + \gamma_{t-x}, \tag{1.17}$$

where $\bar{x}$ denotes the mean age and $\tilde{\sigma}_x^2 = 1/k\sum_x (x-\bar{x})^2$, with k being the number of considered ages. The extension of the CBD predictor leads to additional parameter constraints to ensure identification. The M7 predictor is invariant under the transformation

$$\begin{aligned}\nu : \left(\kappa_t^{(1)}, \kappa_t^{(3)}, \kappa_t^{(3)}, \gamma_{t-x}\right) \mapsto \Big(&\kappa_t^{(1)} + \phi_1 + \phi_2(t-\bar{x}) + \phi_3\left((t-\bar{x})^2 + \tilde{\sigma}^2\right), \kappa_t^{(2)} - \phi_2 - 2\phi_3(t-\bar{x}),\\ &\kappa_t^{(3)} + \phi_3, \gamma_{t-x} - \phi_1 - \phi_2(t-x) - \phi_3(t-x)^2\Big),\end{aligned} \tag{1.18}$$

where ϕ_1, ϕ_2 and ϕ_3 are real constants. Cairns, Blake, Dowd et al. (2009) suggest imposing the following set of constraints

$$\sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c = 0, \qquad \sum_{c=t_1-x_k}^{t_n-x_1} c\gamma_c = 0, \qquad \sum_{c=t_1-x_k}^{t_n-x_1} c^2\gamma_c = 0, \tag{1.19}$$

on the cohort term to achieve parameter identification. The intention of the given constraints is to obtain a cohort term with zero mean and no linear or quadratic trends. Following Haberman and Renshaw (2011), this can be achieved by the transformation of eq. (1.18) using the coefficients of the regression of γ_{t-x} on $t-x$ and $(t-x)^2$, i.e.,

$$\gamma_{t-x} = \phi_1 + \phi_2(t-x) + \phi_3(t-x)^2 + \varepsilon_{t-x},$$

with i.i.d. error term $\varepsilon_{t-x} \sim \mathcal{N}(0,\sigma^2)$. Note, the first constraint of the set given in eq. (1.19) is sufficient to reduce the degrees of freedom and already ensures identifiability. Further constraints impose additional structures on the cohort term γ_{t-x} with the purpose to obtain desirable properties, such that the resulted series of parameter estimates might be able to be projected using stationary processes. Cairns, Blake, Dowd et al. (2009) also investigate simpler predictors such as

$$\eta_{t,x} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \gamma_{t-x} \tag{1.20}$$

and

$$\eta_{t,x} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + (x_c - x)\gamma_{t-x}, \tag{1.21}$$

where x_c is a parameter which has to be estimated. The predictor of eqs. (1.20) and (1.21) are referred to as the models M6 and M8, respectively. Comparing the predictors of eqs. (1.17) and (1.20) reveals that the model M7 nests the model M6, i.e., M6 is a submodel of M7.

Plat Model

The model proposed by Plat (2009) attempts to combine the features of previous models. Similar to the Lee-Carter model, it incorporates a static age function α_x, which makes it applicable to larger age ranges. Furthermore, it uses a cohort term with a pre-specified age-modulation $\beta_x^{(0)}$, like the APC model, and it uses three age-modulating parameters $\beta_x^{(1)} \equiv 1$, $\beta_x^{(2)} = (\bar{x} - x)$, and $\beta_x^{(3)} = (\bar{x} - x)^+ = \max 0, (\bar{x} - x)$. The predictor function, which will be referred to as the general PLAT model, is given by

$$\eta_{t,x} = \alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + (\bar{x} - x)^+\kappa_t^{(3)} + \gamma_{t-x}. \tag{1.22}$$

The original model of Plat (2009) targets the mortality rate $\mu_{t,x}$ using a log link function as the canonical link and Poisson distributed death counts $D_{t,x}$. Like many other mortality models, the PLAT model is not identifiable. The predictor is invariant under the transformations

$$\begin{aligned} \nu_\phi : \left(\alpha_x, \kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \tilde{\gamma}_{t-x}\right) \mapsto \Big(&\alpha_x + \phi_1 - \phi_2 x + \phi_3 x^2, \kappa_t^{(1)} + \phi_2 t + \phi_3(t^2 - 2\bar{x}t), \\ &\kappa_t^{(2)} + 2\phi_3 t, \kappa_t^{(3)}, \gamma_{t-x} - \phi_1 - \phi_2(t - x) - \phi_3(t - x)^2\Big) \end{aligned} \tag{1.23}$$

and

$$\begin{aligned} \nu_c : \left(\alpha_x, \kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_{t-x}\right) \mapsto \Big(&\alpha_x + c_1 + c_2(\bar{x} - x) + c_3(\bar{x} - x)^+, \\ &\kappa_t^{(1)} - c_1, \kappa_t^{(2)} - c_2, \kappa_t^{(3)} - c_3, \gamma_{t-x}\Big), \end{aligned}$$

for some real constants $c_1, c_2, c_3, \phi_1, \phi_2$, and ϕ_3. Parameter identification can be ensured by imposing the following constraints

$$\sum_t \kappa_t^{(1)} = 0, \quad \sum_t \kappa_t^{(2)} = 0, \quad \sum_t \kappa_t^{(3)} = 0, \quad \sum_{c=t_1-x_k}^{t_n-x_1} \gamma_c = 0, \quad \sum_{c=t_1-x_k}^{t_n-x_1} c\gamma_c = 0, \quad \sum_{c=t_1-x_k}^{t_n-x_1} c^2\gamma_c = 0. \tag{1.24}$$

The constraints on the cohort term γ_{t-x} can be achieved by using the transformation ν_ϕ given in eq. (1.23) with coefficients ϕ_i, for $i \in \{1,2,3\}$, obtained by the regression

$$\gamma_{t-x} = \phi_1 + \phi_2(t-x) + \phi_3(t-x)^2 + \varepsilon_{t-x},$$

with an i.i.d. error $\varepsilon_{t-x} \sim \mathcal{N}(0,\sigma^2)$. Note, not all imposed constraints are necessary for identification. The rationale behind the constraints on γ_{t-x} is to ensure that the fitted process will fluctuate around zero with no linear or quadratic trend. This approach aims to force the process γ_{t-x} only to capture the cohort effects and not just to be a compensation for the deficiency of the age-period terms, see Plat (2009). The normalization constraints on the period terms $\kappa_t^{(i)}$, for $i \in \{1,2,3\}$, can be enforced by the transformation ν_c using the constants

$$c_i = \frac{1}{n}\sum_t \kappa_t^{(i)}.$$

Plat (2009) suggests a more parsimonious model if only older ages (above 60) are considered. In that particular case, Plat suggests that the third periodic factor can be excluded from the predictor. Thus,

$$\eta_{t,x} = \alpha_x + \kappa_t^{(1)} + (\bar{x}-x)\kappa_t^{(2)} + \gamma_{t-x},$$

is the resulted predictor of that nested model, which is essentially the M6 model with an additional static age term α_x. Since we are interested in mortality modelling of the elderly population, we will use the reduced predictor in the case studies on Section 1.7.

A modification of the general Plat model, with particular interest of the application to younger ages, has been investigated by O'Hare and Y. Li (2012). The paper suggests the following predictor

$$\eta_{t,x} = \alpha_x + \kappa_t^{(1)} + (\bar{x}-x)\kappa_t^{(2)} + \left((\bar{x}-x)^+ + ((\bar{x}-x)^+)^2\right)\kappa_t^{(3)} + \gamma_{t-x},$$

where the third age-modulation $((\bar{x}-x)^+)^2$ serves as the PLAT model modification.

We conclude the description of popular GAPC models by providing an overview of the predictors together with references to the original papers in Table 1.2. In Table 1.3, we give an overview of suggested parameter constraints associated to these models. Note, the provided approaches to ensure identifiability are not necessarily unique or minimal. In some cases, there are additional constraints on the parameter terms in order to employ stationary processes for forecasting. Figure 1.13 provides an illustration of a structural classification scheme for the introduced mortality models, where the distinct classification layers emerge by structural forms of the predictor terms, cf., Hunt and Blake (2014).

Forecasting Procedure of GAPC Mortality Models

Following the presentation of Villegas, Millossovich and Kaishev (2016), we recall that forecasts of GAPC mortality models are obtained by applying time series methods on the period terms $\kappa_t^{(i)}$, $i = 1,\dots,N$, and the cohort term γ_{t-x}. The following two-stage procedure has emerged in the literature. The initial step is to fit a GAPC model and then treat the estimates of period or cohort terms as time series. The final step is to choose an appropriate time discrete process and estimate its parameters. This process is then used for forecasting.

The standard time series approach for period terms $\boldsymbol{\kappa}_t = \left(\kappa_t^{(1)}, \ldots, \kappa_t^{(N)}\right)$ is to employ a multivariate random walk with drift. The particular assumption for the dynamics of $\boldsymbol{\kappa}_t$ is

$$\boldsymbol{\kappa}_t = \boldsymbol{\delta} + \boldsymbol{\kappa}_{t-1} + \boldsymbol{\varepsilon}_t^{\kappa}, \qquad \boldsymbol{\varepsilon}_t^{\kappa} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{1.25}$$

where $\boldsymbol{\delta}$ denotes the drift term and $\boldsymbol{\Sigma}$ the variance-covariance matrix of the multivariate white noise $\boldsymbol{\varepsilon}_t^{\kappa}$. The approach of eq. (1.25) has been used among others in studies of, e.g., Cairns, Blake and Dowd (2006), Cairns, Blake, Dowd et al. (2011) and Haberman and Renshaw (2011). For the dynamics of the cohort term, univariate autoregressive integrated moving average processes have been considered, see, e.g., Cairns, Blake, Dowd et al. (2011) and Renshaw and Haberman (2006). The common approach for the cohort γ_c is an ARIMA(p,q,d) process of the form

$$\Delta^d \gamma_c = \delta_0 + \phi_1 \Delta^d \gamma_{c-1} + \cdots + \phi_p \Delta^d \gamma_{c-p} + \varepsilon_c + \delta_1 \varepsilon_{c-1} + \cdots + \delta_q \varepsilon_{c-q},$$

where Δ^d is the difference operator of order d, $\phi_1, \ldots, \phi_p$ are the autoregressive coefficients and $\delta_1, \ldots, \delta_q$ the moving average coefficients. δ_0 denotes the drift parameter and ε_c the Gaussian white noise process with variance σ_ε, which is independent of $\boldsymbol{\varepsilon}_t^{\kappa}$. Some concrete approaches for the cohort dynamics include an ARIMA$(1,1,0)$ for the APC and the RH model and an ARIMA$(2,0,0)$ for the M7 and PLAT predictor types, see Villegas, Millossovich and Kaishev (2016). In Section 1.8, we will examine the cohort estimates on a reference dataset in order to determine whether the assumption of ARIMA type time series is indeed justified.

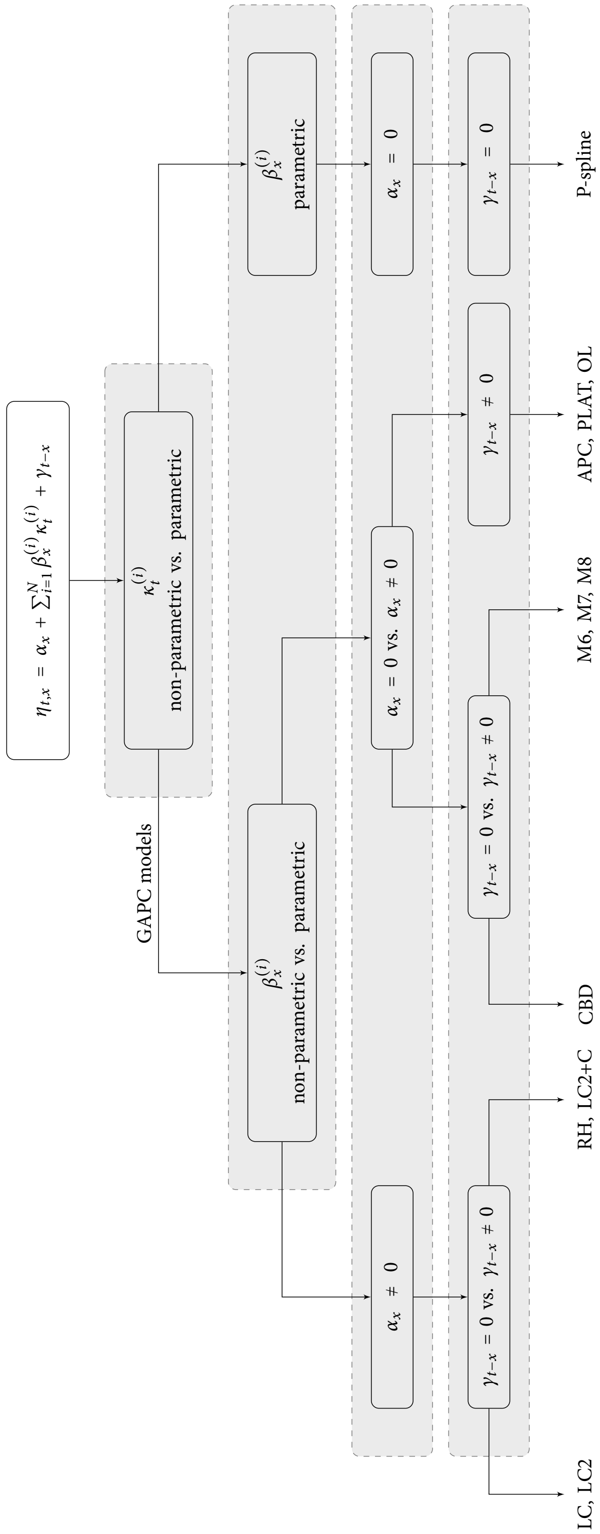


Figure 1.13: Classification layers of mortality models.

Table 1.2: Overview of proposed predictor specifications in the recent literature.

Model	Predictor	Original paper
LC	$\eta_{t,x} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)}$	Lee and Carter (1992)
LC2	$\eta_{t,x} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)}$	Renshaw and Haberman (2003)
LC2+C	$\eta_{t,x} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \gamma_{t-x}$	van Berkum et al. (2014)
RH	$\eta_{t,x} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(0)}\gamma_{t-x}$	Renshaw and Haberman (2006)
APC	$\eta_{t,x} = \alpha_x + \kappa_t^{(1)} + \gamma_{t-x}$	Currie (2006)
CBD	$\eta_{t,x} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}$	Cairns, Blake and Dowd (2006)
M6	$\eta_{t,x} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \gamma_{t-x}$	Cairns, Blake, Dowd et al. (2009)
M7	$\eta_{t,x} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \left((x-\bar{x})^2 - \hat{\sigma}_x^2\right)\kappa_t^{(3)} + \gamma_{t-x}$	Cairns, Blake, Dowd et al. (2009)
M8	$\eta_{t,x} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + (x_c - x)\gamma_{t-x}$	Cairns, Blake, Dowd et al. (2009)
PLAT	$\eta_{t,x} = \alpha_x + \kappa_t^{(1)} + (\bar{x}-x)\kappa_t^{(2)} + (\bar{x}-x)^+\kappa_t^{(3)} + \gamma_{t-x}$	Plat (2009)
OL	$\eta_{t,x} = \alpha_x + \kappa_t^{(1)} + (\bar{x}-x)\kappa_t^{(2)}$ $\left((\bar{x}-x)^+ + ((\bar{x}-x)^+)^2\right)\kappa_t^{(3)} + \gamma_{t-x}$	O'Hare and Y. Li (2012)

Table 1.3: Overview of the suggested parameter constraints to ensure parameter identification.

Model	Constraints				
LC	$\sum_x \beta_x^{(1)} = 1$	$\sum_t \kappa_t^{(1)} = 0$			
LC2	$\sum_x \beta_x^{(1)} = 1$	$\sum_t \kappa_t^{(1)} = 0$	$\sum_x \beta_x^{(2)} = 1$	$\sum_t \kappa_t^{(2)} = 0$	
LC2+C	$\sum_x \beta_x^{(1)} = 1$	$\sum_t \kappa_t^{(1)} = 0$	$\sum_x \beta_x^{(2)} = 1$	$\sum_t \kappa_t^{(2)} = 0$	$\sum_{c=t-x} \gamma_c = 0$
RH	$\sum_x \beta_x^{(1)} = 1$	$\sum_t \kappa_t^{(1)} = 0$	$\sum_x \beta_x^{(0)} = 1$	$\sum_{c=t-x} \gamma_c = 0$	
APC	$\sum_t \kappa_t^{(1)} = 0$	$\sum_{c=t-x} \gamma_c = 0$	$\sum_{c=t-x} c\gamma_c = 0$		
CBD	–				
M6	$\sum_{c=t-x} \gamma_c = 0$	$\sum_{c=t-x} c\gamma_c = 0$			
M7	$\sum_{c=t-x} \gamma_c = 0$	$\sum_{c=t-x} c\gamma_c = 0$	$\sum_{c=t-x} c^2\gamma_c = 0$		
M8	$\sum_{c=t-x} \gamma_c = 0$				
PLAT	$\sum_t \kappa_t^{(i)} = 0$	$\sum_{c=t-x} \gamma_c = 0$	$\sum_{c=t-x} c\gamma_c = 0$	$\sum_{c=t-x} c^2\gamma_c = 0$	
OL	$\sum_t \kappa_t^{(3)} = 0$	$\sum_{c=t-x} \gamma_c = 0$	$\sum_{c=t-x} c\gamma_c = 0$		

1.5 | Parameter Estimation for Generalized Linear Models

Currie (2016) noticed that many mortality models, in particular, the GAPC models, can be expressed as generalized linear models or generalized non-linear models. In this section, we provide the framework for parameter estimation of GAPC models for a general class of distributions, namely the exponential distribution family. As conducted by the seminal publication of McCullagh and Nelder (1989) and Nelder and Wedderburn (1972), if the dependent variable of the regression belongs to the exponential distribution family, then there is a unified procedure to cover those models.

1.5.1 | Exponential Family and its Properties

Suppose $Y_1, \dots, Y_N$ are independent random variables. Let $f_{Y_i}(y_i \mid \theta_i, \phi)$ denote either the probability density function or the probability mass function of the random variable Y_i. The distribution of Y_i belongs to the *exponential family* if $f_{Y_i}(y_i \mid \theta_i, \phi)$ can be written in the form

$$f_{Y_i}(y_i \mid \theta_i, \phi) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)\right), \tag{1.26}$$

for some fixed parameter ϕ, called the dispersion parameter, and θ_i, called the canonical parameter since it primarily determines the expectation. The functions $a_i(\cdot)$, $b(\cdot)$ and $c(\cdot,\cdot)$ are functions specifying distinct members of the exponential family, see, e.g., Table 1.4 for some examples. The function b is also referred to as the cumulant since it determines the moments as will be shown by eqs. (1.36) and (1.37). The primary purpose of the function c is normalization. For many distributions [1] the function $a_i(\phi)$ has the form

$$a_i(\phi) = \frac{\phi}{w_i},$$

where w_i is called the *prior weight*. Thus, eq. (1.26) takes the form

$$f_{Y_i}(y_i \mid \theta_i, \phi, w_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi} w_i + c(y_i, \phi, w_i)\right). \tag{1.27}$$

Many distributions like the normal, Poisson, and binomial distribution belong to the family of exponential distributions. For instance, for a Poisson distributed $Y_i \sim \mathcal{P}(\lambda_i)$ the probability mass function can be written as

$$f_{Y_i}(y_i \mid \lambda_i) = \frac{\lambda_i^{y_i}}{y_i!} e^{\lambda_i} = \exp\left(y_i \ln(\lambda_i) - \lambda_i - \ln(y_i!)\right).$$

For binomial $Y_i \sim \mathcal{B}(n_i, p_i)$, we have

$$f_{Y_i}(y_i \mid n_i, p_i) = \binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i}$$

1 For example, the Poisson, binomial, proportional binomial, and normal distribution, see Table 1.4

$$= \exp\left(y_i \ln\left(\frac{p_i}{1-p_i}\right) + n_i \ln(1-p_i) + \ln\binom{n_i}{y_i}\right),$$

and for normally distributed $Y_i \sim \mathcal{N}(\mu_i,\sigma)$ the density can be written as

$$\begin{aligned} f_{Y_i}(y_i \mid \mu_i,\sigma) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\mu_i)^2}{2\sigma^2}\right) \\ &= \exp\left(\frac{y_i\mu_i - \frac{\mu_i}{2}}{\sigma^2} - \frac{1}{2}\ln(2\pi\phi) - \frac{y^2}{2\phi}\right). \end{aligned}$$

The canonical parameter θ, the dispersion parameter ϕ and the functions $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ for the above examples are compiled in Table 1.4. Often, the parameters θ_i are not of primary interest since for the intended application, Y_i will represent the death counts and there will be as many θ_i as observations. Having as many parameters as observations would lead to a saturated model (see Section 1.6.2). Instead, one is interested in a lower dimensional parameter space, of parameters $\beta_0,\ldots,\beta_K$, where $K < N$. Let μ_i denote the mean of Y_i, i.e.,

$$\mu_i := \mathbb{E}[Y_i \mid \boldsymbol{x}_i],$$

then the key assumption of generalized linear models, as introduced by Nelder and Wedderburn (1972), is that the mean μ_i of the response variable Y_i (member of the exponential family) is coupled to a linear predictor η through a monotone and differentiable link function g, i.e.,

$$g(\mu_i) = \beta_1 x_{i1} + \ldots + \beta_K x_{iK} =: \eta. \tag{1.28}$$

Different predictor types of the GAPC family were introduced Section 1.4.2. For mortality models, the response variable Y_i usually represents either the death counts $D_{t,x}$ or the ratio $D_{t,x}/E_{t,x}$ and $\beta_1,\ldots,\beta_K$ represent a combination of static age, periodic or cohort parameters, while $x_{i1},\ldots,x_{iK}$ are the corresponding common factors age and period. The expression in eq. (1.28) can be abbreviated by using a vector notation which leads to

$$g(\mu_i) = \boldsymbol{x}_i^T\boldsymbol{\beta}, \tag{1.29}$$

where $\boldsymbol{x}_i$ is the i-th column of the model design matrix $\boldsymbol{X}$ and $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_K)^T$, such that the model takes the form

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = g\left(\mathbb{E}[\boldsymbol{Y} \mid \boldsymbol{X}]\right) = \boldsymbol{X}\boldsymbol{\beta}, \tag{1.30}$$

with $\boldsymbol{\eta} = (\eta_1,\ldots,\eta_N)^T$, $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_N)^T$ and $\boldsymbol{Y} = (Y_1,\ldots,Y_N)^T$. Some models, like the Poisson model for the death counts, use a slightly more general setting that extends the relation of eq. (1.29) to

$$g(\mu_i) = \boldsymbol{x}_i^T\boldsymbol{\beta} + \text{offset}_i.$$

The offset term of the Poisson model, as introduced in eq. (1.4), is the central exposure to risk, i.e., the average population number of particular age and period of interest. Note also that for the binomial model, as defined by the random component given in eq. (1.5), the regression is referred to the mean of the distribution $Y_i := X_i/n_i$, with $X_i \sim \mathrm{Bin}(p_i, n_i)$. In the mortality modelling setting, X_i represents the death counts and n_i the initial exposure to risk. In that particular case of a proportional binomial random variable Y_i, we have $\mathbb{E}[Y_i] = p_i$ and $\mathbb{V}[Y_i] = \frac{p_i(1-p_i)}{n_i}$. Furthermore, Y_i does also belong to the exponential family since we have

$$\begin{aligned}\mathbb{P}(Y_i = y_i) &= \binom{n_i}{n_i y_i} p_i^{n_i y_i} (1 - p_i)^{n_i - n_i y_i} \\ &= \exp\left(\frac{y_i \ln\left(\frac{p_i}{1-p_i}\right) + \ln(1 - p_i)}{\frac{1}{n_i}} + \ln\binom{n_i}{n_i y_i} \right).\end{aligned}$$

The above expression of the probability mass function satisfies the form of the exponential family as given in eq. (1.27), with the canonical parameter $\theta_i = \ln\left(p_i/1 - p_i\right)$. The other terms are $w_i = n_i$, $b(\theta_i) = \ln(1 + e^{\theta_i})$ and $c(y_i, \phi, w_i) = \ln\binom{w_i}{w_i y_i}$. Note that n_i are known (number of exposures to risk). To distinguish the binomial model from the proportional binomial model these prior weights are required to be specified while using a software implementation for parameter estimation such as the **R** packages **stats** and **gnm**. For more details see R Core Team (2015) and Turner and Firth (2015).

The following part provides an introduction to established computational approaches of maximum likelihood estimation for generalized linear models. The forthcoming presentation is largely based on the standard reference for GLMs by McCullagh and Nelder (1989) and the books of Dobson and Barnett (2008) and Hardin and Hilbe (2012).

Before we present the maximum-likelihood estimation algorithm, we observe some properties of the exponential family and begin with the log-likelihood function. The log-likelihood function $\mathcal{L}(\theta_i, \phi \mid y_i)$ of Y_i with density $f_{Y_i}(y_i \mid \theta_i, \phi)$, as given in eq. (1.26), has the form

$$\mathcal{L}(\theta_i, \phi \mid y_i) = \frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi), \tag{1.31}$$

where $\mathcal{L}(\theta_i, \phi \mid y_i) = \ln f_{Y_i}(y_i \mid \theta_i, \phi)$. Under suitable regularity conditions, such that the order of integration and differentiation can be interchanged, i.e.,

$$\frac{\partial}{\partial \theta_i} \int_D f_{Y_i}(y_i \mid \theta_i, \phi)\, dy_i = \int_D \frac{\partial}{\partial \theta_i} f_{Y_i}(y_i \mid \theta_i, \phi)\, dy_i,$$

where D is the domain of $f_{Y_i}(y_i \mid \theta_i, \phi)$, one can show that

$$\mathbb{E}\left[\frac{\partial}{\partial \theta_i} \mathcal{L}(\theta_i, \phi \mid y_i)\right] = 0 \tag{1.32}$$

Table 1.4: Parametrization of some distributions from the exponential family. Note that the exponential family also contains the Bernoulli, Geometric, negative binomial, exponential, gamma, Pareto, Weibull, Laplace, and the inverse Gaussian distribution.

Distribution		θ	ϕ	$a(\phi)$	$b(\theta)$	$c(y,\phi)$
normal	$\mathcal{N}(\mu,\sigma)$	μ	σ^2	ϕ	$\frac{\theta^2}{2}$	$-\frac{1}{2}\ln(2\pi\phi) - \frac{y^2}{2\phi}$
Poisson	$\mathcal{P}(\lambda)$	$\ln(\lambda)$	1	1	e^θ	$-\ln(y!)$
binomial	$\mathcal{B}(n,p)$	$\ln\left(\frac{p}{1-p}\right)$	1	1	$n\ln(1+e^\theta)$	$\ln\binom{n}{y}$
prop. binomial	$\mathcal{B}(n,p)/n$	$\ln\left(\frac{p}{1-p}\right)$	1	$\frac{1}{n}$	$\ln(1+e^\theta)$	$\ln\binom{n}{ny}$

and

$$\mathbb{E}\left[\frac{\partial^2}{\partial\theta_i^2}\mathcal{L}(\theta_i,\phi \mid y_i)\right] = -\mathbb{E}\left[\left(\frac{\partial}{\partial\theta_i}\mathcal{L}(\theta_i,\phi \mid y_i)\right)^2\right]. \tag{1.33}$$

Using eqs. (1.31) to (1.33), we can conclude

$$\mathbb{E}\left[\frac{Y_i - \frac{\partial}{\partial\theta_i}b(\theta_i)}{a_i(\phi)}\right] = 0 \tag{1.34}$$

and

$$\mathbb{E}\left[\frac{-\frac{\partial^2}{\partial\theta^2}b(\theta_i)}{a_i(\phi)}\right] = -\mathbb{E}\left[\left(\frac{Y_i - \frac{\partial}{\partial\theta_i}b(\theta_i)}{a_i(\phi)}\right)^2\right] \tag{1.35}$$

Rearranging eqs. (1.34) and (1.35) leads to the following expressions for the mean and variance of Y_i

$$\mathbb{E}[Y_i] = b'(\theta_i) = \mu_i, \tag{1.36}$$

and

$$\mathbb{V}[Y_i] = b''(\theta_i)a_i(\phi). \tag{1.37}$$

Note that the expectation depends only on the canonical parameter θ_i. The second derivative $b''(\theta_i)$ in eq. (1.37) is usually referred to as the variance function of the GLM and is denoted by $V(\mu_i)$, since it can be expressed as a function of the mean μ_i. From eq. (1.36), we have

$$\frac{\partial\mu_i}{\partial\theta_i} = V(\mu_i). \tag{1.38}$$

For the forthcoming discussion of different computational approaches, it will be important to distinguish between canonical and non-canonical link functions. We say that the GLM has a canonical link function if

$$\eta_i = \theta_i,$$

holds, i.e., if the canonical parameter θ_i coincides with the predictor η_i. Alternatively, one could define the canonical link as

$$g = \left(b'\right)^{-1}.$$

This is indeed equivalent, since

$$\eta_i = g(\mu_i) = g(b'(\theta_i)) = \theta_i.$$

Thus, from Table 1.4 we can conclude that for a Poisson distributed response variable the canonical link corresponds to $g = \ln$, and for a normally distributed response, we have $g = \mathrm{id}$, i.e., the identity function as the canonical link. The canonical link of the proportional binomial response variable, with $b(\theta) = \ln(1 + e^{\theta})$, is $g = \mathrm{logit}$, i.e.,

$$g : \mu_i \mapsto \ln\left(\frac{\mu_i}{1 - \mu_i}\right).$$

The exponential family has the useful property that the log-likelihood function $\mathcal{L}$ is concave in the canonical parameter θ. This guarantees the uniqueness of a maximum likelihood estimate. Note that the concavity in the canonical parameter θ does not imply concavity in the parameters $\beta_1, \ldots, \beta_K$. One advantage of using the canonical link functions, where $\theta = \eta$ and η is linear in $\boldsymbol{\beta}$, is that the canonical link preserves concavity of the log-likelihood function for the parameters of interest. GLMs with non-canonical link functions are therefore computationally more challenging.

The choice of the link function as part of a GLM is not limited to the canonical case. In the further course, we will demonstrate that a non-canonical link, such as the logit link for a Poisson response variable, might lead to a better model fit. The objective of the next part is to present two computational approaches for parameter estimation for GLMs. These approaches are both based on the Newton-Raphson (NR) method. However, the algorithms involve different matrices in the iterative scheme. One uses the Hessian matrix, which is closely related to the observed information matrix, while the other involves the Fisher information matrix. These both approaches turn out to be equivalent for a canonical link, as will be shown below.

We begin by assuming independent random variables $Y_1, \ldots, Y_N$ from the exponential family of the same distribution type. Our objective here is to estimate the parameters of the linear predictor $\beta_1, \ldots, \beta_K$, rather than the parameters of the exponential family $\theta_1, \ldots, \theta_N$. The joint probability density function of $Y_1, \ldots, Y_N$ is given by

$$f_{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{\theta}, \phi) = f_{Y_1,\ldots,Y_N}(y_1, \ldots, y_n \mid \theta_1, \ldots, \theta_N, \phi) = \prod_{i=1}^{N} \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)\right).$$

The log-likelihood function based on the independent observations $y_1, \ldots, y_N$ is thus given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}, \phi \mid y_1, \ldots, y_N) &= \ln f_{\boldsymbol{Y}}(y_1, \ldots, y_N \mid \boldsymbol{\theta}, \phi) \\ &= \sum_{i=1}^{N} \ln f_{Y_i}(y_i \mid \theta_i, \phi) \\ &= \sum_{i=1}^{N} \left(\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)\right).\end{aligned} \tag{1.39}$$

The above expression is the sum of the individual log-likelihood functions as given in eq. (1.31). To obtain the maximum likelihood estimates of the parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^T$, the GLM log-likelihood function needs to be maximized with respect to $\boldsymbol{\beta}$ given the observation samples $y_1, \ldots, y_N$. Due to eqs. (1.30) and (1.36), the log-likelihood of the GLM is a function of $\boldsymbol{\beta}$. The first derivative of the log-likelihood function is called *Fisher's score function* or simply the *score function* and is denoted by

$$s(\boldsymbol{\beta}) := \frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = \left(\frac{\partial \mathcal{L}}{\partial \beta_1}, \ldots, \frac{\partial \mathcal{L}}{\partial \beta_K}\right)^T. \tag{1.40}$$

To obtain the maximum likelihood estimator of $\boldsymbol{\beta}$, we need to solve the score function, i.e., we wish to find some $\hat{\boldsymbol{\beta}}$ such that $s(\hat{\boldsymbol{\beta}}) = \boldsymbol{0}$ holds. The standard algorithmic approaches for ML estimation are Newton-Raphson type algorithms that use the first terms of the Taylor series to successively approximate the roots of the score function. Taylor series expansion of the score function at $\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}$ is given by

$$\begin{aligned}0 = {} & s(\beta_1^{(0)}, \cdots, \beta_K^{(0)}) \\ & + \sum_{j=1}^{K} \frac{\partial s(\beta_1^{(0)}, \cdots, \beta_K^{(0)})}{\partial \beta_j}(\beta_j - \beta_j^{(0)}) \\ & + \frac{1}{2!}\sum_{j=1}^{K}\sum_{k=1}^{K} \frac{\partial^2 s(\beta_1^{(0)}, \cdots, \beta_K^{(0)})}{\partial \beta_j \partial \beta_k}(\beta_j - \beta_j^{(0)})(\beta_k - \beta_k^{(0)}) \\ & + \frac{1}{3!}\sum_{j=1}^{K}\sum_{k=1}^{K}\sum_{l=1}^{K} \frac{\partial^3 s(\beta_1^{(0)}, \cdots, \beta_K^{(0)})}{\partial \beta_j \partial \beta_k \partial \beta_l}(\beta_j - \beta_j^{(0)})(\beta_k - \beta_k^{(0)})(\beta_l - \beta_l^{(0)}) + \cdots.\end{aligned}$$

By discarding all super-linear Taylor terms, we obtain the following approximation

$$0 \approx s(\boldsymbol{\beta}^{(0)}) + \frac{\partial s(\boldsymbol{\beta}^{(0)})}{\partial \boldsymbol{\beta}^T}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}).$$

Rewriting that expression leads to

$$\boldsymbol{\beta} \approx \boldsymbol{\beta}^{(0)} - \left(\frac{\partial s(\boldsymbol{\beta}^{(0)})}{\partial \boldsymbol{\beta}^T}\right)^{-1} s(\boldsymbol{\beta}^{(0)}).$$

This approximation may be used iteratively, i.e.,

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} - \left(\frac{\partial s(\boldsymbol{\beta}^{(r)})}{\partial \boldsymbol{\beta}^T}\right)^{-1} s(\boldsymbol{\beta}^{(r)}),$$

for $r \in \mathbb{N}$ and reasonable starting vector $\boldsymbol{\beta}^{(0)}$, see Hardin and Hilbe (2012). The partial derivative in the iterative expression is the Hessian matrix of the log-likelihood function $\mathcal{L}$ which is related to the *observed information matrix*, denoted by $\mathcal{J}(\boldsymbol{\beta})$, through

$$\frac{\partial s(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\mathcal{J}(\boldsymbol{\beta}), \tag{1.41}$$

or in matrix component notation, through

$$\mathcal{J}_{jk} = -\frac{\partial^2 \mathcal{L}}{\partial \beta_j \partial \beta_k}.$$

1.5.2 | Newton-Raphson and Fisher Scoring Algorithm

Analytical solutions of GLM likelihood equations are not available in general. Therefore, they have to be obtained by iterative algorithms. Let $\hat{\boldsymbol{\beta}}^{(r)}$ be the estimate of $\boldsymbol{\beta}$ after r iteration steps, then the $r+1$ update $\hat{\boldsymbol{\beta}}^{(r+1)}$ of the Newton-Raphson method is defined by

$$\hat{\boldsymbol{\beta}}^{(r+1)} = \hat{\boldsymbol{\beta}}^{(r)} + \left(\mathcal{J}(\hat{\boldsymbol{\beta}}^{(r)})\right)^{-1} s(\hat{\boldsymbol{\beta}}^{(r)}). \tag{1.42}$$

The Newton-Raphson algorithm, which depends on the observed information matrix can face some difficulties. As discussed in Hardin and Hilbe (2012) a problem appears if for some $r \in \mathbb{N}$ the matrix of second partial derivatives $\mathcal{J}(\boldsymbol{\beta}^{(r)})$ is not positive definite. This can be avoided by replacing the observed information matrix $\mathcal{J}(\boldsymbol{\beta}^{(r)})$ by its expectation. That particular variation of the Newton-Raphson method was first suggested by Fisher (1935) and is known as the *Fisher scoring* method. The expected information, or the *Fisher information*, is defined as the second moment of the score function. Under some regularity conditions [1] on the log-likelihood function, the Fisher

1 If $f_Y(\boldsymbol{y} \mid \boldsymbol{\theta})$ is twice differentiable in $\boldsymbol{\theta} \in \Theta$ for all $\boldsymbol{y} \in \text{supp}(f)$ almost everywhere and $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \iint f_Y(\boldsymbol{y} \mid \boldsymbol{\theta}) d\boldsymbol{y} = \iint \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} f_Y(\boldsymbol{y} \mid \boldsymbol{\theta}) d\boldsymbol{y}$.

information can be expressed as

$$\mathcal{I} = -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right].$$

Given that expression and eq. (1.41), we have the relation

$$\mathcal{I}(\boldsymbol{\beta}) = \mathbb{E}[\mathcal{J}(\boldsymbol{\beta})]$$

between the Fisher information and the observed information. As indicated above, contrary to the observed information, the Fisher information $\mathcal{I}(\boldsymbol{\beta})$ is positive semidefinite on the entire parameter space. The Fisher scoring, as a variation of the iterative approximation, takes the form

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{(r+1)} &= \hat{\boldsymbol{\beta}}^{(r)} + \left(\mathbb{E}\left[\mathcal{J}(\hat{\boldsymbol{\beta}}^{(r)})\right]\right)^{-1} s(\hat{\boldsymbol{\beta}}^{(r)}) \\
&= \hat{\boldsymbol{\beta}}^{(r)} + \left(\mathcal{I}(\hat{\boldsymbol{\beta}}^{(r)})\right)^{-1} s(\hat{\boldsymbol{\beta}}^{(r)}). \qquad (1.43)
\end{aligned}$$

In the following, we verify that the Newton-Raphson method and the Fisher scoring method coincide for GLMs with a canonical link function and derive an alternative representation of eqs. (1.42) and (1.43), known as *iteratively reweighted least squares algorithm* (IRLS), see, e.g., Charnes, Frome and Yu (1976). Following Hardin and Hilbe (2012), we start by the calculation of the partial derivative using the chain rule for differentiation. Thus, we have

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \beta_j} &= \sum_{i=1}^{N} \left(\frac{\partial \mathcal{L}_i}{\partial \theta_i}\right)\left(\frac{\partial \theta_i}{\partial \mu_i}\right)\left(\frac{\partial \mu_i}{\partial \eta_i}\right)\left(\frac{\partial \eta_i}{\partial \beta_j}\right) \\
&= \sum_{i=1}^{N} \left(\frac{y_i - b'(\theta_i)}{a_i(\phi)}\right)\left(\frac{1}{V(\mu_i)}\right)\left(\frac{\partial \mu_i}{\partial \eta_i}\right)(x_{ij}) \qquad (1.44) \\
&= \sum_{i=1}^{N} \frac{y_i - \mu_i}{a_i(\phi)V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \qquad (1.45) \\
&= \sum_{i=1}^{N} \frac{y_i - \mu_i}{a_i(\phi)V(\mu_i)} \frac{x_{ij}}{g'(\mu_i)}, \qquad (1.46)
\end{aligned}$$

where eq. (1.44) follows since

$$\begin{aligned}
&\frac{\partial \mathcal{L}_i}{\partial \theta_i} \overset{\text{eq. (1.31)}}{=} \frac{\partial}{\partial \theta_i}\left(\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i,\phi)\right) = \frac{y_i - b'(\theta_i)}{a_i(\phi)}, \\
&\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{\frac{\partial \mu_i}{\partial \theta_i}} \overset{\text{eq. (1.38)}}{=} \frac{1}{V(\mu_i)}, \\
&\frac{\partial \eta_i}{\partial \beta_j} \overset{\text{eq. (1.28)}}{=} \frac{\partial}{\partial \beta_j}(\beta_1 x_{i1} + \ldots + \beta_K x_{iK}) = x_{ij}. \qquad (1.47)
\end{aligned}$$

Equation (1.45) is obtained by substitution of eq. (1.36) in eq. (1.44) and for eq. (1.46) we used the following relation between the predictor η_i and the mean μ_i

$$\mu_i = g^{-1}(\eta_i), \tag{1.48}$$

such that

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{g'(\mu_i)}.$$

Note, for the canonical link, i.e., $\eta_i = \theta_i$, we have a simplification of the chain rule

$$\left(\frac{\partial \theta_i}{\partial \mu_i}\right)\left(\frac{\partial \mu_i}{\partial \eta_i}\right) = \left(\frac{\partial \theta_i}{\partial \mu_i}\right)\left(\frac{\partial \mu_i}{\partial \theta_i}\right) = 1$$

or alternatively,

$$V(\mu_i) = \frac{1}{g'(\mu_i)}.$$

Thus, eq. (1.46) simplifies to

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{i=1}^{N} \frac{y_i - \mu_i}{a_i(\phi)} x_{ij}.$$

We continue with the derivation of the observed information matrix by using eqs. (1.45) and (1.47) and applying the chain and the Leibniz rule.

$$\begin{aligned}
\mathcal{J}_{jk} &= -\frac{\partial^2 \mathcal{L}}{\partial \beta_j \partial \beta_k} \\
&= -\sum_{i=1}^{N} \frac{1}{a_i(\phi)} \left(\frac{\partial}{\partial \beta_k}\right)\left[\frac{y_i - \mu_i}{V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}\right] \\
&= -\sum_{i=1}^{N} \frac{1}{a_i(\phi)} \left[\left(\frac{\partial}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_k}\right)\left[\frac{y_i - \mu_i}{V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i}\right] + \frac{y_i - \mu_i}{V(\mu_i)} \left(\frac{\partial}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_k}\right)\left[\frac{\partial \mu_i}{\partial \eta_i}\right]\right] x_{ij} \\
&= -\sum_{i=1}^{N} \frac{1}{a_i(\phi)} \left[\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \left(\frac{\partial}{\partial \mu_i}\right)\left[\frac{y_i - \mu_i}{V(\mu_i)}\right] + \frac{y_i - \mu_i}{V(\mu_i)} \left(\frac{\partial^2 \mu_i}{\partial \eta_i^2}\right)\right] x_{ij} x_{ik} \\
&= \sum_{i=1}^{N} \frac{1}{a_i(\phi)} \left[\frac{1}{V(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right. \\
&\qquad \left. +(\mu_i - y_i)\left(\frac{1}{V(\mu_i)^2}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \frac{\partial V(\mu_i)}{\partial \mu_i} - \frac{1}{V(\mu_i)}\left(\frac{\partial^2 \mu_i}{\partial \eta_i^2}\right)\right)\right] x_{ij} x_{ik}
\end{aligned} \tag{1.49}$$

Using that result, we can conclude that for GLM with a canonical link we have

$$\mathcal{J}(\boldsymbol{\beta}) = -\frac{\partial^2 \mathcal{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right] = \mathcal{I}(\boldsymbol{\beta}), \tag{1.50}$$

i.e., the Newton-Raphson method and the Fisher scoring coincide. This can be obtained by the following. First, for the right-hand side, we have

$$\begin{aligned}
\mathcal{I}_{jk} &= -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}}{\partial \beta_j \partial \beta_k}\right] \\
&\overset{\text{eq. (1.49)}}{=} \mathbb{E}\left[\sum_{i=1}^{N} \frac{1}{a_i(\phi)}\left[\frac{1}{V(\mu_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right.\right. \\
&\qquad \left.\left. +(\mu_i - Y_i)\left(\frac{1}{V(\mu_i)^2}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \frac{\partial V(\mu_i)}{\partial \mu_i} - \frac{1}{V(\mu_i)}\left(\frac{\partial^2 \mu_i}{\partial \eta_i^2}\right)\right)\right] x_{ij} x_{ik}\right] \\
&= \sum_{i=1}^{N} \frac{1}{a_i(\phi)}\left[\frac{1}{V(\mu_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right. \\
&\qquad \left. + \underbrace{\mathbb{E}[\mu_i - Y_i]}_{=0}\left(\frac{1}{V(\mu_i)^2}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \frac{\partial V(\mu_i)}{\partial \mu_i} - \frac{1}{V(\mu_i)}\left(\frac{\partial^2 \mu_i}{\partial \eta_i^2}\right)\right) x_{ij} x_{ik}\right] \\
&= \sum_{i=1}^{N} \frac{1}{a_i(\phi) V(\mu_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 x_{ij} x_{ik}.
\end{aligned} \tag{1.51}$$

For the left-hand side of eq. (1.50) and the canonical case where $\theta_i = \eta_i$, we have

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial \mu_i}{\partial \theta_i} \overset{\text{eq. (1.38)}}{=} V(\mu_i). \tag{1.52}$$

Thus, the last term in eq. (1.49) vanishes, i.e.,

$$\frac{1}{V(\mu_i)^2}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \frac{\partial V(\mu_i)}{\partial \mu_i} - \frac{1}{V(\mu_i)}\left(\frac{\partial^2 \mu_i}{\partial \eta_i^2}\right) = 0. \tag{1.53}$$

This follows by using eq. (1.52) and the substitution of

$$\frac{\partial^2 \mu_i}{\partial \eta_i^2} = \frac{\partial}{\partial \eta_i}\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial}{\partial \eta_i} V(\mu_i) = \frac{\partial V(\mu_i)}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial V(\mu_i)}{\partial \mu_i} V(\mu_i)$$

in eq. (1.53). Therefore, we conclude that for a canonical link eq. (1.49) simplifies to

$$\mathcal{J}_{jk} = \sum_{i=1}^{N} \frac{1}{a_i(\phi) V(\mu_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 x_{ij} x_{ik},$$

and therefore both iterative methods of eqs. (1.42) and (1.43) are equivalent. To obtain the iterative weighted least squares representation of the Fisher scoring algorithm, recall that using the iteration of eq. (1.43), the Fisher information matrix is evaluated at the previous estimate. Let $\hat{\boldsymbol{\beta}}^{(r)}$ be the estimate of $\boldsymbol{\beta}$ after r iterations and

$$\hat{\mu}_i^{(r)} := \mu_i(\hat{\boldsymbol{\beta}}^{(r)}) = g^{-1}\left(\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}^{(r)}\right)$$

the corresponding estimate of the mean $\mu_i = \mathbb{E}[Y_i]$. Furthermore, recall that the involved Fisher information matrix evaluated at $\boldsymbol{\beta}$, with components given in eq. (1.51), has the form

$$\mathcal{I}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{\boldsymbol{x}_i^T\boldsymbol{x}_i}{a_i(\phi)V(\mu_i(\boldsymbol{\beta}))\,(g'(\mu_i(\boldsymbol{\beta})))}. \tag{1.54}$$

Moreover, using eq. (1.46), the score function $s(\boldsymbol{\beta})$ can be expressed as

$$s(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{y_i - \mu_i(\boldsymbol{\beta})}{a_i(\phi)V(\mu_i(\boldsymbol{\beta}))}\frac{\boldsymbol{x}_i}{g'(\mu_i(\boldsymbol{\beta}))}. \tag{1.55}$$

A slide transformation of the Fisher scoring equation of eq. (1.43) to

$$\mathcal{I}(\hat{\boldsymbol{\beta}}^{(r)})\hat{\boldsymbol{\beta}}^{(r+1)} = \mathcal{I}(\hat{\boldsymbol{\beta}}^{(r)})\hat{\boldsymbol{\beta}}^{(r)} + s(\boldsymbol{\beta}^{(r)}), \tag{1.56}$$

and a subsequential substitution of the expressions of eqs. (1.54) and (1.55), leads to

$$\mathcal{I}(\hat{\boldsymbol{\beta}}^{(r)})\hat{\boldsymbol{\beta}}^{(r+1)} = \boldsymbol{X}^T\boldsymbol{W}^{(r)}\boldsymbol{X}\hat{\boldsymbol{\beta}}^{(r+1)},$$

where $\boldsymbol{X}$ is the design matrix with dimensions $N \times K$ and $\boldsymbol{W}^{(r)}$ is an $N \times N$ diagonal matrix with elements

$$w_{ii}^{(r)} = \frac{1}{a_i(\phi)V\left(\hat{\mu}_i^{(r)}\right)\left(g'\left(\hat{\mu}_i^{(r)}\right)\right)^2} = \frac{\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2}{\mathbb{V}[Y_i]}. \tag{1.57}$$

For the right-hand side of eq. (1.56) we get

$$\mathcal{I}(\hat{\boldsymbol{\beta}}^{(r)})\hat{\boldsymbol{\beta}}^{(r)} + s(\boldsymbol{\beta}^{(r)}) = \boldsymbol{X}^T\boldsymbol{W}^{(r)}\boldsymbol{Z}^{(r)}, \tag{1.58}$$

where $\boldsymbol{Z}^{(r)}$ is a vector called the *working variable* with components

$$\begin{aligned} Z_i^{(r)} &= \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}^{(r)} + g'(\hat{\mu}_i^{(r)})(y_i - \hat{\mu}_i^{(r)}) \\ &= g(\hat{\mu}_i^{(r)}) + g'(\hat{\mu}_i^{(r)})(y_i - \hat{\mu}_i^{(r)}) \end{aligned} \tag{1.59}$$

for $i = 1, \ldots, N$. Combination of eqs. (1.58) and (1.59), and a rearrangement of the expression for

$\hat{\boldsymbol{\beta}}^{(r+1)}$ leads to

$$\hat{\boldsymbol{\beta}}^{(r+1)} = \left(\boldsymbol{X}^T \boldsymbol{W}^{(r)} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{W}^{(r)} \boldsymbol{Z}^{(r)}. \tag{1.60}$$

Note that eq. (1.60) has the form of the normal equation of a weighted least squares regression. In other words, the $r+1$ iterative estimate $\hat{\boldsymbol{\beta}}^{(r+1)}$ minimizes the weighted least squares objective function

$$\sum_{i=1}^{N} W_{ii}^{(r)} \left(Z_i^{(r)} - \boldsymbol{x}_i^T \boldsymbol{\beta}\right)^2 .$$

For a consistent proof of the equivalence of the Fisher scoring and the iterative weighted least squares algorithm, see McCullagh and Nelder (1989). The **R** package **StMoMo** which we employ to estimate the parameters of generalized age-period-cohort models, includes a method of the **gnm** package, which contains an implementation of the iterative weighted least squares eq. (1.60).

For the ML estimation, the iterative updating procedure of eq. (1.60) continues until a termination criterion is met. This can be the relative changes of the estimates

$$\frac{\left\|\hat{\boldsymbol{\beta}}^{(r+1)} - \hat{\boldsymbol{\beta}}^{(r)}\right\|}{\left\|\hat{\boldsymbol{\beta}}^{(r+1)}\right\|} \leq \epsilon,$$

or the ratio of the score function and the estimated standard deviation of the coefficients

$$\frac{|s(\hat{\boldsymbol{\beta}}^{(r)})_i|}{\sqrt{\mathcal{I}_{ii}\left(\hat{\boldsymbol{\beta}}^{(r)}\right)}} \leq 10^{-6},$$

for all $i = 1, \ldots, K$, which is the default criterion of the **gnm** package.

Remark 1.5.1 (Weight matrix $\boldsymbol{W}$ and working variable $\boldsymbol{Z}$ for the Poisson model with canonical and non-canonical link function)**.** In the following course of the thesis, we will estimate GAPC models in a Poisson setting. We therefore derive the precise forms of the weight matrix $\boldsymbol{W}^{(r)}$ and working variable $\boldsymbol{Z}^{(r)}$ for the canonical link function $g \equiv \ln$ and a non-canonical link function $g \equiv \text{logit}$, respectively. Recall, in the Poisson setting, we have $Y_i \sim \mathcal{P}(\mu_i)$ and $\mathbb{E}[Y_i] = \mathbb{V}[Y_i] = \mu_i$. Moreover, $a_i(\phi) \equiv 1$ and for the variance function we have $V \equiv \text{id}$. The canonical link $g \equiv \ln$ yields $g'(\mu_i) = 1/\mu_i$. Thus, from eq. (1.57) we can conclude

$$w_{ii}^{(r)} = \frac{1}{a_i(\phi) V\left(\hat{\mu}_i^{(r)}\right)\left(g'\left(\hat{\mu}_i^{(r)}\right)\right)^2} = \hat{\mu}_i^{(r)}$$

and

$$\begin{aligned} Z_i^{(r)} &= \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}^{(r)} + g'(\hat{\mu}_i^{(r)})(y_i - \hat{\mu}_i^{(r)}) \\ &= \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}^{(r)} + \frac{y_i}{\hat{\mu}_i^{(r)}} - 1. \end{aligned}$$

Switching from the GLM notation to the notation of mortality models leads to $w_{ii}^{(r)} = \hat{d}_i^{(r)}$ and $Z_i^{(r)} = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}^{(r)} + d_i/\hat{d}_i^{(r)} - 1$, where d_i are the observed death counts and $\hat{d}_i^{(r)}$ are the fitted death counts after r iterations. For the non-canonical link $g \equiv \text{logit}$, the corresponding weight matrix is given by

$$w_{ii}^{(r)} = \hat{\mu}_i^{(r)}(\hat{\mu}_i^{(r)} - 1)^2, \tag{1.61}$$

since

$$g'(\mu_i) = \frac{1}{\mu_i(1 - \mu_i)}.$$

For the working variable $\boldsymbol{Z}^{(r)}$, we obtain

$$Z_i^{(r)} = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}^{(r)} + \frac{y_i - \mu_i^{(r)}}{\mu_i^{(r)}(1 - \mu_i^{(r)})}. \tag{1.62}$$

Using the notation of GAPC models and the relation of eq. (1.48), then eqs. (1.61) and (1.62) can be rewritten as

$$w_{ii}^{(r)} = \frac{\hat{d}_i^{(r)}}{\left(1 + e^{\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}^{(r)}}\right)^2}$$

and

$$Z_i^{(r)} = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}^{(r)} + \left(1 + e^{\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}^{(r)}}\right)\left(\frac{d_i}{\hat{d}_i^{(r)}} - 1\right).$$

Remark 1.5.2 (Newton-Raphson versus Fisher Scoring)**.** Summarizing the above discussion on the iterative algorithms, it can be noted that the Fisher scoring algorithm is a modification of the Newton-Raphson algorithm, where the observed Fisher information $\mathcal{J}(\boldsymbol{\beta})$ is replaced by the expected Fisher information $\mathcal{I}(\boldsymbol{\beta})$. The crucial difference between those algorithms is that, in general, $\mathcal{J}(\boldsymbol{\beta})$ depends on the observation. Unlike for the expected Fisher information $\mathcal{I}(\boldsymbol{\beta})$, the dependence on the observation might not ensure $\mathcal{J}(\boldsymbol{\beta})$ to be positive definite. Knight (2000) points out that it is difficult to make general statements about the performance of both algorithms. However, according to Knight, the Newton-Raphson algorithm often converges faster if both algorithms converge, but the radius of convergence of the Fisher scoring algorithm is often larger. As illustrated above, the Newton-Raphson algorithm and the Fisher scoring algorithm coincide for generalized linear models with a canonical link function. For non-canonical link functions the concavity property of the GLM log-likelihood function does no longer hold in general. In that case, there is no guarantee for numerical methods to converge to the global maximum. Therefore, further techniques, e.g., perturbations of the starting points are required to get initial estimates sufficiently close to the global maximum. We have also illustrated above that solving the Fisher scoring algorithm is equivalent to solving a sequence of weighted least squares problems. For a detailed discussion on the choice between the Newton-Raphson algorithm and the Fisher scoring, we refer to Efron and Hinkley (1978).

1.5.3 | Asymptotic Properties of MLEs

Under quite general regularity conditions, we have strong consistency and asymptotic normality of the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, i.e.,

$$\hat{\boldsymbol{\beta}}_N \xrightarrow[N\to\infty]{\text{a.s.}} \boldsymbol{\beta}$$

and

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}_K(\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\beta})), \tag{1.63}$$

where N is the sample size. For proofs on consistency and asymptotic normality and more details on conditions, in particular, for the GLM framework with canonical or non-canonical link functions, we refer to Fahrmeir and Kaufmann (1985). The asymptotic covariance matrix for the estimator in eq. (1.63) is the inverse Fisher information

$$\mathcal{I}(\boldsymbol{\beta}) = \text{Cov}(s(\boldsymbol{\beta})) = \mathbb{E}\left[\frac{\partial \mathcal{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial \mathcal{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\right].$$

Note that $\mathcal{I}(\boldsymbol{\beta})$ depends on the unknown parameters $\boldsymbol{\beta}$. Common practice is to replace the unknown $\boldsymbol{\beta}$ by the estimated value $\hat{\boldsymbol{\beta}}$ and using

$$\hat{\mathcal{I}}_{ij}(\hat{\boldsymbol{\beta}}_N, N) = \frac{1}{N}\sum_{i=1}^{N} \frac{x_{ij}x_{ik}}{a_i(\phi)V\left(\hat{\mu}_i\right)\left(g'\left(\hat{\mu}_i\right)\right)^2} \tag{1.64}$$

as the approximated covariance matrix of the estimator. Substituting the weight components of eq. (1.57), reduces eq. (1.64) to matrix expression

$$\hat{\mathcal{I}} = \boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X}.$$

The asymptotic standard error estimator $\hat{\text{se}}(\hat{\beta}_i)$ for parameter β_i can be obtained as the square root of the corresponding diagonal element of the inverse Fisher matrix, i.e.,

$$\hat{\text{se}}(\hat{\beta}_i) = \sqrt{\hat{\mathcal{I}}_{ii}^{-1}} = \sqrt{\left(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X}\right)_{ii}^{-1}},$$

using the estimated weight matrix of the final iteration step of the IRLS procedure $\hat{\boldsymbol{W}}$, with components w_i given by eq. (1.57). Note, in the special case of a GLM with a multivariate normally distributed response variable and the canonical link (identity function), the weight matrix is constant and coincides with the identity matrix. Furthermore, the working variable Z_i coincides with the original response variable Y_i and the IRLS converges after the first iteration step. The asymptotic distribution of $\hat{\boldsymbol{\beta}}$ is multivariate normal with the covariance matrix $\sigma^2 \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$. This is a well-known result for standard linear regressions with normally distributed response variables.

1.6 | Hypothesis Tests, Goodness-of-fit Measures and Model Selection Criteria

In this section, we provide a brief introduction to some common tests of statistical hypothesis and goodness-of-fit measures which will be used in the forthcoming case studies to assess and compare different GAPC models. First, we will consider the Wald test, the likelihood ratio test, and the Lagrange multiplier test for hypothesis testing. Then, we will discuss the deviance statistic as goodness-of-fit measure and provide the concrete forms for Poisson and binomial distributed response variables. Finally, we briefly present information bases criteria for model selection and discuss different types of residuals, which are used to detect model misspecifications and are therefore of key importance for model assessment.

The following presentation is based on Lütkepohl (2007) and Rodríguez (2008). More details and proofs of the asymptotic properties of the involved statistics can be found in Knight (2000).

1.6.1 | Wald Test

Generally, regression modelling encompasses several procedures for model selection and validation. In particular, one is interested in determining the significance of particular parameters by performing statistical tests. By omitting non-significant coefficients from the regression one obtains a more parsimonious model. Consider a hypothesis test of the form

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \qquad \text{vs.} \qquad H_A : \boldsymbol{\beta}_j \neq \boldsymbol{\beta}_0,$$

for some fixed $\boldsymbol{\beta}_0$, and an asymptotically normal estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_K)$, then the *Wald statistic*, defined by

$$T_W^2 = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \Sigma_{\hat{\boldsymbol{\beta}}}^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),$$

is approximately chi-squared distributed with K degrees of freedom, where $\Sigma_{\hat{\boldsymbol{\beta}}}$ is the non-singular asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$. If the objective is to test the significance of a single parameter, say β_j, with the hypothesis pair

$$H_0 : \beta_j = 0 \qquad \text{vs.} \qquad H_A : \beta_j \neq 0,$$

then commonly the Wald statistic T_W^2 is replaced by its square root. This leads to the statistic given by the ratio

$$T_W = \frac{\hat{\beta}_j}{\hat{\text{se}}(\hat{\beta}_j)}. \tag{1.65}$$

Remark 1.6.1 (*t*-value & *z*-value). In statistical software, the test value of the Wald statistic is either denoted as the t-value or z-value. The reason for the different notations is the following. For ordinary

Gaussian regression, the standard error of $\hat{\beta}_j$

$$\hat{\text{se}}(\hat{\beta}_j) = \sqrt{\sigma^2 \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}_{jj}}$$

depends on an unknown variance σ^2. In practice, σ^2 is replaced by an unbiased estimator $\hat{\sigma}$ which is based on the residual sum of squares, see Rodríguez (2008). Under the null hypothesis, the resulting statistic then follows asymptotically a Student-t distribution with $N - K$ degrees of freedom. In the other case, where σ^2 is known, the statistic T_W follows asymptotically a normal distribution, if the null hypothesis holds. This is also the case for the Poisson and binomial regression models, where the variance is a function of the expectation which does not depend on an additional free parameter. Thus, when the Wald statistic is normally distributed, then the test value is denoted by the z-value, and in the other case, where the statistic is Student-t distributed, the test value is denoted by the t-value. In order to decide whether to reject the hypothesis H_0 at the level $\alpha \in (0,1)$, we need to compare the absolute value of the test with the $(1 - \alpha/2)$-quantile of the corresponding limiting distribution. For the Poisson, binomial, or Gaussian regression with a known variance, a size α Wald test rejects H_0 when $|T_W| > z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $1 - \alpha/2$-quantile of the standard normal distribution. For the ordinary Gaussian regression with an unknown variance, the α Wald test rejects H_0 when $|T_W| > t_{1-\alpha/2,N-K}$, where $t_{1-\alpha/2,N-K}$ is the $1 - \alpha/2$-quantile of the Student-t distribution with $N - K$ degrees of freedom.

In summary, whether a t-test or z-test is used for testing significance of individual parameters of a GLM depends on whether the dispersion parameter of the exponential family is known or has to be estimated. For large sample sizes, the inference of both tests tends to correspond since the Student-t distribution is asymptotically normal with growing number of degrees of freedom. The Wald statistic can also be used to express the confidence interval for $\hat{\beta}_j$. The following interval

$$C = \left[\hat{\beta}_j - z_{1-\alpha/2}\,\hat{\text{se}}(\hat{\beta}_j), \hat{\beta}_j + z_{1-\alpha/2}\,\hat{\text{se}}(\hat{\beta}_j)\right] \tag{1.66}$$

contains the true parameter β_j with the confidence level of $100(1 - \alpha)\%$. As Wasserman (2013) points out, the size α Wald test rejects $H_0 : \beta_j = \tilde{\beta}_j$ versus $H_\text{A} : \beta_j \neq \tilde{\beta}_j$ if and only if $\tilde{\beta}_j \notin C$, where the interval C is given in eq. (1.66). Thus, testing the hypothesis is equivalent to checking whether the confidence interval contains the null value. Finally, it should be noted that the Wald test can also be used for testing joint significance of model parameters, i.e., testing $H_0 : \beta_i = \beta_j = \beta_l = 0$, or jointly multiple hypotheses, e.g., testing M hypotheses on K parameters by $H_0 : \boldsymbol{M\beta} = \boldsymbol{m}$ versus $H_A : \boldsymbol{M\beta} \neq \boldsymbol{m}$, for some $M \times K$ matrix $\boldsymbol{M}$ and M-dimensional vector $\boldsymbol{m}$. If $H_0 : \boldsymbol{M\beta} = \boldsymbol{m}$ is true and if $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}$, the consistent estimator of the covariance matrix, is invertible, then the corresponding Wald statistic

$$N(\boldsymbol{M}\hat{\boldsymbol{\beta}}_N - \boldsymbol{m})^T \left(\boldsymbol{M}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\boldsymbol{M}^T\right)^{-1} (\boldsymbol{M}\hat{\boldsymbol{\beta}}_N - \boldsymbol{m}) \xrightarrow[N\to\infty]{\mathcal{D}} \chi^2_M$$

follows asymptotically a chi-squared distribution with M degrees of freedom. For more details on this topic, see, e.g., Harrell (2015).

Remark 1.6.2 (*p*-value)**.** Checking the test statistic T of an α sized test with critical values c_α gives only a binary type information, which is, either to reject H_0 or to retain H_0. If a test rejects H_0 at level α, it will also reject H_0 for all higher levels $\tilde{\alpha} \in (\alpha, 1)$. The smallest level at which H_0 is rejected based on the available observation $(x_1, \ldots, x_N)$ is also known as the p-value, i.e.,

$$p\text{-value} = \inf\{\alpha : T(x_1, \ldots, x_N) \in R_\alpha\}$$

where R_α is the rejection region of the test of size α. As Wasserman (2013) points out, for a size α test of the form

$$\text{reject } H_0 : \theta \in \Theta_0 \quad \text{if and only if} \quad T(X_1, \ldots, X_n) \geq c_\alpha,$$

the p-value coincides with the probability, under H_0, of observing a test value more extreme than for the available observation. More precisely,

$$p\text{-value} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta \left[T(X_1, \ldots, X_n) \geq T(x_1, \ldots, x_n) \right].$$

The p-value can be seen as a measure of evidence against the null hypothesis. The smaller the p-value, the stronger the evidence against H_0. On the other hand, a large p-value does not necessarily support the null hypothesis, since it can also occur if H_0 does not hold, but the power of the test is too low to detect that.

1.6.2 | Likelihood Ratio Test

Another important type of statistical tests is the so-called *likelihood ratio test* (LR). The idea of the LR test is based on the comparison of the maximized likelihoods of *nested models*, these are models, where one model can be obtained from the other by imposing some constraints on the parameters. For instance, let Θ_r be the parameter state space of the restricted model (submodel) and Θ the state space of the larger model, i.e., $\Theta_r \subset \Theta$. The simplest example of nested models is where the submodel emerges from the larger model by restricting some parameters to zero. In the following, we consider tests of the form

$$H_0 : \boldsymbol{\theta}_0 \in \Theta_r \qquad \text{vs.} \qquad H_\text{A} : \boldsymbol{\theta}_0 \in \Theta \setminus \Theta_r. \tag{1.67}$$

Let $\boldsymbol{y} = (y_1, \ldots, y_N)$ be a sample of size N and

$$\sup_{\boldsymbol{\theta} \in \Theta_r} L(\boldsymbol{\theta}, \boldsymbol{y}) = L(\hat{\boldsymbol{\theta}}_{\Theta_r}, \boldsymbol{y}) \tag{1.68}$$

be the maximized likelihood function of the submodel, where $\hat{\boldsymbol{\theta}}_{\Theta_r}$ denotes MLE of $\boldsymbol{\theta}_0$ under the restrictions stated by the null hypothesis. Let further

$$\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}, \boldsymbol{y}) = L(\hat{\boldsymbol{\theta}}_{\Theta}, \boldsymbol{y}) \tag{1.69}$$

be the maximized likelihood function of the unconstrained model with the MLE $\hat{\boldsymbol{\theta}}_\Theta$. For the ratio of eqs. (1.68) and (1.69)

$$\lambda = \frac{L(\hat{\boldsymbol{\theta}}_{\Theta_r}, \boldsymbol{y})}{L(\hat{\boldsymbol{\theta}}_\Theta, \boldsymbol{y})},$$

we have $\lambda \in [0,1)$. The lower bound of the interval follows from the non-negativity of likelihood functions, and the upper from the fact that for nested models the maximized likelihood of the submodel is always smaller than the maximized likelihood of the full model. Note that a low λ indicates that the observation of a particular sample is far less likely under the submodel than under the full model. Whereas, $\lambda \approx 1$ indicates that the observation under the submodel is almost as likely as under the full model. The likelihood ratio test is based on the statistic λ_{LR} defined by

$$\lambda_{\mathrm{LR}} := -2 \ln \lambda.$$

Thus, we have

$$\begin{aligned} \lambda_{\mathrm{LR}} &= 2 \ln L(\hat{\boldsymbol{\theta}}_\Theta) - 2 \ln L(\hat{\boldsymbol{\theta}}_{\Theta_r}) \\ &= 2 \left(\mathcal{L}(\hat{\boldsymbol{\theta}}_\Theta) - \mathcal{L}(\hat{\boldsymbol{\theta}}_{\Theta_r}) \right), \end{aligned} \tag{1.70}$$

where $\mathcal{L}$ denotes the log-likelihood function. Under H_0, as stated in eq. (1.67), and some suitable regularity conditions (see, e.g., Knight, 2000, Theorem 7.5), we have

$$\lambda_{\mathrm{LR}} = \lambda_{\mathrm{LR}}(Y_1, \ldots, Y_N) \xrightarrow[N\to\infty]{\mathcal{D}} \chi^2_\kappa, \tag{1.71}$$

where the number of degrees of freedom, κ, is the difference of the state space dimensions, i.e., $\kappa = \dim(\Theta) - \dim(\Theta_r)$. Usually, submodels are constructed in that way that the number of degrees of freedom is simply the difference of the number of parameters. The likelihood ratio test rejects the null Hypothesis on level α when

$$\lambda_{\mathrm{LR}}(y_1, \ldots, y_N) > \chi^2_{\kappa,1-\alpha} \tag{1.72}$$

where $\chi^2_{\kappa,1-\alpha}$ denotes the $1-\alpha$ quantile of the chi-squared distribution with κ degrees of freedom. In terms of number of parameters, there are two extreme cases of statistical models. The first is the so-called *saturated model*, which has as many parameters as observations. On the other side the *null model* in regression analysis includes only one parameter corresponding to a constant factor. The main purpose of considering the null model is to dismiss the null hypothesis that all regression parameters but one are zero, which would imply that the best description of the data is obtained by the mean. Considering the saturated model on the other hand, with as many parameters as observations, does not yield any simplification. However, the concept of a saturated model as the largest reasonable model is useful for goodness-of-fit analysis of any model of interest.

1.6.3 | Lagrange Multiplier / Score Test

Another statistic for hypothesis tests is based on the score function and is therefore called the *Score Test*, or alternatively the *Lagrange multiplier test* (LM). The LM test statistic for testing a hypothesis of the form as in eq. (1.67) is given by the quadratic form

$$\lambda_{\mathrm{LM}} = s(\hat{\boldsymbol{\theta}}_{\Theta_r})^T \mathcal{I}(\hat{\boldsymbol{\theta}}_{\Theta_r})^{-1} s(\hat{\boldsymbol{\theta}}_{\Theta_r}),$$

where $s(\cdot)$ denotes the score function (see eq. (1.40)), $\mathcal{I}(\cdot)$ the information matrix and $\hat{\boldsymbol{\theta}}_{\Theta_r}$ the maximum likelihood estimate under the restricted model Θ_r. Under the null hypothesis $H_0 : \boldsymbol{\theta} \in \Theta_r$ and some regularity conditions, the λ_{LM} statistic is asymptotically χ^2_N distributed, see Lütkepohl (2007).

Note that although the three introduced tests have equivalent asymptotic distributions under the null hypothesis, they might have different small sample properties, see, e.g., Lütkepohl (2007). Our choice among the tests will be based on the objective of the application. For the comparison of nested GAPC models that differ by categorical effects, e.g., models with a cohort effect vs. restricted models without a cohort effect, we will use the LR statistic. For significance tests of particular ages, cohorts or other predictors, we employ the Wald test.

1.6.4 | Deviance Statistic for the Exponential Family

A special case of the likelihood ratio statistic, as provided in eq. (1.70), is to consider the saturated model as the unrestricted model and compare it to the model of interest in order to determine how appropriate the proposed model fits the data. In that particular case, the likelihood ratio statistic λ_{LR} is called the *deviance* and is denoted by D.

The objective of the following is to derive the deviance statistic for the exponential family and to show how the likelihood ratio criterion for the comparison of nested models can be expressed in terms of the deviance. Let Θ denote the model of interest and Θ_S the saturated model. Further, let $\hat{\theta}_i$ denote the estimated canonical parameters and $\hat{\mu}_i$ the fitted values under the model Θ, where $\hat{\mu}_i = b'(\hat{\theta}_i)$ (see eq. (1.36)). Alternatively, let $\tilde{\theta}_i$ and $\tilde{\mu}_i$ denote the corresponding estimates under the saturated model Θ_S. Recall, the log-likelihood function for the exponential family is given by

$$\mathcal{L}(\theta_i, \phi \mid y_i) = \frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi).$$

Thus, the likelihood ratio criterion applied to those models leads to

$$\begin{aligned}\lambda_{\mathrm{LR}} &= 2\left(\mathcal{L}(\tilde{\boldsymbol{\theta}}) - \mathcal{L}(\hat{\boldsymbol{\theta}})\right) \\ &= 2\sum_{i=1}^{N} \frac{y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)}{a_i(\phi)}.\end{aligned} \tag{1.73}$$

As noted before in Section 1.5, for the distributions of interest, we have $a_i(\phi) = \phi / w_i$ with known prior weights w_i. Using that simplification, we can rewrite eq. (1.73) as

$$\phi\,\lambda_{\mathrm{LR}} = 2\sum_{i=1}^{N} w_i\left(y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\right). \tag{1.74}$$

The right-hand side of this expression is called the deviance and is denoted by $D(\boldsymbol{y},\hat{\boldsymbol{\mu}})$. The deviance can be seen as the generalization of the residual sum of squares. With Table 1.4, it can be easily verified that the deviance for the normal distributed response variables is indeed given by $D(\boldsymbol{y},\hat{\boldsymbol{\mu}}) = \sum_{i=1}^{N}(y_i - \hat{\mu}_i)^2$. Since GAPC stochastic mortality models employ Poisson and binomial distributed response variables, the corresponding deviances are derived in the following. Recall that for the Poisson distribution we have, $\theta_i = \ln\mu_i$, $b(\theta_i) = e^{\theta_i}$, $a_i(\phi) = \phi$ with $\phi = 1$ and prior weights $w_i = 1$. Thus, the deviance is given by

$$\begin{aligned} D(\boldsymbol{y},\hat{\boldsymbol{\mu}}) &= 2\sum_{i=1}^{N} w_i\left(y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\right) \\ &= 2\sum_{i=1}^{N}\left(y_i\ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right). \end{aligned} \tag{1.75}$$

Note that for the saturated model, the means $\mu_i = b(\theta_i)$ are estimated by the corresponding observations y_i. Alternatively, for the binomial distribution, we have $b(\theta) = n\ln(1 + e^{\theta})$ and $a(\phi) \equiv 1$, such that $\tilde{\theta}_i = (b')^{-1}(y_i)$ and $\hat{\theta}_i = (b')^{-1}(\hat{\mu}_i)$, and thus

$$D(\boldsymbol{y},\hat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{N}\left(y_i\ln\left(\frac{y_i}{\hat{\mu}_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i - \hat{\mu}_i}\right)\right).$$

For both, the Poisson and the binomial distribution, the dispersion parameter is $\phi = 1$ and therefore we can conclude from eqs. (1.71) and (1.74) that for large sample sizes the deviance is approximately chi-squared distributed with $N - K - 1$ degrees of freedom, where N is the number of observations and K the number of parameters of the proposed model Θ.

To see that the likelihood ratio test can be expressed in terms of the deviance, observe that for the increasing model setting, e.g., $\Theta_r \subset \Theta \subset \Theta_S$, we have

$$\begin{aligned} \lambda_{\mathrm{LR}}(\Theta_r,\Theta) &= 2\left(\mathcal{L}(\hat{\boldsymbol{\theta}}_{\Theta}) - \mathcal{L}(\hat{\boldsymbol{\theta}}_{\Theta_r})\right) \\ &= 2\left(\mathcal{L}(\hat{\boldsymbol{\theta}}_{\Theta_S}) - \mathcal{L}(\hat{\boldsymbol{\theta}}_{\Theta_r})\right) - 2\left(\mathcal{L}(\hat{\boldsymbol{\theta}}_{\Theta_S}) - \mathcal{L}(\hat{\boldsymbol{\theta}}_{\Theta})\right) \\ &= \frac{D(\boldsymbol{y},\hat{\boldsymbol{\mu}}_{\Theta_r}) - D(\boldsymbol{y},\hat{\boldsymbol{\mu}}_{\Theta})}{\phi}. \end{aligned}$$

In the forthcoming numerical case studies, we will only consider distributions where ϕ is known, like the Poisson or the binomial distribution ($\phi = 1$). Thus, in order to test $H_0 : \boldsymbol{\theta} \in \Theta_r$ versus $H_A : \boldsymbol{\theta} \in \Theta \setminus \Theta_r$, we follow the condition of eq. (1.72) and compare the difference of the deviances to a corresponding quantile of a chi-squared distribution.

1.6.5 | Information based Criteria for Model Selection

As we have seen above, the likelihood ratio test provides a method for model selection of nested models. The LR method is based on the asymptotic distribution of the test statistic and does not directly account model complexity. Alternative approaches for model selection, of not necessarily nested models, are established in terms of penalized likelihood functions. A well-known representative of that class is the *Akaike's information criterion* (see, Akaike, 1973)

$$\mathrm{AIC} = -2\mathcal{L}(\hat{\boldsymbol{\mu}}_\Theta \mid \boldsymbol{y}) + 2K \tag{1.76}$$

The criterion is based on the maximum likelihood estimate under model Θ and an additional complexity penalizing term of $2K$, where $K = \dim(\Theta)$. Another penalized maximum likelihood-based measure of goodness-of-fit is the *Bayesian information criterion* (see, Schwarz, 1978)

$$\mathrm{BIC} = -2\mathcal{L}(\hat{\boldsymbol{\mu}}_\Theta \mid \boldsymbol{y}) + K \ln N, \tag{1.77}$$

where the penalization term incorporates both, the model complexity K and sample size N. Equation (1.77) shows, that for $N > 8$, the Bayesian information criterion penalizes the model complexity higher than the Akaike's information criterion. Both criteria are used for comparison of competing statistical models, where models with lower values are preferable over those with larger values. Other information based criteria will be introduced in Section 2.3.1.

1.6.6 | GLM Residuals

Residuals are important for GLM assessment and diagnostic purposes. The accuracy of the model fit can be analysed with respect to all components of a GLM, namely, the choice of the response variable distribution, the linear predictor, and the link function. Possible weaknesses of a model can be identified through residual patterns. In general, residuals measure the deviation between the observed and the fitted values. In the following part, we will present the definition of *deviance residuals* which are a generalization of raw response residuals used for Linear Models, see Pierce and Schafer (1986) for more details and an overview of further alternative approaches to residuals for GLMs.

The raw *response residuals* are primarily used for diagnostics of LMs and are defined as the difference between the observation and the fitted value

$$r_i^R = y_i - \hat{\mu}_i, \qquad i = 1, \ldots, N.$$

In GLMs, however, the variance of Y_i depends on the covariates $\boldsymbol{x}_i$. Therefore, the raw response residual r_i^R would be an inappropriate choice since it does not cover heterogeneous variance structures. That issue can be avoided by using *Pearson residuals*

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}, \qquad i = 1, \ldots, N,$$

as scaled modifications of the response residuals, where $V(\cdot)$ is the variance function of the GLM (see eq. (1.38)). In the forthcoming case studies we will employ deviance residuals which are usually preferred over Pearson residuals due to their properties (see Hardin and Hilbe (2012)). The *deviance residuals* are defined by

$$r_i^{\bar{D}} = \operatorname{sign}(y_i - \hat{\mu}_i)\sqrt{\widehat{\text{dev}}_i}, \qquad i = 1, \ldots, N, \tag{1.78}$$

where $\widehat{\text{dev}}_i$ corresponds to the partial deviation of the i-th observation. By using eq. (1.75), we have

$$\widehat{\text{dev}}_i = w_i \left(y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \right) \tag{1.79}$$

as individual contributions to the deviance for a response variable from the exponential family. From the definition of eq. (1.78), we see that the total deviance satisfies

$$D(\boldsymbol{y}, \hat{\boldsymbol{\mu}}) = \sum_{i=1}^{N} \left(r_i^{\bar{D}} \right)^2 = \sum_{i=1}^{N} \widehat{\text{dev}}_i.$$

For Poisson distributed response variables, we conclude from eq. (1.79) and Table 1.4 that

$$\widehat{\text{dev}}_i = \begin{cases} 2\,\hat{\mu}_i & \text{if } y_i = 0 \\ 2\left(y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i) \right) & \text{otherwise.} \end{cases}$$

For the particular forms of $\widehat{\text{dev}}_i$ for the binomial and other distributions, we refer to Hardin and Hilbe (2012, Table A.11).

The **StMoMo** package, which will be used for a comparative case study of several GAPC models (see Figures 1.16 to 1.24), employs a standardized form of the deviance residuals. These *standardized deviance residuals* are defined as

$$r_i^D = \operatorname{sign}(y_i - \hat{\mu}_i)\sqrt{\frac{(N-K)\,\widehat{\text{dev}}_i}{D(y_i, \hat{\mu}_i)}}, \tag{1.80}$$

where $D(y_i, \hat{\mu}_i)$ is the total deviance, N the number of observations and K the number of model parameters. The standardized deviance residuals are the most commonly used residuals for GLM assessment. It is worth noting that for normally distributed response variables, the introduced generalizations, such as the Poisson and the deviance residuals, coincide with the raw response residuals r_i^R.

1.6.7 | Notations for GAPC Models under the GLM Framework

In this section, we leave the general notation of the previous sections and consider the concrete quantities which are required for stochastic mortality modelling. As presented in Section 1.4.3, the key quantity of interest in mortality modelling is the death count $D_{t,x}$ at period t and individuals aged x. As in Villegas, Kaishev and Millossovich (2015), the first component of the GAPC model determines the distribution of the discrete random variable $D_{t,x}$, which takes the role of the response variable within the GLM framework. As previously presented, some plausible distributions for $D_{t,x}$, which have been primarily considered in the literature, are the Poisson and the binomial distribution, i.e., $D_{t,x} \sim \text{Poi}(\mu_{t,x}E^c_{t,x})$ or $D_{t,x} \sim \text{Bin}(q_{t,x}, E^0_{t,x})$, where $E^c_{t,x}$ or $E^0_{t,x}$ denote the known central or initial exposures to risk, respectively. The aim of the GAPC models is to regress a particular function of the expected death count on some linear combination of covariates, such as age, period, and cohort terms. In case of a Poisson distributed death counts, the force of mortality $\mu_{t,x}$, as the unknown term of $\mathbb{E}[D_{t,x}] = \mu_{t,x}E^c_{t,x}$ is modelled by

$$\eta_{t,x} = g(\mu_{t,x}) = g\left(\frac{\mathbb{E}[D_{t,x}]}{E^c_{t,x}}\right),$$

where g denotes the link function and $\eta_{t,x}$ a general age-period-cohort predictor function of form

$$\eta_{t,x} = \alpha_x + \sum_{i=1}^{N} \beta^{(i)}_x \kappa^{(i)}_t + \beta^{(0)}_x \gamma_{t-x}. \tag{1.81}$$

The terms α_x, β_x in eq. (1.81) account age-related effects, whereas κ_t, γ_{t-x} represent periodic and cohort effects, respectively. An overview of popular predictor functions is provided in Table 1.2. The form of the log-likelihood function for the Poisson and the binomial model follows directly from the general exponential family log-likelihood function, as given in eq. (1.39). Thus, we have

$$\mathcal{L}(\hat{d}_{t,x} \mid d_{t,x}) = \sum_{t,x} \left(d_{t,x} \ln \hat{d}_{t,x} - \hat{d}_{t,x} - \ln\left(d_{t,x}!\right)\right) \tag{1.82}$$

for Poisson distributed death counts, and

$$\mathcal{L}(\hat{d}_{t,x} \mid d_{t,x}) = \sum_{t,x} \left(d_{t,x} \ln\left(\frac{\hat{d}_{t,x}}{E^0_{t,x}}\right) + (E^0_{t,x} - d_{t,x}) \ln\left(\frac{E^0_{t,x} - \hat{d}_{t,x}}{E^0_{t,x}}\right) + \binom{E^0_{t,x}}{d_{t,x}} \right),$$

for binomial distributed death counts, where $d_{t,x}$ denotes the observed death counts of individuals aged x at period t, and $\hat{d}_{t,x}$ denotes the model prediction obtained by MLE. The predicted numbers of deaths $\hat{d}_{t,x}$ are related to the estimated model parameters through

$$\hat{d}_{t,x} = E^c_{t,x}\hat{\mu}_{t,x} = E^c_{t,x}\, g^{-1}\left(\hat{\alpha}_x + \sum_{i=1}^{N} \hat{\beta}^{(i)}_x \hat{\kappa}^{(i)}_t + \hat{\beta}^{(0)}_x \hat{\gamma}_{t-x}\right)$$

in the Poisson case, and through

$$\hat{d}_{t,x} = E^0_{t,x}\hat{q}_{t,x} = E^0_{t,x}\, g^{-1}\left(\hat{\alpha}_x + \sum_{i=1}^{N} \hat{\beta}^{(i)}_x \hat{\kappa}^{(i)}_t + \hat{\beta}^{(0)}_x \hat{\gamma}_{t-x}\right)$$

in the binomial case, where g^{-1} denotes the inverse of the link function g. For goodness-of-fit and misspecification analysis of GAPC models, standardized deviance residuals will be used. Rewriting the general notion of standardized deviance residuals, as provided in eq. (1.80), leads to

$$r^D_{t,x} = \text{sign}(d_{t,x} - \hat{d}_{t,x})\sqrt{\frac{(N-K)\,\widehat{\text{dev}}(t,x)}{D(d_{t,x},\hat{d}_{t,x})}}, \tag{1.83}$$

where the individual contributions $\widehat{\text{dev}}(t,x)$ to the total deviance $D(d_{t,x},\hat{d}_{t,x}) = \sum_{t,x}\widehat{\text{dev}}(t,x)$ are given as

$$\widehat{\text{dev}}(t,x) = \begin{cases} 2\,\hat{d}_{t,x} & \text{if } d_{t,x} = 0 \\ 2\left(d_{t,x}\ln\left(\frac{d_{t,x}}{\hat{d}_{t,x}}\right) - (d_{t,x} - \hat{d}_{t,x})\right) & \text{otherwise,} \end{cases} \tag{1.84}$$

in the Poisson case, and as

$$\widehat{\text{dev}}(t,x) = \begin{cases} 2\,E^0_{t,x}\ln\left(\frac{E^0_{t,x}}{E^0_{t,x}-\hat{d}_{t,x}}\right) & \text{if } d_{t,x} = 0 \\ 2\,d_{t,x}\ln\left(\frac{d_{t,x}}{\hat{d}_{t,x}}\right) + 2(E^0_{t,x} - d_{t,x})\ln\left(\frac{E^0_{t,x}-d_{t,x}}{E^0_{t,x}-\hat{d}_{t,x}}\right) & \text{if } 0 < d_{t,x} < E^0_{t,x} \\ 2\,E^0_{t,x}\ln\left(\frac{E^0_{t,x}}{\hat{d}_{t,x}}\right) & \text{if } d_{t,x} = E^0_{t,x}, \end{cases}$$

for the binomial random component, respectively. The deviance statistic, standardized deviance residuals and information criteria, such as the Akaike's (AIC) and Bayesian (BIC), will be employed in the following section for a goodness-of-fit analysis of some GAPC stochastic mortality models as presented in Section 1.4.3.

1.7 | GAPC Case Study

In the following section, we provide a comparative analysis of some GAPC stochastic mortality models as introduced in Section 1.4.2. The objective of the case study is in the first place to assess the ability of the models to reflect the observed mortality rates and to capture the effects of the time evolution. In the second place, we will analyse how suitable these models are to obtain forecasts of future rates and identify the advantages and disadvantages of the distinct models. For a composition of the distinct predictor functions of the models, see Table 1.2. Parameter estimation for GAPC models is obtained by maximization of the log-likelihood function which is achieved by numerical methods as introduced in Section 1.5. In the following section, we first provide an analysis comparing the accuracy of some GAPC models from Section 1.4.3 using a Poisson random component and the canonical link function. Subsequently, we continue our analysis with likelihood ratio tests for nested model pairs to determine the significance of periodic or cohort-related effects. Afterwards, in Section 1.7.2, we analyse how these models compare to models with the same random component and same predictor function but a non-canonical link function. Subsequently, in Section 1.8, we provide an in-depth discussion on common weaknesses of widely used GAPC models.

Reference Dataset

For the following comparative analysis, we use the annual central mortality rates of the Swedish female population as our reference dataset. The analysis is restricted to the periods between 1900 and 2014, and ages between 60 and 106. The observed mortality rates of the corresponding dataset are visualized in Figure 1.14 on a linear scale and in Figure 1.15 on a logarithmic scale. From these illustrations one can observe that for each period the mortality rates are strongly increasing with the age. Moreover, the growth of the mortality rates for fixed periods shows high regularity, see also Figures 1.5 and 1.6. This indicates that models with parametric age-related effects $\beta_x^{(i)}$ might perform well, while having the advantage of being more parsimonious. Another observation is that despite the fluctuation in time there is a general decline in the mortality rate across all ages. However, the decline is not constant and is higher for lower ages (60-80) compared to higher ages (90+). GAPC models aim to capture these effects by age-periodic terms $\beta_x^{(i)}\kappa_t^{(i)}$. Cohort-related effects are usually not detectable from plots containing crude mortality rates. Hence, one has to use relative changes in the mortality in time to reveal these effects. As already reviewed in Section 1.3, Figure 1.7 shows the mortality improvements of the Swedish females (see Section 2.5.4 for the definition). From that illustration, cohort effects are not clearly detectable as they are for the males or the UK-Wales population as shown in Figures 1.8 to 1.10. In the upcoming analysis, likelihood ratio tests are used to determine whether general models with a cohort effect γ_{t-x} are favourable over nested models with no cohort term.

Software Considerations

The fitting methodology of Section 1.5.2 will be applied to several GAPC models. The computational analysis is performed within the **R** software environment. In particular, we use the **demography** package to obtain the corresponding dataset from the Human Mortality Database, and the packages **StMoMo** and **gnm** for parameter estimation. The **StMoMo** package accommodates a great collection of functions and routines which are useful for model design and validation of stochastic mortality models. It provides a very comfortable way for model specification, and handles the connection of the model, the historical data, and the routines of **gnm** package. The **gnm** package itself does the heavy workload estimating several hundred parameters by using an implementation of the Fisher scoring algorithm (see Section 1.5.2). The results of the iterative estimation scheme are then again passed to the routines of **StMoMo** package, which ensure a proper representation of the estimates and also provides some relevant statistics for goodness-of-fit analysis. For the following quantitative analysis of several GAPC models, we use our own modification of the **StMoMo** package in order to employ non-canonical link functions. [1]

1.7.1 | GAPC Model Analysis (Setting 1)

The primary objective of the following analysis is to investigate the fitting performance for distinct predictor types and study the significance of age, period, and cohort terms on the quality-of-fit.

1 At the time of writing, the current version **StMoMo** v0.3.1 can only handle canonical link functions.

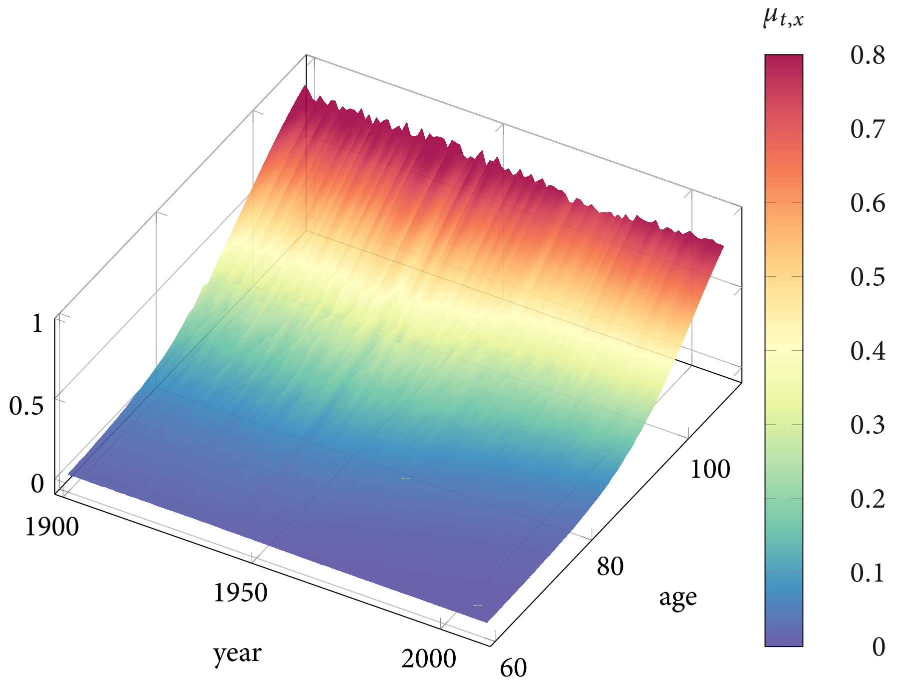


Figure 1.14: Historical central mortality rates of the Swedish female population aged 60 to 106 during the periods between 1900 and 2014.

Specifications for the first Case Study

For the first part of our quantitative analysis of several commonly used mortality models, we choose the following GAPC specification. The random component follows a Poisson distribution with mean $\mu_{t,x}E^c_{t,x}$, i.e., $D_{t,x} \sim \mathrm{Poi}(\mu_{t,x}E^c_{t,x})$. For the systematic components, the predictor functions of the types LC, LC2, CBD, APC, RH, M6, M7, and PLAT will be used, see Table 1.2 for an overview of the distinct functions.[1] The canonical Poisson link function $g = \ln$ will be used as well as the parameter constraints from Table 1.3 to ensure identification. These four components provide an unambiguous specification of a GAPC model. These models are compared on our reference dataset of Swedish females aged 60-106 in the periods between 1900 and 2014.

The estimates of the model coefficients α_x, $\beta^{(i)}_x$, $\kappa^{(i)}_t$ and γ_{t-x} are obtained by maximizing the Poisson type log-likelihood function of eq. (1.82) using the iterative Fisher scoring algorithm. Subsequently the corresponding transformations on the estimates are applied to satisfy the imposed parameter constraints.

1 Originally, the predictor functions of CBD, M6, and M7 were applied to the initial mortality rates $q_{t,x}$ and not as specified here to the central mortality rates $\mu_{t,x}$. In order to model $q_{t,x}$, an adaption from central $E^c_{t,x}$ to initial exposures $E^0_{t,x}$ is required. Since the mortality database does typically not provide both exposures, the initial exposure rates $E^0_{t,x}$ are then generated using the approximation of eq. (1.1). However, this transformation generates a new dataset. By doing that, we are no longer capable to compare different models for the dataset, but rather different models for different datasets. To avoid that circumstance and provide an even comparison, we apply all predictor functions to the central mortality rates.

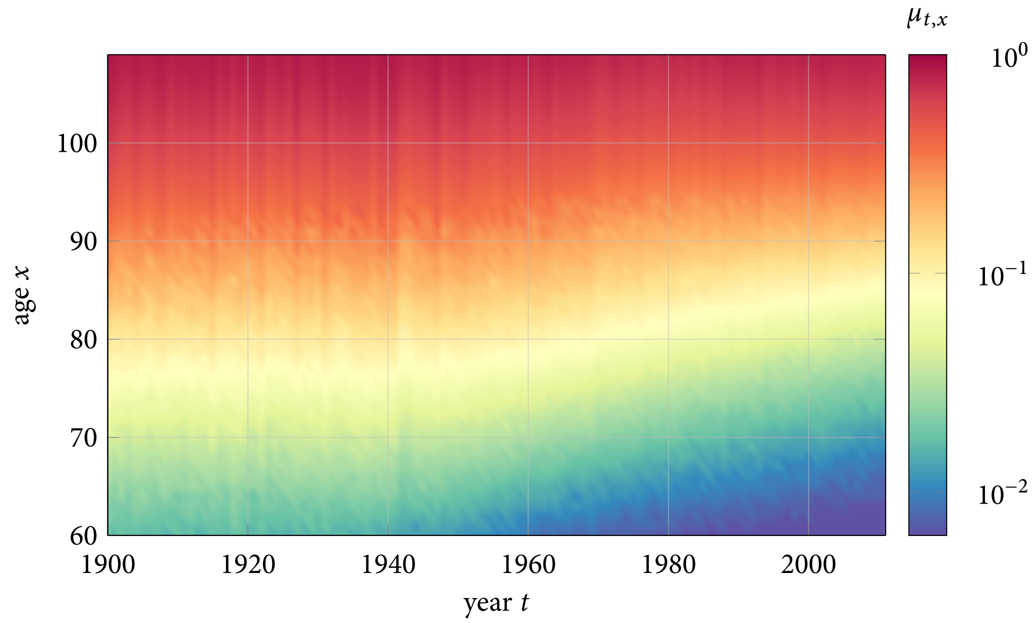


Figure 1.15: Historical central mortality rates of the Swedish female population aged 60 to 106 during the periods between 1900 and 2014 (logarithmic scale).

Fitting Results and Goodness-of-fit Analysis

The estimation results of the specified set of candidate models are provided in Table 1.5. The table contains the number of used parameters, the maximum value of the likelihood function, the total deviance, as well as the information criteria AIC and BIC. The total deviance for the Poisson case can be obtained by using eq. (1.84) which provides a measure fitting accuracy between the observed $d_{t,x}$ and the fitted deaths counts $\hat{d}_{t,x}$. Boldface numbers indicate the most favourable models across the measures: deviance, AIC, and BIC. The best fit according to the total deviance is achieved by the M7 predictor. Note that M7 predictor with 3 periodic terms $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, $\kappa_t^{(3)}$ and a cohort γ_{t-x} requires 498 parameters for the overall 5103 degrees of freedom of the underlying dataset, which is almost a ratio of 1/10. The PLAT:3 model, which is the second according to the deviance ranking, has a slightly higher deviance despite the fact that it incorporates even more parameters through the additional age-specific term α_x. The LC, APC, and the CBD predictor types are least favourable according to the deviance ranking.

To provide a comparison of the fitting performance relative to the number of used parameters, information criteria are employed. To avoid over-parametrization the AIC and BIC criteria are considered to provide a trade-off between fit accuracy and model parsimoniousness. Recall from eqs. (1.76) and (1.77) that models with a lower AIC or BIC value are preferable over the models with higher values. The M7 model turns out to be the most favourable also according to the AIC ranking. Regardless the penalizing term of the AIC criterion, the models with more parameters (M7, PLAT:2, PLAT:3) tend to dominate the models with lower number of parameters. Using the Bayesian information criterion for model selection, which has a higher penalization for more parameters,

Table 1.5: Number of parameters, log-likelihood, deviance, AIC, and BIC for the specified set of candidate models fitted to the Sweden's female population for ages 60-106 and the periods between 1900 and 2014 (5103 available observations). PLAT:2 and PLAT:3 denote the PLAT model (see eq. (1.22)) with 2 or 3 periodic terms, respectively. For predictor functions of the corresponding models, see Table 1.2.

Model	# of parm.	$\ln L$	Deviance	AIC	BIC
LC	207	−24516	9535	49446	50799
LC2	365	−22538	5579	45806	48193
CBD	230	−27543	15589	55546	57050
APC	315	−24884	10270	50397	52456
RH	362	−22726	5954	46175	48542
M6	384	−24674	9850	50115	52626
M7	498	−22157	**4817**	**45310**	48566
PLAT:2	428	−22257	5018	45371	**48169**
PLAT:3	542	−22168	4839	45420	48964

leads to the PLAT:2 model as the best performing model. The most favourable candidates, M7 and PLAT:2, have in common that both incorporate a cohort term γ_{t-x} and parametric age-related effects $\beta_x^{(i)}$.

Before we address the significance of the cohort term, as well as its interpretability, we provide a better insight into the fitting performance of these models by inspecting the standardized deviance residuals, as defined in eq. (1.83). The lack of the ability to describe essential features of the dataset will be indicated by regular residual patterns. Figures 1.16 to 1.24 illustrate heat plots of the standardized deviance residuals, plotted versus calendar year and age. Note, the reddish areas (positive residuals) on the heat maps indicate an underestimation of the death counts by the particular model, whereas bluish regions (negative residuals) indicate an overestimation of the death counts, compared to the actual observations. From Figures 1.16, 1.18, 1.19 and 1.21, we see that the models LC, CBD, APC, and M6 display significant residual patterns, whereas the models LC2, M7, and the both PLAT models, in Figures 1.17 and 1.22 to 1.24, appear quite random. The residual plot of the APC model in Figure 1.19 indicates the lack of the model to capture non-uniform age-related improvements, since this is the only model without any age-adaptive terms $\beta_x^{(i)}$. The CBD and the M6 show both strong clustering across particular age bands. That might indicate a missing non-linear modulation term $\beta_x^{(2)}$ in order to capture the progressive curvature of the mortality rates, as illustrated in Figure 1.6. None of the residual plots shows sharp diagonal patterns which would indicate the inability to capture the cohort effect. The absence of diagonal patterns for models without cohort terms is an indication that there are no substantial cohort effects in the corresponding dataset. This is in accordance with the inference from the mortality improvements, as shown in Figure 1.7, which also do not provide an indication of cohort effects.

In the following part of our quantitative analysis, we explore whether certain extensions of the predictor function are justified. Note that some proposed models are nested within other models, for instance, the APC model is a submodel of the more general PLAT models. The APC predictor function of eq. (1.14) emerges from the PLAT type predictor function, as defined in eq. (1.22), by imposing the setting $\kappa_t^{(2)} = \kappa_t^{(3)} = 0$. For nested models, likelihood ratio tests (see Section 1.6.2) are more appropriate for model assessment than the information criteria AIC and BIC. In the following, we use the LR test to check the null hypothesis that the parameters lie in the subspace of the nested model against the alternative that the more general parameter space is required. In other words, whether the restricted model can be rejected in favour of the more general one or not. Recall from eqs. (1.70) and (1.72), that the null hypothesis is rejected if the test statistic λ_{LR} exceeds the $1 - \alpha$ quantile of the chi-squared distribution with κ degrees of freedom, where α is the significance level and κ is the number of the additional parameters incorporated in the general model.

For the likelihood ratio tests, we consider the nine models from the previous analysis (see Table 1.5) joined by a LC2+C type predictor function, which was given in eq. (1.13), and a reduced M7 model without a cohort term. This collection of models leads to 13 nested pairs. The testing results are presented in Table 1.6, where the columns contain the submodel, the full model, the corresponding restriction, followed by the test value, the degrees of freedom, and the p-value. As the results show, for all nested pairs the null hypothesis is rejected in favour of the full model on any reasonable level. [1] As the p-value column shows, there is a clear statistical evidence for the justification of additional terms for each nested pair. For instance, the rejection of null hypothesis for the pairs LC vs. RH, LC2 vs. LC2+C, CBD vs. M6, and M7:sub vs. M7, justifies the existence of cohort term. Consequently, there is an indication for cohort-related mortality effects in the corresponding dataset. However, in Section 1.8.2, we will discuss the validity of that implication by analysing whether the cohort term truly captures the cohort effects of the dataset.

Other consequences of the previous analysis are that the rejection of the null hypothesis for the pairs APC vs. PLAT:2 and APC vs. PLAT:3, indicates the presence of a non-trivial correlation structure. This is coherent with the observation of age-related mortality improvements as illustrated in Figures 1.7 and 1.15.

Conclusions for the GAPC Model Analysis (Setting 1)

In the previous analysis of various GAPC models, we have investigated the fitting performance for distinct predictor types. The primary objective was to provide better insights on the impacts of the age, period, and cohort terms on the quality-of-fit. The fitting results, as summarized in Table 1.5, suggest that the models M7 and PLAT:2 provide the most favourable predictor structures to describe the essential features of our reference dataset. The overall conclusion of significance tests, as depicted in Table 1.6, is that model selection based on the likelihood ratio tests would justify additional terms in favour of a less parsimonious model, even though, the larger model contains up to several hundreds more parameters. Note that the LR test checks the significance for the entire parameter groups such

1 For any significance level $\alpha > 0.001$.

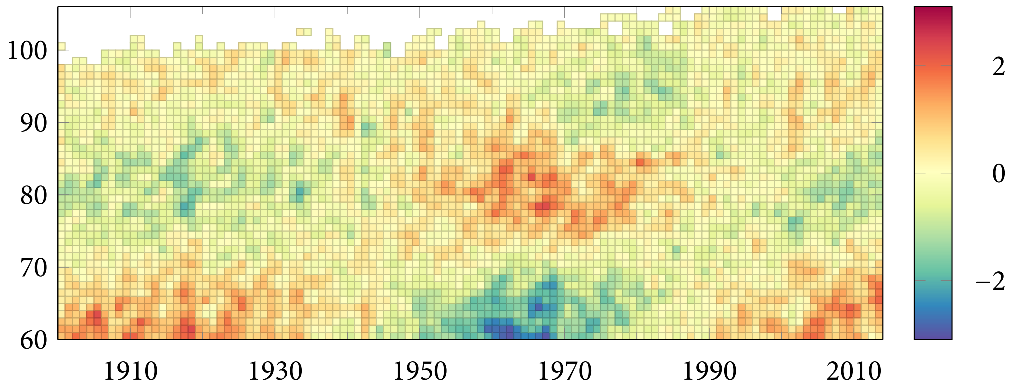


Figure 1.16: Standardized deviance residuals of the Poisson LC model with log link function.

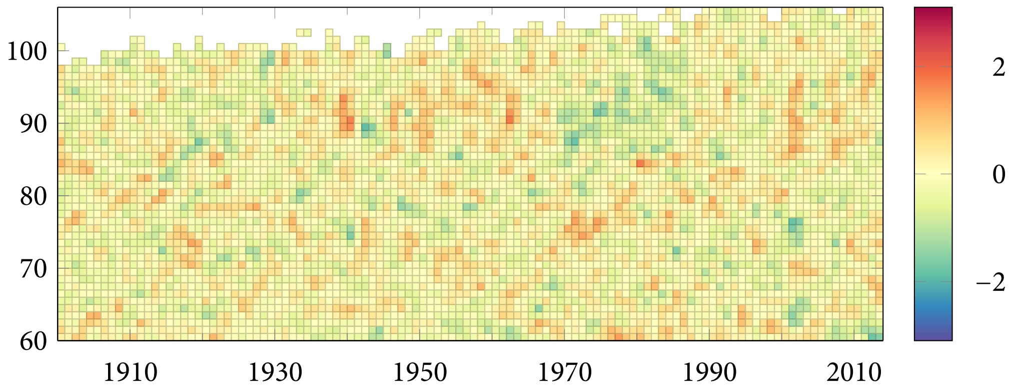


Figure 1.17: Standardized deviance residuals of the Poisson LC2 model with log link function.

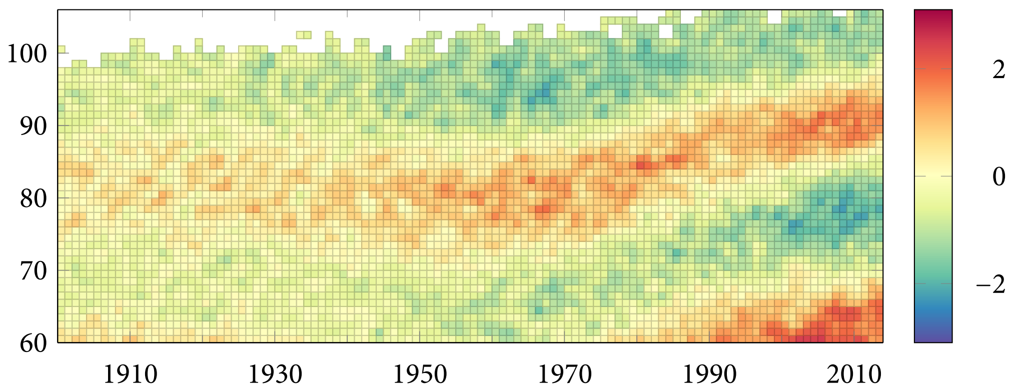


Figure 1.18: Standardized deviance residuals of the Poisson CBD model with log link function.

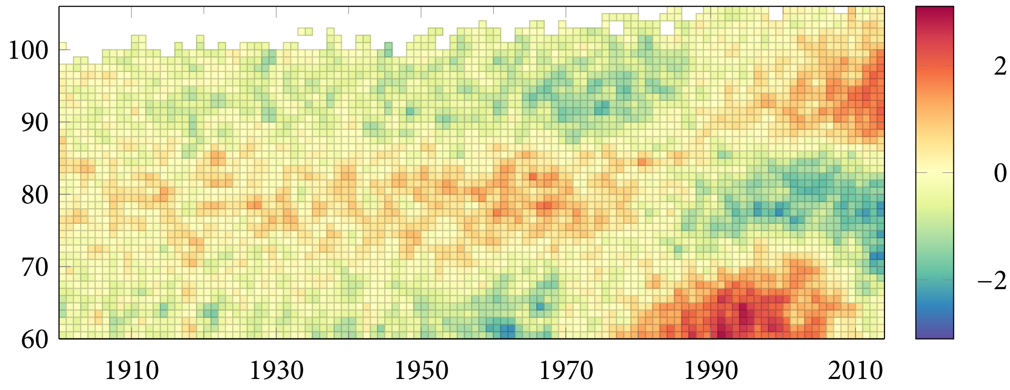


Figure 1.19: Deviance residuals of the Poisson APC model with log link function.

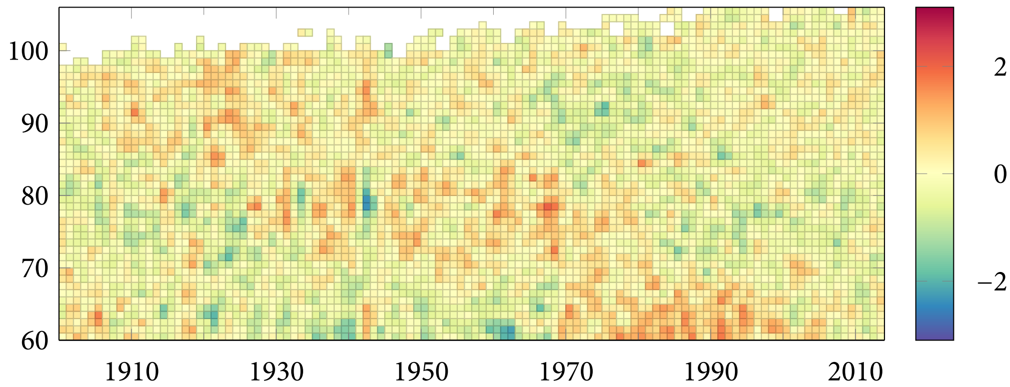


Figure 1.20: Standardized deviance residuals of the Poisson RH model with log link function.

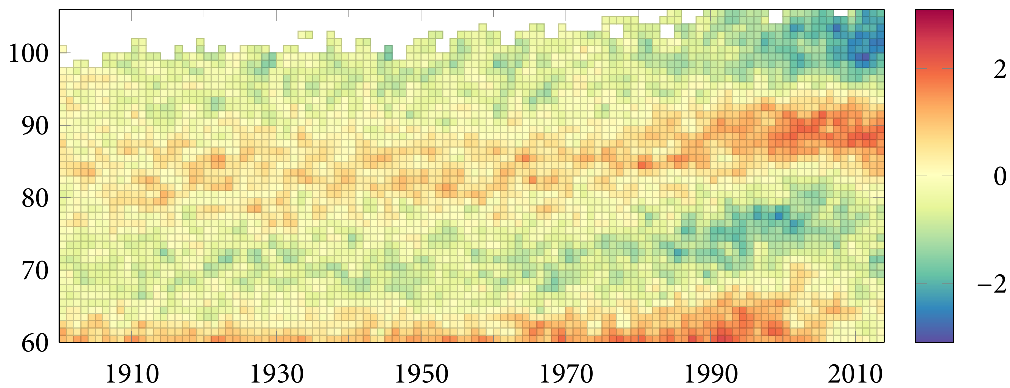


Figure 1.21: Standardized deviance residuals of the Poisson M6 model with log link function.

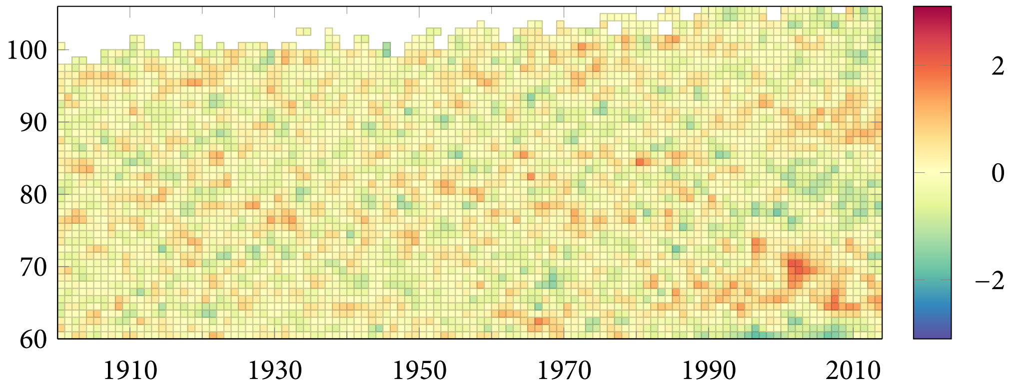


Figure 1.22: Standardized deviance residuals of the Poisson M7 model with log link function.

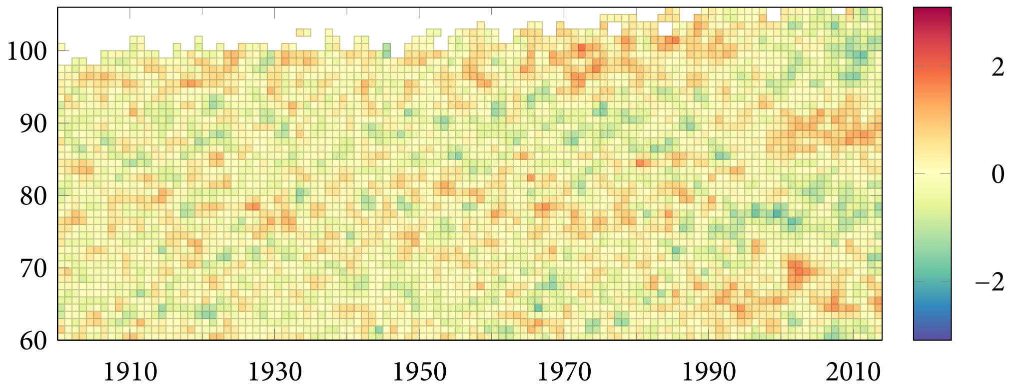


Figure 1.23: Standardized deviance residuals of the Poisson PLAT:2 model with log link function.

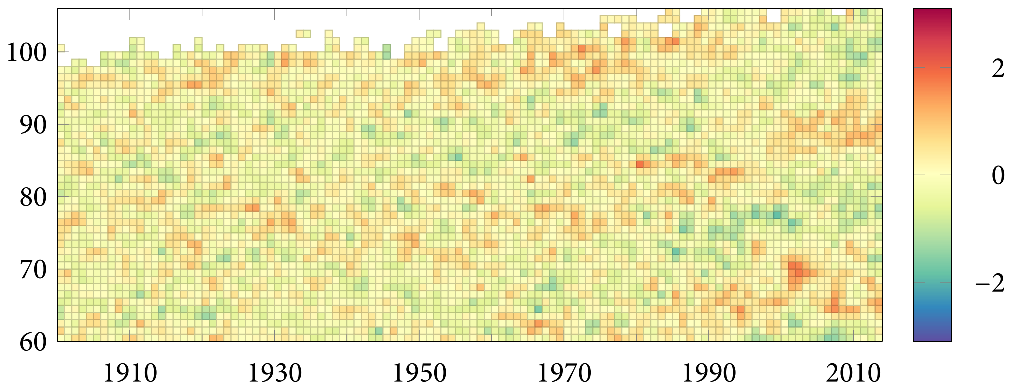


Figure 1.24: Standardized deviance residuals of the Poisson PLAT:3 model with log link function.

Table 1.6: Results of the likelihood ratio tests for various pairs of general and restricted models. The columns λ_{LR} and d.f. contain the LR test statistic and the degrees of freedom. The latter is equal to the difference in the dimensionality of the general and the restricted model. For all test pairs, the null hypothesis is rejected at any reasonable significance level in favour of the general model. Predictor functions of these models can be found in Table 1.2.

H_0 (nested)	H_A (general)	Restriction	λ_{LR}	d.f.	p-value
LC	RH	$\gamma_{t-x} = 0$	3580	155	< 0.0001
LC	LC2	$\beta_x^{(2)}\kappa_t^{(2)} = 0$	3956	158	< 0.0001
LC	LC2+C	$\beta_x^{(2)}\kappa_t^{(2)} = \gamma_{t-x} = 0$	5000	313	< 0.0001
LC2	LC2+C	$\gamma_{t-x} = 0$	1045	155	< 0.0001
CBD	M6	$\gamma_{t-x} = 0$	5738	154	< 0.0001
CBD	M7	$\kappa_t^{(3)} = \gamma_{t-x} = 0$	10772	268	< 0.0001
CBD	PLAT:2	$\alpha_x = \gamma_{t-x} = 0$	10572	198	< 0.0001
CBD	PLAT:3	$\alpha_x = \kappa_t^{(3)} = \gamma_{t-x} = 0$	10750	312	< 0.0001
APC	PLAT:2	$\kappa_t^{(2)} = 0$	5254	113	< 0.0001
APC	PLAT:3	$\kappa_t^{(2)} = \kappa_t^{(3)} = 0$	5432	227	< 0.0001
M6	M7	$\kappa_t^{(3)} = 0$	5034	114	< 0.0001
M7:sub	M7	$\gamma_{t-x} = 0$	7148	153	< 0.0001
PLAT:2	PLAT:3	$\kappa_t^{(3)} = 0$	178	114	0.00012

as cohort effects or period-related effects. Before we continue with an analysis on the significance of individual elements of these categorical groups in Section 1.8, we investigate how the accuracy of the fit changes by using an alternative link function.

1.7.2 | GAPC Model Analysis (Setting 2)

Many of the predictor types of the previous analysis were proposed before it become aware that these mortality modelling approaches could be unified by the GLM framework. The GLM framework has an additional degree of freedom, which can be utilized for modelling, and, for traditional reasons, has been mostly ignored in the actuary literature. Apart from a few exceptions (see, e.g., (Currie, 2016)), most GAPC models only use the canonical link function. In the following case study, we demonstrate the impact of a non-canonical link function to the quality-of-fit.

Specification for the second Case Study

For the second quantitative analysis of GAPC model, the distribution type of the response variable, the predictor functions, and the imposed parameter constraints remain the same as in the first case study of Section 1.7.1. The only modified component is the link function, where in addition to the Poisson canonical ($g = \ln$), a logit link ($g = \text{logit}$) is considered. In Remark 1.5.1, we have provided the explicit forms of the weighting matrix W and working variable Z for the IRLS algorithm, which are employed here for parameter estimation. Note, to apply the logit link to a Poisson distributed response variable, we have to ensure that the underlying data is within the domain of the logit function, which is the interval $(0,1)$. However, for the human mortalities, $0 < \mu < 1$ holds even for very high ages.

To examine whether this setting leads to a higher quality fit, we provide an analysis based on 6 datasets including the female mortality rates of the countries Denmark, Finland, France, Sweden, Switzerland, and UK-Wales. The data from HMD are considered on the same periods from 1900 to 2014 and ages from 60 to 106, as in the first setting. The objective of the upcoming analysis is to conclude whether the non-canonical logit link is preferable for a set of different predictors.

Goodness-of-fit Analysis of Canonical Poisson Link versus Logit Link

Tables 1.7 to 1.12 summarize the fitting results of both link functions. Note that despite taking the same period and the same age range, the number of observations differs for individual countries. The discrepancy in the number comes due to missing data points in the HMD, which are present, in particular, at higher ages at the first part of the 20th century. The results of Tables 1.7 to 1.9, 1.11 and 1.12 lead to the conclusion, that for all countries, the logit link function provides a better fit for the LC, CBD, APC, M7, and PLAT models, with Denmark being the only exception for the CBD model. The only model which shows a different behaviour, favouring the canonical link, is the RH model. This exception might be the result of the already discussed convergence issues of the RH predictor, see Section 1.4.3, and Hunt and Villegas (2015) and Macdonald, Gallop, Miller et al. (2007). As pointed out in Section 1.5.1, in addition to the already problematic behaviour for the canonical case, using the logit link, leads to a not necessarily concave function. Thus further complications can potentially arise since numerical approximation of the estimates might converge to a local and not to the global maximum.

The accumulation of the results in Tables 1.7 to 1.12 shows that for our reference dataset the predictor preference is mostly preserved across the used link functions. For instance, similar to Sweden, we observe the preference of the M7 and the PLAT models for Denmark, Finland, France, Switzerland, and UK-Wales. For poorly performing predictors there are a few changes in the predictor rankings, see, e.g., the models LC and CBD for Denmark.

Conclusions for the GAPC Model Analysis (Setting 2)

As we have demonstrated, a non-canonical link function can have a positive impact on the quality-of-fit. It is important to stress out that most mortality studies within the GAPC framework have been conducted using only the canonical choice, see among others, e.g., Cairns, Blake, Dowd et al. (2009). Since, from the GLM perspective, there is no a priori reason for omitting this degree of freedom for model selection, we assume that this issue arises due to the inability to handle alternative link functions in current software environments. For further results, advocating the application of non-canonical link functions see, Currie (2016), who also demonstrated the advantage of non-canonical link functions for the binomial random component.

Table 1.7: Number of observations, number of parameters, and the total deviance for the Poisson LC model with log and logit link functions. The lower deviance is shown in bold.

Country	# of obs.	# of parm.	Dev (g = ln)	Dev (g = logit)
Denmark	4958	207	8529	**8433**
Finland	4757	205	9143	**8955**
France	5176	206	43594	**39120**
Sweden	5103	207	9533	**9334**
Switzerland	4774	204	8173	**7646**
UK-Wales	5273	206	40450	**38872**

Table 1.8: Number of observations, number of parameters, and the total deviance for the Poisson CBD model with log and logit link functions. The lower deviance is shown in bold.

Country	# of obs.	# of parm.	Dev (g = ln)	Dev (g = logit)
Denmark	4958	230	**8215**	10573
Finland	4757	226	11765	**9151**
France	5176	228	96433	**88916**
Sweden	5103	230	15637	**14901**
Switzerland	4774	224	12743	**12130**
UK-Wales	5273	228	52628	**43405**

Table 1.9: Number of observations, number of parameters, and the total deviance for the Poisson APC model with log and logit link functions. The lower deviance is shown in bold.

Country	# of obs.	# of parm.	Dev (g = ln)	Dev (g = logit)
Denmark	4958	314	7866	**6324**
Finland	4757	308	7450	**6840**
France	5176	316	46454	**32602**
Sweden	5103	315	10282	**8799**
Switzerland	4774	305	8782	**7417**
UK-Wales	5273	316	37037	**27956**

Table 1.10: Number of observations, number of parameters, and the total deviance for the Poisson RH model with log and logit link functions. The lower deviance is shown in bold.

Country	# of obs.	# of parm.	Dev (g = ln)	Dev (g = logit)
Denmark	4958	361	**4992**	5051
Finland	4757	355	**6040**	6200
France	5176	363	**13729**	15307
Sweden	5103	362	**5953**	6475
Switzerland	4774	352	**4870**	5061
UK-Wales	5273	363	**16299**	16398

Table 1.11: Number of observations, number of parameters, and the total deviance for the Poisson M7 model with log and logit link functions. The lower deviance is shown in bold.

Country	# of obs.	# of parm.	Dev (g = ln)	Dev (g = logit)
Denmark	4958	497	4248	**4087**
Finland	4757	487	4893	**4882**
France	5176	497	10096	**9483**
Sweden	5103	498	4817	**4709**
Switzerland	4774	482	4562	**4439**
UK-Wales	5273	497	15285	**14446**

Table 1.12: Number of observations, number of parameters, and the total deviance for the Poisson PLAT:2 model with log and logit link functions. The lower deviance is shown in bold.

Country	# of obs.	# of parm.	Dev (g = ln)	Dev (g = logit)
Denmark	4958	427	4280	**4239**
Finland	4757	419	5049	**5015**
France	5176	428	11809	**10310**
Sweden	5103	428	5013	**4788**
Switzerland	4774	415	4798	**4543**
UK-Wales	5273	428	12114	**11781**

The fact that models with a logit link function perform better than their competitors with a logarithmic link indicates that human mortality rates for higher ages (60-106) obey a logistic-type growth rather than an exponential. Note that the logit function is the inverse of a logistic function. The summarized results in Tables 1.7 to 1.12 lead to the conclusion that the logit-transformed mortality rates appear more linear compared to the log-transformed case.

1.8 | Common Issues and Problematic Properties of GAPC Models

In this section, we provide a critical reflection on GAPC models based on their general properties and the case studies of Sections 1.7.1 and 1.7.2. The discussion with respect to the quality criteria for mortality models by Cairns, Blake and Dowd (2008), as listed on page 20, also aims to examine to which extent the best explanatory models are also well-suited for forecasting purposes.

In the following, we identify and discuss some areas of main difficulties related to GAPC models. Section 1.8.1 focuses on issues which arise due to additional parameter constraints in order to ensure model identification. In Section 1.8.2, we discuss the weaknesses of the cohort term and its ability to capture or reveal cohort effects. Subsequently, in Section 1.8.3, we focus on the significance test of individual parameters in contrast to the analysis on whole categorical terms, as considered in Section 1.7.2.

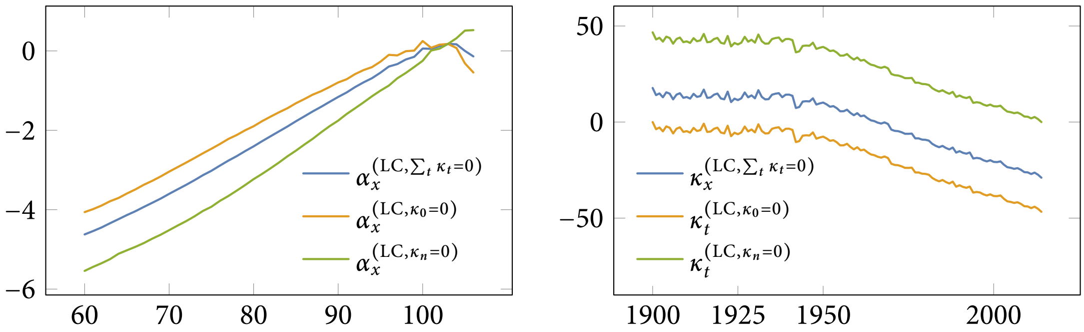
1.8.1 | Identification Issues and the Implication on Robustness, Interpretability, and Internal Dependencies

First, we start with the issue of parameter identification, which was already discussed in Section 1.4.3. Recall, that except the CBD, all predictor types require additional parameter constraints to ensure identification and that these constraints can be chosen arbitrarily. This has versatile implications on the concepts of interpretation, comparability, and robustness.

First, in Figure 1.25, we illustrate how different constraints influence the parameter estimates. The figure depicts the estimates for the LC predictor for three distinct constraints. In the first case we use the suggested constraints, i.e., $\sum_x \beta_x = 1$ and $\sum_t \kappa_t = 0$ (see eq. (1.8)). Next, we consider two constraint modifications of the periodic term, namely, $\kappa_0 = 0$ and $\kappa_n = 0$. Figure 1.25 clearly shows that the parameter estimates obey some structural changes, only due to arbitrarily chosen constraints. Note that all pairs of estimates (blue, orange, green) in Figure 1.25 lead to the same model predictor and thus describe the same mortality structure. The different behaviour of the term α_x for high ages is only the result of different constraints and not an indicator of particular mortality effects in the dataset. The LC model is one of the simplest models within the GAPC family. However, as the observation from above demonstrates, the paths of the estimates are difficult to interpret. Neither the absolute level, as shown in Figure 1.25(b), nor the trend changing behaviour, as shown in Figure 1.25(a), are solely caused by the underlying data. The decision whether the path behaviour is purely data-driven or artificial is a challenging task. For related studies on how different identifiability constraints lead to different patterns in the estimates, we refer to Hunt and Villegas (2015).

To highlight further problematic behaviour of the most GAPC models, we refer to Figures 1.26 and 1.27, where the parameter estimates of the most favourable models from Section 1.7.1 are presented. Figure 1.26 illustrates the estimates of the periodic terms $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, $\kappa_t^{(3)}$ and the cohort term γ_{t-x} of the M7 model. Figure 1.27 shows the estimates of α_x, $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and γ_{t-x} of the PLAT:2 model. The solid lines present the estimates on the full reference dataset of Swedish females aged 60-106 in between 1900 to 2014, whereas the dashed lines show the estimates on the reduced dataset, omitting the period 1900-1949. Several effects are apparent in these illustrations. Apparently, omitting the first part of the dataset, influences the estimates not only at the beginning of the period, around the 50s, but also in very recent periods. That means, that the representation of the mortality in the year 2014 depends not only on the data from that year but also on the mortality data from decades ago. This is common for models which require additional parameter constraints.

Furthermore, it is important to underline that in order to represent the mortality of, e.g., 2014 for the ages 60-106, it requires 96 parameters for the PLAT:2 model. This is indeed more than the degrees of freedom of these observations. This appears to be paradoxical, however, we have 47 observations $d_{2014,60}, \ldots, d_{2014,106}$, which are explained by the PLAT:2 predictor (see eq. (1.22)) with 47 age terms $\alpha_{60}, \ldots, \alpha_{106}$, 2 periodic terms $\kappa_{2014}^{(1)}$, $\kappa_{2014}^{(2)}$, and 47 cohort terms $\gamma_{1908}, \ldots, \gamma_{1954}$. The M7 model requires 50 parameters (3 periodic and 47 cohort variables) to explain the 47 observations. The reason that these over-parameterized models do not explain the observations perfectly is of course that the same parameter terms as α_x or γ_{t-x} are also intended to explain the mortality for various years. The CBD is the only model which does not share this behaviour. To explain the mortality of a given year t it requires only 2 parameters $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. The CBD model, however, showed a



(a) Estimates of the static age functions α_x for the LC predictor.

(b) Estimates of the periodic function κ_t for the LC predictor.

Figure 1.25: Estimated parameters α_x and κ_t of the LC model under different identification constraints. The blue lines represent the standard choice of the Lee Carter model, i.e., $\sum_t \kappa_t = 0$. The orange lines represent the estimates obtained by the constraint $\kappa_0 = 0$. The green lines represent the estimates for $\kappa_n = 0$. Note, while the κ_t estimates only differ up to affine linear transformation, the static age functions α_x, however, differ by a non-linear transformation.

relatively poor quality-of-fit, as emerged in the case studies of Sections 1.7.1 and 1.7.2.

According to the mentioned quality criteria (see page 20) for mortality models, the parameter estimates should be robust relative to the period of data and range of ages employed. As we have noticed, models which require parameter constraints are not robust since the range of data has an influence on the constraints and those have direct impact on the estimates. For instance, simple dataset updates, such as adding the most recent period of mortality data, changes the values of any previously estimated parameters. From that perspective, most GAPC models do not show any robustness.

Furthermore, we would like to draw the attention to the dependence structure of the estimates. In contrast to the estimates of the CBD model, where $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ only depend on data of period t and not on estimates of other parameters, the situation for all other models is much more complicated. Non-parametric age or cohort terms and additional parameter constraints lead to an entanglement of these parameters. As we discussed above, a parameter which is indexed by t or x does not only rely on the mortality data from the particular period or age, but also on the data of other periods or ages. Thus, there is an internal complex connection between the estimates. This dependence structure appears to be too difficult to be modelled by a simple random walk process, as commonly assumed in the literature, see, e.g., Cairns, Blake, Dowd et al. (2009). Even more questionable appears to be the standard assumption of the independence of the cohort term and the period terms. From Figure 1.26, we see that the regularity of the estimates $(\kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)})$ of model M7 can hardly be explained by a three-dimensional random walk. It is quite challenging to find an appropriate time series process to capture the behaviour of the period terms of the illustrated models. Based on the estimated paths, random walk processes are hardly suitable to express the progress of the shown paths $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. Further results on the assumption of independence between the predictor terms and the role of the constraints on the estimation can be found in Currie (2012).

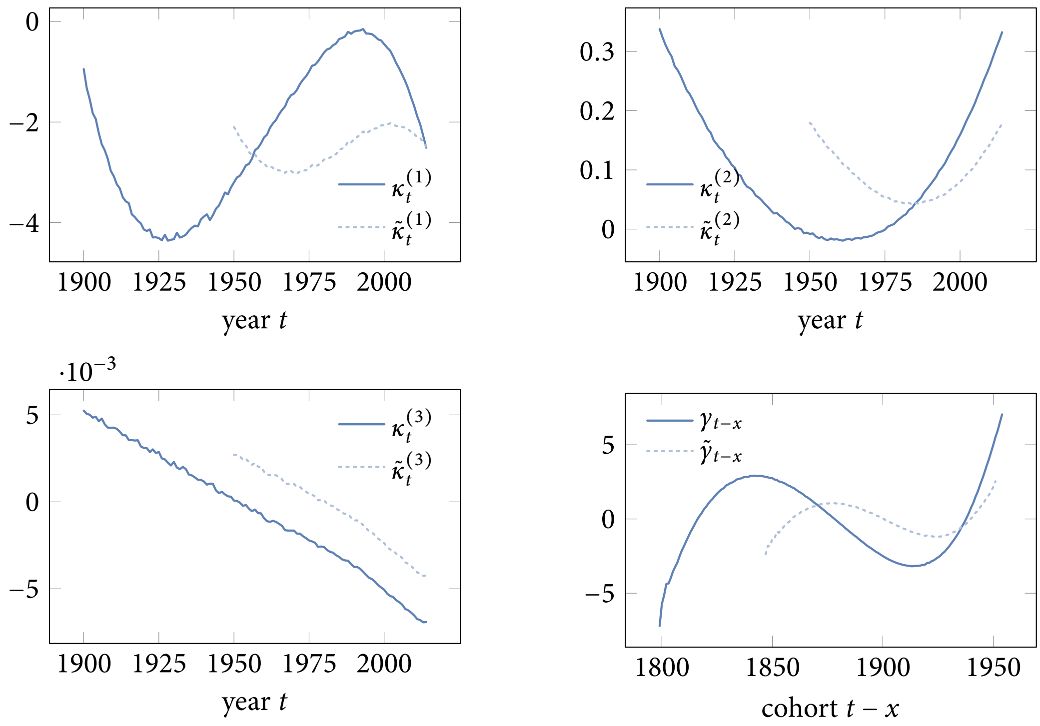


Figure 1.26: Parameter estimates of the M7 model fitted to the Sweden's female population for ages 60-106 and the periods between 1900 and 2014 in solid lines. The dotted lines show the estimates of the same model for the periods between 1950 and 2014.

1.8.2 | Issues and Weaknesses of the Cohort Term

The second part of the critical assessment is devoted to the modelling approach of the cohort effect. Recall from Section 1.4.3 that all considered approaches use a non-parametric cohort term γ_{t-x}. Moreover, additional to identifiability constraints some structural constraints are imposed on the cohort term. The purpose of these postulations is mainly to obtain desirable properties, such that the resulted time series of parameter estimates might be modelled by stationary processes. For instance, the constraints of eq. (1.19) for the M7 model, ensure that the cohort term vary around zero with no linear or quadratic trends. The idea behind this proposal is to obtain a cohort effect which is coherent with the intuition regarding some desired properties of the cohort effect. According to Hunt and Blake (2014), these properties include, e.g., the absence of any systematic trends in value or variability. Furthermore, averaged across all cohorts, the effect should be zero and represent deviations from a reference level, rather than compensate discrepancies of other terms. Finally, the effect should be mean reverting and show positive autocorrelation across successive cohorts.

The bottom right panels of Figures 1.26 and 1.27 show the parameter estimates of the cohort term of the models M7 and PLAT:2 from the case study of Section 1.7.1. Recall that models with a cohort term outperformed their competitors, see the goodness-of-fit results in Table 1.5 and the likelihood ratio tests in Table 1.6. From the paths of the estimates, we see that the most desired property does not occur. The paths do not show deviations from a reference level for particular cohorts, indicating

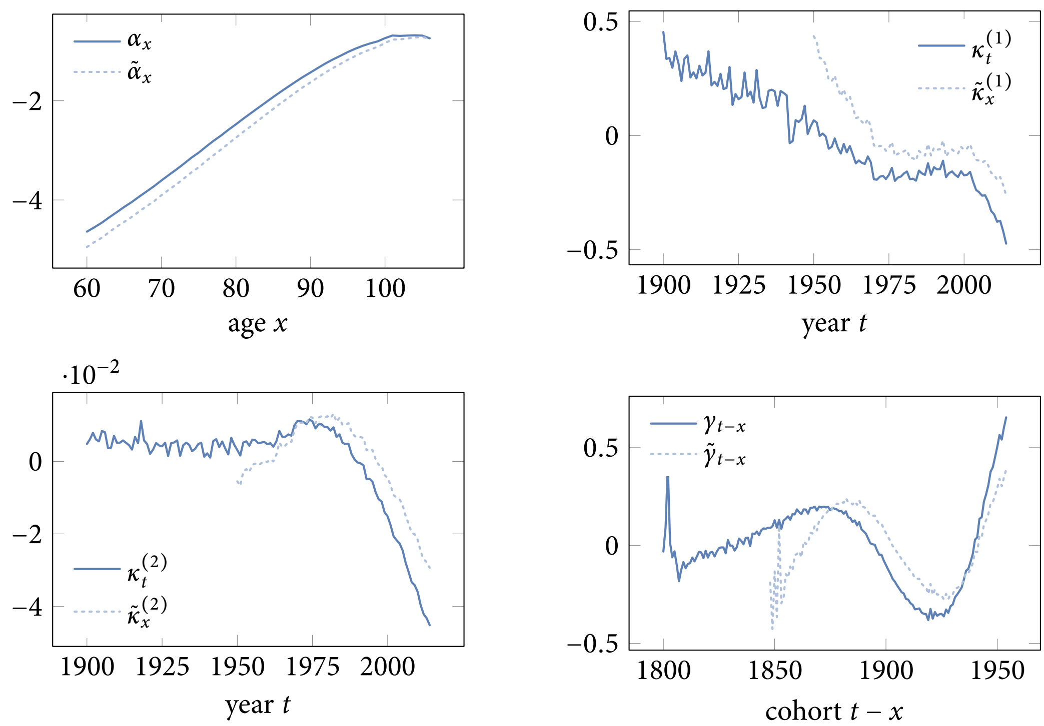


Figure 1.27: Parameter estimates of the PLAT:2 model fitted to the Sweden's female population for ages 60-106 and the periods between 1900 and 2014 in solid lines. The dotted lines show the parameter estimates of the same model on the periods between 1950 and 2014.

higher or lower cohort-related mortality effects, but rather reveal that the paths structures are mainly determined by constraints of eqs. (1.19) and (1.24). As already criticized above, the patterns of the paths are not only data-driven but also heavily influenced by the imposed constraints. Even though the cohort terms improve the accuracy of the fit, it is challenging to ensure interpretability of these parameters and attribute cohort effects to the values of the cohort term.

To underline the lack of parameters interpretability, we provide another example. The mortality data of the UK-Wales females often serves as a prime example for the cohort effects. The mortality improvements of UK-Wales females, which are provided in Figure 1.10 on page 18, clearly show the impact of the "golden generations" born around the year 1935. To detect whether this considerable effect is captured by the cohort term, we fit several models to the UK-Wales female population aged 60-106 in the period 1900 to 2014. Figure 1.28 illustrates the parameter estimates of the cohort term using the predictors of the models APC, RH, M6, M7, and PLAT. The results reveal that none of these models is able to provide evidence of the observed cohort effect. As for the Swedish females, the patterns of the paths are primary dominated by the imposed constraints. Furthermore, the regularity of the paths in Figure 1.28 and the examples of Figures 1.26 and 1.27 does not justify the common assumption that the cohort term follows a stationary ARIMA process.

For another example, which demonstrates that the GAPC cohort term misses the intended purpose, we consider a comparison of the reduced M7:sub ($\gamma_{t-x} = 0$) and the full M7 predictor. As shown

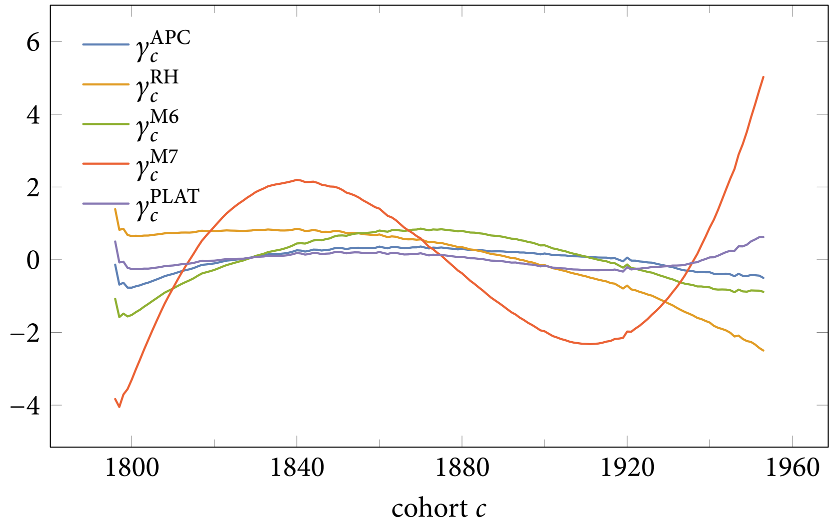


Figure 1.28: Estimated cohort parameters of the models APC, RH, M6, M7, and PLAT for UK-Wales female population. The annual mortality improvements, as visualized in Figure 1.10, demonstrate a clear effect of cohort-related mortality improvements around the 1935 birth cohorts. None of these models is able to provide evidence of this cohort effect.

in Table 1.6, the null hypothesis $H_0 : \gamma_{t-x} = 0$ is rejected in favour of the full M7 model. Thus, the fitting performance of the full models justifies the cohort term. As introduced in Section 1.4.3 the purpose of the cohort term is to reveal the cohort-related features of the dataset rather than capture the fitting discrepancies of the remaining period or age-related terms. Figures 1.29 and 1.30 show scatter plots of the standardized deviance residuals by age, period, and cohort for the M7:sub and the M7 models. For the reduced M7:sub model Figure 1.29(a) reveals the difficulties of the model to capture age-related mortality effects. This is recognizable by the appearance of systematic patterns of the residuals when plotted against the involved ages. For instance, we observe some systematic underestimations of the death counts for ages around 60-65 and 85-90 and some overestimations of the death counts for ages around 75-80 and 100-105. The residuals plotted versus period and cohort, as illustrated in Figures 1.29(b) and 1.29(c), do not show substantial systematic deviations. Adding the cohort term and considering the residuals scatter plots of the full M7 model in Figure 1.30, shows that the age-related systematic deviations have disappeared. The direct comparison of Figures 1.29(a) and 1.30(a) illustrates that the cohort term of the full model does increase the fitting performance. However, this is achieved by the compensation of the fitting discrepancy across ages rather than cohorts. The conclusion of this analysis and the results from both case studies are, that the cohort term is indeed useful, in terms of providing additional degrees of freedom, but it does not serve the intended purpose to explain or reveal cohort effects.

1.8.3 | Significance of Individual GAPC Parameters

In Section 1.7.2 likelihood ratio tests have been considered to examine whether age, period, or cohort terms provide significant improvements of the quality-of-fit. In the following section, we focus on the significance of individual parameters from these categorical groups, since we have not found

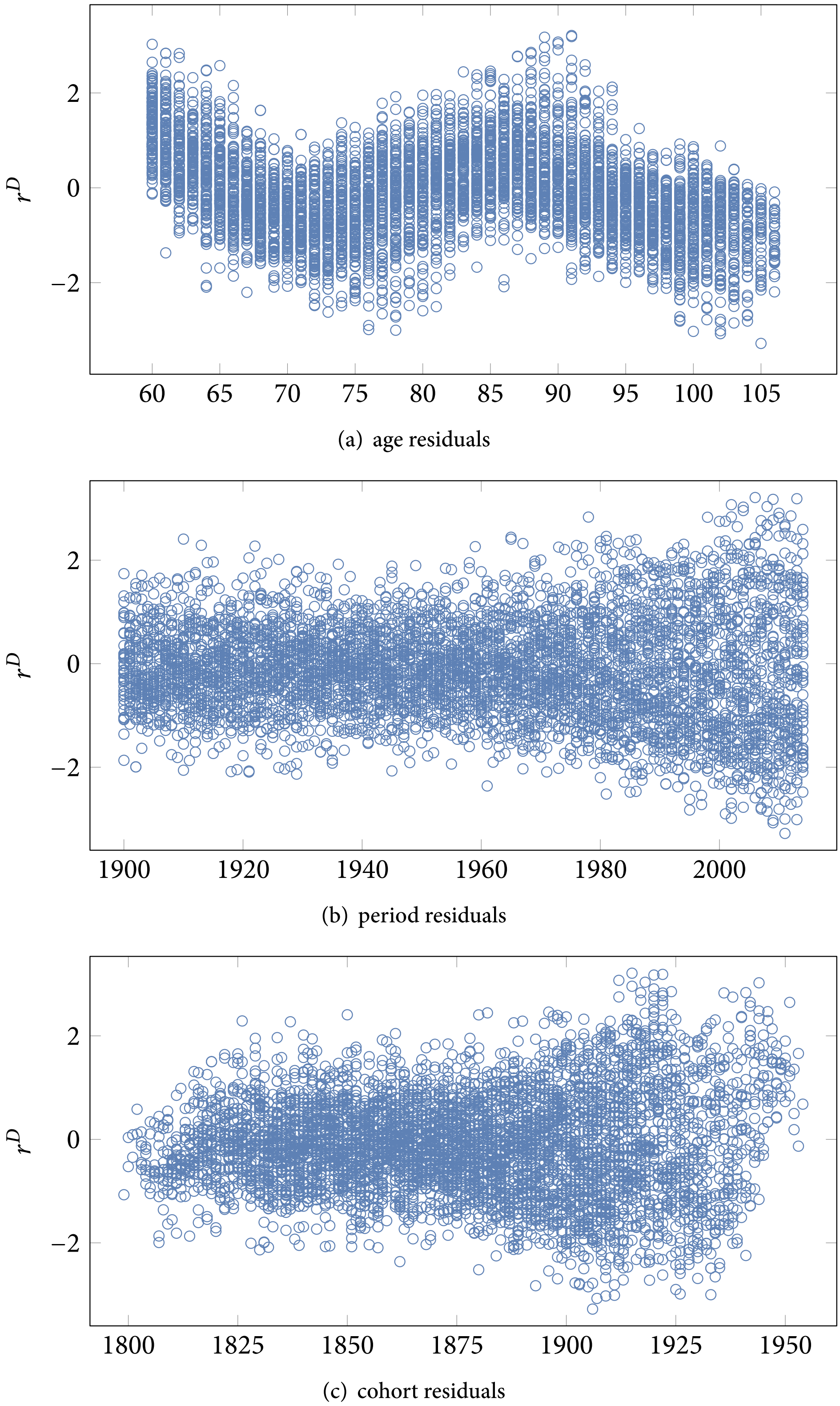
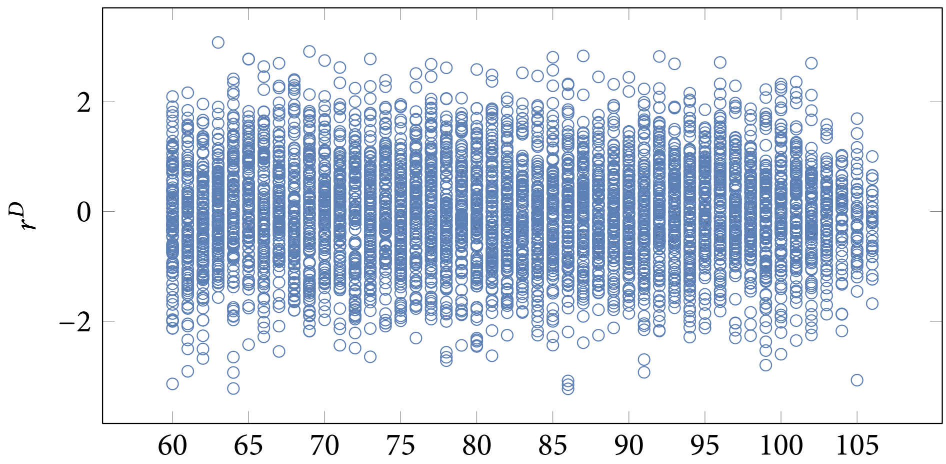
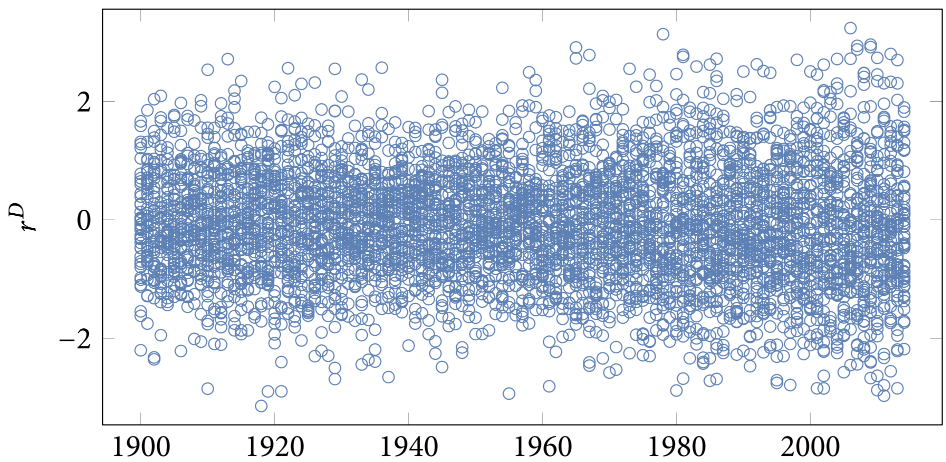


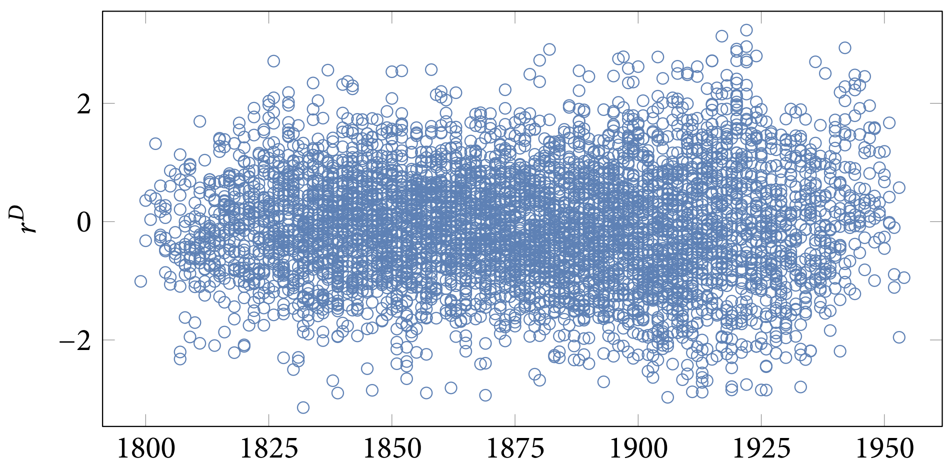
Figure 1.29: Scatter plots of deviance residuals $r^D_{t,x}$ for the Poisson model with log link and a reduced M7 predictor function ($\gamma_{t-x} = 0$) fitted to Swedish female population for ages 60-106 and the periods between 1900 and 2014.



(a) age residuals



(b) period residuals



(c) cohort residuals

Figure 1.30: Scatter plots of deviance residuals $r_{t,x}^D$ for the Poisson model with log link and M7 predictor function fitted to Swedish female population for ages 60-106 and the period 1900-2014.

any other considerations on this topic in the literature. Commonly, the fitting procedure of linear or generalized linear models always includes a significance analysis of the model coefficients to determine which predictors have a statistically relevant influence on the response variable. For GAPC models this analysis is rather done for whole parameter groups (α_x, β_x, κ_x, γ_{t-x}) since it is not a priori clear why certain ages, periods or cohorts should be included in the model while others can be removed. We do not intend to criticize this approach, since we consider that to be reasonable to treat these factors equally, however, we want to stress out our findings on the significance of individual GAPC model parameters.

To determine the significance of individual model parameters, we consider Wald type tests as introduced in Section 1.6.1. More specifically, we intend to test the hypothesis $H_0 : \delta = 0$ versus $H_A : \delta \neq 0$, for δ being one parameter of the corresponding set of model parameters. For instance, taking the APC model and the setting of the first case study, we have

$$\delta \in \{\alpha_{60}, \ldots, \alpha_{106}, \kappa_{1900}, \ldots, \kappa_{2014}, \gamma_{1800}, \ldots, \gamma_{1952}\},$$

i.e., δ is one of the 315 APC parameters. Table B.1 on page 259 shows the results of the regression analysis for the APC predictor with the canonical link. The columns contain the coefficient names, the estimates, the standard errors, the test values, the p-values and the standardized significance codes for better readability. Recall from eq. (1.65), that the z-value denotes the test value of the Wald test, which is given by

$$T_W = \frac{\hat{\delta}}{\hat{\mathrm{se}}(\hat{\delta})}.$$

Recall also that the p-value is the probability (under H_0) of obtaining a test value equal to or even more extreme than the observed test value. Thus, the p-value is the smallest level at which H_0 can be rejected. For instance, for $p < 0.01$, we have a very strong evidence against H_0, which means, we have a strong evidence to suggest that the corresponding parameter is not zero. Table B.1 shows that for the APC model only 143 of 315 parameters ($\approx$ 45%) are significant at the 5% level. For the other 172 parameters, the evidence is not strong to reject the null hypothesis at the 5% level. Among them are 12 of 47 from the α_x term, 102 of 114 from the κ_t term and 58 of 154 from the γ_{t-x} term. Despite the fact, that the APC model is relatively parsimonious, about 55% of its parameters are not individually significant at the 5% level. From the modelling perspective, it appears not to be justifiable to include that amount of non-significant parameters.

This observation is not unique as we point out in the following. Recall from Section 1.7.1 that the PLAT:2 model turned out to be among the most favourable models. Furthermore, likelihood ratio tests of the pairs CBD vs. PLAT:2 and APC vs. PLAT:2 showed that the parameter groups α_x or γ_{t-x} contribute significantly to the fitting accuracy in the first case and the parameter group $\kappa_t^{(2)}$ leads to a significant improvement compared to the APC model, see Table 1.6. Significance analysis for the PLAT:2 parameters on the individual level can be found in Table B.2. The results show that even fewer parameters are significant on the individual level. Only 26 of the 428 parameters ($\approx$ 6%) showed significance at the 5% level. The null hypothesis could not be rejected for all γ_{t-x} and $\kappa_t^{(2)}$

parameters and for all but one $\kappa_t^{(1)}$ parameters.

The fact that the whole parameter group is significant, but none of the individual parameters is, could be an indicator for over-parametrization. It certainly shows that due to the complicated dependence structure among the parameters, imposed by identifiability constraints, it appears to be difficult to identify the most influencing factors.

Not all models share this behaviour of too many non-significant parameters. For instance, the M7 model, with results reported in Table B.3, show a better ratio of significant parameters. The null hypothesis was rejected for 438 of 498 parameters of the 5% level. Subsequently, we will see that this ratio can even be improved by considering predictors which do not have any identifiability issues. The simplest of our proposed models in Section 1.9, will only have significant parameters.

1.8.4 | Summary of the Identified Issues Related to GAPC Models

In Sections 1.8.1 to 1.8.3, we stressed out many fundamental issues and conceptional difficulties of the well-established GAPC mortality modelling family. As pointed out in Section 1.8.1, many of these issues are direct consequences of non-identifiable predictor functions and corresponding parameter constraints. For a better overview, we provide a summary of the identified problems.

poor interpretability of parameters: Due to non-identifiable parameters, it is difficult to determine, whether particular patterns in the parameter paths are solely data-driven or artificial due to the imposed constraints. Modifications of the arbitrarily chosen constraints yield alternative representations of the same mortality structure.

excessive number of parameters: To describe the observed age-related mortality of a particular period, most models typically require more parameters than the actual degrees of freedom in the observation.

poor robustness: Simple database updates, such as adding the most recent period or omitting the oldest, leads to changes in all parameter estimates.

complex dependence structure: Despite a very complicated dependency structure between the parameters due to their entanglement, the common assumption, for forecasting purposes, is independence between period terms and the cohort term.

meaningless cohort term: The cohort term provides additional degrees of freedom, which often improve the quality-of-fit, but it does not serve the intended purpose to capture or reveal cohort effects. Path structures of the cohort estimates are heavily influenced by the imposed constraints. Moreover, given the obtained estimates, the assumption of ARIMA driven cohort terms is questionable.

poor significance of individual parameters: Wald type tests revealed that some models have only a few parameters (6% for PLAT) that are significant at the individual level. There are instances where likelihood ratio tests for nested models showed that a particular group of parameters is significant, however, none of the individual parameters of this group appeared to be significant.

Regarding the quality criteria for stochastic mortality models by Cairns, Blake and Dowd (2008), as listed on page 20, we can draw the following conclusion. All GAPC models produce positive mortality rates which are consistent with historical data. The fitting accuracy tends to increase with predictor complexity. Likelihood ratio tests also showed that relatively parsimonious models were less preferable. All models with identifiability issues are to some extent not robust against changes of the employed periods or ages. The numerical implementation is straightforward since these models fall in the GLM framework. There are further criteria mentioned by Cairns, Blake and Dowd (2008). However, they involve model forecasts, which cannot be assessed here since we have not considered model forecasts yet. As we will see later, the common modelling approach of random walk driven coefficient series does not lead to desirable forecasts in terms of the level of uncertainty and consistency with historical developments. We will therefore study more sophisticated time series approaches in Chapter 2, which involve long-run dependencies, called cointegration relations, between the components of multivariate time series.

1.9 | Kannisto Model

In the previous Section 1.8, we reviewed the GAPC class of well-established stochastic mortality models, and revealed several problematic properties and fundamental issues of these models. The case studies of Section 1.7 showed that models with more extensive predictors (e.g., M7 or PLAT) performed well in describing historical observations. However, this came at the cost of parameter identifiability. That means, all predictors, except the CBD predictor, require additional parameter constraints that cause many further issues.

In the following section, we propose a family of stochastic mortality models, which also belongs to the GAPC class, but avoids the identified disadvantages, as highlighted in Section 1.8.4. Furthermore, the new model family incorporates the key conclusion of the second case study Section 1.7.2, which showed that for various predictors a non-canonical link often leads to a better fit than the canonical link, in particular, for mortality rates of high ages (60+). The proposed models will be called the *Kannisto* family. Our choice of that name relies on the fact that in the simplest form, the model has a parametric logistic-type hazard rate. This form has been originally considered by the demographer Väinö Kannisto in 1992. Kannisto studied historical mortality rates with a focus on higher ages, see Kannisto (1992) and Thatcher, Kannisto and Vaupel (1998). However, similar to the Lee-Carter model, the authors did not consider their model within the broader framework of generalized linear models, but rather used the ordinary least squares fitting method on logit transformed mortality rates.

Recent contributions addressing the Kannisto model can be found be in Pitacco (2016) and Pitacco and Rroji (2016). Based on a semi-parametric bootstrap technique Pitacco and Rroji (2016) investigates the impact of uncertainty in parameter estimation for the Gompertz and Kannisto-type mortality rates. In contrast to our model, their approach uses the canonical link function for both predictor functions. Furthermore, projections of future mortality rates are obtained by bootstrapped ARIMA processes, rather then a multidimensional cointegrated processes, as will be proposed in Section 2.4. Pitacco (2016) provides an overview of parametric models representing the age-specific of mortality. The main focus of Pitacco (2016) lies on mortality of high ages, in particular, on the observed effect of decelerated mortality increase in the age-pattern. Pitacco (2016) provides various hypothesis for the main causes of a sub-exponential increase of the age-specific mortality (cf., Section 1.3). The proposed explanations include heterogeneity of populations as well as of individuals.

This connection to frailty models and mixtures of distributions will be discussed in the upcoming Chapter 3 in Proposition 3.3.7 and Remark 3.3.8.

1.9.1 | Specification of the Kannisto Model Family

For the specification of the Kannisto model we combine the insights from empirical observations and the case studies from Sections 1.3 and 1.7. The primary objective of the following specification is to obtain a model with no identification issues. Secondly, we employ a non-canonical link function and, finally, we utilize the fact of strong regularities in the mortalities of elderly populations by including parametric age terms.

GAPC Components of the Kannisto Model Family

As stated in Section 1.4.3, the specification of a GAPC model requires 4 components. For the Kannisto model, we consider a Poisson distributed response variable, i.e.,

$$D_{t,x} \sim \mathrm{Poi}(\mu_{t,x} E^c_{t,x}).$$

The systematic component of the Kannisto model of order p is defined by the predictor function

$$\eta_{t,x} = \sum_{i=1}^{p} \beta_x^{(i)} \kappa_t^{(i)} = \sum_{i=1}^{p} (x - x_{\min})^{i-1} \kappa_t^{(i)}. \tag{1.85}$$

The connection of the random component and the predictor is established by the logit link function g, i.e.,

$$g : \mu_{t,x} \mapsto \ln\left(\frac{\mu_{t,x}}{1 - \mu_{t,x}}\right). \tag{1.86}$$

This non-canonical choice for a Poisson distributed response is based on the results of the second case study of Section 1.7.2. Since the parameters of the Kannisto predictor in eq. (1.85) are entirely identifiable, the set of the required parameter constraints is empty.

Remarks on the Kannisto Model

The form of the Kannisto predictors in eq. (1.85) shows that the Kannisto family neither incorporate a static age function α_x nor a cohort term γ_{t-x}. Since we focus on mortality modelling of the elderly population, a static age function is expendable, considering the regular structure of the mortality rates, as shown in Figures 1.5 and 1.6. A cohort term is not incorporated in eq. (1.85) for two reasons. Firstly, to ensure identifiability without additional parameter constraints and, secondly because our case studies showed that the cohort term misses its actual purpose, see Section 1.8.2. Note that the Kannisto predictor has parametric age-modulation terms $\beta_x^{(i)} = (x - x_{\min})^{i-1}$, where $x_{\min}$ denotes the lower bound of the considered ages. To ensure comparability to the previous case studies, we will consider $x_{\min} = 60$ in the further course. In contrast to other predictors, such as those of the models LC, LC2, LC2+C, and RH (see, Table 1.2) the age, as a whole, is a model factor. For models with non-parametric age-modulation terms, each individual age of the considered range is a model factor. This conceptual difference causes an extensive amount of parameters in latter case.

In the further course, we will consider 3 models of the Kannisto family. These are characterized by the polynomial order of the age-modulating term $\beta_x^{(i)}$. The predictor

$$\eta_{t,x} = \kappa_t^{(1)} + (x - x_{\min})\kappa_t^{(2)}$$

will be denoted as the KAN predictor. Models with a Kannisto predictor that incorporates quadratic or cubic terms will be referred to as the KAN:2 and KAN:3 models.

Kannisto Model in the context of Logistic Hazard Rate Models

In the following, we provide some remarks on how the Kannisto model is related to logistic hazard rate models. Recall that in generalized linear models the expectation of the variables of interest (death counts $D_{t,x}$), is connected to a linear predictor $\eta_{t,x}$ through a link function g. Thus, with a logit link and the Kannisto predictor, we have the relation

$$\mu_{t,x} = \mathbb{E}\left[\frac{D_{t,x}}{E^c_{t,x}}\right] = g^{-1}(\eta_{t,x}) = \text{logit}^{-1}\left(\kappa_t^{(1)} + (x - x_{\min})\kappa_t^{(2)}\right) = \frac{e^{\kappa_t^{(1)} + \kappa_t^{(2)}(x - x_{\min})}}{1 + e^{\kappa_t^{(1)} + \kappa_t^{(2)}(x - x_{\min})}}, \tag{1.87}$$

where the mortality rates $\mu_{t,x}$ are logistic functions in the age x. Note that the logistic function is the inverse of the logit function.

$$\text{logit}^{-1}(x) = \text{logistic}(x) = \frac{\exp(x)}{1 + \exp(x)}.$$

Therefore, postulating the KAN model is equivalent to proposing a logistic-type growth of the mortality rates. Hazard rate models, which will be introduced in Chapter 3, provide a conceptual extension from a discrete to a continuous setting.

Distinction of the Kannisto and the CBD Model (Logistic vs. Exponential Growth)

Note that the CBD predictor, as defined in eq. (1.9) is very similar to the KAN predictor. The difference between the models lies not only in the shifting term ($\bar{x}$ vs. $x_{\min}$) of the $\beta_x^{(1)}$ term but also in the link function. The CBD model uses a canonical link and the KAN model a non-canonical logit link. A similar consideration for the CBD model as in eq. (1.87), shows that mortality rates in the CBD model follow an exponential growth. An exponential type hazard rate corresponds to the so-called Gompertz lifetime distribution as will be introduced in Section 3.2.7. The key conceptional distinction of those models is therefore, how age-related mortalities increase. For the Kannisto model, the rates increase logistically (saturated growth), while for the Gompertz model, the rates grow exponentially.

1.9.2 | Kannisto Model Parameter Estimation and Goodness-of-fit Analysis

The following section provides the estimation results and goodness-of-fit analysis for the Kannisto models. To ensure comparability with the other models, we consider the same reference dataset, as for the first two case studies of Sections 1.7.1 and 1.7.2. This means, we consider the dataset of the Swedish female population on the age range 60 to 106 and the periods between 1900 and 2014.

Kannisto Fitting Accuracy in terms of Deviance, AIC, and BIC

The fitting results of the KAN, KAN:2 and KAN:3 models are presented in Table 1.13. The table presents the deviance as a measure of quality-of-fit and the values of AIC and BIC information criteria. The first column shows the total number of parameters of the corresponding model. Note, to represent the mortality for a given period, the Kannisto models require only 2-4 parameters, depending on the polynomial order of the predictor. Recall, that for other model predictors the number of parameters exceeded the actual degrees of freedom, as discussed in Section 1.8. Since we consider nested models, the deviance is declining by the incorporation of higher polynomial terms in the predictor. Note that the simplest KAN model has a better fitting accuracy by having lower deviance, AIC, and BIC values, compared to the CBD model, which has the same number of parameters (see Table 1.5). The comparison with the best performing models of the first case study shows that the KAN:3 model has a 8% higher deviance than the M7 model (498 parameters), or a 4% higher deviance compared to the PLAT:2 (428 parameters) model. This slightly higher model fit discrepancy of the Kannisto model is the trade-off for having identifiable parameters, where the estimates depend only on the data and not on arbitrarily imposed constraints.

Likelihood Ratio Tests for Nested Kannisto Pairs

Table 1.14 summarizes the results of the likelihood ratio test for the nested Kannisto pairs. The LR tests check whether extensions to higher polynomial age-modulation terms are justified. Similar to the results of the first case study, as shown Table 1.6, for all nested pairs the null hypothesis is rejected. Consequently, additional age-modulation terms provide significant improvements of the quality-of-fit. Nevertheless, as we will demonstrate below, even the simplest KAN predictor provides a very good approximation of the lifetime characteristic.

Residuals Analysis of the Kannisto Predictors

Figures 1.31 to 1.33 illustrate the standardized deviance residuals, as defined in eq. (1.83), plotted against the calendar year and age. The reddish regions (positive residuals) on the heat maps indicate an underestimation of the death counts by the particular model, whereas bluish areas (negative residuals) indicate an overestimation of the death counts compared to the actual observations. Figure 1.31 displays substantial residual patterns for the simplest Kannisto model. There is a clear evidence that the KAN model is not capable to capture non-linear age-modulation effects of the dataset, since there is a considerable reduction of the residual patterns for the predictors with additional quadratic and cubic β_x terms, as displayed in Figures 1.32 and 1.33. The patterns for the KAN:3 model are qualitatively very similar to those of the M7 and PLAT models as illustrated in Figures 1.22 to 1.24.

Significance of Individual Kannisto Parameters

In Section 1.8.3, we have criticized that some traditional GAPC models have too many non-significant parameters. For instance, in the first case study, only 45% of the 315 parameters of the APC predictor were individually significant at a 5% level (see Table B.1). The PLAT:2 model, which was one of the most favourable models, was even less satisfying with a significance for only 6% of the 428 parameters at the 5% level (see Table B.2). Significance analysis for the parameters of the Kannisto model family can be found in Tables B.4 to B.6. For the simplest KAN model with 2 parameters for

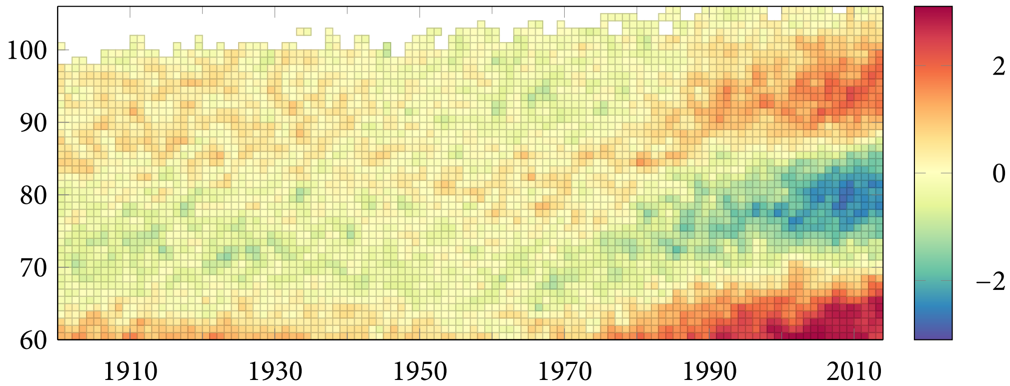


Figure 1.31: Standardized deviance residuals of the KAN model.

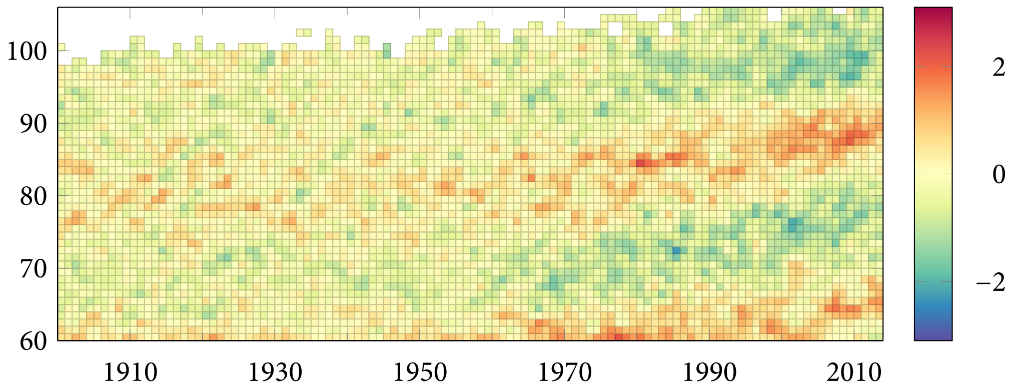


Figure 1.32: Standardized deviance residuals of the KAN:2 model.

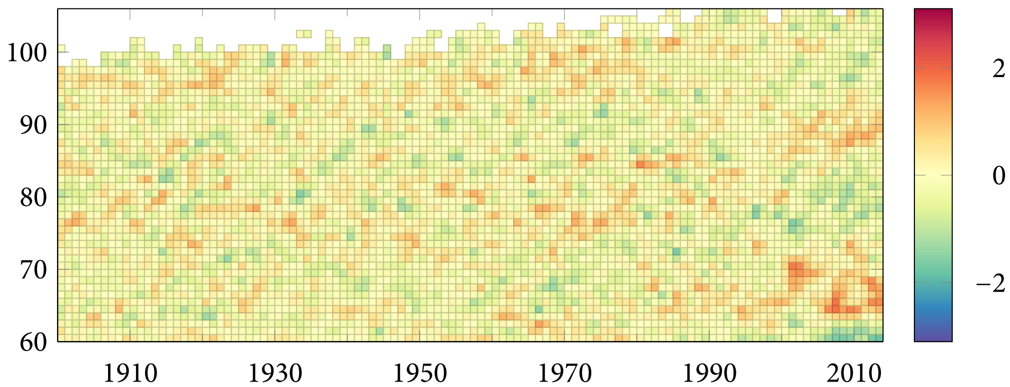


Figure 1.33: Standardized deviance residuals of the KAN:3 model.

Table 1.13: Number of parameters, log-likelihood, deviance, AIC, and BIC for the Kannisto models fitted to the Swedish female population for ages 60-106 and the period between 1900 and 2014 (5103 available observations).

Model	# of parm.	$\ln L$	Deviance	AIC	BIC
KAN	230	−27200	14903	54860	56364
KAN:2	345	−23075	6654	46841	49096
KAN:3	460	−22358	5219	45636	48643

Table 1.14: Results of the likelihood ratio test for the Kannisto models. The columns λ_{LR} and d.f. contain the LR test statistic and the degrees of freedom.

H_0 (nested)	H_{A} (general)	Restriction	λ_{LR}	d.f.	p-value
KAN	KAN:2	$\kappa_t^{(2)} = 0$	8250	115	< 0.0001
KAN:2	KAN:3	$\kappa_t^{(3)} = 0$	1434	115	< 0.0001
KAN	KAN:3	$\kappa_t^{(2)} = \kappa_t^{(3)} = 0$	9684	230	< 0.0001

each period, Wald type tests showed that all 230 parameters are significant at the 5% level. For KAN:2, the model with a quadratic age-modulation term, we obtain that 92% of parameters are individually significant at the 5% level. Some parameters associated to the quadratic term in the middle of the past century are not significant, see Table B.5. As pointed out in Section 1.8.3, most of the traditional GAPC models need an excessive number of parameters to represent the mortality structure of a certain period. However, here we have a simple 3 parameters model, where the non-significance of some quadratic terms can easily be explained by a sufficient linear structure of the logit transformed mortality rates, see Figure 1.6 for an illustration. For the KAN:3 model the Wald tests show that 84% of all parameters are individually significant at the 5% level, see Table B.6. Similar to the KAN:2 model, we can explain the non-significance of some parameters of higher polynomial terms by observing that the logit transformed mortality rates are sufficient linear. Conclusively, we can point out, that not only the Kannisto models do have a better significance rate of their parameters, but throughout the uncomplicated and identifiable model structure we can also explain why certain parameters turn out to be non-significant.

Fitting Accuracy in Terms of Lifetime Characteristics

The deviance is a common measure of the fitting discrepancy for GLMs and the deviance residual plots are eligible for identifying particular domains where the models fail to describe the essential features of the dataset. However, they are not suitable to assess the impact of the discrepancy on relevant actuarial quantities such as the life expectancy. The objective of the following exploration is to show that even the simplest KAN model provides a remarkably good approximation of the lifetime characteristics. A lifetime is a continuous non-negative random variable modelling the time to death of individuals of particular groups. A rigorous focus on the topic of survival analysis is provided in Chapter 3. For the current purpose it is sufficient to know, that the age-specific mortality rates $\mu_{t,x}$ are representatives of the lifetime distribution. Equivalently to probability densities, distributions can be represented by hazard rates, which are continuous versions of discrete central mortality rates. A positive continuous random variable, with a 2 parameter logistic hazard rate function of the type

$$\mu_t(x) = \frac{e^{\alpha_t+\beta_t x}}{1 + e^{\alpha_t+\beta_t x}}, \tag{1.88}$$

will be defined to be the *Kannisto distribution* and denoted be $\mathcal{K}(\alpha_t,\beta_t)$ (see Section 3.2.7), where $\alpha_t \in \mathbb{R}$ and $\beta_t \in \mathbb{R}_+$ are distribution parameters at the period t. Note that the hazard rates in eq. (1.88) are of the same type as in eq. (1.87), where we showed that the mortality rates obey a logistical function given the GLM setting with a KAN predictor and a logit link function. For the benefit of a simpler notation, we will alternatively denote the periodic terms $\kappa_t^{(1)},\kappa_t^{(2)},\kappa_t^{(3)},\kappa_t^{(4)}$ in the Kannisto predictor as α_t (constant term), β_t (linear term), γ_t (quadratic term), δ_t (cubic term).

To provide a better insight on the fitting performance of the KAN model regarding lifetime related characteristics, we present a comparison between some empirical quantities and the corresponding quantities derived from the estimated lifetime distributions. Table 1.15 shows parameter estimates, the mean, the standard deviation, the skewness and the kurtosis of both the empirical data and the Kannisto distribution using the estimates of the KAN model. Note that in this context, the empirical mean and the expectation of the lifetime distribution correspond to the expected remaining lifetime of individuals. Since we investigate the mortality rates on the age range 60-106, the remaining lifetime has to be added to the starting age of 60 to obtain the life expectancy of the underlying population group. Note furthermore, since the HMD does not contain the exact times of death events, but rather just their total numbers, we set the occurrence of the events to the middle of the corresponding period in order to calculate the empirical moments. To summarize the results of Table 1.15, we point out that the accuracy of the KAN model fit on several actuarial relevant quantities is remarkably good. For instance, between 1900 and 2014, the relative error of the mean (remaining life expectancy) is on average about 0.1%. In absolute values, this is a discrepancy of about only 9 days on the remaining lifetimes with values reaching between 17 to 26 years. The relativ error of the standard deviation is about 0.5% and for the kurtosis about 1.4% on average. The absolute error of the skewness is on average about 0.05. Overall, we observe a very good approximation of the empirical lifetime characteristics by the Kannisto distribution by using the estimated parameters of the KAN model. An illustration of the time evolution of the discussed characteristics is provided in Figure A.1.

Another way to demonstrate the great fitting capabilities of the Kannisto model is provided via a series of quantile-quantile-plots (Q-Q-plots) in Figure 1.34. In a Q-Q-plot quantiles of two distributions are plotted against each other to compare these distributions visually. For our purpose, we plot the empirical quantiles (abscissa) against the quantiles of the estimated Kannisto distribution, denoted by theoretical quantiles (ordinate). Each plot of Figure 1.34 shows different periods ranging from 1940 to 2010 with 10 years apart. The dashed orange lines in each plot represent the identity function, where the empirical quantiles coincide with the estimated Kannisto quantiles. For values below the identity, there is an overestimation of the survival probability by the $\mathcal{K}(\alpha_t,\beta_t)$ distribution

Table 1.15: Distribution characteristics of the empirical vs. the estimated Kannisto distribution for Swedish females with age above 60 between 1902 and 2014.

year	est. parameter		mean		stand. deviation		skewness		kurtosis	
	$\hat{\alpha}$	$\hat{\beta}$	data	$\mathcal{K}(\hat{\alpha},\hat{\beta})$	data	$\mathcal{K}(\hat{\alpha},\hat{\beta})$	data	$\mathcal{K}(\hat{\alpha},\hat{\beta})$	data	$\mathcal{K}(\hat{\alpha},\hat{\beta})$
1902	−4.210	0.107	16.98	17.00	8.45	8.42	0.007	0.054	2.329	2.372
1906	−4.277	0.106	17.59	17.61	8.66	8.63	−0.024	0.035	2.303	2.371
1910	−4.291	0.109	17.37	17.39	8.50	8.44	−0.032	0.022	2.324	2.377
1914	−4.288	0.110	17.28	17.30	8.45	8.39	−0.020	0.021	2.336	2.378
1918	−4.145	0.102	17.01	17.03	8.67	8.63	0.040	0.089	2.290	2.367
1922	−4.247	0.112	16.85	16.86	8.27	8.23	−0.010	0.031	2.343	2.378
1926	−4.302	0.108	17.52	17.53	8.53	8.50	−0.009	0.020	2.345	2.376
1930	−4.294	0.108	17.51	17.52	8.57	8.52	−0.008	0.024	2.321	2.375
1934	−4.337	0.110	17.61	17.64	8.51	8.47	−0.037	0.005	2.329	2.380
1938	−4.323	0.111	17.44	17.45	8.40	8.39	−0.014	0.008	2.371	2.381
1942	−4.479	0.110	18.60	18.60	8.68	8.68	−0.069	−0.042	2.379	2.394
1946	−4.466	0.116	17.96	17.96	8.34	8.32	−0.076	−0.049	2.400	2.403
1950	−4.531	0.119	18.11	18.11	8.25	8.25	−0.086	−0.075	2.429	2.418
1954	−4.629	0.120	18.71	18.71	8.35	8.34	−0.122	−0.107	2.452	2.436
1958	−4.727	0.122	19.24	19.23	8.39	8.39	−0.140	−0.139	2.488	2.459
1962	−4.841	0.126	19.55	19.55	8.27	8.28	−0.175	−0.179	2.526	2.496
1966	−4.932	0.125	20.30	20.30	8.46	8.46	−0.200	−0.204	2.551	2.517
1970	−4.983	0.122	21.03	21.03	8.72	8.71	−0.221	−0.213	2.562	2.522
1974	−5.055	0.122	21.46	21.47	8.76	8.76	−0.245	−0.234	2.561	2.544
1978	−5.151	0.124	21.97	21.98	8.79	8.78	−0.273	−0.263	2.626	2.579
1982	−5.216	0.123	22.50	22.53	8.91	8.90	−0.331	−0.279	2.620	2.597
1986	−5.329	0.126	22.93	22.97	8.88	8.83	−0.378	−0.313	2.674	2.645
1990	−5.420	0.128	23.26	23.33	8.87	8.78	−0.417	−0.339	2.705	2.686
1994	−5.523	0.128	23.97	24.05	9.00	8.88	−0.462	−0.365	2.767	2.725
1998	−5.591	0.130	24.25	24.34	8.99	8.85	−0.489	−0.382	2.790	2.757
2002	−5.682	0.134	24.30	24.39	8.85	8.66	−0.530	−0.408	2.832	2.808
2006	−5.743	0.132	24.99	25.02	8.97	8.80	−0.575	−0.421	2.899	2.829
2010	−5.787	0.133	25.33	25.33	8.98	8.83	−0.601	−0.431	2.936	2.848
2014	−5.896	0.134	25.85	25.85	8.98	8.81	−0.626	−0.457	2.983	2.901

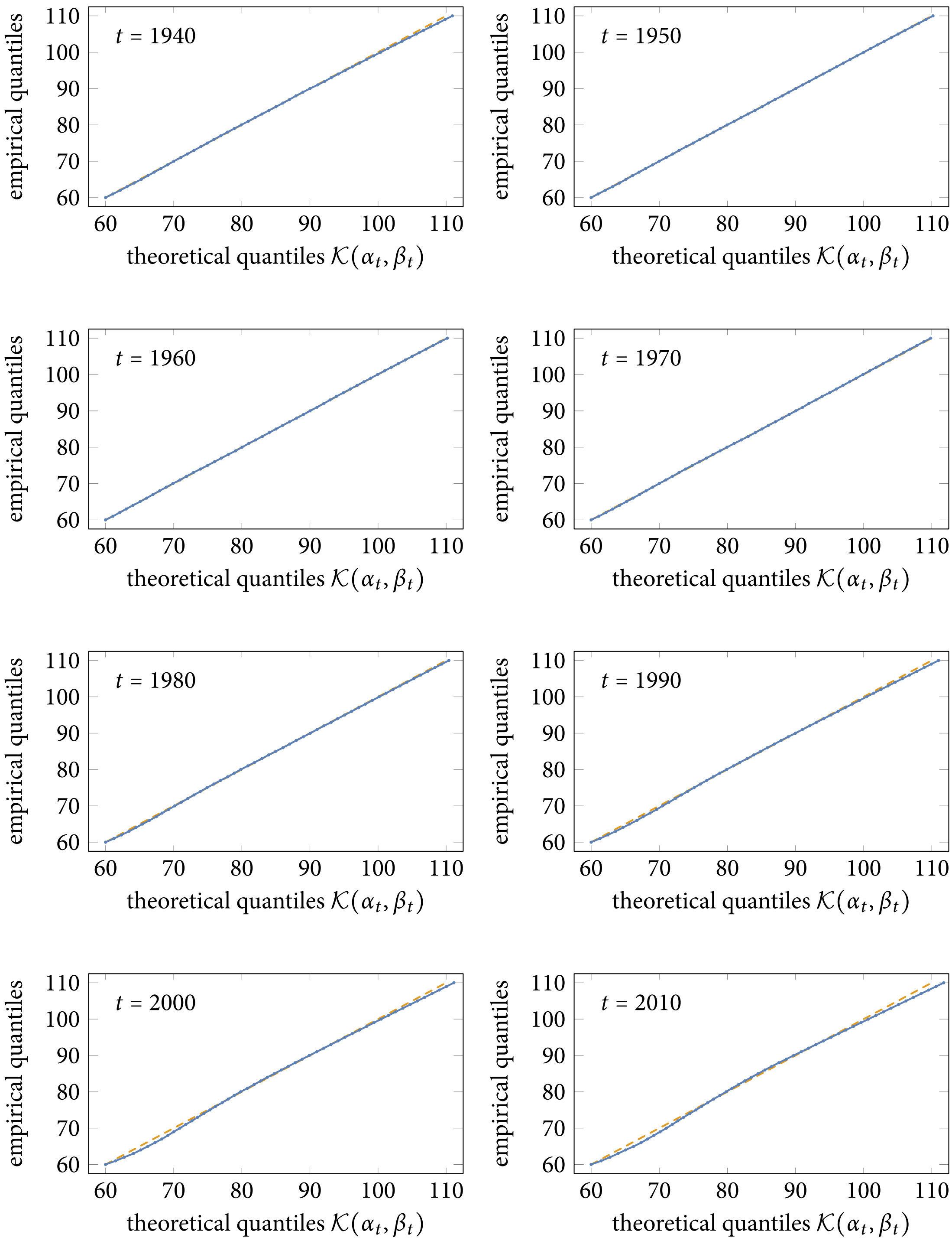


Figure 1.34: Empirical quantiles vs. theoretical quantiles of the Kannisto distribution

with estimated parameters (α_t,β_t). Values above the dashed line indicate an underestimation of survival probability of the theoretical distribution compared to the empirical distribution. As the Q-Q-plots in Figure 1.34 show, there is a very close correspondence of the empirical distribution and the Kannisto distribution. For the periods 2000 and 2010, there is a slight overestimation of the survival probabilities by the Kannisto distribution between the ages 60-75 and above 100. However, these quantile deviations have only marginal influences on the moments as Table 1.15 shows.

As a concluding remark, we point out that although the elementary KAN predictor has in terms of the deviance measure the highest fitting discrepancy among the Kannisto model family, it captures the empirical properties of the lifetime very well, which from a practitioner's perspective might be a better scale for model assessment.

1.9.3 | Interpretability and Comparability of the Kannisto Parameters

In Section 1.8.1, we stressed out some interpretability and comparability challenges of many GAPC models with identifiability issues. Recall, that due to additional imposed parameter constraints, required to achieve identifiability, the parameter estimates do not only depend on the observed data anymore but also on those constraints, see, for instance, the illustration in Figure 1.25. Moreover, adding recent or removing old mortality data, does change all estimates, not only those of the corresponding periods, see the examples in Figures 1.26 and 1.27.

In the following application-oriented example, we demonstrate how practitioners benefit, in the sense of model interpretability & comparability, by considering predictors with no identifiability issues. In Figure 1.35, we display trajectories of the KAN parameter estimates (α_t,β_t) for the periods $t \in \{1900,\ldots,2011\}$ of the female populations for Sweden (magenta), Switzerland (green) and France (cyan). These parameter trajectories are embedded in a contour plot illustrating constant slices of the expected remaining lifetime as a function of estimated Kannisto parameters. The trajectories of all countries show a general trend towards a higher life expectancy. The given plot provides a possibility of comparing mortalities of different populations either for fixed time frames or for particular parameter regions. For instance, in 1989 the parameter values of Switzerland and France almost coincide and indicate high similarities of the population's mortality at this time. The path of Sweden parameters reaches this region with a delay of 10 years. The delay of the mortality improvements for Sweden appears to increase. The region of parameter values, which has been obtained by France and Switzerland in 1977, was reached four years later by Sweden. From 1990 the paths of France and Switzerland drift apart. The remaining life expectancy at age 60 of France in 2011 is 27.4 years and is 0.8 years higher compared to Switzerland. Note that for 2011 the α values for France and Switzerland almost coincide, while the β values of France and Sweden almost coincide in 2011. In the first case, we have $\alpha^{\mathrm{FR}}_{2011} \approx \alpha^{\mathrm{CH}}_{2011}$ and $\beta^{\mathrm{FR}}_{2011} < \beta^{\mathrm{CH}}_{2011}$. From this situation, we can conclude that the higher life expectancy of France compared to Switzerland is mainly due to lower mortality rates for higher ages. In the second case, where $\beta^{\mathrm{FR}}_{2011} \approx \beta^{\mathrm{SE}}_{2011}$ and $\alpha^{\mathrm{FR}}_{2011} < \alpha^{\mathrm{SE}}_{2011}$, the higher remaining life expectancy of France, in comparison to Sweden, benefits from the overall lower mortality rates, especially at ages 60-95.

Consequently, by having identifiable parameters, we gain the ability for a meaningful interpretation of the estimates, and compare mortalities based on these value. That holds, since other than for non-identifiable models, the parameters depend only on the underlying data, rather than on additional constraints.

1.9.4 | Kannisto Model Forecasts obtained by a Multivariate Random Walk with Drift

In Section 1.4.3, we gave a brief introduction to the standard forecasting approach of GAPC models. In the upcoming section, we follow this approach of modelling the periodic $\boldsymbol{\kappa}_t$ components by a multivariate random walk with drift, illustrate the forecasting results, and also reveal the limitations of the underlying concept.

Figures 1.36 to 1.38 illustrate the trajectories of the estimates $\boldsymbol{\kappa}_t$ for the Kannisto predictors KAN, KAN:2, and KAN:3, which will be used to calibrate a random walk with drift. Note, in the illustration, we use an alternative notation for the components of $\boldsymbol{\kappa}_t$, namely, α_t, β_t, γ_t, and δ_t for the constant, linear, quadratic, and the cubic term. Note that a stochastic process is called stationary if its first and second moments are time invariant, see Definition 2.2.2 or Lütkepohl (2007). A visual inspection of the trajectories reveals that all paths show global or local trends. Therefore, only non-stationary processes appear to be eligible to model the underlying periodic terms. A random walk with drift is an integrated process of order one (see Definition 2.2.7) and is the standard modelling approach in the GAPC literature, see, e.g., Cairns, Blake and Dowd (2006), Cairns, Blake, Dowd et al. (2011) and Haberman and Renshaw (2011). In order to provide comparability to a later proposed alternative modelling approach which will utilize cointegration relations between the components, we initially demonstrate the standard modelling approach.

Let $\boldsymbol{\kappa}_t$, with $t = 0,1,2,\dots,t_N$, denote a multivariate time series of estimated parameters. We now assume that $\boldsymbol{\kappa}_t$ follows a multivariate random walk with drift, i.e.,

$$\boldsymbol{\kappa}_t = \boldsymbol{\kappa}_{t-1} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}_t, \tag{1.89}$$

where $\boldsymbol{\delta}$ is the drift vector and $\boldsymbol{\varepsilon}_t$ a white noise process (see the explanations of Definition 2.2.2) with variance $\boldsymbol{\Sigma}$. Let $\boldsymbol{y}_t := \Delta\boldsymbol{\kappa}_t = \boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1}$ denote the time series of the first-order differences. Substituting $\boldsymbol{y}_t$ in eq. (1.89) yields

$$\boldsymbol{y}_t = \boldsymbol{\delta} + \boldsymbol{\varepsilon}_t. \tag{1.90}$$

Following Haberman and Renshaw (2011), the estimates of the drift components can be obtained by

$$\hat{\delta}_i = \frac{1}{N}\sum_{j=1}^{N} y_{i,j} = \frac{1}{N}\left(\kappa_{i,N} - \kappa_{i,0}\right). \tag{1.91}$$

With $\hat{\varepsilon}_{i,t} = r_{i,t} := \left(y_{i,t} - \hat{\delta}_i\right)$, the estimator of the white noise covariance matrix is given by

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-1}\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}' = \frac{1}{N-1}\begin{bmatrix} \sum_t r_{1,t}^2 & \sum_t r_{1,t}r_{2,t} & \sum_t r_{1,t}r_{3,t} & \cdots \\ \sum_t r_{2,t}r_{1,t} & \sum_t r_{2,t}^2 & \sum_t r_{2,t}r_{3,t} & \cdots \\ \sum_t r_{3,t}r_{1,t} & \sum_t r_{3,t}r_{2,t} & \sum_t r_{3,t}^2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{1.92}$$

The right-hand side of the above expression is the sample covariance matrix of $\Delta\boldsymbol{\kappa}_t$ which is an

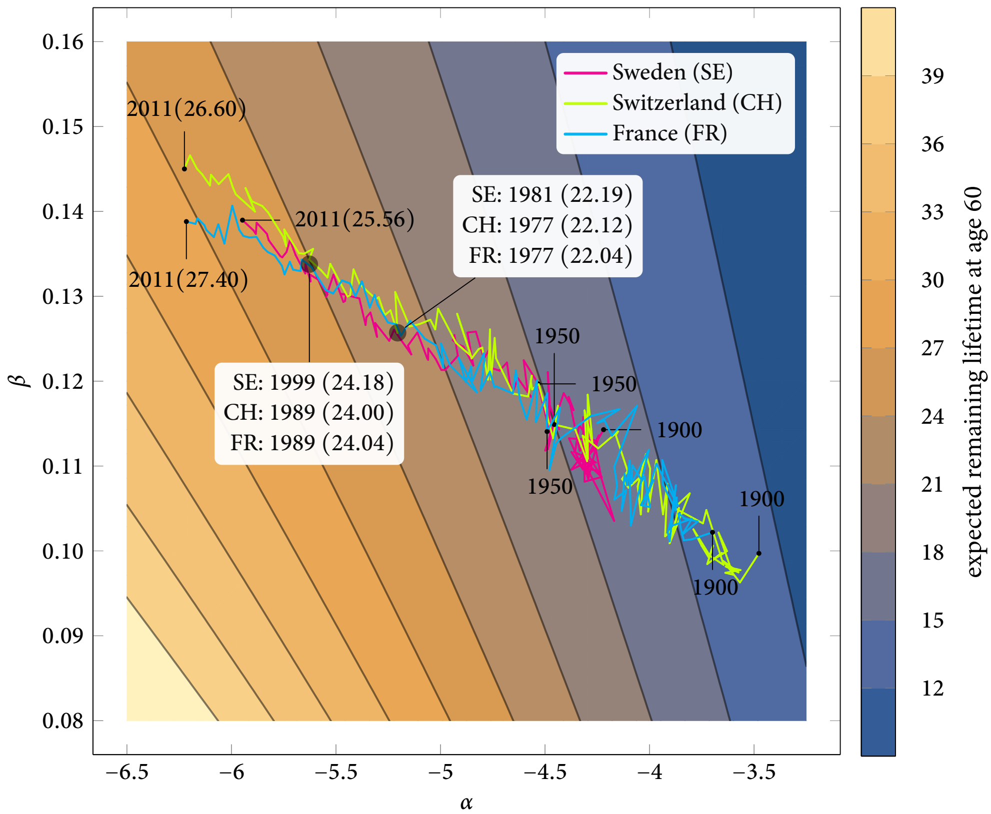


Figure 1.35: Trajectories of the KAN model parameters (α_t, β_t) for $t \in \{1900, \ldots, 2011\}$ of the female population of Sweden (magenta), Switzerland (green) and France (cyan).

unbiased estimator of the white noise covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t']$. Notice, that successive substitution for lagged $\boldsymbol{\kappa}_t$ terms in eq. (1.89) leads to

$$\boldsymbol{\kappa}_{t_N+j} = \boldsymbol{\kappa}_{t_N} + j\boldsymbol{\delta} + \boldsymbol{\varepsilon}_{t_N+j} + \boldsymbol{\varepsilon}_{t_N+j-1} + \ldots + \boldsymbol{\varepsilon}_{t_N+1}. \tag{1.93}$$

Taking the expectation of eq. (1.93) and using the white noise property $\mathbb{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$, for all t, yields

$$\boldsymbol{\kappa}_{t_N}(j) := \mathbb{E}[\boldsymbol{\kappa}_{t_N+j}] = \boldsymbol{\kappa}_{t_N} + j\boldsymbol{\delta}, \tag{1.94}$$

where $\boldsymbol{\kappa}_{t_N}(j)$ denotes the j-step ahead forecast of $\boldsymbol{\kappa}_{t_N+j}$ at origin t_N. To obtain prediction intervals of the forecast, notice that

$$\boldsymbol{\kappa}_{t_N+j} - \boldsymbol{\kappa}_{t_N}(j) = \boldsymbol{\varepsilon}_{t+j} + \boldsymbol{\varepsilon}_{t+j-1} + \ldots + \boldsymbol{\varepsilon}_{t+1},$$

and thus, the j-step forecast *mean squared error* (MSE), a quantity which reflects the forecast uncertainty, takes the form

$$\boldsymbol{\Sigma}_{\boldsymbol{\kappa}}(j) := \mathrm{MSE}[\boldsymbol{\kappa}_{t_N}(j)] = \mathbb{E}\left[\left(\boldsymbol{\kappa}_{t_N+j} - \boldsymbol{\kappa}_{t_N}(j)\right)\left(\boldsymbol{\kappa}_{t_N+j} - \boldsymbol{\kappa}_{t_N}(j)\right)'\right] = j\boldsymbol{\Sigma},$$

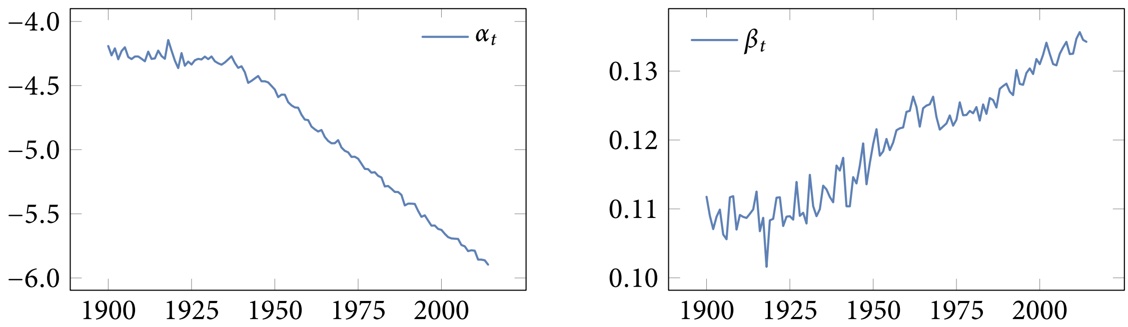


Figure 1.36: Parameters for the KAN model fitted to the Sweden's female population aged 60-106 in the periods between 1900 and 2014.

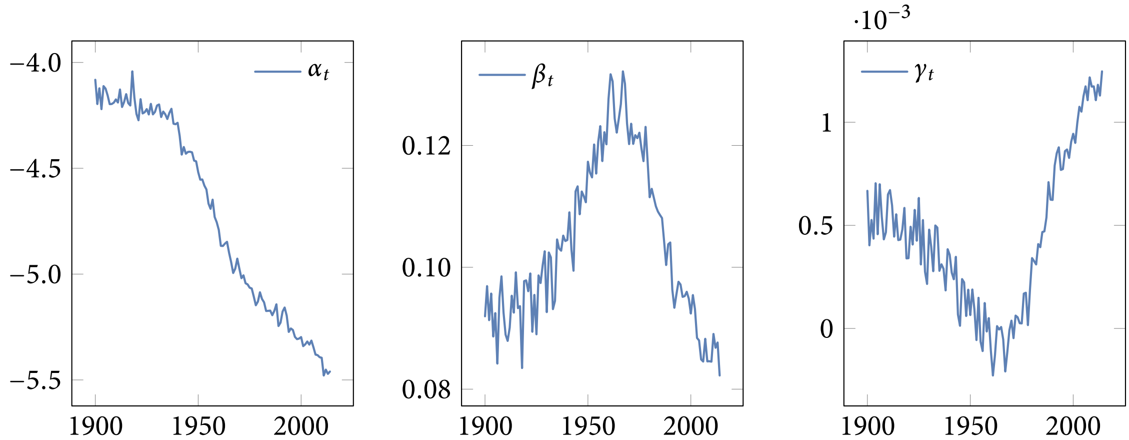


Figure 1.37: Parameters for the KAN:2 model fitted to the Sweden's female population aged 60-106 the periods between 1900 and 2014.

since $\mathbb{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] = \boldsymbol{\Sigma}$ and $\mathbb{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s'] = 0$ for $t \neq s$. Thus, using the fact that the forecast errors of the individual components are normally distributed, an $(1-\alpha)100\%$ forecast prediction interval (P.I.), j periods ahead is given by

$$\left[\kappa_{i,t_N}(j) - z_{(\alpha/2)}\sigma_i(j), \kappa_{i,t_N}(j) + z_{(\alpha/2)}\sigma_i(j)\right], \tag{1.95}$$

where $\sigma_i(j)$ is the square root of the i-th diagonal element of $\boldsymbol{\Sigma}_{\boldsymbol{\kappa}}(j)$ and $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution, see Lütkepohl (2007).

Using eqs. (1.91) and (1.92), we obtain the following estimates

$$\hat{\boldsymbol{\delta}} = \begin{bmatrix} -1.495 \cdot 10^{-2} \\ 1.973 \cdot 10^{-4} \end{bmatrix} \qquad \hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 1.498 \cdot 10^{-3} & -1.544 \cdot 10^{-5} \\ -1.544 \cdot 10^{-5} & 6.534 \cdot 10^{-6} \end{bmatrix} \tag{1.96}$$

for the drift and the covariance matrix of the two-dimensional random walk which drives the period

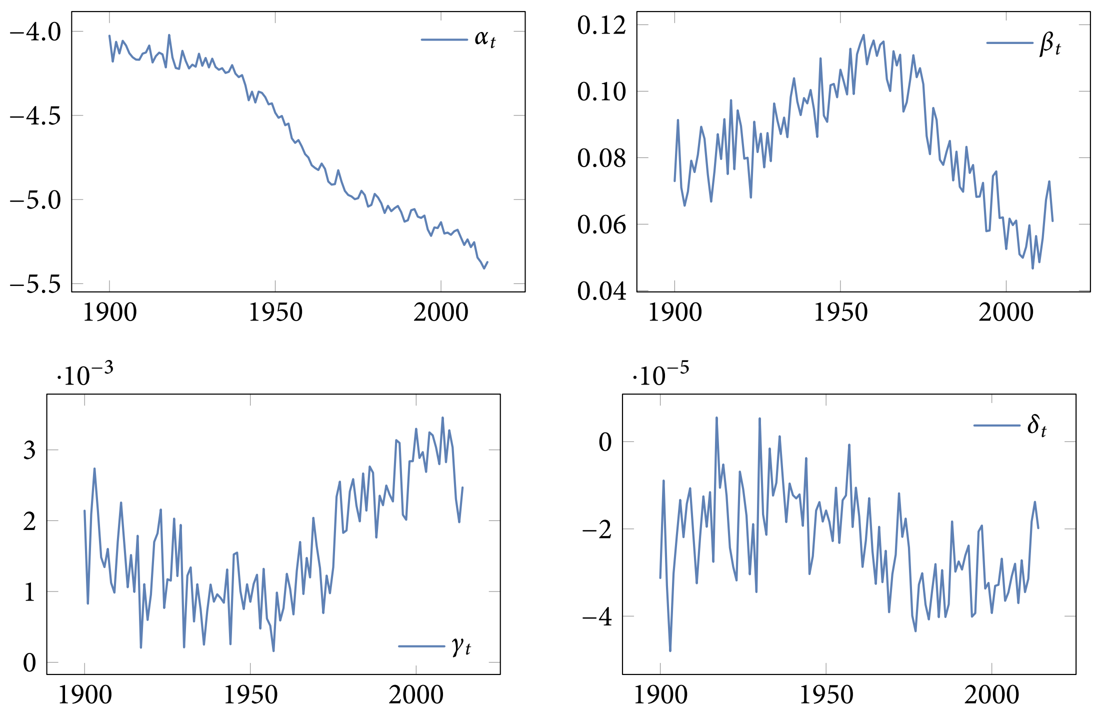


Figure 1.38: Parameters for the KAN:3 model fitted to the Sweden's female population aged 60-106 the periods between 1900 and 2014.

terms of the KAN model. An estimation for the KAN:2 time series yields

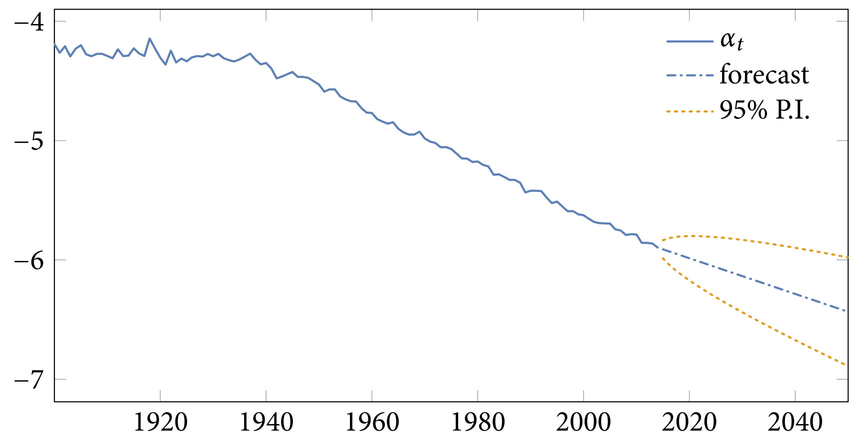
$$\hat{\boldsymbol{\delta}} = \begin{bmatrix} -1.209 \cdot 10^{-2} \\ -8.521 \cdot 10^{-5} \\ 5.081 \cdot 10^{-6} \end{bmatrix} \qquad \hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 2.008 \cdot 10^{-3} & -1.072 \cdot 10^{-4} & 3.047 \cdot 10^{-6} \\ -1.072 \cdot 10^{-4} & 2.273 \cdot 10^{-5} & -5.243 \cdot 10^{-7} \\ 3.047 \cdot 10^{-6} & -5.243 \cdot 10^{-7} & 1.692 \cdot 10^{-8} \end{bmatrix}. \tag{1.97}$$

In the same way, we obtain the estimates for the KAN:3 model as

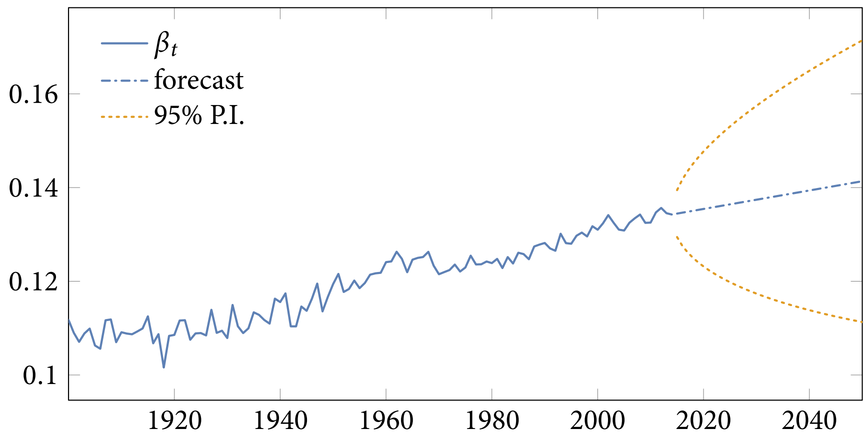
$$\hat{\boldsymbol{\delta}} = \begin{bmatrix} -1.180 \cdot 10^{-2} \\ -1.055 \cdot 10^{-4} \\ 2.876 \cdot 10^{-6} \\ 1.002 \cdot 10^{-7} \end{bmatrix} \qquad \hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 2.658 \cdot 10^{-3} & -3.301 \cdot 10^{-4} & 1.882 \cdot 10^{-5} & -3.132 \cdot 10^{-7} \\ -3.301 \cdot 10^{-4} & 9.448 \cdot 10^{-5} & -5.517 \cdot 10^{-6} & 9.752 \cdot 10^{-8} \\ 1.882 \cdot 10^{-5} & -5.517 \cdot 10^{-6} & 3.667 \cdot 10^{-7} & -6.859 \cdot 10^{-9} \\ -3.132 \cdot 10^{-7} & 9.752 \cdot 10^{-8} & -6.859 \cdot 10^{-9} & 1.347 \cdot 10^{-10} \end{bmatrix}. \tag{1.98}$$

Figures 1.39 to 1.41 illustrate the parameter estimates of the models KAN, KAN:2 and KAN:3 along with the corresponding central forecasts $\boldsymbol{\kappa}_{2014}(36)$ and their 95% prediction intervals. The 36 years ahead central forecasts (until 2050) are obtained by eq. (1.94) using random walks with drift and covariance matrices as in eqs. (1.96) to (1.98). Prediction intervals are obtained by a consecutive application of eq. (1.95).

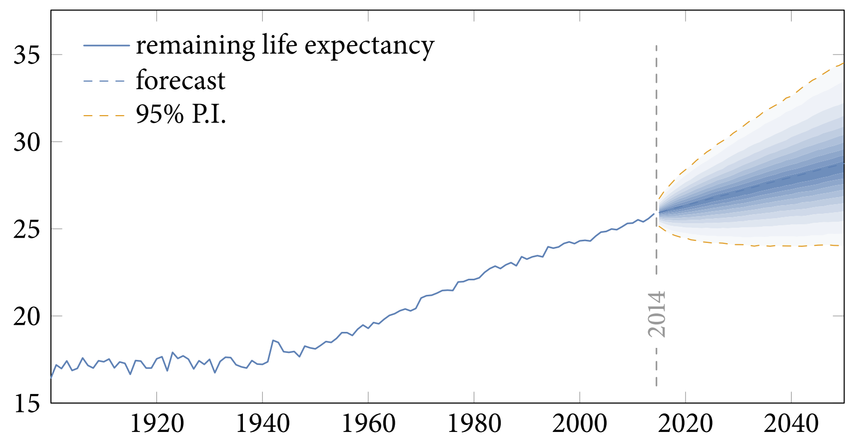
In Figure 1.39(c), we illustrate the historical and the projected remaining life expectancy for the



(a) Trajectories and projections of α_t.

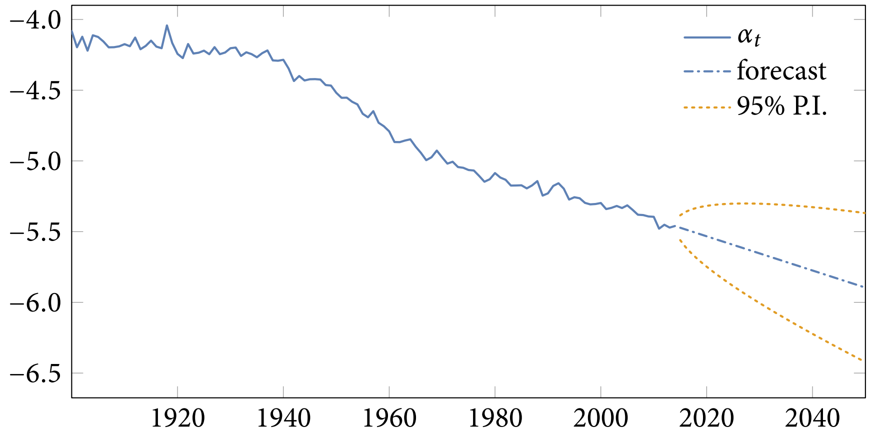


(b) Trajectories and projections of β_t.

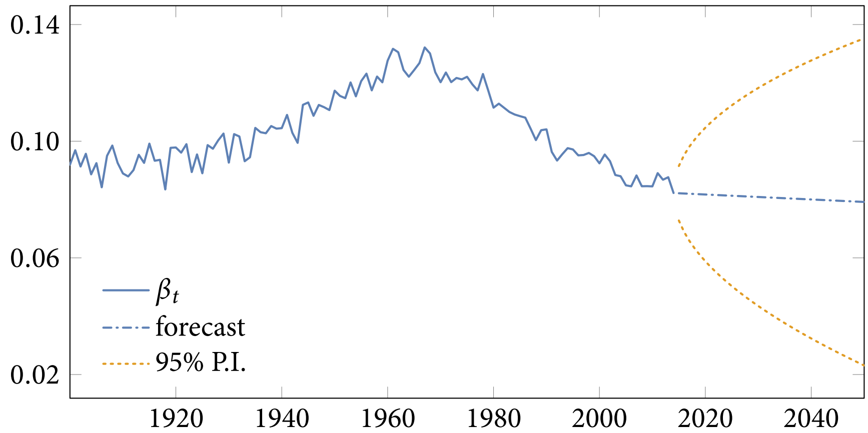


(c) Historical and projected remaining life expectancy for Swedish females aged 60.

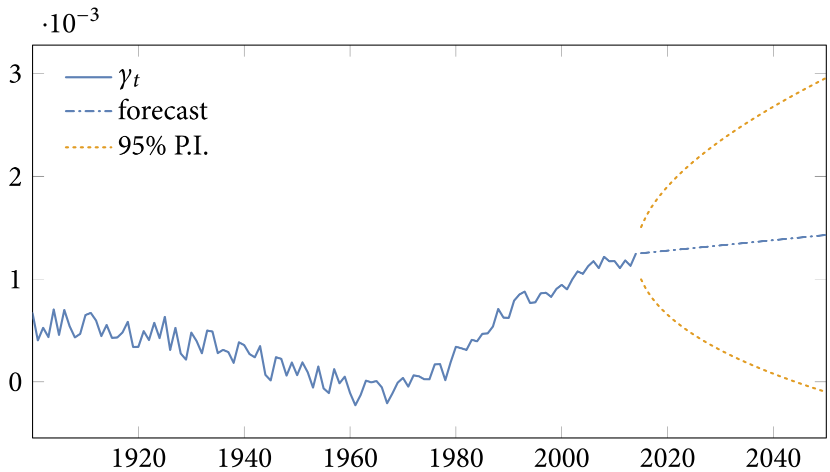
Figure 1.39: Projections of the KAN model coefficients with 95% prediction intervals.



(a) Estimates and projections of α_t.



(b) Estimates and projections of β_t.



(c) Estimates and projections of γ_t.

Figure 1.40: Random walk driven projections of the KAN:2 coefficients. Dashed lines represent the central forecasts and dotted lines show the 95% prediction intervals.

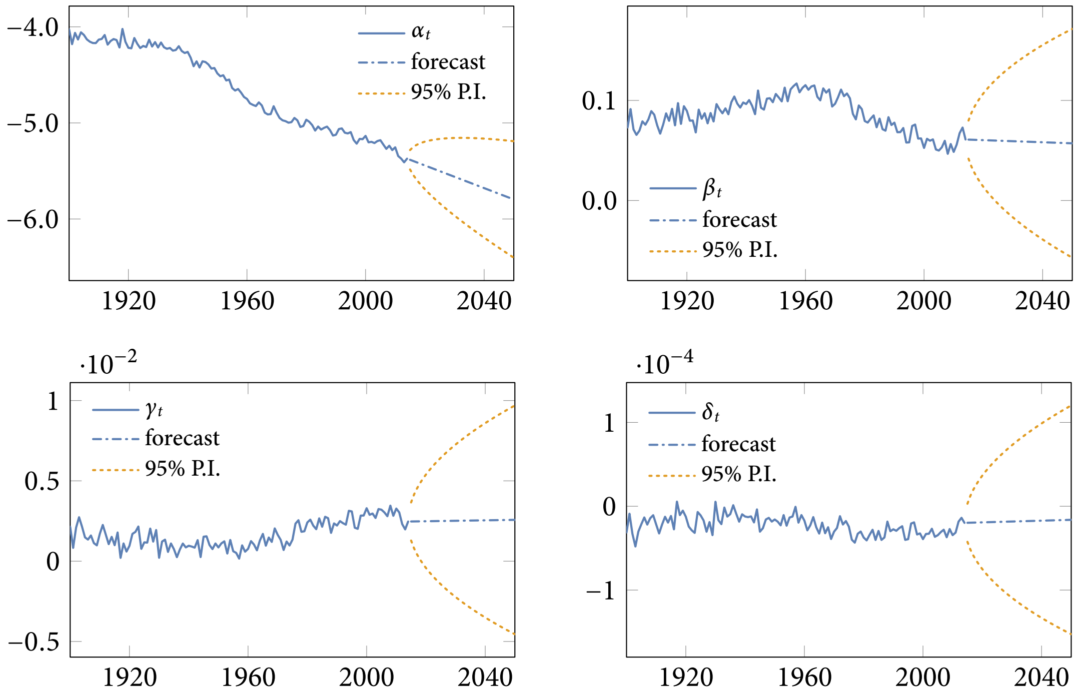


Figure 1.41: Random walk driven projections of the KAN:3 coefficients. Dashed lines represent the central forecasts and dotted lines show the 95% prediction intervals.

Swedish female population at the age of 60. The projections of the remaining life expectancy and the corresponding prediction intervals are simulated using the Monte Carlo method with a sample size of 10^5, by first sampling from the fitted random walk and then calculating the remaining life expectancy based on the sampled mortality structure. The projected life expectancy is an important actuarial quantity and also serves here as a plausibility assessment since it is modelled indirectly through the parameters of the KAN model. Large deviations from the historical experience in values or slope would raise doubts in the quality of the underlying model.

The 36 years ahead forecast of the KAN model shows an average increase of the remaining life expectancy of about one month (29.17 days) per year. The overall increase until 2050 is about 3 years, raising from 25.85 to 28.79 years. The width of the 95% prediction interval at the period 2050 is about 10.5 years, ranging from 24 to 34.5 years. Notice, that the slope of the forecast in Figure 1.39(c) is slightly lower than the historical value. The observed averaged increase of the remaining life expectancy over the previous 36 years is 38.12 days per year. It is also worth mentioning that, based on the historical trajectories, the 95% prediction intervals appear to be wider than anticipated. For instance, the lower 95% prediction interval is located below the current value. A decrease of the life expectancy or an increase of almost 9 years until 2050 appears to be not plausible, even for a 95% prediction interval.

For the sake of completeness, we add that the forecasts of the remaining life expectancy for the models KAN:2 and KAN:3 are almost identical to those of the KAN model, as illustrated in

Figure 1.39(c). We observe a relative discrepancy below 0.1% for both, the central projections and the prediction intervals.

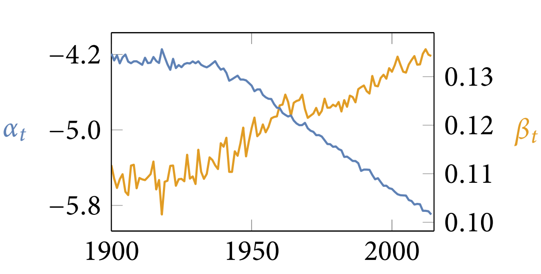
To conclude, the standard approach of treating the periodic term $\boldsymbol{\kappa}_t$ as a multivariate random walk with drift does lead to decent central forecasts. However, the forecasts have unsatisfying prediction uncertainties. These properties are inherited by the deduced actuarial quantities, such as remaining life expectancy. As demonstrated, the central forecasts of the life expectancy are plausible, but the uncertainty levels are difficult to work with in real-world applications. Mortality based contracts priced according to that level of uncertainty might not be competitive.

1.9.5 | Indications for the Presence of Cointegration Relations between the Periodic Components $\boldsymbol{\kappa}_t$ of the Kannisto Model

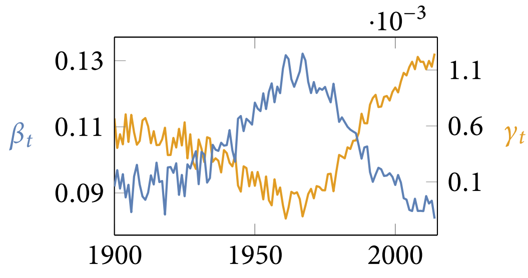
As illustrated in Figures 1.39 to 1.41, the forecasts of the periodic components of the Kannisto model show that the prediction intervals increase for further-reaching forecasts. This is typical for integrated processes where the MSEs are generally unbounded over time. This implies that, as the forecast horizon increases the forecast uncertainty also increases, see Lütkepohl and Krätzig (2004). For stationary processes, however, the MSEs are bounded, such that the forecast uncertainty does not become arbitrarily large. From that perspective, stationary processes do have better properties. However, as suggested in Figures 1.36 to 1.38 the time series of the estimated Kannisto parameters, taken individually, are likely non-stationary and therefore cannot be modelled directly by stationary processes on an individual level. However, even if individual variables are non-stationary, there might be a linear combination which leads to stationarity. This topic is presented in the seminal papers by Engle and Granger (1987) and Granger (1981), who formalized the concept of cointegration. According to the authors, two integrated processes are called cointegrated if there exists a linear combination of them which is stationary. This concept has become very important in research on equilibrium relationships between economic variables and their long-run trends. The contributions by Engle and Granger have been awarded the Nobel Prize in 2003.

To demonstrate some indications on the existence of cointegration relations between the individual components of the Kannisto model, we refer to Figure 1.42, where every sub-figure displays a pair of components, plotted on two different scales. For instance, Figure 1.42(a) illustrates the estimated components α_t and β_t of the KAN model. We can observe the following similarities between the trajectories. In the first part, where α_t holds a certain level, β_t also does. However, as soon α_t starts trending downwards β_t starts showing an upward trend. The visual impression of the trajectories is that a linear combination of them might lead to a trajectory without trends. This observation is even more valid for other coefficient pairs as illustrated in Figures 1.42(b) to 1.42(d). The paths of β_t and γ_t of the KAN:2 model demonstrate the contradictory behaviour of those components, not only globally but also locally, where positive increments of β_t are reflected as negative increments of γ_t and vice verse. For all presented pairs of coefficients, there is some evidence that particular linear combinations might be non-trending and that it is worth checking these variables for cointegration relations, which will be done in Chapter 2. Note that although it is possible to model non-stationary variables by first differencing them and using an appropriate stationary process afterwards, however, this procedure might not be optimal since differencing can distort the relationship between the original variables if cointegration relations exist, see Lütkepohl (2007). The loss of the long-run equilibrium relations, caused by differencing, can negatively affect the modelling capabilities.

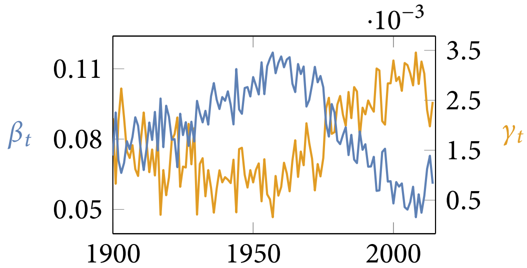
Further evidence for existing cointegration relations can also be found Figure 1.35, where the trajectories of the KAN model for three different countries show a long-term linear relation between them. We made this observation not only for the countries shown in that example, but also in many other cases which are presented Figures A.16 to A.27 in Appendix A.3.



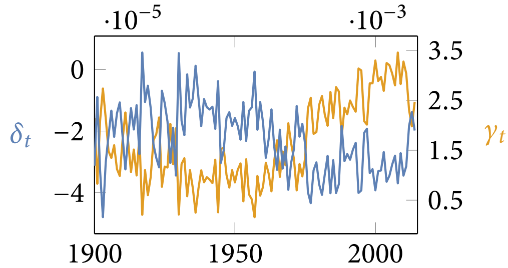
(a) Trajectories α_t and β_t of the KAN model.



(b) Trajectories β_t and γ_t of the KAN:2 model.



(c) Trajectories β_t and γ_t of the KAN:3 model.



(d) Trajectories δ_t and γ_t of the KAN:3 model.

Figure 1.42: Evidence on cointegration relations between Kannisto coefficients.

1.10 | Conclusion and Outlook

This chapter presents an important class of stochastic mortality models. The key aspect of the Generalized Age-Period-Cohort model family is that mortality rates are decomposed across the dimensions of age, period, and cohort. Even though this modelling approach was followed by practitioners and actuarial researchers for more than two decades, recent contributions showed that most Age-Period-Cohort mortality models can be expressed in the framework of generalized linear and non-linear models.

Throughout this chapter, we introduce the building blocks of GAPC models, provide an overview of the most popular mortality models and review numerical methods for parameter estimation. To assess various mortality models, we review some model selection criteria and common statistical tests. A quantitative analysis, with the focus on elderly populations, is provided for distinct predictor functions to compare their abilities to capture historical mortality changes. The results identify the M8 and the PLAT model as the most favourable models. In general, less parsimonious models are preferred over their sub-models as likelihood ratio tests demonstrate. In the second case study, we show that by using a non-canonical link function we obtain, for all but one predictor function, a better fitting performance. This is an important observation since the vast majority of the literature does not consider this particular degree of freedom, which is a key component of generalized linear models.

In the further course, we stress out many issues of the most popular GAPC models which are direct consequences of the imposed parameter constraints in order to ensure identifiability. The central points of criticisms are presented in Section 1.8. Considering the results of the case studies and the identified issues of current models, we propose a model class which is using a non-canonical logit link and a predictor function which does not require additional constraints for parameters to be identifiable. Our proposed model class is named after the demographer Väinö Kannisto who originally studied logistic hazard rate models. In a further case study, we provide a goodness-of-fit analysis to show how the Kannisto models compare to the established models and demonstrate its accuracy to reflect main characteristics of the lifetime distribution.

For forecasting purposes, the common approach of GAPC models is to treat periodic and cohort parameters as stochastic factors and use discrete multivariate stochastic processes to model the dynamics of those parameters. This approach uses a multivariate random walk with drift which we employ here in order to have a comparison to a more sophisticated modelling approach, which will be provided in Chapter 2. The observation of the Kannisto trajectories suggests that the individual components do not move independently but rather follow common patterns. An alternative modelling approach is to use particular vector autoregressive processes, which are capable of capturing long-run equilibrium relations between the individual components. An analysis of cointegration relations between the Kannisto coefficients is covered by the following chapter.

While Chapter 2 describes the dynamic part of our proposed stochastic mortality model, Chapter 3 is largely independent of the previous. It is dedicated to an extensive characterization of the distribution, implied by a logistic hazard rate function, as proposed by the KAN predictor. We will study the properties of the Kannisto distribution and show its relation to other well-known distributions.

Chapter 2 | Cointegration Analysis for the Kannisto Model

Contents

2.1 | Introduction

The historical evolution of human mortality, as presented in Section 1.3, shows that improvements of mortality rates are driven by many factors. The GAPC family of mortality models, which is formulated in the framework of generalized linear models, aims to capture relevant changes by a decomposition of the mortality across the dimensions of age, period, and cohort. Many models of this class introduce multiple factors that allow to capture mortality changes at different ages to different extents. As Gaille and Sherris (2011) point out, the consequence of that approach is that several factors of the model often follow common stochastic trends.

In the following chapter, we focus on the dynamics of the system of Kannisto variables. Recall that the Kannisto model, as proposed in Section 1.9, is a parametric logistic-type model for age-related mortality rates, including two, three or four periodic terms. In the following, we treat the Kannisto parameters as stochastic factors and model their dynamics by a multivariate process. Based on the time series of Kannisto parameters, we aim to find an appropriate discrete time stochastic process which is able to capture their characteristics.

A VAR (vector autoregressive) process is considered to be a popular approach for modelling dynamic interactions between multivariate variables. This process is capable of capturing dependencies through time and between variables. However, without further restrictions, it cannot capture long-run relations between the components. In the case of common stochastic trends between the variables, a VECM (vector error correction model) is a better suited time series model.

In the following chapter, a cointegration analysis is performed for the Kannisto model family. The result will reveal whether common stochastic trends are present in the considered time series. The analysis will provide deeper insights in the relations between those factors that drive the structure of mortality. By comparing the more sophisticated VECM/VAR models with the standard choice of a random walk, we demonstrate the impact of common trends on the forecasting performance. The implementation of these methods by practitioners in the life insurance industry might beneficially affect the risk managing strategies.

First, we provide a short overview of other related mortality modelling approaches that involve cointegration methods. Subsequently, in Sections 2.2 and 2.3, we provide a brief introduction to the theoretical background of multivariate time series and the specification procedure for VECM/VAR processes. Note that these sections serve as preparation and are largely based on Johansen (1995), Lütkepohl (2007) and Pfaff (2008). They do not contain original research. Our contribution continues in Section 2.4, where we apply the described methods to the multivariate series of Kannisto parameters. In Section 2.5, we consider projections of the Kannisto parameters under VECMs and compare them to the standard approach, as discussed in Section 1.9.4. Section 2.6 concludes.

Related Studies on Mortality Modelling with Cointegration Methods

The cointegration approach for stochastic mortality models has been considered in several academic studies. Among others, see, e.g., Salhi and Loisel (2011) for a study on long-run equilibrium relations between mortality rates of the insured and the total national population. See also Lazar and Denuit (2009) for an analysis of cointegration relations between the mortality rates of multiple age ranges. The study of Gaille and Sherris (2012) reveals long-run equilibrium relations between the five main causes of death.

H. Li and Lu (2017) proposes a first order spatial-temporal autoregressive model for the mortality surface, where the mortality rates of each age depend on their historical values and, in addition, on the rates of the neighbouring ages. Their approach implies co-integrated mortality rates at different ages and thus prevents the long run divergence of the mortality forecast at different ages, which is a common issue, such as for the Lee-Carter model. Furthermore, the approach by H. Li and Lu (2017) also captures the cohort effect without imposing additional identification constraints on the parameters. In particular, the proposed model avoids a widespread issue of arbitrarily chosen constraints, as discussed in Section 1.8. In contrast to our model, which will be specified in Section 2.4, the model by H. Li and Lu (2017) is non-parametric, where the smoothness of the mortality surface is achieved by penalized least square estimation. In comparison, the non-parametric model is less parsimonious than ours, however, it can be applied to wider age ranges. A further distinction is that in H. Li and Lu (2017) the autoregressive model is applied to logarithmic transformed mortality rates, where in our approach we consider a logistic-type transformation, which is, in particular, more suitable for higher ages.

The closest consideration to our approach is the study of Gaille and Sherris (2011), where the authors use a parametric mortality rate model as introduced by Heligman and Pollard (1980). The so-called Heligman-Pollard model attempts to cover the mortality effects on the entire age range using nine terms for each year. The study of Gaille and Sherris (2011) identifies common stochastic trends between the Heligman-Pollard model parameters.

2.2 | Basic Concepts of Multivariate Time Series Analysis

We start by providing some basic definitions and useful terminologies for the subsequent sections. The presented material is standard and is based on Lütkepohl (2007). The books of Hamilton (1994) and Brockwell and Davis (2013) are further standard references on time series.

Definition 2.2.1 (Multivariate Stochastic Process). Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and Z a countable index set, such as $\mathbb{Z}$ or $\mathbb{N}_0$. A K-dimensional (discrete) *multivariate stochastic process* is a map

$$y : Z \times \Omega \to \mathbb{R}^K,$$

where, for each fixed $t \in Z$, $y(t,\cdot)$ is a K-dimensional random vector. In other words, a multivariate stochastic process is a parameterized collection $y = (y_t)_{t\in Z}$ of random vectors $y_t : \Omega \to \mathbb{R}^K$. To keep the notation simple, it is common to denote the stochastic process by y_t.

For a fixed $\omega \in \Omega$ the map $Z \to \mathbb{R}^K$ with $t \mapsto y_t(\omega)$ is called a *realization* or *path* of a stochastic process. A *time series* is regarded as a finite part of a realization, i.e., as a set of values

$$y_1(\omega),\ldots,y_T(\omega).$$

Since in practice the stochastic process which generates the underlying set of observations is generally unknown, the process itself is referred to as the *data generation process* (DGP). A time series generated by y_t will usually be denoted by $y_1,\ldots,y_T$, where T is called the *sample size*.

Definition 2.2.2 (Stationary Process). A stochastic process y_t is said to be *stationary* if

(a) $\mathbb{E}[|y_t|^2] < \infty$ for all $t \in Z$,

(b) $\mathbb{E}[y_t] = \mu$ for all $t \in Z$,

(c) $\mathbb{E}[(y_t - \mu)(y_{t-h} - \mu)'] = \Gamma_y(h) = \Gamma_y(-h)'$ for all $t \in Z$ and $h = \mathbb{N}_0$.

The first condition states that all first and second moments are finite. The second condition states that all random vectors y_t have the same constant mean. The last condition requires that the *autocovariance*, defined by Γ_y, only depends on the distance h between two variables and not on t. Note that the last condition also ensures that the covariance matrices are invariant under t. If the stochastic process is not stationary, it is said to be *non-stationary*. The above concept of stationarity must not be confused with a stricter form of stationarity, where a process y_t is said to be *strictly stationary* if

$$(y_{t_1},\ldots,y_{t_n}) \overset{d}{=} (y_{t_1+h},\ldots,y_{t_n+h}),$$

for all $n \in \mathbb{N}$ and for all $t_1,\ldots,t_n,h \in Z$, i.e., if the joint distribution of n consecutive vectors is time invariant for all n. Strict stationarity implies stationarity but not the other way around. A stochastic process u_t is called a *white noise* process if the conditions $\mathbb{E}[u_t] = 0$ for all t, $\mathbb{E}[u_t u_t'] = \Sigma_u$, and $\mathbb{E}[u_t u_s'] = 0$ for $t \neq s$ are satisfied. If not stated otherwise, the covariance matrix Σ_u is assumed to be non-singular. If, in addition, u_t is assumed to be normally distributed, i.e., $u_t \sim \mathcal{N}(0,\Sigma_u)$, then the process is called a *Gaussian white noise*.

Definition 2.2.3 (Trend Stationary Process). A stochastic process y_t is said to be *trend stationary* if it has the following decomposition

$$y_t = f(t) + z_t,$$

where z_t is a stationary process according to Definition 2.2.2 and $f(t)$ is a deterministic trend function with values in $\mathbb{R}^K$. We will mainly consider stationarity around a linear trend, such that the trend function will have the form $f(t) = a + bt$ for some fixed parameters a and b.

Definition 2.2.4 (Vector Autoregressive Process (VAR)). A K-dimensional vector autoregressive process of order p, denoted by VAR(p), is defined by

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + CD_t + u_t, \tag{2.1}$$

where the A_i are $(K \times K)$ coefficient matrices for $i = 1, \ldots, p$, C is the $(K \times M)$ coefficient matrix of potentially included deterministic terms, which are represented by a $(M \times 1)$ vector D_t, and u_t is a K-dimensional white noise process.

Thus, a VAR(p) process describes the evolution of K endogenous variables as a regression on a deterministic term as well as on p of their own lags perturbed by a white noise process. An important property of a VAR(p) process is stability, which is the subject of the next definition.

Definition 2.2.5 (Stable Process). A K-dimensional VAR(p) process is said to be *stable* if the condition

$$\det\left(\mathbb{1}_K - A_1 z - \ldots - A_p z^p\right) \neq 0 \quad \text{for} \quad |z| \leq 1. \tag{2.2}$$

is satisfied. Thus, the process is stable if the polynomial of VAR coefficient matrices has no roots in and on the complex circle. The condition of eq. (2.2) is also referred to as the *stability condition* and the polynomial as the *reverse characteristic polynomial* of a VAR(p) process (see Lütkepohl (2007)).

Realizations of stable processes differ qualitatively from realizations of unstable processes. While stable processes generate trajectories which typically fluctuate around constant means with time-invariant variance, trajectories of unstable processes usually show trends or strong seasonal fluctuations. See Lütkepohl (2007) for a detailed discussion on that topic.

Proposition 2.2.6 (Stationarity Condition). A stable VAR(p) process of the form

$$y_t = \nu + A_1 y_{t-1} + \cdots A_p y_{t-p} + u_t, \quad \text{for } t \in \mathbb{Z}$$

is stationary.

Proof. The result can be obtained by the moving average representation of the VAR process, see Lütkepohl, 2007, Proposition 2.1. □

The previous result states that stability of a VAR(p) process, which has been initiated in the infinite past, implies stationarity. See Lütkepohl (2007) for possible generalizations of this result for VAR(p)

processes starting at some finite time t_0. In preparation for the next definition, we introduce the notation of the *lag operator* L which shifts the index of the process by one period, i.e., $Ly_t = y_{t-1}$ and the *difference operator* Δ such that $\Delta y_t = (1-L)y_t = y_t - y_{t-1}$. Consequently, $\Delta^d y_t = (1-L)^d y_t$ will denote the difference of order d.

Definition 2.2.7 (Integrated & Cointegrated Stochastic Processes; Lütkepohl (2007)). A K-dimensional stochastic process y_t is called *integrated* of order d, denoted by $y_t \sim I(d)$, if $\Delta^d y_t$ is stable and $\Delta^{d-1} y_t$ is not stable. An integrated process is sometimes also called a *unit root process*. An $I(d)$ process y_t is said to be *cointegrated* if there exists a linear combination $\boldsymbol{\beta}' y_t$ with $\boldsymbol{\beta} \neq 0$ which is integrated of order less than d. The vector $\boldsymbol{\beta}$ is called the *cointegrating vector*.

In the following, we will only consider processes that are at most integrated of order one. Thus, cointegration relations are necessarily stationary or trend stationary, i.e., for a K-dimensional process y_t and cointegrating vector $\boldsymbol{\beta}$, we have

$$\boldsymbol{\beta}' y_t = \beta_1 y_{Kt} + \cdots + \beta_K y_{Kt} = z_t, \tag{2.3}$$

where z_t is stationary or trend stationary. The process z_t in eq. (2.3) is considered to be a deviation from the long-run equilibrium $\boldsymbol{\beta}' y_t$.

Definition 2.2.8 (Random Walk with Drift). The process y_t is called a *random walk with drift* if it has the form

$$y_t = \nu + y_{t-1} + u_t, \tag{2.4}$$

where ν is a non-zero constant vector and u_t is a white noise process. For $\nu = 0$, the process is simply called a *random walk*.

A random walk with drift is non-stationary and integrated of order one. To see that, let the univariate process start in t_0 with value y_0 and consider the successive substitution for lagged y_t's, i.e.,

$$y_t = \nu + y_{t-1} + u_t = \nu + (\nu + y_{t-2} + u_{t-1}) + u_t = \cdots = y_0 + \nu t + \sum_{i=1}^{t} u_t.$$

Thus, we can directly conclude by $\mathbb{E}[y_t] = y_0 + \nu t$ and $\mathbb{V}[y_t] = \mathbb{V}[u_t] = t\sigma_u^2$ that the random walk is non-stationary. Furthermore, checking the stability condition of Definition 2.2.5 reveals that the reverse characteristic polynomial of the random walk has a root on the unit circle. It is important to point out that constant terms in VAR processes have different influences on stable and unstable processes. Note that the constant term in eq. (2.4) corresponds to the slope of the deterministic trend. Consider a univariate first-order autoregressive process defined by

$$y_t = \nu + \phi y_{t-1} + u_t, \tag{2.5}$$

with $|\phi| < 1$ and other terms as in eq. (2.4). This process is indeed stable and it can be shown that the mean and autocovariance of eq. (2.5) are given by

$$\mathbb{E}[y_t] = \frac{\phi}{1-\phi} =: \mu$$

and

$$\Gamma_y(h) = \mathbb{E}\left[(y_t - \mu)(y_{t-h} - \mu)\right] = \frac{\phi^h}{1-\phi^2}\sigma_u^2.$$

In contrast, for stable processes, the constant term ν does not correspond to a linear trend, as it does for the random walk, but rather determines the mean of the process. As illustrated in this example, stable and non-stable processes differ significantly. To be able to distinguish whether a given sample is generated by stable and non-stable processes, testing strategies have been developed. The so-called unit root or stationarity tests will be discussed in Section 2.3.2.

2.3 | Specification Procedure for Multivariate Time Series

In the following section, we provide an overview of common methods and procedures which are applied to multivariate time series in order to find an appropriate DGP among the family of VAR processes. The specification steps of the general procedure are described in the following remark.

Remark 2.3.1 (VAR/VECM Specification). Following Lütkepohl (2007, Chapter 8), the specification steps include:

(1) The lag order selection of an unrestricted VAR(p) is obtained through several information criteria, such as the Akaike's Information Criterion (AIC), the Hannan-Quinn Criterion (HQ), the Schwarz Criterion (SC) or the Final Prediction Error (FPE).

(2) Unit root and stationary tests are applied to each univariate time series. The considered tests include the Augmented Dickey-Fuller test (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS). These tests have contrary null hypotheses. The ADF test checks the null hypothesis of a unit root, while the null hypothesis of KPSS is level or trend stationarity.

(3.a) If all univariate variables are stationary, then a VAR(p) is an appropriate DGP for the underlying observation. Parameters of VAR processes with Gaussian noise term can be consistently estimated using the OLS regression. See Lütkepohl (2007, Chapter 3) for the concrete form of the estimator and its asymptotic properties.

(3.b) If some univariate variables are integrated of order one, then the Johansen test procedure should be applied in order to detect the presence of cointegration relations among the variables. If there are some cointegration relations, then the representation of a VAR as given in eq. (2.1) is not optimal, since there are restrictions on the VAR coefficients imposed by cointegration relations. These restrictions are only covered implicitly by the VAR representation. A superior representation of a VAR with cointegration relations is given by the so-called vector error correction model (VECM). The VECM representation covers the imposed restrictions by cointegration relations explicitly as rank restrictions on a particular coefficient matrix. The VECM methodology will be introduced in Section 2.3.3.

(3.c) If some univariate variables are integrated of order one, but there are no cointegration relations, then a VAR$(p-1)$ process can be applied to the difference process Δy_t.

(4) The specification terminates with a validation procedure, where several tests are applied to the residuals to detect model misspecification through exceptional autocorrelations or non-normality, see Section 2.3.6.

2.3.1 | Lag Order Selection

In the upcoming section, we provide a brief introduction to some widely used criteria for lag order selection. These include the information criteria as defined by Akaike (1981), abbreviated by AIC, the Hannan and Quinn (1979) criterion, denoted by HQ, the Schwarz (1978) criterion, abbreviated SC, and the *final prediction error* (FPE) criterion by Lütkepohl (2007). The measures for the length of the lag determination are defined as

$$\mathrm{AIC}(m) = \ln\left|\tilde{\Sigma}_u(m)\right| + \frac{2}{T}mK^2, \tag{2.6}$$

$$\mathrm{HQ}(m) = \ln\left|\tilde{\Sigma}_u(m)\right| + \frac{2\ln\ln T}{T}mK^2, \tag{2.7}$$

$$\mathrm{SC}(m) = \ln\left|\tilde{\Sigma}_u(m)\right| + \frac{\ln T}{T}mK^2, \tag{2.8}$$

$$\mathrm{FPE}(m) = \left(\frac{T+Km+1}{T-Km-1}\right)^K \left|\tilde{\Sigma}_u(m)\right|, \tag{2.9}$$

where $\tilde{\Sigma}_u(m) = T^{-1}\sum_{t=1}^T \hat{u}_t\hat{u}_t'$ is the estimator of Σ_u obtained by fitting a VAR(m) to the K-dimensional time series. Note that while having distinct penalization terms, all above criteria are functions of the determinant of the residual covariance matrix. Let Cr denote one of the criteria defined in eqs. (2.6) to (2.9), then the lag order p is chosen as

$$p = \underset{0\le m\le M}{\arg\min}\,\mathrm{Cr}(m),$$

where $M \in \mathbb{N}$ is a pre-specified maximal considered lag order. For a detailed discussion on the presented criteria, such as their derivation, small sample properties, and their consistency as lag order estimators, see Lütkepohl (2007).

2.3.2 | Unit Root and Stationary Tests

In this section, we present two common statistical tests which are used to analyse whether the DGP of the observed time series is a stationary or a non-stationary unit root process.

Augmented Dickey-Fuller Test

The most common test is the *Augmented Dickey-Fuller* (ADF) unit root test which is based on the regression

$$\Delta\theta_t = \xi_0 + \xi_1 t + \pi\,\theta_{t-1} + \sum_{i=1}^k \gamma_i \Delta\theta_{t-i} + \epsilon_t, \tag{2.10}$$

where the error term ϵ_t is a Gaussian white noise process. The number of lagged differences k in the regression can be either determined by information criteria, by a significance t-type test of the regression parameter or by checking the residuals for the absence of serial correlation, see Pfaff (2008). Based on the regression of eq. (2.10), the objective is to test the hypothesis pair

$$H_0 : \pi = 0 \quad \text{vs.} \quad H_1 : \pi \neq 0,$$

which is given by a t-statistic of the OLS estimated parameter π. The test statistic does not have an asymptotic normal distribution. Critical values have to be obtained by simulation and are provided by Dickey and Fuller (1981) and Fuller (1976). Furthermore, the limiting distribution does also depend on the type of the included deterministic term. The null hypothesis $\pi = 0$ implies that θ_t is an $I(1)$ process. This can be seen by adding y_{t-1} to both sides of eq. (2.10) which yields

$$\theta_t = \xi_0 + \xi_1 t + (1 + \pi)\theta_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta\theta_{t-i} + \epsilon_t.$$

For $\pi = 0$, the coefficient of the term y_{t-1} is equal to one, thus, the stability condition of eq. (2.2) is violated. Since we do not consider processes of a higher order of integration, the term $\sum_{i=1}^{k} \gamma_i \Delta\theta_{t-i}$ is necessarily a stationary process. Thus, for $\pi = 0$, the process θ_t can be essentially decomposed as

$$\theta_t \cong \text{deterministic term} + \text{random walk} + \text{stationary error process}.$$

As mentioned above, the limiting distribution of the ADF test depends on whether the regression in eq. (2.10) only includes a constant ξ_0 or a linear $\xi_0 + \xi_1 t$ term. Therefore, there exists a series of possible tests for $\pi = 0$. In the more general case, where $\xi_0 + \xi_1 t$ is included, we have a t-type test statistic

$$\tau_3 = \frac{\hat{\pi}}{\text{s.e.}(\hat{\pi})},$$

for the hypothesis pair

$$H_0 : \pi = 0 \quad \text{vs.} \quad H_1 : \pi \neq 0 \tag{2.11}$$

and two joined null F-type test statistics ϕ_3 and ϕ_2 for the pairs

$$H_0 : (\xi_1 = 0 \wedge \pi = 0) \quad \text{vs.} \quad H_1 : (\xi_1 \neq 0 \vee \pi \neq 0) \tag{2.12}$$

and

$$H_0 : (\xi_0 = 0 \wedge \xi_1 = 0 \wedge \pi = 0) \quad \text{vs.} \quad H_1 : (\xi_0 \neq 0 \vee \xi_1 \neq 0 \vee \pi \neq 0).$$

The null hypothesis, corresponding to the statistic τ_3, is that the considered process is essentially a random walk with drift and a deterministic time trend. The corresponding null hypothesis for the statistic ϕ_3 is a random walk with drift and for ϕ_2 the null hypothesis is a random walk without drift.

In the second case, where only a constant term ξ_0 is included in eq. (2.10), we have a t-type statistic, denoted by τ_2, testing $H_0 : \pi = 0$ versus $H_1 : \pi \neq 0$, and a joined F-type test for the hypothesis pair $H_0 : \xi_0 = 0 \wedge \pi = 0$ versus $H_1 : \xi_0 \neq 0 \vee \pi \neq 0$. These null hypotheses correspond to a random walk with and without drift. Critical values of the involved non-standard statistics can be found in Dickey and Fuller (1981) and Fuller (1976).

In Section 2.4, we will provide a unit root analysis for time series given by the coefficients $\boldsymbol{\kappa}_t$ of the Kannisto model. To verify the assumption that the underlying DGPs are indeed $I(1)$ processes, the following procedure is applied. First, the ADF test is applied to each individual component of $\boldsymbol{\kappa}_t$. If the null hypotheses cannot be rejected, the ADF tests are reapplied to the components of the first-order differences $\Delta\boldsymbol{\kappa}_t$. If the latter tests turn out to be significant, then we conclude that $\Delta\boldsymbol{\kappa}_t \sim I(0)$ and therefore, $\boldsymbol{\kappa}_t \sim I(1)$. For a comprehensive introduction to the testing procedure see, e.g., Pfaff (2008).

Kwiatkowski-Phillips-Schmidt-Shin Test

As a confirmatory analysis for unit roots, we use the KPSS test proposed by Kwiatkowski, Phillips, Schmidt and Shin (1992). The KPSS test has the null hypothesis of stationarity, which is contrary to the ADF test with the null hypothesis of a unit root. Kwiatkowski, Phillips, Schmidt and Shin (1992) derived their test by considering a decomposition of the process into a deterministic component $\beta' D_t$, a pure random walk v_t with innovation variance σ_ϵ^2, and a stationary error process u_t given by

$$\begin{aligned} \theta_t &= CD_t + v_t + u_t \\ v_t &= v_{t-1} + \epsilon_t . \end{aligned}$$

Obviously, for $\sigma_\epsilon^2 = 0$, θ_t is trend stationary and for $\sigma_\epsilon^2 > 0$, the process is integrated. Thus, the null hypothesis of $\theta_t \sim I(0)$ is formulated as $H_0 : \sigma_\epsilon^2 = 0$, and the alternative is $H_1 : \sigma_\epsilon^2 > 0$. The KPSS test uses a one-sided Lagrange multiplier statistic (see Section 1.6.3) which is given by

$$\text{KPSS} = \frac{1}{T^2} \frac{\sum_{t=1}^{T} \hat{S}_t^2}{\hat{\sigma}_u^2(l)},$$

where $\hat{S}_t = \sum_{j=1}^{t} \hat{u}_j$ is the partial sum of the residual obtained by a regression of θ_t on the deterministic trend CD_t and

$$\hat{\sigma}_u^2(l) = \frac{1}{T} \sum_{i=1}^{T} \hat{u}_t^2 + \frac{2}{T} \sum_{s=1}^{l} \left(1 - \frac{s}{l+1}\right) \sum_{t=s+1}^{T} \hat{u}_t \hat{u}_{t-s} \tag{2.13}$$

is a consistent estimator of the long-run variance of u_t using l as a length of a spectral window in the so-called Barlett weighting function. Suitable choices for l might be $l = 4\sqrt[4]{T/100}$ or $l = 12\sqrt[4]{T/100}$ for T being the sample size, see Pfaff (2008). Kwiatkowski, Phillips, Schmidt and Shin (1992) showed that under H_0 the KPSS statistic converges in distribution to a function of standard Brownian motion that depends on the form D_t, but not on C. Critical values of the KPSS statistic for $D_t = 1$ and $D_t = (1,t)'$ are obtained by simulation and can be found in Kwiatkowski, Phillips, Schmidt and Shin (1992). In Section 2.4, η_τ will denote the KPSS statistic when $D_t = (1,t)'$ is used to test trend-stationarity, and

η_μ will denote the KPSS statistic for $D_t = 1$ to test stationarity.

Similar to the ADF unit root testing procedure, we will apply the KPSS test to $\boldsymbol{\kappa}_t$ and $\Delta\boldsymbol{\kappa}_t$. If the underlying DGP of $\boldsymbol{\kappa}_t$ is $I(1)$, then we expect that KPSS tests will reject the null hypothesis for components of $\boldsymbol{\kappa}_t$ but not for $\Delta\boldsymbol{\kappa}_t$.

2.3.3 | Vector Error Correction Model

The section is devoted to *vector error correction models* (VECM), or alternatively called *vector equilibrium correction models*. VECMs provide an alternative representation of VAR processes which is more convenient if cointegration relations between variables exist. The topics presented in this section are considered as known and are based on Lütkepohl (2007).

VECM Methodology

Recall from Definition 2.2.4 that omitting any deterministic terms, a K-dimensional vector autoregressive process y_t with order p is defined as

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{2.14}$$

where $A_1, \ldots, A_p$ denote the coefficient matrices $(K \times K)$ and u_t is a white noise process of dimension K. Following Lütkepohl (2007), subtracting y_{t-1} from eq. (2.14) and subsequently rearranging terms, we obtain the VECM form

$$\Delta y_t = \mathbf{\Pi} y_{t-1} + \mathbf{\Gamma}_1 \Delta y_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \Delta y_{t-p} + u_t, \tag{2.15}$$

which is an equivalent representation of the VAR standard form, as given in eq. (2.14). In eq. (2.15), that specifies a VECM$(p-1)$ process, we have

$$\mathbf{\Pi} := -(\mathbb{1}_K - A_1 - \ldots - A_p)$$

and

$$\mathbf{\Gamma}_i := -(A_{i+1} + \cdots + A_p),$$

for $i = 1, \ldots, p-1$. If we now assume that all variables of y_t are at most $I(1)$, we see that the left-hand side of eq. (2.15), Δy_t, must be $I(0)$. On the right-hand side, Πy_{t-1} is the only term with $I(1)$ variables, but since it must also be $I(0)$, it has to contain the potential cointegration relations. If we now consider a VAR(p) process with unit roots, then by definition we have

$$|\mathbb{1}_K - A_1 z - \ldots - A_p z^p| = 0,$$

for $z = 1$. This implies that $\mathbf{\Pi} = -(\mathbb{1}_K - A_1 - \ldots - A_p)$ in eq. (2.15) does not have a full rank K. Let $\operatorname{rk} \mathbf{\Pi} = r < K$, then $\mathbf{\Pi}$ can be decomposed into a matrix product $\boldsymbol{\alpha\beta}'$ with $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}'$ being $(K \times r)$ matrices of rank r, see Lütkepohl (2007). The matrix $\boldsymbol{\alpha}$ is called a *loading matrix*. Since $\boldsymbol{\beta}' y_{t-1}$ contains the cointegration relations, $\boldsymbol{\beta}$ is referred to as a *cointegration matrix*, and the rank of $\mathbf{\Pi}$ as the *cointegration rank*. As Lütkepohl (2007) points out, the decomposition $\mathbf{\Pi} = \boldsymbol{\alpha\beta}'$ is not unique.

However, by a suitable rearrangement of variables it is always possible to achieve a unique normalized form, e.g.,

$$\boldsymbol{\beta}' = \left[\mathbb{1}_r : \tilde{\boldsymbol{\beta}}'_{(K-r)}\right], \tag{2.16}$$

where $\mathbb{1}_r$ is the $(r \times r)$ identity matrix and $\tilde{\boldsymbol{\beta}}'_{(K-r)}$ is an $(r \times K - r)$ matrix. This normalized form ensures a unique cointegration matrix and will be used from this point on.

In a VECM as in eq. (2.15), three essential cases for the cointegration rank can be distinguished. If $r = K$, then all variables are stationary and the process y_t has a stable VAR(p) representation. This corresponds to the case (3.a) at page 106. If $r = 0$, then there are no cointegration relations among the variables, $\boldsymbol{\Pi} y_{t-1}$ vanishes in eq. (2.15), and therefore Δy_t has as stable representation as a VAR$(p-1)$ process, see case (3.c) on page 107. Assuming now that all variables are $I(1)$, then $0 < r < K$ implies that there are r cointegration relations among the variables such that $\boldsymbol{\beta}' y_t \sim I(0)$, or equivalently, the variables are driven by $K - r$ common trends. This situation describes the case (3.b), where the system of variables is represented as a VECM$(p-1)$, as given in eq. (2.15) or equivalently by

$$\Delta y_t = \boldsymbol{\alpha}\boldsymbol{\beta}' y_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p} + u_t.$$

This VECM$(p-1)$ representation can also be transformed to a VAR(p) representation by the following reorganization of coefficient matrices

$$\begin{aligned} A_1 &= \mathbb{1}_K + \boldsymbol{\alpha}\boldsymbol{\beta}' + \boldsymbol{\Gamma}_1, \\ A_i &= \boldsymbol{\Gamma}_i - \boldsymbol{\Gamma}_{i-1}, \quad \text{for } i = 2, \ldots, p-1 \\ A_p &= -\boldsymbol{\Gamma}_{p-1}. \end{aligned}$$

However, the advantage of the VECM representation is that the hypothesis of cointegration relations among the variables can be formulated in terms of a reduced rank tests of the matrix $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, i.e.,

$$H_0 : \operatorname{rk} \boldsymbol{\Pi} = r \quad \text{vs.} \quad H_1 : \operatorname{rk} \boldsymbol{\Pi} > r, \tag{2.17}$$

for $r = 0, 1, \ldots, K-1$. Likelihood ratio based testing procedures for the hypothesis pair of eq. (2.17), as well as reduced rank ML estimation methods, have been developed by Johansen (1988, 1995) and will be introduced in Section 2.3.5

2.3.4 | VECM Parameter Estimation

In this section, we introduce two estimation methods for VECMs. The presentation of the following is based on Lütkepohl (2007, Section 7.2). We begin by providing some notations needed to express the estimators. As the underlying process, we consider a VECM with deterministic terms of the form

$$\begin{aligned} \Delta y_t &= \boldsymbol{\alpha}[\boldsymbol{\beta}' : \eta'] \begin{bmatrix} y_{t-1} \\ D^{\text{co}}_{t-1} \end{bmatrix} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p+1} + \Psi D_t + u_t \\ &= \boldsymbol{\Pi}^+ y^+_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p+1} + \Psi D_t + u_t, \end{aligned} \tag{2.18}$$

where y_t is a K-dimensional process, $\mathrm{rk}(\boldsymbol{\alpha}\boldsymbol{\beta}') = r$, with $0 < r < K$, where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are $K \times r$ matrices with full rank r, $\boldsymbol{\Gamma}_{i\in 1,\ldots,p-1}$ are $K \times K$ parameter matrices, and $u_t \sim (0,\Sigma_u)$ is a white noise process. Further, we use the notation

$$\boldsymbol{\Pi}^+ := \boldsymbol{\alpha}[\boldsymbol{\beta}' : \eta'] = \boldsymbol{\alpha}\boldsymbol{\beta}^{+\prime} \qquad \text{and} \qquad y_{t-1}^+ := \begin{bmatrix} y_{t-1} \\ D_{t-1}^{\mathrm{co}} \end{bmatrix},$$

where η' represents a vector of coefficients of the deterministic term D_{t-1}^{co}, which is restricted to the cointegration relation. The coefficients Ψ, on the other side, correspond to the unrestricted deterministic D_t terms. To avoid an over-determined system of equations, we assume that specific deterministic terms appear either in D_{t-1}^{co} or in D_t. We also assume that y_t is an integrated process of order one and that

$$\boldsymbol{\alpha}_\perp' \left(\mathbb{1}_K - \sum_{i=1}^{p-1} \Gamma_i \right) \boldsymbol{\beta}_\perp \tag{2.19}$$

is invertible, where $\boldsymbol{\alpha}_\perp'$ and $\boldsymbol{\beta}_\perp$ denote orthogonal complements of the $(r \times K)$ matrices $\boldsymbol{\alpha}'$ and $\boldsymbol{\beta}$. The assumption in eq. (2.19) is required for the Granger representation theorem, which is a result on the decomposition of y_t into integrated and stationary components, see Engle and Granger (1987) for further details.

Following the notation of Lütkepohl (2007), for $t = 1,\ldots,T$ the VECM in eq. (2.18) can be written in matrix notation as

$$\Delta Y = \boldsymbol{\Pi}^+ Y_{-1}^+ + \boldsymbol{\Gamma}^+ \Delta X^+ + U,$$

where

$$\Delta Y := [\Delta y_1,\ldots,\Delta y_T], \tag{2.20}$$
$$Y_{-1}^+ := [y_0^+,\ldots,y_{T-1}^+], \tag{2.21}$$
$$\boldsymbol{\Gamma}^+ := [\boldsymbol{\Gamma}_1,\ldots,\boldsymbol{\Gamma}_{p-1},\Psi], \tag{2.22}$$
$$\Delta X^+ := [\Delta X_0^+,\ldots,\Delta X_{T-1}^+], \tag{2.23}$$

with

$$\Delta X_{t-1}^+ := \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \\ D_t \end{bmatrix} \tag{2.24}$$

and

$$U := [u_1,\ldots,u_T].$$

LS Estimator for VECM

The multivariate least squares (LS) estimator, where no rank restriction on $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ is taken into account, is given by (Lütkepohl, 2007)

$$[\hat{\mathbf{\Pi}}^+ : \hat{\mathbf{\Gamma}}^+] = \left[\Delta Y \Delta Y_{-1}^{+\prime} : \Delta Y \Delta X^{+\prime}\right] \begin{bmatrix} Y_{-1}^+ Y_{-1}^{+\prime} & Y_{-1}^+ \Delta X^{+\prime} \\ \Delta X^+ Y_{-1}^{+\prime} & \Delta X^+ \Delta X^{+\prime} \end{bmatrix}^{-1}. \tag{2.25}$$

A consistent estimator of the white noise covariance matrix can be obtained as

$$\hat{\Sigma}_u = (T - Kp)^{-1}(\Delta Y - \hat{\mathbf{\Pi}}^+ Y_{-1}^+ - \hat{\mathbf{\Gamma}}^+ \Delta X^+)(\Delta Y - \hat{\mathbf{\Pi}}^+ Y_{-1}^+ - \hat{\mathbf{\Gamma}}^+ \Delta X^+)'. \tag{2.26}$$

The asymptotic properties of the LS estimator are derived in Lütkepohl (2007, Proposition 7.1). Based on that, the estimator as given eq. (2.25) is consistent and asymptotically normal, i.e.,

$$\sqrt{T}\,\mathrm{vec}\left([\hat{\mathbf{\Pi}}^+ : \hat{\mathbf{\Gamma}}^+] - [\mathbf{\Pi}^+ : \mathbf{\Gamma}^+]\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_{co}), \tag{2.27}$$

where

$$\Sigma_{\mathrm{co}} = \Lambda \otimes \Sigma_u, \tag{2.28}$$

with

$$\Lambda = \left(\begin{bmatrix} \boldsymbol{\beta}^{+\prime} & 0 \\ 0 & \mathbb{1}_{Kp-K} \end{bmatrix} \Omega^{-1} \begin{bmatrix} \boldsymbol{\beta}^{+\prime} & 0 \\ 0 & \mathbb{1}_{Kp-K} \end{bmatrix}\right),$$

and

$$\Omega = \operatorname{plim} \frac{1}{T} \begin{bmatrix} \boldsymbol{\beta}^{+\prime} Y_{-1}^+ Y_{-1}^{+\prime} \boldsymbol{\beta}^+ & \boldsymbol{\beta}^{+\prime} Y_{-1}^+ \Delta X^{+\prime} \\ \Delta X^+ Y_{-1}^{+\prime} \boldsymbol{\beta}^+ & \Delta X^+ \Delta X^{+\prime} \end{bmatrix}, \tag{2.29}$$

where $\otimes$ in eq. (2.28) denotes the Kronecker product and plim in eq. (2.29) the probability limit. The Matrix Λ can be consistently estimated by

$$\hat{\Lambda} = T \begin{bmatrix} Y_{-1}^+ Y_{-1}^{+\prime} & Y_{-1}^+ \Delta X^{+\prime} \\ \Delta X^+ Y_{-1}^{+\prime} & \Delta X^+ \Delta X^{+\prime} \end{bmatrix}^{-1}.$$

For a known $\boldsymbol{\beta}$, the LS estimator of $[\boldsymbol{\alpha}, \mathbf{\Gamma}^+]$ can be derived as

$$[\hat{\boldsymbol{\alpha}}, \hat{\mathbf{\Gamma}}^+] = \left[\Delta Y \Delta Y_{-1}^{+\prime} \boldsymbol{\beta}^+ : \Delta Y \Delta X^{+\prime}\right] \begin{bmatrix} \boldsymbol{\beta}^{+\prime} Y_{-1}^+ Y_{-1}^{+\prime} \boldsymbol{\beta}^+ & \boldsymbol{\beta}^{+\prime} Y_{-1}^+ \Delta X^{+\prime} \\ \Delta X^+ Y_{-1}^{+\prime} \boldsymbol{\beta}^+ & \Delta X^+ \Delta X^{+\prime} \end{bmatrix}^{-1}. \tag{2.30}$$

The LS estimator of eq. (2.30) has an asymptotic normal distribution, i.e.,

$$\sqrt{T}\,\mathrm{vec}\left([\hat{\boldsymbol{\alpha}} : \hat{\mathbf{\Gamma}}^+] - [\boldsymbol{\alpha} : \mathbf{\Gamma}^+]\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\boldsymbol{\alpha},\mathbf{\Gamma}}), \tag{2.31}$$

where

$$\Sigma_{\boldsymbol{\alpha},\Gamma} = \Omega^{-1} \otimes \Sigma_u = \operatorname{plim} \frac{1}{T} \begin{bmatrix} \boldsymbol{\beta}^{+\prime} Y_{-1}^{+} Y_{-1}^{+\prime} \boldsymbol{\beta}^{+} & \boldsymbol{\beta}^{+\prime} Y_{-1}^{+} \Delta X^{+\prime} \\ \Delta X^{+} Y_{-1}^{+\prime} \boldsymbol{\beta}^{+} & \Delta X^{+} \Delta X^{+\prime} \end{bmatrix}^{-1} \otimes \Sigma_u.$$

As noted in Lütkepohl (2007), the asymptotic distributions of eq. (2.31) and eq. (2.27), reduced to the coefficients $[\boldsymbol{\alpha},\boldsymbol{\Gamma}^{+}]$, coincide and do no depend on whether $\boldsymbol{\beta}$ has been estimated or assumed to be known. This fact is a consequence of the faster convergence rate for the estimation of $\boldsymbol{\beta}$.

ML Estimator for VECM

Under the assumption that the process y_t is Gaussian, i.e., $u_t \sim \mathcal{N}(0,\Sigma_u)$, a maximum likelihood estimator (ML) can be used for the estimation of a VECM. The ML estimator for VECM is derived in Johansen (1995, Theorem 6.1) and considers explicitly the rank restriction of the matrix $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$. Using the notations of eqs. (2.20) to (2.24), the log-likelihood function of VECM with sample size T is given by

$$\ln L(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\Gamma},\Sigma_u) = -\frac{KT}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma_u| - \frac{1}{2} \operatorname{tr}\left[(\Delta Y - \boldsymbol{\alpha}\boldsymbol{\beta}' Y_{-1} - \boldsymbol{\Gamma}\Delta X)' \Sigma_u^{-1} (\Delta Y - \boldsymbol{\alpha}\boldsymbol{\beta}' Y_{-1} - \boldsymbol{\Gamma}\Delta X) \right], \tag{2.32}$$

where tr denotes the trace operator. In the following, we use the above notation and outline the ML estimator for a VECM as can be found in Lütkepohl (2007, Proposition 7.3). The ML estimator given below is an estimator for a VECM with no deterministic terms. However, a corresponding estimator for eq. (2.18) can be analogously obtained by replacing $(Y_{-1},\Delta X,\Gamma)$ with the extended variables $(Y_{-1}^{+},\Delta X^{+},\Gamma^{+})$.

Proposition 2.3.2 (ML Estimators for VECMs; Lütkepohl (2007))**.** Let y_t be a VECM driven by Gaussian white noise u_t and let $M := \mathbb{1}_T - \Delta X(\Delta X \Delta X')^{-1}$, $R_0 := \Delta Y M$ and $R_1 := Y_{-1}M$, where $\Delta Y := [\Delta y_1,\ldots,\Delta y_T]$ as before, $Y_{-1} := [y_0,\ldots,y_{T-1}]$, $\boldsymbol{\Gamma} := [\boldsymbol{\Gamma}_1,\ldots,\boldsymbol{\Gamma}_{p-1}]$, and $\Delta X := [\Delta X_0,\ldots,\Delta X_{T-1}]$ with $\Delta X_{t-1} := [\Delta y_{t-1},\ldots,\Delta y_{t-p+1}]'$ and define

$$S_{ij} := \frac{1}{T} R_i R_j', \quad i,j \in \{0,1\}.$$

Let $\lambda_1 \geq \cdots \geq \lambda_K$ be the ordered eigenvalues of the generalized eigenvalue equation

$$|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0, \tag{2.33}$$

with corresponding orthonormal eigenvectors $v_1,\ldots v_K$, then the log-likelihood function as given in eq. (2.32) is maximized for

$$\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}} := [v_1 \ldots,v_r]' S_{11}^{-1/2}, \tag{2.34}$$

$$\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}} := \Delta Y M Y_1' \tilde{\boldsymbol{\beta}} (\tilde{\boldsymbol{\beta}}' Y_{-1} M Y_{-1}' \tilde{\boldsymbol{\beta}})^{-1} = S_{01} \tilde{\boldsymbol{\beta}} (\tilde{\boldsymbol{\beta}}' S_{11} \tilde{\boldsymbol{\beta}})^{-1}, \tag{2.35}$$

$$\boldsymbol{\Gamma} = \tilde{\boldsymbol{\Gamma}} := (\Delta Y - \tilde{\boldsymbol{\alpha}} \tilde{\boldsymbol{\beta}}' Y_{-1}) \Delta X' (\Delta X \Delta X')^{-1}, \tag{2.36}$$

$$\Sigma_u = \tilde{\Sigma}_u := \frac{1}{T}(\Delta Y - \tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{\beta}}' Y_{-1} - \tilde{\boldsymbol{\Gamma}}\Delta X)(\Delta Y - \tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{\beta}}' Y_{-1} - \tilde{\boldsymbol{\Gamma}}\Delta X)'. \tag{2.37}$$

Proof. See Lütkepohl (2007, Proposition 7.3) or Johansen (1995, Theorem 6.1). □

The maximum of the log-likelihood function obtained by plugging in the estimators of eqs. (2.34) to (2.37) is given by

$$\max \ln L = \ln L(\tilde{\boldsymbol{\alpha}},\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\Gamma}},\tilde{\Sigma}_u) = -\frac{KT}{2}\ln 2\pi - \frac{T}{2}\left[\ln|S_{00}| + \sum_{i=1}^{r}\ln(1-\lambda_i)\right] - \frac{KT}{2}. \tag{2.38}$$

These estimators are consistent and jointly asymptotically normal, i.e.,

$$\sqrt{T}\,\mathrm{vec}\left([\tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{\beta}}' : \tilde{\boldsymbol{\Gamma}}] - [\boldsymbol{\Pi} : \boldsymbol{\Gamma}]\right) \xrightarrow{d} \mathcal{N}(0,\Sigma_{\mathrm{co}}),$$

where Σ_{co} is defined in eq. (2.28). The estimator $\tilde{\Sigma}_u$ of the covariance matrix Σ_u is also asymptotically normal, see Lütkepohl (2007, Proposition 7.4) for the asymptotic covariance matrix, and Johansen (1995) for the proof. Note that without further restrictions, only the product $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ and therefore the cointegration space can be estimated consistently, but not the parameters of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. To obtain uniqueness, it is necessary to impose further restrictions. This can be done by normalization of $\boldsymbol{\beta}$ in accordance to eq. (2.16). That means, for $r > 1$, $\boldsymbol{\beta}$ is normalized to the form

$$\boldsymbol{\beta} = \begin{bmatrix} \mathbb{1}_r \\ \boldsymbol{\beta}_{(K-r)} \end{bmatrix},$$

where $\beta_{(K-r)}$ is a $K \times (K-r)$ matrix denoting the last $K-r$ rows. As mentioned in Lütkepohl (2007, Remark 3 of Proposition 7.4), the ML estimator $\tilde{\tilde{\boldsymbol{\beta}}}_{(K-r)}$ of $\boldsymbol{\beta}_{(K-r)}$ can be obtained as the last $K-r$ rows of $\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}_{(r)}^{-1}$, where $\tilde{\boldsymbol{\beta}}_{(r)}$ consists of the first r rows of $\tilde{\boldsymbol{\beta}}$, i.e.,

$$\tilde{\tilde{\boldsymbol{\beta}}}_{(K-r)} := \left(\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}_{(r)}^{-1}\right)_{(K-r)}.$$

For $r = 1$, the normalization is directly obtained through division by the first component of the vector $\tilde{\boldsymbol{\beta}} = v_1' S_{11}^{-1/2}$. Johansen (1995) shows that the estimators are consistent and both $T(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $\sqrt{T}(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ converge in distribution. The latter case has the same asymptotic distribution as in eq. (2.31). The faster convergence rate of the estimator of $\boldsymbol{\beta}$ is also said to be *superconsistent*. For the sake of completeness, the asymptotics of the normalized coefficients follow

$$\mathrm{vec}\left[\left(\tilde{\tilde{\boldsymbol{\beta}}}'_{(K-r)} - \boldsymbol{\beta}'_{(K-r)}\right)\left(R_{1_{(K-r)}}R'_{1_{(K-r)}}\right)^{\frac{1}{2}}\right] \xrightarrow{d} \mathcal{N}\left(0,\mathbb{1}_{K-r} \otimes (\boldsymbol{\alpha}'\Sigma_u^{-1}\boldsymbol{\alpha})^{-1}\right), \tag{2.39}$$

where $R_{1_{(K-r)}}$ denotes the last $K-r$ rows of R_1. For a more detailed discussion and proofs, we refer to Lütkepohl (2007) and Ahn and Reinsel (1990, Theorem 4). The asymptotic covariance matrices as given in eq. (2.27) or in eq. (2.31) and eq. (2.39) are used to derive standard deviations and t-ratios of the estimated coefficients.

2.3.5 | Johansen Tests for Cointegration Rank

For the following, we consider a VECM driven by Gaussian white noise. Under this setting, a ML estimator is available as a closed-form expression as presented in Proposition 2.3.2. Based on this result, the likelihood ratio statistic can be applied for testing hypothesis pairs involving a particular cointegration rank of the system. For the hypothesis pair

$$H_0 : \operatorname{rk} \mathbf{\Pi} = r_0 \quad \text{vs.} \quad H_1 : r_0 < \operatorname{rk} \mathbf{\Pi} \leq r_1, \tag{2.40}$$

the likelihood ratio statistic is given by

$$\begin{aligned} \lambda_{\mathrm{LR}}(r_0,r_1) &= 2\,(\ln L(r_1) - \ln L(r_0)) \\ &\overset{\text{eq. (2.38)}}{=} T\left(-\sum_{i=1}^{r_1} \ln(1-\lambda_i) + \sum_{i=1}^{r_0} \ln(1-\lambda_i) \right) \\ &= -T \sum_{i=r_0+1}^{r_1} \ln(1-\lambda_i), \end{aligned}$$

where $\ln L(r_i)$ denotes the maximum of the log-likelihood function using the cointegration rank r_i. The statistic $\lambda_{\mathrm{LR}}(r_0,r_1)$ is not normal, i.e., the limiting null distribution is not χ^2 and it depends on the number of common trends, $K - r_0$, and the type of the deterministic term included in the VECM, see Lütkepohl (2007).

Johansen Trace & Maximum Eigenvalue Test

Two particular choices for the alternative hypothesis in eq. (2.40) are commonly known as Johansen cointegration tests. First, the LR statistic $\lambda_{\mathrm{LR}}(r_0,K)$ for testing the hypothesis pair

$$H_0 : \operatorname{rk} \mathbf{\Pi} = r_0 \quad \text{vs.} \quad H_1 : r_0 < \operatorname{rk} \mathbf{\Pi} \leq K,$$

that is, testing that there are at most r_0 cointegration relations, is also known as the *trace statistic*. The trace statistic is given by

$$\lambda_{\mathrm{LR}}(r_0,K) = -T \sum_{i=r_0+1}^{K} \ln(1-\lambda_i), \tag{2.41}$$

where $\lambda_{r_0+1} > \ldots > \lambda_K$ are the $K - r_0$ smallest eigenvalues of the generalized eigenvalue equation eq. (2.33). The statistic $\lambda_{\mathrm{LR}}(r_0,r_0+1)$ for checking the existence of r_0 against r_0+1 cointegration relation by the hypothesis pair

$$H_0 : \operatorname{rk} \mathbf{\Pi} = r_0 \quad \text{vs.} \quad H_1 : \operatorname{rk} \mathbf{\Pi} = r_0 + 1,$$

is referred to as the *maximum eigenvalue statistic* and takes the form

$$\lambda_{\mathrm{LR}}(r_0,r_0+1) = -T \ln(1-\lambda_{r_0+1}). \tag{2.42}$$

The maximum eigenvalue statistic has been proposed by Johansen and Juselius (1990). The limiting distributions of the trace statistic and the maximum eigenvalue statistic under the null hypothesis have been studied by Johansen (1988, 1995) and Johansen and Juselius (1990). In the simplest case, where the VECM does not include any deterministic terms, the test statistics of eqs. (2.41) and (2.42) have the following asymptotics (Lütkepohl, 2007)

$$\lambda_{\mathrm{LR}}(r_0,K) \xrightarrow{d} \operatorname{tr}(\mathcal{W})$$

and

$$\lambda_{\mathrm{LR}}(r_0,r_0+1) \xrightarrow{d} \lambda_{\max}(\mathcal{W}), \tag{2.43}$$

where $\mathcal{W}$ is defined by

$$\mathcal{W} := \left(\int_0^1 \boldsymbol{W}\,d\boldsymbol{W}'\right)' \left(\int_0^1 \boldsymbol{W}\boldsymbol{W}'dt\right)^{-1} \left(\int_0^1 \boldsymbol{W}\,d\boldsymbol{W}'\right),$$

with $\boldsymbol{W}$ being a $(K-r_0)$-dimensional standard Brownian motion process and $\lambda_{\max}(\mathcal{W})$ in eq. (2.43) denotes the maximum eigenvalue of $\mathcal{W}$. The asymptotic distributions of both test statistics are also available for more general VECMs with deterministic terms. Critical values are obtained by simulation and can be found in Johansen and Juselius (1990, Appendix A). The procedure for the determination of the cointegration rank consists of a sequence of tests with null hypotheses given by

$$H_0 : \operatorname{rk}\boldsymbol{\Pi} = 0, \quad H_0 : \operatorname{rk}\boldsymbol{\Pi} = 1, \quad \ldots \quad H_0 : \operatorname{rk}\boldsymbol{\Pi} = K-1.$$

The testing procedure terminates at the first time when the test statistic is not significant (see Lütkepohl (2007)). The corresponding null hypothesis at the termination point determines the choice of the cointegration rank. This procedure is also described as the 'top → bottom' approach in Juselius (2006, Section 8.1).

2.3.6 | VAR/VECM Model Diagnostics

The following section covers some frequently used diagnostic tests which are employed on the residuals in order to detect model misspecification. These tests are based on the residuals $\hat{u}_t := y_t - \hat{y}_t$, with $t = 1,\ldots,T$, and include testing procedures for autocorrelation, conditional heteroscedasticity and non-normality. If the model is wrongly specified in terms of the lag order, deterministic components, or the cointegration rank, the residuals will significantly differ from a white noise process. Further discussions on misspecification tests can be found in Juselius (2006), Lütkepohl and Krätzig (2004) and Pfaff (2008).

Before introducing some formal tests, it should be noted that several graphical tools are also available for diagnostic purposes. These include, in particular, the *autocorrelation function* (ACF) and the *cross-correlation function* of the model residuals. The ACF is defined by $h \mapsto \operatorname{Corr}(\hat{u}_{i,t},\hat{u}_{i,t-h})$, with $i = 1,\ldots,K$, $h = 1,\ldots,h_{\max} > p$, and the cross-correlation function by $h \mapsto \operatorname{Corr}(\hat{u}_{i,t},\hat{u}_{j,t-h})$

for $i \neq j$. These graphical methods are helpful to detect model misspecification. As Lütkepohl and Krätzig (2004) point out, autocorrelation and cross-correlation functions can reveal remaining serial dependence in the residuals, while autocorrelations of squared residuals may detect conditional heteroscedasticity. The assessment, whether an estimated correlation coefficient is significant or not, is conducted using its standard error, which is asymptotically $1/\sqrt{T}$, see Juselius (2006). Thus, an unusual number of autocorrelation or cross-correlations outside the 95% confidence region, given by the band $\pm 1.96/\sqrt{T}$, are indications for model misspecification. The following introduction to formal residual tests is based on Lütkepohl and Krätzig (2004).

Portmanteau Test for Residual Autocorrelation

The potential presence of residual autocorrelation can be assessed by the *portmanteau* or *adjusted portmanteau statistic*. The portmanteau test checks the null hypothesis of no residual autocorrelation up to lag h, which can be stated as

$$H_0 : \mathbb{E}[\hat{u}_t \hat{u}_{t-i}'] = 0, \qquad i = 1, \ldots, h,$$

versus the alternative that at least one autocorrelation up to lag h differs from zero. The test statistic is given by

$$Q_h = T \sum_{i=1}^{h} \operatorname{tr}(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}), \tag{2.44}$$

where

$$\hat{C}_i = T^{-1} \sum_{t=i+1}^{T} \hat{u}_t \hat{u}_{t-i}'$$

is the empirical autocovariance matrix. The limiting distribution of the statistic Q_h under the null hypothesis is the χ^2 distribution with $K^2(h-p)$ degrees of freedom, where K is the dimension of the time series and p is the order of the VAR process. The number of degrees of freedom of the limiting χ^2 distribution depends on the number of free model parameters. Thus, for a VECM, the number of independent parameters reduces due to a rank restriction to $hK^2 - K^2(p-1) - Kr$, where r is the cointegration rank, see Lütkepohl (2007). Note that deterministic terms in the VAR/VECM must also be considered for the calculation of the degrees of freedom. A modified version of the portmanteau statistic of eq. (2.44), referred to as the *adjusted portmanteau statistic*, is defined by

$$Q_h^* = T^2 \sum_{i=1}^{h} (T-i)^{-1} \operatorname{tr}(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1}).$$

This adjusted statistic is potentially more suitable for small sample sizes.

Breusch-Godfrey Test for Residual Autocorrelation

An alternative test for residual autocorrelation, the so-called *Breusch-Godfrey test*, is a Lagrange multiplier type test proposed by Breusch (1978) and Godfrey (1978). The test for residual autocorrelation up to order h is based on the auxiliary regression

$$\hat{u}_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + CD_t + B_1 \hat{u}_{t-1} + \cdots + B_h \hat{u}_{t-h} + e_t$$

for a VAR model, and in analogy on the regression

$$\hat{u}_t = \boldsymbol{\alpha\beta}' y_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p} + CD_t + B_1 \hat{u}_{t-1} + \cdots + B_h \hat{u}_{t-h} + e_t$$

for a VECM. The test checks the hypothesis pair

$$H_0 : B_1 = \cdots = B_h = 0 \qquad \text{vs.} \qquad H_1 : \exists B_i \neq 0 \text{ for } i = 1, \ldots, h.$$

The Breusch-Godfrey test statistic is defined as

$$\lambda_{\text{LM}}(h) = T\left[K - \operatorname{tr}\left(\tilde{\Sigma}_e \tilde{\Sigma}_R^{-1}\right)\right],$$

where $\tilde{\Sigma}_e$ and $\tilde{\Sigma}_R$ denote the estimated covariance matrices of the auxiliary regressions for the unrestricted case and for the case with imposed restrictions $B_1 = \cdots = B_h = 0$, i.e.,

$$\tilde{\Sigma}_e = \frac{1}{T} \sum_{t=1}^{T} \hat{e}_t \hat{e}_t' \quad \text{and} \quad \tilde{\Sigma}_R = \frac{1}{T} \sum_{t=1}^{T} \hat{e}_t^R \hat{e}_t^{R\prime}.$$

The limiting distribution of the $\lambda_{\text{LM}}(h)$ under the null hypothesis is $\chi^2(hK^2)$. As Lütkepohl (2007) points out, the Breusch-Godfrey test is useful, in particular, for testing low order residual autocorrelation, where the χ^2 approximation of the portmanteau statistic might be insufficient. On the other hand, a portmanteau test is superior for larger h. Thus, usually both tests are used complementary to check the residuals for autocorrelation.

Non-normality Tests for Residuals

The following non-normality tests are based on skewness and kurtosis of the standardized residuals. Let $\tilde{\Sigma}_u$ denote the empirical residual covariance matrix

$$\tilde{\Sigma}_u = \frac{1}{T} \sum_{t=1}^{T} (\hat{u}_t - \bar{\hat{u}})(\hat{u}_t - \bar{\hat{u}})',$$

where $\bar{\hat{u}} := T^{-1} \sum_{t=1}^{T} \hat{u}_t$ is the residual mean, then the standardized residuals are obtained by

$$\hat{u}_t^s = (\hat{u}_{1,t}^s, \ldots, \hat{u}_{K,t}^s)' = \tilde{\Sigma}_u^{-\frac{1}{2}} (\hat{u} - \bar{\hat{u}}).$$

Let the following auxiliary quantities $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ be defined as

$$\boldsymbol{b}_1 = (b_{11}, \ldots, b_{1K}) \quad \text{with} \quad b_{1k} = \frac{1}{T} \sum_{t=1}^{T} \left(\hat{u}_{kt}^s\right)^3$$

and

$$\boldsymbol{b}_2 = (b_{21}, \ldots, b_{2K}) \quad \text{with} \quad b_{2k} = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{u}_{kt}^{s}\right)^4,$$

then it can be shown, see, e.g., Lütkepohl (2007, Proposition 4.9) that b_1 and b_2 are asymptotically independent and normally distributed, i.e.,

$$\sqrt{T}\begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 - \mathbf{3}_K \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} 6\,\mathbb{1}_K & 0 \\ 0 & 24\,\mathbb{1}_K \end{bmatrix}\right),$$

where $\mathbf{3}_K := (3, \ldots, 3)'$. This result implies that

$$s_3^2 := \frac{T}{6}\boldsymbol{b}_1'\boldsymbol{b}_1 \xrightarrow{d} \chi^2(K)$$

and

$$s_4^2 := \frac{T}{24}(\boldsymbol{b}_2 - \mathbf{3}_K)'(\boldsymbol{b}_2 - \mathbf{3}_K) \xrightarrow{d} \chi^2(K).$$

The statistics s_3^2 and s_4^2 can be used to check the hypothesis pairs

$$H_0 : \mathbb{E}[(\hat{u}^s)^3] = 0 \qquad \text{vs.} \qquad H_1 : \mathbb{E}[(\hat{u}^s)^3] \neq 0 \tag{2.45}$$

and

$$H_0 : \mathbb{E}[(\hat{u}^s)^4] = \mathbf{3}_K \qquad \text{vs.} \qquad H_1 : \mathbb{E}[(\hat{u}^s)^4] \neq \mathbf{3}_K. \tag{2.46}$$

A joint test for the null hypotheses of eqs. (2.45) and (2.46) can be checked by the statistic

$$\lambda_{\text{LJB}_{\text{multi}}} := s_3^2 + s_4^2,$$

which is a multivariate generalization of the univariate *Lomnicki-Jarque-Bera test* for non-normality, proposed by Jarque and Bera (1987) and Lomnicki (1961). Using the asymptotic properties of s_3^2 and s_4^2 we can conclude that

$$\lambda_{\text{LJB}_{\text{multi}}} \xrightarrow{d} \chi^2(2K).$$

In the upcoming case study in Section 2.4, we will report the multivariate test statistics s_3^2, s_4^2 and $\lambda_{\text{LJB}_{\text{multi}}}$ as well as the univariate statistic λ_{LJB} for each time series component.

(M)ARCH Test for Residual Heteroscedasticity

Some frequently used tests for conditional heteroscedasticity in the residuals are the Lagrange multiplier type tests, denoted by ARCH in the univariate case, and its generalization to a multivariate conditional heteroscedasticity test, denoted by MARCH. Following Lütkepohl and Krätzig (2004), the multivariate q-th order test is based on the regression

$$\operatorname{vech}(\hat{u}_t\hat{u}_t') = \beta_0 + B_1 \operatorname{vech}(\hat{u}_{t-1}\hat{u}_{t-1}') + \cdots + B_q \operatorname{vech}(\hat{u}_{t-q}\hat{u}_{t-q}') + \varepsilon_t, \tag{2.47}$$

where $\operatorname{vech}(\cdot)$ is the *half-vectorization operator* which converts a symmetric matrix to a vector by vectorizing only the lower triangular part of the matrix. The corresponding coefficient matrices B_j for $j = 1, \ldots, q$ are therefore $({}^1\!/\!{}_2K(K+1) \times {}^1\!/\!{}_2K(K+1))$-dimensional. Note that in the univariate case, the regression of eq. (2.47) reduces to

$$\hat{u}_t^2 = \beta_0 + B_1\hat{u}_{t-1}^2 + \cdots + B_q\hat{u}_{t-q}^2 + \varepsilon_t,$$

The MARCH test checks the hypothesis pair

$$H_0 : B_1 = \cdots = B_q = 0 \qquad \text{vs.} \qquad H_1 : \exists B_i \neq 0 \quad \text{for } i = 1, \ldots, q$$

through the multivariate LM statistic

$$\mathrm{MARCH_{LM}}(q) := \frac{1}{2}TK(K+1) - T\operatorname{tr}(\hat{\Sigma}_{\text{vech}}\hat{\Sigma}_0^{-1}), \tag{2.48}$$

where $\hat{\Sigma}_{\text{vech}}$ is the residual covariance matrix obtained by the regression of eq. (2.47) and $\hat{\Sigma}_0$ is the alternative residual covariance matrix obtained by using the regression with $q = 0$. Under the null hypothesis of no conditional heteroscedasticity, the statistic has an asymptotic $\chi^2(qK^2(K+1)^2/4)$ distribution, see Lütkepohl (2007). For the univariate case, the statistic of eq. (2.48) reduces to

$$\mathrm{ARCH_{LM}}(q) := T\left(1 - \frac{\hat{\sigma}}{\hat{\sigma}_0}\right).$$

The limiting distribution of the $\mathrm{ARCH_{LM}}(q)$ statistic is consequently given by the $\chi^2(q)$ distribution, see Engle (1982).

2.4 | VECM Specification for the Kannisto Model

In the following, we return to the Kannisto model as specified in Section 1.9.1. We investigate whether the presence of cointegration relations among the multivariate time series, given by Kannisto model estimates, can be confirmed as it was suggested in Section 1.9.5. In the subsequent sections, we will analyse all three proposed models KAN, KAN:2, and KAN:3 following the specification procedure described in Remark 2.3.1.

2.4.1 | VECM Specification for the KAN Model

The basis for the upcoming analysis is the 2-dimensional time series of the KAN model parameters estimated in Section 1.9.2. The trajectories of the KAN coefficients $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, as well as their first-order differences $\Delta\kappa_t^{(1)}$ and $\Delta\kappa_t^{(2)}$, are provided in Figure 2.1. Visual inspection of the time series indicates a structural change of the trajectories in the 1940's. This period corresponds to the end of the Second World War that brought significant improvements of mortality rates. That also applies to the non-directly involved Swedish female population, as discussed in Section 1.3 and illustrated in Figure 1.4. For all following analyses, we will therefore only consider the corresponding time series in the period between 1946 and 2014.

Lag Order Selection for the KAN Model

Following the specification procedure of Remark 2.3.1, we start by selecting the lag order p of an unrestricted VAR(p) model by using the information criteria from Section 2.3.1. Table 2.1 shows the results for an unrestricted VAR model with both a constant and a deterministic trend and a maximum lag order of $p = 7$. For the criteria AIC, HQ, SC, and FPE see eqs. (2.6) to (2.9).

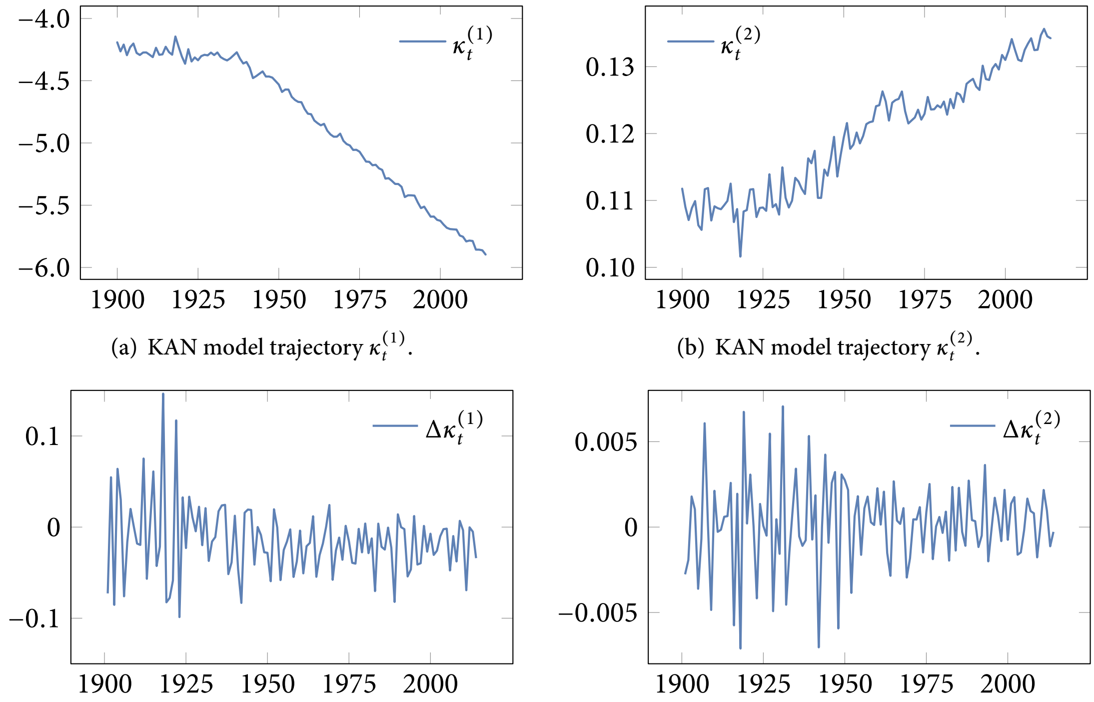
According to the AIC and FPE criteria, the optimal VAR lag is $p = 3$. On the other hand, the HQ and SC criteria suggest the optimal lag order to be $p = 1$. Note that choosing the lag order according to information criteria does not necessarily imply that the residuals of that model will pass the standard diagnostic tests. Using the methods of Section 2.3.6, we detect significant autocorrelations in the residual of the VAR(1) model and therefore choose $p = 3$ as the VAR lag order as suggested by the AIC and FPE criteria.

Unit Root Tests for the KAN Model

Next, we investigate whether the DGP of the underlying KAN time series contains unit roots or otherwise can be considered as stationary. The upper panel of Figure 2.1 suggests that based on visual judgment $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ do not have a stationary DGP. Nevertheless, they could have been generated by a trend-stationary DGP. Formal tests, as introduced in Section 2.3.2, will be considered to assess stationarity of those time series.

The objective of the forthcoming testing procedure is to confirm that both time series $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ have an $I(1)$ DGP. First, we apply the ADF test on each component of $\boldsymbol{\kappa}_t$ to test whether we can reject the unit null hypothesis. Additionally, the KPSS test is applied to check stationarity as the null hypothesis. If the null hypotheses of the ADF tests cannot be rejected and the null hypotheses of the conformational KPSS tests are rejected, then both tests are reapplied to individual components of the first-order differences $\Delta\boldsymbol{\kappa}_t$ in order to confirm that the unit root null hypothesis can be rejected, while the stationarity null hypothesis cannot. Since the trajectories of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ have a drift, we include a constant and a linear trend term in the ADF regression (see eq. (2.10)) and use the t-type τ_3 and the F-type ϕ_3 statistics to test the hypothesis pairs of eqs. (2.11) and (2.12). The corresponding KPSS tests employ the statistic η_τ to check trend-stationarity. The testing procedure continues with the first-order differences. Due to the absence of a drift, the ADF t-type τ_2 and the F-type ϕ_1 statistics are employed for testing the presence of a unit root and the KPSS statistic η_μ for checking stationarity. The ADF and KPSS tests are evaluated in **R** using the **urca** package (Pfaff, Zivot and Stigler, 2016). The results are summarized in Table 2.2. The second column of the table contains the information about the deterministic term used for the corresponding test and the third column either contains the number of lagged differences k in eq. (2.10), or the length l of the spectral window in the Barlett weighting function, as given in eq. (2.13).

The ADF test results imply that we cannot reject the null hypothesis of a unit root for $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. The KPSS test results of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, on the other hand, are significant at the 5% level. The results of the first-order differences $\Delta\kappa_t^{(1)}$ and $\Delta\kappa_t^{(2)}$ show that the ADF test suggests to reject the



(a) KAN model trajectory $\kappa_t^{(1)}$.

(b) KAN model trajectory $\kappa_t^{(2)}$.

(c) First difference $\Delta\kappa_t^{(1)} = \kappa_t^{(1)} - \kappa_{t-1}^{(1)}$.

(d) First difference $\Delta\kappa_t^{(2)} = \kappa_t^{(2)} - \kappa_{t-1}^{(2)}$.

Figure 2.1: Trajectories of the KAN model estimates and the corresponding first-order differences.

Table 2.1: Information criteria based VAR lag order selection for the KAN model.

lags p	AIC(p)	HQ(p)	SC(p)	FPE(p)
1	−21.049	**−20.941**	**−20.775**	$7.22 \cdot 10^{-10}$
2	−21.066	−20.904	−20.654	$7.11 \cdot 10^{-10}$
3	**−21.081**	−20.865	−20.532	$\mathbf{7.01 \cdot 10^{-10}}$
4	−21.039	−20.769	−20.353	$7.34 \cdot 10^{-10}$
5	−21.067	−20.743	−20.243	$7.16 \cdot 10^{-10}$
6	−21.062	−20.685	−20.102	$7.24 \cdot 10^{-10}$
7	−21.065	−20.634	−19.967	$7.28 \cdot 10^{-10}$

null hypothesis of unit root for both time series. Furthermore, the stationarity null of the KPSS test is accepted at the 5% level. Thus, based on that testing procedure, we can conclude that both time series are integrated of order one.

Cointegration Tests for the KAN Model

Stationarity tests of the previous analysis confirmed that both components of $\boldsymbol{\kappa}_t$ are integrated of order one. Next, we will conduct Johansen cointegration tests to determine whether a VAR model for $\Delta\boldsymbol{\kappa}_t$ is appropriate, if no cointegration relation exist, or otherwise a VECM can be applied to $\boldsymbol{\kappa}_t$, if a long-run equilibrium relation can be confirmed. These cases are outlined in (3.c) and (3.b) of

Table 2.2: ADF and KPSS tests for the components of $\boldsymbol{\kappa}_t$ and $\Delta\boldsymbol{\kappa}_t$ of the KAN model.

time series	deterministic trend	lags / length	statistic	test value	critical values		
					1%	5%	10%
$\kappa_t^{(1)}$	constant, trend	$k = 3$	ADF : τ_3	−2.39	−4.04	−3.45	−3.15
$\kappa_t^{(1)}$	constant, trend	$k = 3$	ADF : ϕ_3	3.66	8.73	6.49	5.47
$\kappa_t^{(1)}$	constant, trend	$l = 4$	KPSS : η_τ	0.168	0.216	0.146	0.119
$\kappa_t^{(2)}$	constant, trend	$k = 3$	ADF : τ_3	−2.15	−4.04	−3.45	−3.15
$\kappa_t^{(2)}$	constant, trend	$k = 3$	ADF : ϕ_3	2.34	8.73	6.49	5.47
$\kappa_t^{(2)}$	constant, trend	$l = 4$	KPSS : η_τ	0.159	0.216	0.146	0.119
$\Delta\kappa_t^{(1)}$	constant	$k = 3$	ADF : τ_2	−7.43	−3.51	−2.89	−2.58
$\Delta\kappa_t^{(1)}$	constant	$k = 3$	ADF : ϕ_1	27.60	6.70	4.71	3.86
$\Delta\kappa_t^{(1)}$	constant	$l = 4$	KPSS : η_μ	0.075	0.739	0.463	0.347
$\Delta\kappa_t^{(2)}$	constant	$k = 3$	ADF : τ_2	−5.93	−3.51	−2.89	−2.58
$\Delta\kappa_t^{(2)}$	constant	$k = 3$	ADF : ϕ_1	17.61	6.70	4.71	3.86
$\Delta\kappa_t^{(2)}$	constant	$l = 4$	KPSS : η_μ	0.064	0.739	0.463	0.347

Remark 2.3.1.

Tables 2.3 and 2.4 report the results of the Johansen cointegration tests, as introduced in Section 2.3.5, with the trace statistic given in eq. (2.41) and the maximum eigenvalue statistic in eq. (2.42), respectively. Recall that the procedure of cointegration rank selection involves a series of Johansen tests until a rejection of the null hypothesis arises for the first time. Since the corresponding time series has the dimension $K = 2$, there are 2 null hypotheses to be checked here. As for the unit root test, we use the **urca** package for the cointegration rank tests. Critical values used by this implementation have been taken from Osterwald-Lenum (1992). An alternative source for critical values can also be found in MacKinnon, Haug and Michelis (1999). The LR type trace and the maximum eigenvalue statistics are obtained by estimating pairs of restricted VAR(p) models using Proposition 2.3.2. For that estimation, we use the suggested lag order of $p = 3$ and a deterministic term containing a trend ξ_{co} restricted only to appear in the cointegrating relations and an unrestricted constant $\boldsymbol{\xi}$, i.e.,

$$\begin{aligned}\Delta\boldsymbol{\kappa}_t &= \boldsymbol{\alpha}[\boldsymbol{\beta}' : \xi_{\text{co}}]\begin{bmatrix}\boldsymbol{\kappa}_{t-1}\\ t-1\end{bmatrix} + \boldsymbol{\Gamma}_1\Delta\boldsymbol{\kappa}_{t-1} + \boldsymbol{\Gamma}_2\Delta\boldsymbol{\kappa}_{t-2} + \boldsymbol{\xi} + \boldsymbol{u}_t,\\ &= \boldsymbol{\Pi}^+\boldsymbol{\kappa}_{t-1}^+ + \boldsymbol{\Gamma}_1\Delta\boldsymbol{\kappa}_{t-1} + \boldsymbol{\Gamma}_{p-1}\Delta\boldsymbol{\kappa}_{t-2} + \boldsymbol{\xi} + \boldsymbol{u}_t.\end{aligned}$$

It should be noted here that an unrestricted linear trend would allow a quadratic growth of the variables. However, the above specification allows a linear trend in the cointegration relation as well as in the variables. For a detailed discussion on the selection of deterministic terms and the implication of their restrictions, we refer to Juselius (2006).

The trace test shows that the null hypothesis of no cointegration relation, i.e., $r = 0$, is rejected at the 5% level. The maximum eigenvalue test confirms this result by also rejecting the null hypothesis of $r = 0$ at the 5% level. For $K = 2$ and $r = 1$, both statistics coincide and accept the null hypothesis $r = 1$. Conclusively, there is statistical evidence that both periodic components of the KAN model $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, which shape the mortality curve in the period t, have a long-run equilibrium relation. As outlined in Remark 2.3.1, the appropriate model for this system of variables is a VECM of order 2, which will be estimated in the following section.

VECM Estimation for the KAN Model

In the following, the maximum likelihood estimator of Proposition 2.3.2 is used to estimate a VECM of the form

$$\Delta\boldsymbol{\kappa}_t = \boldsymbol{\alpha}[\boldsymbol{\beta}' : \xi_{co}]\begin{bmatrix}\boldsymbol{\kappa}_{t-1}\\ t-1\end{bmatrix} + \boldsymbol{\Gamma}_1\Delta\boldsymbol{\kappa}_{t-1} + \boldsymbol{\Gamma}_2\Delta\boldsymbol{\kappa}_{t-2} + \boldsymbol{\xi} + \boldsymbol{u}_t, \tag{2.49}$$

where $\boldsymbol{\alpha}$ denotes the loading vector, $\boldsymbol{\beta}$ the cointegration vector, ξ_{co} the linear trend parameter in the cointegration relation, and $\boldsymbol{\xi}$ is the parameter vector corresponding to a constant term. $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are the coefficient matrices of the VECM, and $\boldsymbol{u}_t$ is a Gaussian white noise process with covariance matrix Σ_u. Using the KAN time series $\boldsymbol{\kappa}_t$ for $t = 1946, \ldots, 2014$, and a VECM of the form eq. (2.49) yields the following estimates

$$\begin{aligned}\begin{bmatrix}\Delta\kappa_{1,t}\\ \Delta\kappa_{2,t}\end{bmatrix} &= \begin{bmatrix}\underset{(\mathbf{-3.569})}{-0.359}\\ \underset{(\mathbf{2.965})}{0.022}\end{bmatrix}\begin{bmatrix}1.00 : & \underset{(-1.713)}{-6.805} & \underset{(\mathbf{22.922})}{0.023}\end{bmatrix}\begin{bmatrix}\kappa_{1,t-1}\\ \kappa_{2,t-1}\\ t-1\end{bmatrix} + \begin{bmatrix}\underset{(\mathbf{-2.379})}{-0.311} & \underset{(\mathbf{-2.907})}{-4.835}\\ \underset{(-1.057)}{-0.010} & \underset{(-1.856)}{-0.229}\end{bmatrix}\begin{bmatrix}\Delta\kappa_{1,t-1}\\ \Delta\kappa_{2,t-1}\end{bmatrix}\\ &+ \begin{bmatrix}\underset{(\mathbf{-2.146})}{-0.255} & \underset{(-0.577)}{-0.919}\\ \underset{(-1.575)}{-0.014} & \underset{(\mathbf{-2.418})}{-0.286}\end{bmatrix}\begin{bmatrix}\Delta\kappa_{1,t-2}\\ \Delta\kappa_{2,t-2}\end{bmatrix} + \begin{bmatrix}\underset{(\mathbf{-3.644})}{-1.919}\\ \underset{(\mathbf{2.975})}{0.116}\end{bmatrix} + \begin{bmatrix}\hat{u}_{1,t}\\ \hat{u}_{2,t}\end{bmatrix}.\end{aligned} \tag{2.50}$$

Table 2.3: Trace test for cointegration rank of the KAN model.

H_0	H_1	trace statistic	critical values		
			10%	5%	1%
$r = 0$	$0 < r \leq 2$	28.07	22.76	25.32	30.45
$r = 1$	$1 < r \leq 2$	9.07	10.49	12.25	16.26

Table 2.4: Maximum eigenvalue test for cointegration rank of the KAN model.

H_0	H_1	max eig. statistic	critical values		
			10%	5%	1%
$r = 0$	$r = 1$	19.00	16.85	18.96	23.65
$r = 1$	$r = 2$	9.07	10.49	12.25	16.26

The t-values of the coefficient estimates are given in parentheses. Coefficient estimates which are significant at the 5% level are denoted in bold. Equation (2.50) shows that all VECM matrices have significant coefficients. Thus, it is unlikely that a model reduction to a lower lag order, as suggested by the HQ or SC information criteria, is possible. A VAR representation of the VECM form in eq. (2.50) is given by

$$\begin{aligned}\boldsymbol{\kappa}_t &= A_1\boldsymbol{\kappa}_{t-1} + A_2\boldsymbol{\kappa}_{t-2} + A_3\boldsymbol{\kappa}_{t-2} + CD_t + \hat{\boldsymbol{u}}_t \\ &= \begin{bmatrix} 0.330 & -2.392 \\ 0.012 & 0.620 \end{bmatrix}\begin{bmatrix} \kappa_{1,t-1} \\ \kappa_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0.056 & 3.915 \\ -0.004 & -0.057 \end{bmatrix}\begin{bmatrix} \kappa_{1,t-2} \\ \kappa_{2,t-2} \end{bmatrix} \\ &+ \begin{bmatrix} 0.255 & 0.919 \\ 0.014 & 0.286 \end{bmatrix}\begin{bmatrix} \kappa_{1,t-3} \\ \kappa_{2,t-3} \end{bmatrix} + \begin{bmatrix} -1.919 & -0.008 \\ 0.0116 & 0.001 \end{bmatrix}\begin{bmatrix} 1 \\ t-1 \end{bmatrix} + \begin{bmatrix} \hat{u}_{1,t} \\ \hat{u}_{2,t} \end{bmatrix},\end{aligned} \tag{2.51}$$

and is obtained using the transformations $A_1 = \boldsymbol{\Pi} + \boldsymbol{\Gamma}_1 + \mathbb{1}_2$, $A_2 = \boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_1$, and $A_3 = -\boldsymbol{\Gamma}_2$. The coefficient matrix of $D_t = (1, t-1)'$ is given by $C = \left[\boldsymbol{\xi} : \boldsymbol{\alpha}\xi_{\text{co}}\right]$. The estimated residual covariance and correlation matrices can be obtained using the estimator of eq. (2.26) which leads to

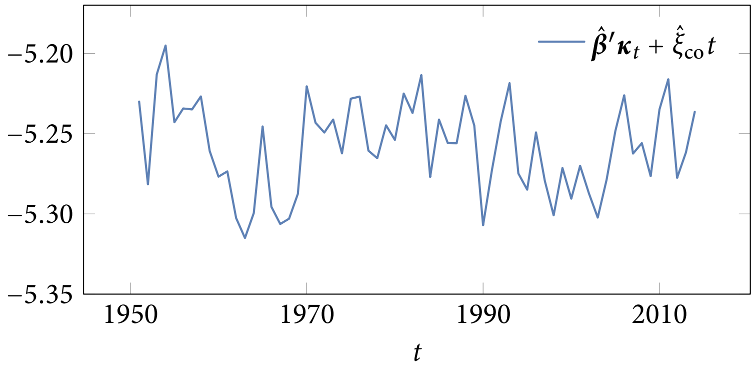
$$\widetilde{\Sigma}_u = \begin{bmatrix} 3.357 & -0.117 \\ -0.117 & 0.018 \end{bmatrix} \times 10^{-4} \qquad \widetilde{\text{Corr}}(\boldsymbol{u}_t) = \begin{bmatrix} 1 & -0.471 \\ -0.471 & 1 \end{bmatrix}. \tag{2.52}$$

In Figure 2.2(a), we illustrate the estimated long-run equilibrium relation $\hat{\boldsymbol{\beta}}'\boldsymbol{\kappa}_t + \hat{\xi}_{\text{co}}t$ of the KAN model coefficients. Recall that the existence of cointegration relations implies that while the DGPs of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are non-stationary unit root processes, there exists a linear combination of them, such that the resulting process is stationary. Alternatively, by omitting the drift term $\xi_{\text{co}}t$ in the cointegration relation, we illustrate the trend stationarity of $\hat{\boldsymbol{\beta}}'\boldsymbol{\kappa}_t$ in Figure 2.2(b).

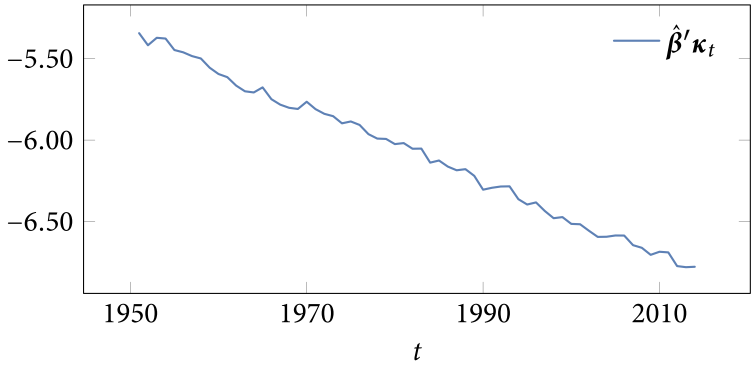
VECM Validation for the KAN Model

In this section, the diagnostic tests from Section 2.3.6 are performed to check the residuals for autocorrelation, non-normality and heteroscedasticity. We begin the analysis by visual inspection of the residuals of the estimation as given by eq. (2.50). The standardized residuals, residual autocorrelations, cross-correlations, and Gaussian kernel density estimators for the residuals are illustrated in Figure 2.3. The ACF and the cross-correlation plots show that although most correlations are below the significance bounds of $\pm 1.96/\sqrt{T}$ there are few correlations outside the bounds which might cause problems in formal validation tests. Notice that the sample size for the time series from 1946 to 2014 is $T = 66$ since 3 presample values are used in the VECM(2)/VAR(3) model. The correlation significance bounds are therefore given by ± 0.2413. Further inspection of the autocorrelations of the squared residuals gives no indication of ARCH effects. From Figure 2.3(g), we can observe by the Gaussian kernel density estimators that the dispersion of both standardized residuals is slightly higher than the standard normal distribution. This potential indication of non-normality will be assessed by multivariate/univariate Lomnicki-Jarque-Bera tests.

Table 2.5 summarizes the results of employed diagnostic tests for the VECM(2) model. The table contains the test values and the p-values, which are derived according to the corresponding χ^2 limiting distributions. Recall from Remark 1.6.2 that a test will reject the null hypothesis if the p-value



(a) Level stationary equilibrium relation $\hat{\boldsymbol{\beta}}'\boldsymbol{\kappa}_t + \hat{\xi}_{co}t$ of the KAN model.



(b) Trend stationary equilibrium relation $\hat{\boldsymbol{\beta}}'\boldsymbol{\kappa}_t$ of the KAN model.

Figure 2.2: Estimated equilibrium relation of the KAN model.

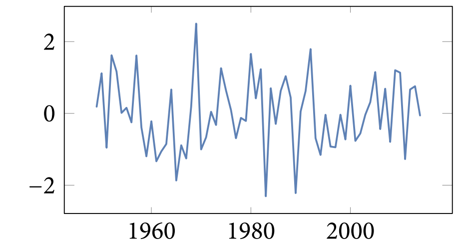
is smaller than the chosen significance level. Since all p-values in Table 2.5 are above 0.05, none of the null hypotheses is rejected at the 5% level. The absence of autocorrelations in the residuals is accepted by the Portmanteau test, by checking high order autocorrelations, as well as, by the Breusch-Godfrey test for low order autocorrelations. Normality of the residuals is also accepted using multivariate and univariate tests. According to the results of the MARCH/ARCH tests, there is no indication for the presence of ARCH effects in the residuals.

In summary, none of the autocorrelation, non-normality or heteroscedasticity tests indicates problems with the VECM specified in eq. (2.49). We can conclude that the presented model provides an appropriate representation of the DGP of the KAN model coefficients and thus the estimated cointegration relation captures the long-run behaviour of the mortality structure.

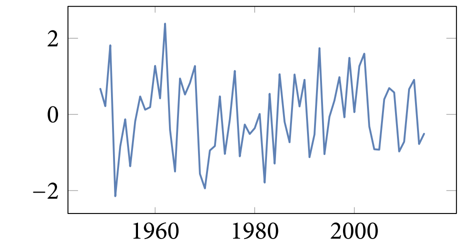
2.4.2 | VECM Specification for the KAN:2 Model

In this section, we proceed our analysis of the multivariate time series given by the KAN:2 model as presented in Section 1.9.2. Recall that the predictor function of the KAN:2 model

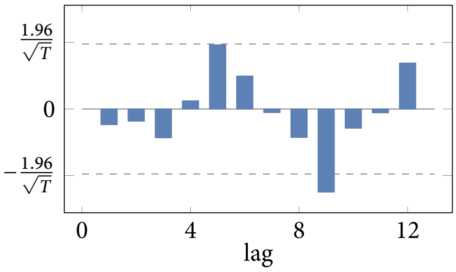
$$\eta_{t,x} = \sum_{i=1}^{3}(x - x_{\min})^{i-1}\kappa_t^{(i)}$$



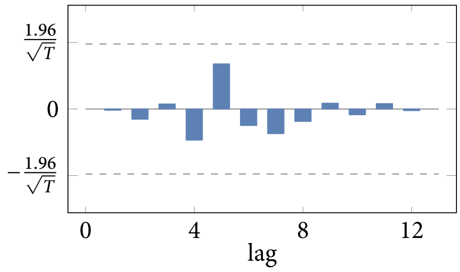
(a) standardized residuals $u_{1,t}$



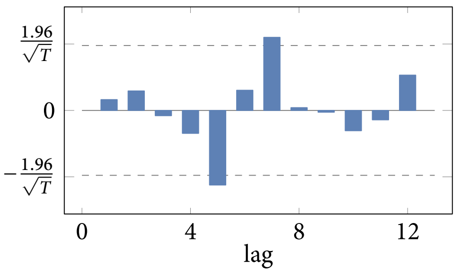
(b) standardized residuals $u_{2,t}$



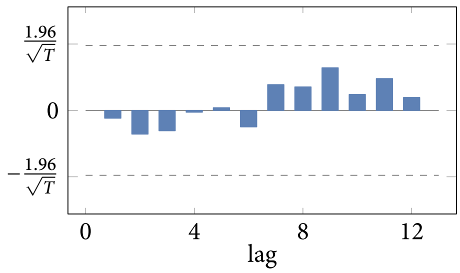
(c) ACF of residuals $u_{1,t}$



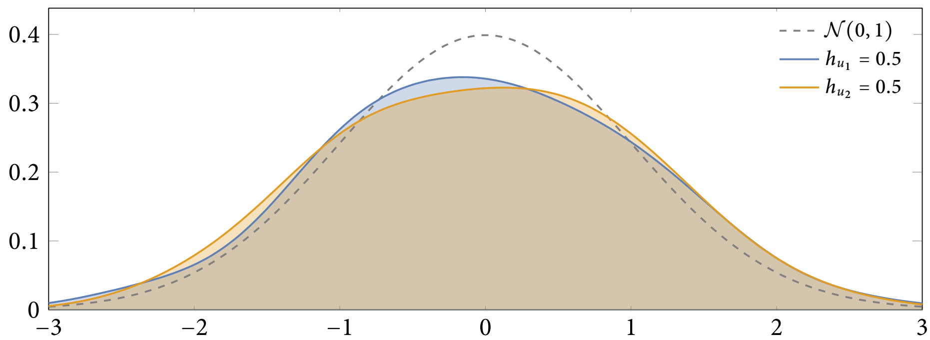
(d) ACF of residuals $u_{2,t}$



(e) Cross-correlation of $u_{1,t}$ against $u_{2,t}$



(f) Cross-correlation of $u_{2,t}$ against $u_{1,t}$



(g) Gaussian kernel density estimators of the residuals $u_{1,t}$ with selected bandwidth of $h_{u_{1,t}} = 0.5$ and with bandwidth $h_{u_{2,t}} = 0.5$ for $u_{2,t}$.

Figure 2.3: Residual autocorrelation, cross-correlations and Gaussian kernel density estimators.

Table 2.5: Diagnostics of the VECM for the KAN time series.

diagnostic type	test name	test statistic	test value	appr. dist.	p-value
autocorrelation	Portmanteau	Q_{16}	59.585	$\chi^2(54)$	0.280
	adjusted Portmanteau	Q^*_{16}	69.276	$\chi^2(54)$	0.079
	Breusch-Godfrey	$\lambda_{LM}(5)$	28.069	$\chi^2(20)$	0.108
non-normality	multivariate	$\lambda_{LJB_{multi}}$	0.629	$\chi^2(4)$	0.960
	skewness only	s_3^2	0.094	$\chi^2(2)$	0.954
	kurtosis only	s_4^2	0.535	$\chi^2(2)$	0.765
	univariate u_1	λ_{LJB}	0.279	$\chi^2(2)$	0.870
	univariate u_2	λ_{LJB}	0.988	$\chi^2(2)$	0.610
heteroscedasticity	multivariate	$\text{MARCH}_{LM}(5)$	60.131	$\chi^2(45)$	0.065
	univariate u_1	$\text{ARCH}_{LM}(16)$	5.739	$\chi^2(16)$	0.991
	univariate u_2	$\text{ARCH}_{LM}(16)$	18.516	$\chi^2(16)$	0.295

has, compared to the KAN model, an additional quadratic term, such that there are altogether 3 coefficients, namely $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, and $\kappa_t^{(3)}$, which determine the mortality curve for the period t. Using the time series of the KAN:2 estimates, we again follow the specification procedure for the VECM/VAR models as described in Remark 2.3.1. For the plots of the KAN:2 trajectories, see Figure 1.37. The objective of the following is to analyse whether we can formally confirm the observation of long-run relations between the coefficients, as made in Section 1.9.5 and shown in Figure 1.42(b) on page 98.

Lag Order Selection for the KAN:2 Model

Analogous to Section 2.4.1, we start by selecting the lag order p of an unrestricted VAR(p) model, with a constant and a deterministic trend, by using the information criteria from Section 2.3.1. The results are presented in Table 2.6 and show that the AIC, HQ, and the FPE information criteria are minimized for $p = 3$. On the other hand, the SC criterion prefers the lag order $p = 1$. In this case, we will follow the suggestion made by the majority and choose the order $p = 3$ for the KAN:2 time series.

Table 2.6: Information criteria based VAR lag order selection for the KAN:2 model.

lags p	AIC(p)	HQ(p)	SC(p)	FPE(p)
1	−39.745	−39.543	**−39.230**	$5.49 \cdot 10^{-18}$
2	−39.819	−39.495	−38.995	$5.12 \cdot 10^{-18}$
3	**−40.043**	**−39.598**	−38.911	$\mathbf{4.12 \cdot 10^{-18}}$
4	−39.946	−39.380	−38.505	$4.59 \cdot 10^{-18}$
5	−39.907	−39.220	−38.157	$4.87 \cdot 10^{-18}$
6	−39.876	−39.068	−37.818	$5.17 \cdot 10^{-18}$
7	−39.781	−38.851	−37.414	$5.91 \cdot 10^{-18}$

Unit Root Tests for the KAN:2 Model

In this section, we investigate whether the underlying DGPs of the three KAN:2 components have unit roots or can be considered as stationary. From visual inspection of the trajectories as given in Figure 1.37 the DGP is unlikely to be stationary. The formal analysis by ADF and KPSS tests is shown in Table 2.7. The results show that the unit root null hypotheses of the ADF tests are accepted for all components of $\boldsymbol{\kappa}_t$ at the 5% level. Simultaneously, the KPSS null hypotheses of stationarity are rejected for all components of $\boldsymbol{\kappa}_t$ at the 5% level. Furthermore, we observe an opposite behaviour for the first-order differences $\Delta\boldsymbol{\kappa}_t$. The unit root null hypotheses are rejected by the ADF tests at the 5% level, and simultaneously, stationarity is accepted by the KPSS tests at the 5% level. Based on the results of this testing procedure, we can conclude that the underlying univariate DGPs of $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$ are integrated of order one.

Cointegration Tests for the KAN:2 Model

Unit root and stationarity tests from confirmed that components of $\boldsymbol{\kappa}_t$ are integrated of order one. Next, we will conduct Johansen cointegration tests to determine whether an unrestricted VAR or a VECM is better suited for the KAN:2 time series. The results of the Johansen trace and maximum eigenvalue test are reported in Tables 2.8 and 2.9. Applying the 'top → bottom' approach for the trace test, we find that the null hypothesis $r = 0$ is rejected, but the null hypothesis $r = 1$ is accepted at the 5% level. The same holds for the maximum eigenvalue test. The null hypothesis $r = 0$ is rejected at the 5% level in favour for the alternative $r = 1$, while at the next step the null hypothesis $r = 1$ is accepted at the 5% level. Thus, we conclude that there are at least two unit roots and at most one stationary relation.

Table 2.7: ADF and KPSS tests for the components of $\boldsymbol{\kappa}_t$ and $\Delta\boldsymbol{\kappa}_t$ of the KAN:2 model.

time series	deterministic trend	lags / length	statistic	test value	critical values		
					1%	5%	10%
$\kappa_t^{(1)}$	constant, trend	$k = 3$	ADF : τ_3	−2.63	−4.04	−3.45	−3.15
$\kappa_t^{(1)}$	constant, trend	$l = 2$	KPSS : η_τ	0.511	0.216	0.146	0.119
$\kappa_t^{(2)}$	constant, trend	$k = 3$	ADF : τ_3	−3.10	−4.04	−3.45	−3.15
$\kappa_t^{(2)}$	constant, trend	$l = 2$	KPSS : η_τ	0.444	0.216	0.146	0.119
$\kappa_t^{(3)}$	constant, trend	$k = 3$	ADF : τ_3	−2.83	−4.04	−3.45	−3.15
$\kappa_t^{(3)}$	constant, trend	$l = 2$	KPSS : η_τ	0.466	0.216	0.146	0.119
$\Delta\kappa_t^{(1)}$	constant	$k = 3$	ADF : τ_2	−3.53	−3.51	−2.89	−2.58
$\Delta\kappa_t^{(1)}$	constant	$l = 2$	KPSS : η_μ	0.329	0.739	0.463	0.347
$\Delta\kappa_t^{(2)}$	constant	$k = 3$	ADF : τ_2	−3.68	−3.51	−2.89	−2.58
$\Delta\kappa_t^{(2)}$	constant	$l = 2$	KPSS : η_μ	0.327	0.739	0.463	0.347
$\Delta\kappa_t^{(3)}$	constant	$k = 3$	ADF : τ_2	−3.47	−3.51	−2.89	−2.58
$\Delta\kappa_t^{(3)}$	constant	$l = 2$	KPSS : η_μ	0.237	0.739	0.463	0.347

Table 2.8: Trace test for the cointegration rank of the KAN:2 model.

H_0	H_1	trace statistic	critical values		
			10%	5%	1%
$r = 0$	$0 < r \leq 3$	51.79	39.06	42.44	48.45
$r = 1$	$1 < r \leq 3$	19.50	22.76	25.32	30.45
$r = 2$	$2 < r \leq 3$	8.39	10.49	12.25	16.26

Table 2.9: Maximum eigenvalue test for the cointegration rank of the KAN:2 model.

H_0	H_1	max eig. statistic	critical values		
			10%	5%	1%
$r = 0$	$r = 1$	32.29	23.11	25.54	30.34
$r = 1$	$r = 2$	11.10	16.85	18.96	23.65
$r = 2$	$r = 3$	8.39	10.49	12.25	16.26

VECM Estimation for the KAN:2 Model

Analogously to the KAN model, the maximum likelihood estimator of Proposition 2.3.2 is used to estimate a VECM with cointegration rank $r = 1$. The concrete form is given by

$$\Delta\boldsymbol{\kappa}_t = \boldsymbol{\alpha}[\boldsymbol{\beta}' : \xi_{co}]\begin{bmatrix}\boldsymbol{\kappa}_{t-1} \\ t-1\end{bmatrix} + \boldsymbol{\Gamma}_1\Delta\boldsymbol{\kappa}_{t-1} + \boldsymbol{\Gamma}_2\Delta\boldsymbol{\kappa}_{t-2} + \boldsymbol{\xi} + \boldsymbol{u}_t,$$

where $\boldsymbol{\alpha}$ denotes the loading vector, $\boldsymbol{\beta}$ the cointegration vector, ξ_{co} the linear trend parameter in the cointegration relation, and $\boldsymbol{\xi}$ the parameter of the unrestricted deterministic term. The ML estimator yields

$$\begin{bmatrix}\Delta\kappa_{1,t} \\ \Delta\kappa_{2,t} \\ \Delta\kappa_{3,t}\end{bmatrix} = \begin{bmatrix} \underset{(\mathbf{-3.364})}{-2.45\times10^{-1}} \\ \underset{(\mathbf{2.848})}{2.23\times10^{-2}} \\ \underset{(\mathbf{-1.429})}{-2.98\times10^{-4}} \end{bmatrix} \begin{bmatrix} 1.00 : & \underset{(\mathbf{-2.994})}{-3.16\times10^{1}} & \underset{(\mathbf{-3.562})}{-1.52\times10^{3}} & \underset{(\mathbf{9.268})}{2.51\times10^{-2}} \end{bmatrix} \begin{bmatrix}\kappa_{1,t-1} \\ \kappa_{2,t-1} \\ \kappa_{3,t-1} \\ t-1\end{bmatrix}$$

$$+ \begin{bmatrix} \underset{(-0.170)}{-3.193\times10^{-2}} & \underset{(\mathbf{-2.917})}{-8.738} & \underset{(\mathbf{-3.892})}{-4.160\times10^{2}} \\ \underset{(\mathbf{-3.025})}{-6.113\times10^{-2}} & \underset{(0.236)}{7.599\times10^{-2}} & \underset{(\mathbf{2.744})}{3.148\times10^{1}} \\ \underset{(\mathbf{3.355})}{1.810\times10^{-3}} & \underset{(0.264)}{-2.263\times10^{-3}} & \underset{(\mathbf{-3.404})}{-1.043} \end{bmatrix} \begin{bmatrix}\Delta\kappa_{1,t-1} \\ \Delta\kappa_{2,t-1} \\ \Delta\kappa_{3,t-1}\end{bmatrix}$$

$$+ \begin{bmatrix} \underset{(\mathbf{-0.325})}{-5.986\times10^{-2}} & \underset{(-0.749)}{-1.908} & \underset{(\mathbf{-2.276})}{-1.999\times10^{2}} \\ \underset{(\mathbf{-2.014})}{-3.977\times10^{-2}} & \underset{(-0.506)}{-1.384\times10^{-1}} & \underset{(1.744)}{1.644\times10^{1}} \\ \underset{(\mathbf{2.176})}{1.147\times10^{-3}} & \underset{(-0.122)}{-8.873\times10^{-4}} & \underset{(\mathbf{-2.574})}{-6.476\times10^{-1}} \end{bmatrix} \begin{bmatrix}\Delta\kappa_{1,t-2} \\ \Delta\kappa_{2,t-2} \\ \Delta\kappa_{3,t-2}\end{bmatrix}$$

$$+\begin{bmatrix} \underset{(-3.395)}{-2.038} \\ \underset{(2.815)}{1.814 \times 10^{-1}} \\ \underset{(-1.383)}{-2.379 \times 10^{-3}} \end{bmatrix} + \begin{bmatrix} \hat{u}_{1,t} \\ \hat{u}_{2,t} \\ \hat{u}_{3,t} \end{bmatrix}. \tag{2.53}$$

In eq. (2.53) the t-values of parameter estimates are given in parentheses. The estimated innovation covariance and correlation matrices are given by

$$\widetilde{\Sigma}_u = \begin{bmatrix} 7.909 \times 10^{-4} & -7.002 \times 10^{-5} & 1.759 \times 10^{-6} \\ -7.002 \times 10^{-5} & 9.112 \times 10^{-6} & -2.235 \times 10^{-7} \\ 1.759 \times 10^{-6} & -2.235 \times 10^{-7} & 6.494 \times 10^{-9} \end{bmatrix}$$

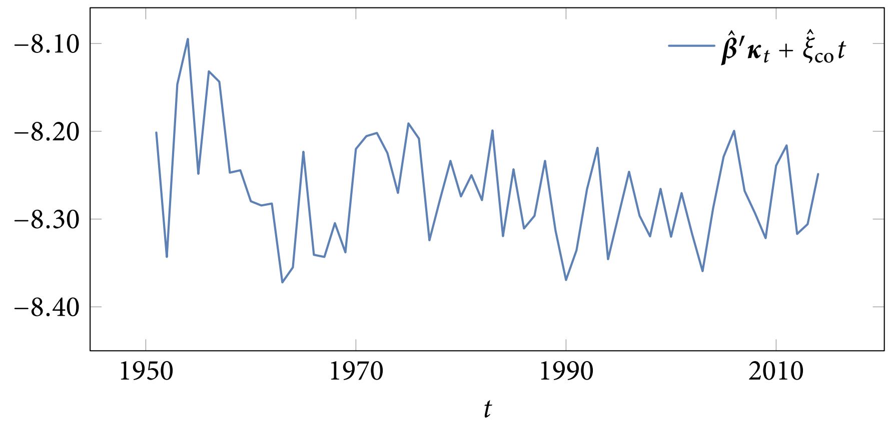
and

$$\widetilde{\text{Corr}}(\boldsymbol{u}_t) = \begin{bmatrix} 1 & -0.825 & 0.776 \\ -0.825 & 1 & -0.919 \\ 0.776 & -0.919 & 1 \end{bmatrix}. \tag{2.54}$$

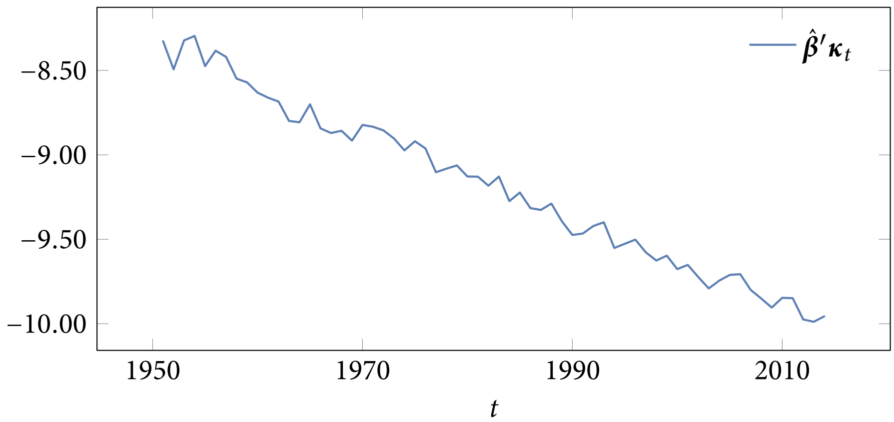
The VECM shows very high instantaneous correlations between the variables. For instance, there is a correlation of −0.825 between the first and second variable and an instantaneous correlation of −0.919 between the second and third component. Furthermore, there is a significant positive correlation of 0.776 between the first and third variable. Figure 2.4 illustrates the estimated long-run equilibrium relation of the KAN:2 model by showing the expressions $\hat{\boldsymbol{\beta}}'\boldsymbol{\kappa}_t + \hat{\xi}_{\text{co}}t$ and $\hat{\boldsymbol{\beta}}'\boldsymbol{\kappa}_t$. From Figure 2.4(b), we see that the complex and non-stationary evolutions of the individual trajectories of the KAN:2 model, as presented Figure 1.37, follow a trend stationary process given by $\hat{\boldsymbol{\beta}}'\boldsymbol{\kappa}_t$. This also implies that the essential degrees of freedom of the KAN:2 model are reduced from three to only one dimension.

By using the same transformation as for eq. (2.51), we obtain a VAR representation of eq. (2.54) as

$$\begin{aligned}
\boldsymbol{\kappa}_t &= A_1\boldsymbol{\kappa}_{t-1} + A_2\boldsymbol{\kappa}_{t-2} + A_3\boldsymbol{\kappa}_{t-2} + CD_t + \hat{\boldsymbol{u}}_t \\
&= \begin{bmatrix} 7.23 \times 10^{-1} & -9.85 \times 10^{-1} & -4.39 \times 10^{1} \\ -3.88 \times 10^{-2} & 3.71 \times 10^{-1} & -2.33 \\ 1.51 \times 10^{-3} & 7.18 \times 10^{-3} & 4.10 \times 10^{-1} \end{bmatrix} \begin{bmatrix} \kappa_{1,t-1} \\ \kappa_{2,t-1} \\ \kappa_{3,t-1} \end{bmatrix} \\
&+ \begin{bmatrix} -2.79 \times 10^{-2} & 6.83 & 2.16 \times 10^{2} \\ 2.14 \times 10^{-2} & -2.14 \times 10^{-1} & -1.50 \times 10^{1} \\ -6.62 \times 10^{-4} & 1.38 \times 10^{-3} & 3.95 \times 10^{-1} \end{bmatrix} \begin{bmatrix} \kappa_{1,t-2} \\ \kappa_{2,t-2} \\ \kappa_{3,t-2} \end{bmatrix} \\
&+ \begin{bmatrix} 5.99 \times 10^{-2} & 1.91 & 2.00 \times 10^{2} \\ 3.98 \times 10^{-2} & 1.38 \times 10^{-1} & -1.64 \times 10^{1} \\ -1.15 \times 10^{-3} & 8.87 \times 10^{-4} & 6.48 \times 10^{-1} \end{bmatrix} \begin{bmatrix} \kappa_{1,t-3} \\ \kappa_{2,t-3} \\ \kappa_{3,t-3} \end{bmatrix} \\
&+ \begin{bmatrix} -2.06 & -6.16 \times 10^{-3} \\ 1.83 \times 10^{-1} & 5.60 \times 10^{-4} \\ -2.40 \times 10^{-3} & -7.50 \times 10^{-6} \end{bmatrix} \begin{bmatrix} 1 \\ t-1 \end{bmatrix} + \begin{bmatrix} \hat{u}_{1,t} \\ \hat{u}_{2,t} \\ \hat{u}_{3,t} \end{bmatrix}.
\end{aligned}$$



(a) Level stationary equilibrium relation $\hat{\beta}'\kappa_t + \hat{\xi}_{co} t$ of the KAN:2 model.

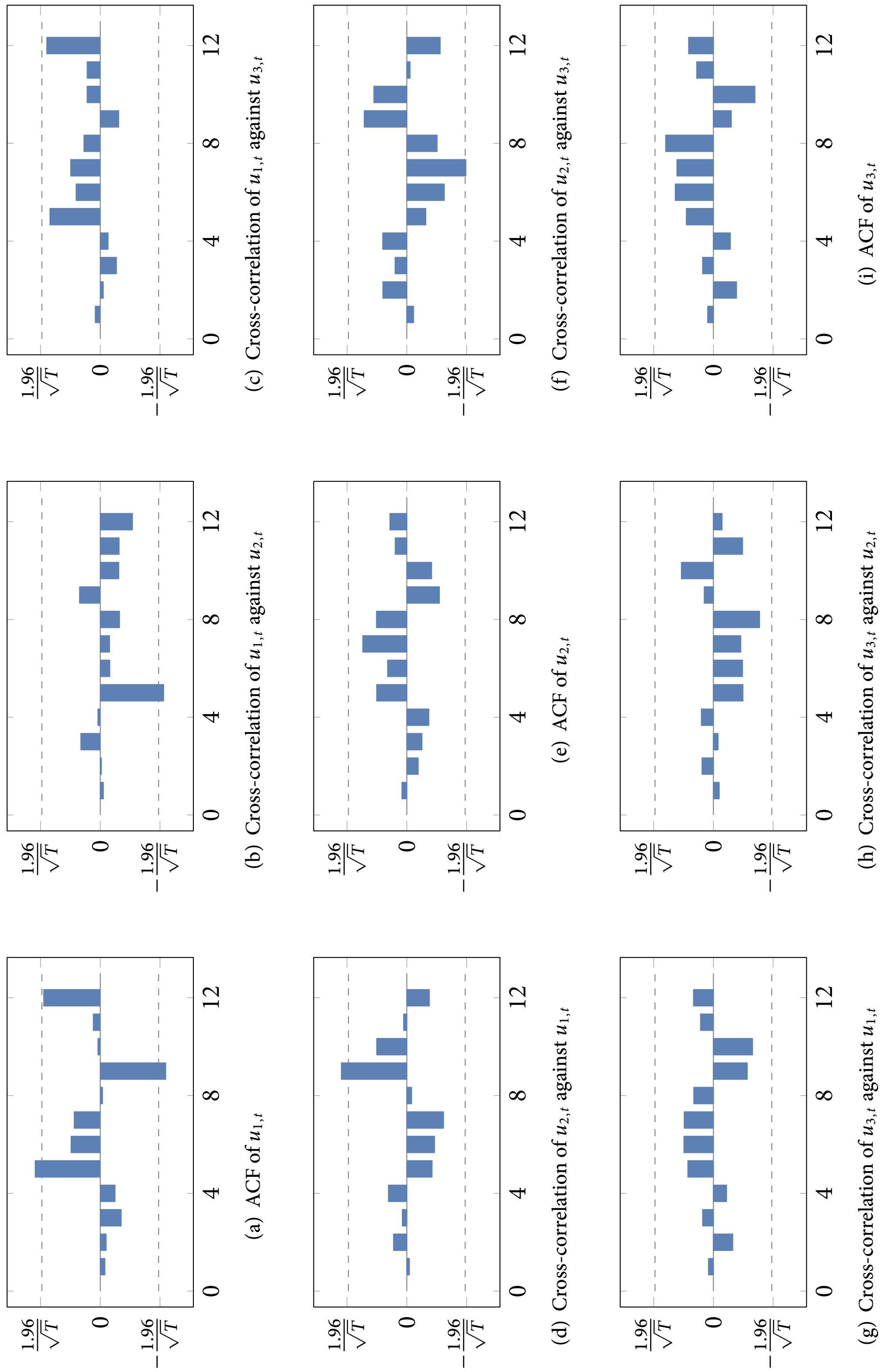


(b) Trend stationary equilibrium relation $\hat{\beta}'\kappa_t$ of the KAN:2 model.

Figure 2.4: Estimated equilibrium relation of the KAN:2 model.

VECM Validation for the KAN:2 Model

We proceed the analysis by performing diagnostic tests to check the residuals of the VECM in eq. (2.53) for autocorrelation, non-normality and heteroscedasticity. Visual inspection of the residual ACF and cross-correlation functions, as shown in Figure 2.5, does not reveal any serious issues with the fitted VECM(2). Apart from only a few exceptions, the estimated residual auto and cross-correlation are below the significance bounds. The results of the formal diagnostic tests are summarized in Table 2.10 and show that none of the autocorrelation, non-normality or heteroscedasticity tests indicates problems with the fitted model of eq. (2.53). All null hypotheses are accepted at the 5% level. Therefore, we conclude that the estimated VECM(2) with cointegration rank $r = 1$ provides a good approximation of the KAN:2 DGP.



(a) ACF of $u_{1,t}$
(b) Cross-correlation of $u_{1,t}$ against $u_{2,t}$
(c) Cross-correlation of $u_{1,t}$ against $u_{3,t}$
(d) Cross-correlation of $u_{2,t}$ against $u_{1,t}$
(e) ACF of $u_{2,t}$
(f) Cross-correlation of $u_{2,t}$ against $u_{3,t}$
(g) Cross-correlation of $u_{3,t}$ against $u_{1,t}$
(h) Cross-correlation of $u_{3,t}$ against $u_{2,t}$
(i) ACF of $u_{3,t}$

Figure 2.5: Residual autocorrelation and cross-correlations of the VECM for KAN:2.

Table 2.10: Diagnostics of the VECM for the KAN:2 time series.

diagnostic type	test name	test statistic	test value	appr. dist.	p-value
Autocorrelation	Portmanteau	Q_{16}	127.299	$\chi^2(123)$	0.377
	adjusted Portmanteau	Q^*_{16}	148.221	$\chi^2(123)$	0.060
	Breusch-Godfrey	$\lambda_{\mathrm{LM}}(5)$	55.508	$\chi^2(45)$	0.136
Non-normality	multivariate	λ_{LJB}	7.320	$\chi^2(6)$	0.292
	skewness only	s_3^2	5.635	$\chi^2(3)$	0.131
	kurtosis only	s_4^2	1.685	$\chi^2(3)$	0.640
	univariate u_1	λ_{LJB}	2.829	$\chi^2(2)$	0.243
	univariate u_2	λ_{LJB}	2.724	$\chi^2(2)$	0.256
	univariate u_3	λ_{LJB}	2.349	$\chi^2(2)$	0.309
Heteroscedasticity	multivariate	$\mathrm{MARCH}_{\mathrm{LM}}(5)$	206.438	$\chi^2(180)$	0.086
	univariate u_1	$\mathrm{ARCH}_{\mathrm{LM}}(16)$	10.877	$\chi^2(16)$	0.817
	univariate u_2	$\mathrm{ARCH}_{\mathrm{LM}}(16)$	9.960	$\chi^2(16)$	0.869
	univariate u_3	$\mathrm{ARCH}_{\mathrm{LM}}(16)$	12.099	$\chi^2(16)$	0.737

2.4.3 | VECM Specification for the KAN:3 Model

The objective of this section is to analyse the multivariate time series given by the KAN:3 model presented in Section 1.9.2. The predictor function of the KAN:3 model is given by

$$\eta_{t,x} = \sum_{i=1}^{4} (x - x_{\min})^{i-1} \kappa_t^{(i)}.$$

The 4 coefficients, namely $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, $\kappa_t^{(3)}$, and $\kappa_t^{(4)}$ determine the mortality curve at the period t. For the plots of the KAN:3 trajectories, see Figure 1.38 on page 92. As for the KAN and KAN:2 models, we follow the specification procedure for VAR/VECM as described in Remark 2.3.1 to determine whether we can formally confirm the observation of long-run relations between the coefficients, as made in Section 1.9.5 and shown in Figures 1.42(c) and 1.42(d). Since the specification steps are very similar to those described in Sections 2.4.1 and 2.4.2, we provide only a summarized presentation of the procedure.

Lag Order Selection for KAN:3

By using the information criteria from Section 2.3.1, we begin by selecting the lag order p of an unrestricted VAR(p) model, with an included constant and a deterministic trend. The results are summarized in Table 2.11. Similar to the previous analysis of the KAN and KAN:2, two different lag orders are suggested. The AIC information criterion is minimized for $p = 3$, whereas the other criteria prefer the lag order $p = 1$. Based on diagnostic tests and significance analysis of coefficients corresponding to higher lag order, we choose the lag order $p = 3$ for the further analysis.

Table 2.11: Information criteria based VAR lag order selection for the KAN:3 model.

lags p	AIC(p)	HQ(p)	SC(p)	FPE(p)
1	−63.060	**−62.737**	**−62.237**	$\mathbf{4.11 \cdot 10^{-28}}$
2	−62.971	−62.432	−61.599	$4.54 \cdot 10^{-28}$
3	**−63.086**	−62.332	−61.165	$4.13 \cdot 10^{-28}$
4	−62.884	−61.914	−60.414	$5.24 \cdot 10^{-28}$
5	−62.940	−61.755	−59.921	$5.27 \cdot 10^{-28}$
6	−62.755	−61.354	−59.187	$6.94 \cdot 10^{-28}$
7	−62.794	−61.177	−58.677	$7.63 \cdot 10^{-28}$

Unit Root Tests for the KAN:3 Model

In Table 2.12, we summarize the results of the ADF unit root tests and the KPSS stationary tests for the KAN:3 time series. The results show that the unit root null hypotheses of the ADF tests are accepted for all components of $\boldsymbol{\kappa}_t$ at the 5% level. Simultaneously, the KPSS null hypotheses of stationarity are rejected for the components $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, and $\kappa_t^{(4)}$ at the 5% level. Stationarity of $\kappa_t^{(3)}$ could only be rejected at the 10% level. Furthermore, for the first-order differences $\Delta\boldsymbol{\kappa}_t$ the unit root null hypotheses are rejected by the ADF tests at the 5% level and simultaneously stationarity is accepted by the KPSS tests at the 5% level. Based on the results of this testing procedure, we can conclude that the underlying univariate DGPs of $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, $\kappa_t^{(3)}$ and $\kappa_t^{(4)}$ are integrated of order one.

Cointegration Tests for the KAN:3 Model

Next, Johansen cointegration tests will be performed to assess the number of cointegration relations between the KAN:3 time series. The results of the trace and maximum eigenvalue test are reported in Tables 2.13 and 2.14. The trace test rejects the null hypotheses of $r = 0$ and $r = 1$ and accepts the null hypothesis of $r = 2$ at the 5% level. The maximum eigenvalue test rejects the null hypothesis of $r = 0$ and accepts the null hypotheses of $r = 1$ at the 5% level. This means that the trace test suggests $r = 2$ and the maximum eigenvalue test $r = 1$ as the cointegration rank. As a consequence of this ambiguous result, we estimated two VECMs using both suggested cointegration ranks. The estimation results for $r = 2$ showed that the loading matrix $\boldsymbol{\alpha}$ only contained non-significant coefficients for the second cointegration relation. Therefore, we will only consider the more restrictive case and present the estimation results of a VECM with cointegration rank $r = 1$.

VECM Estimation for the KAN:3 Model

The maximum likelihood estimator of Proposition 2.3.2 is used to estimate a Gaussian noise driven VECM in cointegration rank $r = 1$. Analogously to the previous two models, we estimate a VECM of the form

$$\Delta\boldsymbol{\kappa}_t = \boldsymbol{\alpha}[\boldsymbol{\beta}' : \xi_{\mathrm{co}}]\begin{bmatrix}\boldsymbol{\kappa}_{t-1}\\ t-1\end{bmatrix} + \boldsymbol{\Gamma}_1\Delta\boldsymbol{\kappa}_{t-1} + \boldsymbol{\Gamma}_2\Delta\boldsymbol{\kappa}_{t-2} + \boldsymbol{\xi} + \boldsymbol{u}_t,$$

Table 2.12: ADF and KPSS tests for the components of $\boldsymbol{\kappa}_t$ and $\Delta\boldsymbol{\kappa}_t$ of the KAN:3 model.

time series	deterministic trend	lags / length	statistic	test value	critical values		
					1%	5%	10%
$\kappa_t^{(1)}$	constant, trend	$k = 3$	ADF : τ_3	−2.95	−4.04	−3.45	−3.15
$\kappa_t^{(1)}$	constant, trend	$l = 2$	KPSS : η_τ	0.441	0.216	0.146	0.119
$\kappa_t^{(2)}$	constant, trend	$k = 3$	ADF : τ_3	−2.70	−4.04	−3.45	−3.15
$\kappa_t^{(2)}$	constant, trend	$l = 2$	KPSS : η_τ	0.253	0.216	0.146	0.119
$\kappa_t^{(3)}$	constant, trend	$k = 3$	ADF : τ_3	−2.54	−4.04	−3.45	−3.15
$\kappa_t^{(3)}$	constant, trend	$l = 2$	KPSS : η_τ	0.137	0.216	0.146	0.119
$\kappa_t^{(4)}$	constant, trend	$k = 3$	ADF : τ_3	−2.05	−4.04	−3.45	−3.15
$\kappa_t^{(4)}$	constant, trend	$l = 2$	KPSS : η_τ	0.254	0.216	0.146	0.119
$\Delta\kappa_t^{(1)}$	constant	$k = 3$	ADF : τ_2	−5.38	−3.51	−2.89	−2.58
$\Delta\kappa_t^{(1)}$	constant	$l = 2$	KPSS : η_μ	0.216	0.739	0.463	0.347
$\Delta\kappa_t^{(2)}$	constant	$k = 3$	ADF : τ_2	−5.62	−3.51	−2.89	−2.58
$\Delta\kappa_t^{(2)}$	constant	$l = 2$	KPSS : η_μ	0.112	0.739	0.463	0.347
$\Delta\kappa_t^{(3)}$	constant	$k = 3$	ADF : τ_2	−5.54	−3.51	−2.89	−2.58
$\Delta\kappa_t^{(3)}$	constant	$l = 2$	KPSS : η_μ	0.055	0.739	0.463	0.347
$\Delta\kappa_t^{(4)}$	constant	$k = 3$	ADF : τ_2	−5.50	−3.51	−2.89	−2.58
$\Delta\kappa_t^{(4)}$	constant	$l = 2$	KPSS : η_μ	0.047	0.739	0.463	0.347

Table 2.13: Trace test for cointegration rank of the KAN:3 model.

H_0	H_1	trace statistic	critical values		
			10%	5%	1%
$r = 0$	$0 < r \leq 4$	78.01	59.14	62.99	70.05
$r = 1$	$1 < r \leq 4$	43.47	39.06	42.44	48.45
$r = 2$	$2 < r \leq 4$	18.79	22.76	25.32	30.45
$r = 3$	$3 < r \leq 4$	7.75	10.49	12.25	16.26

Table 2.14: Maximum eigenvalue test for cointegration rank of the KAN:3 model.

H_0	H_1	max eig. statistic	critical values		
			10%	5%	1%
$r = 0$	$r = 1$	34.54	29.12	31.46	36.65
$r = 1$	$r = 2$	24.68	23.11	25.54	30.34
$r = 2$	$r = 3$	11.04	16.85	18.96	23.65
$r = 3$	$r = 4$	7.75	10.49	12.25	16.26

where $\boldsymbol{\alpha}$ denotes the loading vector, $\boldsymbol{\beta}$ the cointegration vector, ξ_{co} the linear trend parameter in the cointegration relation, and $\boldsymbol{\xi}$ the coefficient of the unrestricted deterministic term. The ML estimator leads to

$$
\begin{aligned}
\begin{bmatrix} \Delta\kappa_{1,t} \\ \Delta\kappa_{2,t} \\ \Delta\kappa_{3,t} \\ \Delta\kappa_{4,t} \end{bmatrix}
&= \begin{bmatrix} \underset{(\mathbf{-3.13})}{-7.87\times 10^{-1}} \\ \underset{(\mathbf{2.68})}{1.32\times 10^{-1}} \\ \underset{(\mathbf{-3.39})}{-9.74\times 10^{-3}} \\ \underset{(\mathbf{-3.47})}{1.82\times 10^{-4}} \end{bmatrix}
\begin{bmatrix} 1.00: & \underset{(1.56)}{4.77} & \underset{(-1.07)}{-1.3\times 10^{2}} & \underset{(\mathbf{-2.75})}{-1.12\times 10^{4}} & \underset{(\mathbf{21.28})}{1.96\times 10^{-2}} \end{bmatrix}
\begin{bmatrix} \kappa_{1,t-1} \\ \kappa_{2,t-1} \\ \kappa_{3,t-1} \\ \kappa_{4,t-1} \\ t-1 \end{bmatrix} \\
&+ \begin{bmatrix}
\underset{(1.75)}{4.56\times 10^{-1}} & \underset{(0.35)}{1.09} & \underset{(-1.32)}{-1.35\times 10^{2}} & \underset{(-1.77)}{-7.35\times 10^{3}} \\
\underset{(\mathbf{-3.44})}{-1.72\times 10^{-1}} & \underset{(\mathbf{-2.62})}{-1.61} & \underset{(-0.35)}{-7.11} & \underset{(0.42)}{3.39\times 10^{2}} \\
\underset{(\mathbf{3.47})}{1.04\times 10^{-2}} & \underset{(\mathbf{2.39})}{8.58\times 10^{-2}} & \underset{(0.07)}{7.59\times 10^{-2}} & \underset{(-0.66)}{-3.13\times 10^{1}} \\
\underset{(\mathbf{-3.07})}{-1.67\times 10^{-4}} & \underset{(\mathbf{-2.37})}{-1.55\times 10^{-3}} & \underset{(-0.64)}{-1.35\times 10^{-2}} & \underset{(-0.04)}{-3.58\times 10^{-2}}
\end{bmatrix}
\begin{bmatrix} \Delta\kappa_{1,t-1} \\ \Delta\kappa_{2,t-1} \\ \Delta\kappa_{3,t-1} \\ \Delta\kappa_{4,t-1} \end{bmatrix} \\
&+ \begin{bmatrix}
\underset{(1.07)}{2.52\times 10^{-1}} & \underset{(0.98)}{2.92} & \underset{(-0.51)}{-5.05\times 10^{1}} & \underset{(-1.16)}{-4.43\times 10^{3}} \\
\underset{(\mathbf{-2.83})}{-1.31\times 10^{-1}} & \underset{(\mathbf{-2.06})}{-1.20} & \underset{(-0.31)}{-6.03} & \underset{(0.27)}{2.03\times 10^{2}} \\
\underset{(\mathbf{3.09})}{8.32\times 10^{-3}} & \underset{(1.84)}{6.25\times 10^{-2}} & \underset{(0.05)}{5.81\times 10^{-2}} & \underset{(-0.50)}{-2.21\times 10^{1}} \\
\underset{(\mathbf{-2.89})}{-1.42\times 10^{-4}} & \underset{(\mathbf{-1.97})}{-1.22\times 10^{-3}} & \underset{(-0.48)}{-9.9\times 10^{-3}} & \underset{(0.05)}{4.04\times 10^{-2}}
\end{bmatrix}
\begin{bmatrix} \Delta\kappa_{1,t-2} \\ \Delta\kappa_{2,t-2} \\ \Delta\kappa_{3,t-2} \\ \Delta\kappa_{4,t-2} \end{bmatrix} \\
&+ \begin{bmatrix} \underset{(\mathbf{-3.15})}{-3.01} \\ \underset{(\mathbf{2.66})}{5.00\times 10^{-1}} \\ \underset{(\mathbf{-3.37})}{-3.69\times 10^{-2}} \\ \underset{(\mathbf{3.46})}{6.92\times 10^{-4}} \end{bmatrix}
+ \begin{bmatrix} \hat{u}_{1,t} \\ \hat{u}_{2,t} \\ \hat{u}_{3,t} \\ \hat{u}_{4,t} \end{bmatrix}.
\end{aligned}
\tag{2.55}
$$

The covariance and correlation estimates of the Gaussian innovation process are given by

$$
\widetilde{\Sigma}_u = \begin{bmatrix}
1.15\times 10^{-3} & -1.81\times 10^{-4} & 8.2\times 10^{-6} & -1.15\times 10^{-7} \\
-1.81\times 10^{-4} & 4.47\times 10^{-5} & -2.46\times 10^{-6} & 4.1\times 10^{-8} \\
8.2\times 10^{-6} & -2.46\times 10^{-6} & 1.51\times 10^{-7} & -2.69\times 10^{-9} \\
-1.15\times 10^{-7} & 4.1\times 10^{-8} & -2.69\times 10^{-9} & 5.05\times 10^{-11}
\end{bmatrix}
\tag{2.56}
$$

and

$$
\widetilde{\mathrm{Corr}}(\boldsymbol{u}_t) = \begin{bmatrix}
1 & -0.80 & 0.62 & -0.48 \\
-0.80 & 1 & -0.95 & 0.86 \\
0.62 & -0.95 & 1 & -0.97 \\
-0.48 & 0.86 & -0.97 & 1
\end{bmatrix}.
$$

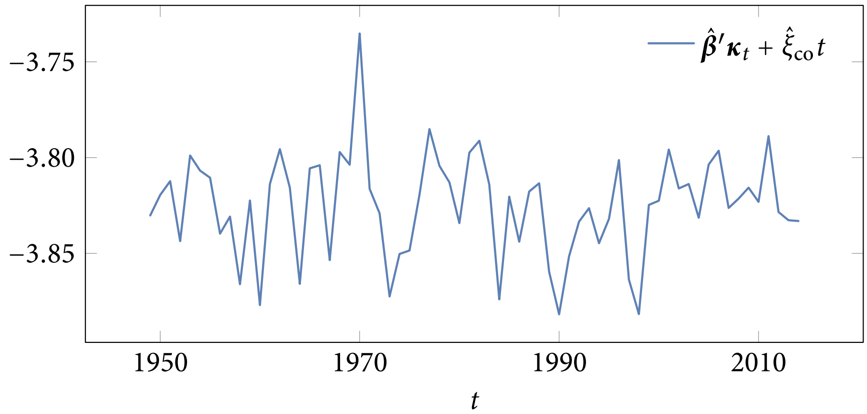
Similar to both previous cases, the VECM shows high instantaneous correlations between particular variables. For instance, there is a high negative correlation between the first and second variable, the second and third, and the third and fourth component. The VAR representation of the VECM of eq. (2.55) is given by

$$
\begin{aligned}
\boldsymbol{\kappa}_t &= A_1\boldsymbol{\kappa}_{t-1} + A_2\boldsymbol{\kappa}_{t-2} + A_3\boldsymbol{\kappa}_{t-2} + CD_t + \hat{\boldsymbol{u}}_t \qquad (2.57)\\
&= \begin{bmatrix} 6.69\times10^{-1} & -2.67 & -3.23\times10^{1} & 1.48\times10^{3} \\ -3.91\times10^{-2} & 2.03\times10^{-2} & -2.44\times10^{1} & -1.15\times10^{3} \\ 6.34\times10^{-4} & 3.92\times10^{-2} & 2.35 & 7.79\times10^{1} \\ 1.51\times10^{-5} & -6.83\times10^{-4} & -3.73\times10^{-2} & -1.08 \end{bmatrix} \begin{bmatrix} \kappa_{1,t-1} \\ \kappa_{2,t-1} \\ \kappa_{3,t-1} \\ \kappa_{4,t-1} \end{bmatrix}\\
&+ \begin{bmatrix} -2.04\times10^{-1} & 1.83 & 8.45\times10^{1} & 2.92\times10^{3} \\ 4.05\times10^{-2} & 4.1\times10^{-1} & 1.08 & -1.35\times10^{2} \\ -2.06\times10^{-3} & -2.33\times10^{-2} & -1.78\times10^{-2} & 9.25 \\ 2.51\times10^{-5} & 3.33\times10^{-4} & 3.65\times10^{-3} & 7.62\times10^{-2} \end{bmatrix} \begin{bmatrix} \kappa_{1,t-2} \\ \kappa_{2,t-2} \\ \kappa_{3,t-2} \\ \kappa_{4,t-3} \end{bmatrix}\\
&+ \begin{bmatrix} -2.52\times10^{-1} & -2.92 & 5.05\times10^{1} & 4.43\times10^{3} \\ 1.31\times10^{-1} & 1.2 & 6.03 & -2.03\times10^{2} \\ -8.32\times10^{-3} & -6.25\times10^{-2} & -5.81\times10^{-2} & 2.21\times10^{1} \\ 1.42\times10^{-4} & 1.22\times10^{-3} & 9.9\times10^{-3} & -4.04\times10^{-2} \end{bmatrix} \begin{bmatrix} \kappa_{1,t-3} \\ \kappa_{2,t-3} \\ \kappa_{3,t-3} \\ \kappa_{4,t-3} \end{bmatrix}\\
&+ \begin{bmatrix} -3.01 & -1.54\times10^{-2} \\ 5.\times10^{-1} & 2.59\times10^{-3} \\ -3.69\times10^{-2} & -1.91\times10^{-4} \\ 6.92\times10^{-4} & 3.57\times10^{-6} \end{bmatrix} \begin{bmatrix} 1 \\ t-1 \end{bmatrix} + \begin{bmatrix} \hat{u}_{1,t} \\ \hat{u}_{2,t} \\ \hat{u}_{3,t} \\ \hat{u}_{4,t} \end{bmatrix}.
\end{aligned}
$$

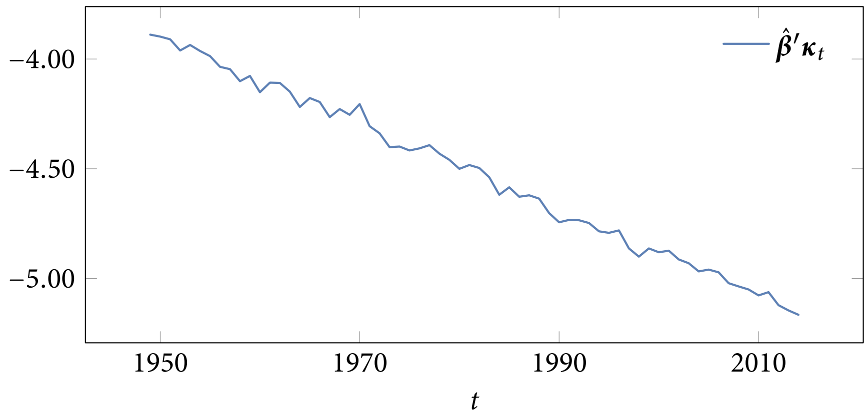
Figure 2.6 illustrates the estimated long-run equilibrium relation of the KAN:3 model. The upper panel shows the non-trending relation $\hat{\boldsymbol{\beta}}'\boldsymbol{\kappa}_t + \hat{\xi}_{\text{co}}t$, while the lower panel illustrates the linear trending relation $\hat{\boldsymbol{\beta}}'\boldsymbol{\kappa}_t$. By these plots, one can visually assess stationarity or trend stationarity of the corresponding processes. The trajectories of Figures 2.6(a) and 2.6(b) do not indicate problems with the VECM. Formal model validation is conducted in the following section.

VECM Validation for the KAN:3 Model

We proceed the specification by performing diagnostic tests to check the residuals of the VECM(2) with cointegration rank $r = 1$, as given in eq. (2.55), for autocorrelation, non-normality and heteroscedasticity. Figure 2.7 illustrates the residual ACF and cross-correlation functions. We observe some significant correlations at the lag orders 4 and 9, see, for instance, Figures 2.7(a) to 2.7(e), 2.7(i) and 2.7(m). However, formal tests for autocorrelation, non-normality or heteroscedasticity do not indicate problems with the fitted VECM. The testing results, which are summarized in Table 2.15, show that all null hypotheses are accepted at the 5% level. Therefore, we conclude that the estimated VECM(2) with cointegration rank $r = 1$ provides a good approximation of the KAN:3 DGP.

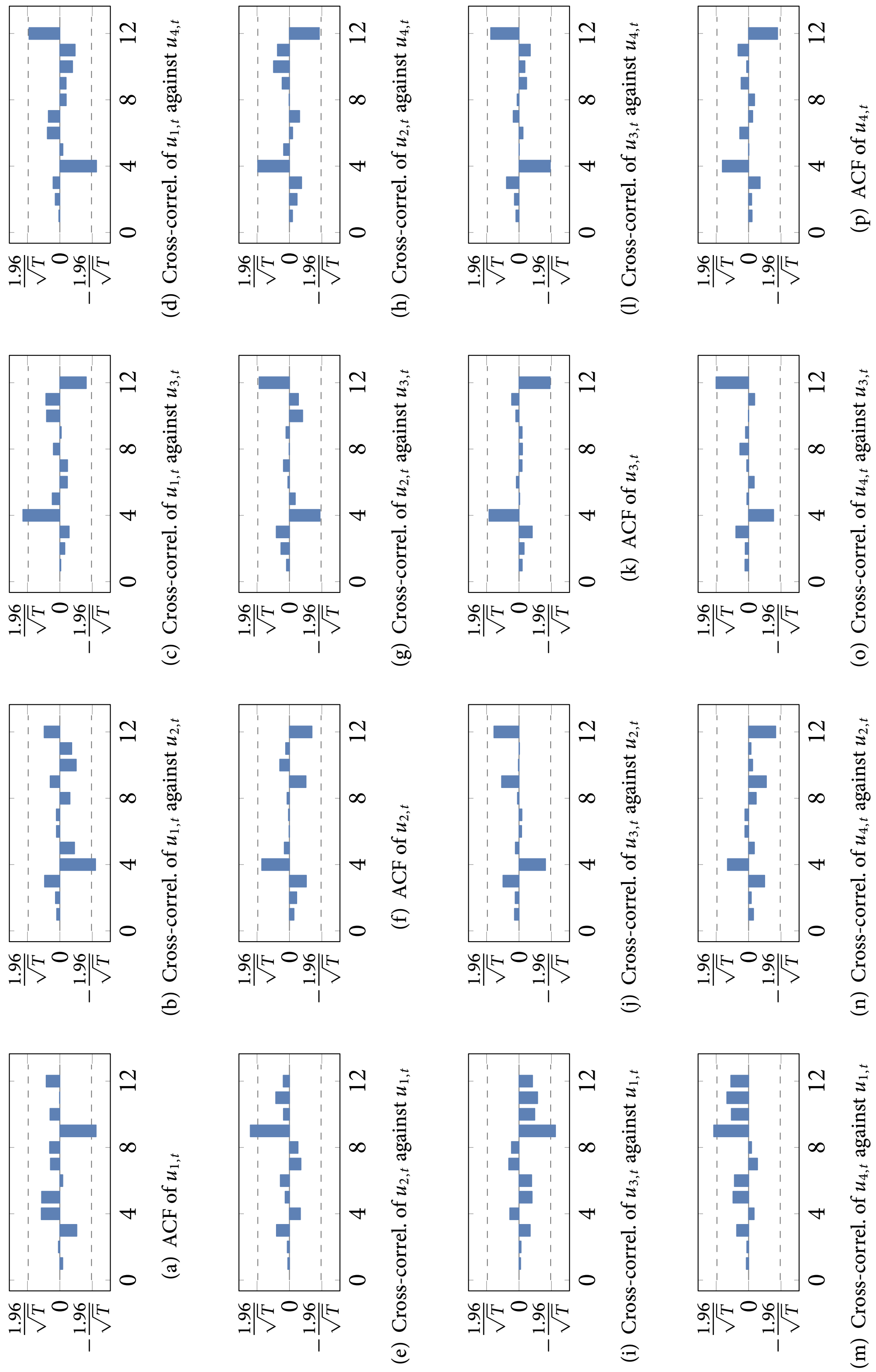


(a) Level stationary equilibrium relation $\hat{\beta}'\kappa_t + \hat{\xi}_{co}t$ of the KAN:3 model.



(b) Trend stationary equilibrium relation $\hat{\beta}'\kappa_t$ of the KAN:3 model.

Figure 2.6: Estimated equilibrium relation of the KAN:3 model.



(a) ACF of $u_{1,t}$
(b) Cross-correl. of $u_{1,t}$ against $u_{2,t}$
(c) Cross-correl. of $u_{1,t}$ against $u_{3,t}$
(d) Cross-correl. of $u_{1,t}$ against $u_{4,t}$
(e) Cross-correl. of $u_{2,t}$ against $u_{1,t}$
(f) ACF of $u_{2,t}$
(g) Cross-correl. of $u_{2,t}$ against $u_{3,t}$
(h) Cross-correl. of $u_{2,t}$ against $u_{4,t}$
(i) Cross-correl. of $u_{3,t}$ against $u_{1,t}$
(j) Cross-correl. of $u_{3,t}$ against $u_{2,t}$
(k) ACF of $u_{3,t}$
(l) Cross-correl. of $u_{3,t}$ against $u_{4,t}$
(m) Cross-correl. of $u_{4,t}$ against $u_{1,t}$
(n) Cross-correl. of $u_{4,t}$ against $u_{2,t}$
(o) Cross-correl. of $u_{4,t}$ against $u_{3,t}$
(p) ACF of $u_{4,t}$

Figure 2.7: Residual autocorrelation and Cross-correlations of the VECM for KAN:3.

Table 2.15: Diagnostics of the VECM for the KAN:3 time series.

diagnostic type	test name	test statistic	test value	appr. dist.	p-value
Autocorrelation	Portmanteau	Q_{16}	197.012	$\chi^2(220)$	0.865
	adjusted Portmanteau	Q^*_{16}	231.211	$\chi^2(220)$	0.289
	Breusch-Godfrey	$\lambda_{\mathrm{LM}}(5)$	88.357	$\chi^2(80)$	0.245
Non-normality	multivariate	λ_{LJB}	2.298	$\chi^2(8)$	0.971
	skewness only	s_3^2	1.523	$\chi^2(4)$	0.823
	kurtosis only	s_4^2	0.775	$\chi^2(4)$	0.942
	univariate u_1	λ_{LJB}	0.928	$\chi^2(2)$	0.629
	univariate u_2	λ_{LJB}	1.757	$\chi^2(2)$	0.415
	univariate u_3	λ_{LJB}	0.829	$\chi^2(2)$	0.661
	univariate u_4	λ_{LJB}	0.912	$\chi^2(2)$	0.634
Heteroscedasticity	multivariate	$\mathrm{MARCH}_{\mathrm{LM}}(5)$	496.19	$\chi^2(500)$	0.540
	univariate u_1	$\mathrm{ARCH}_{\mathrm{LM}}(16)$	20.489	$\chi^2(16)$	0.199
	univariate u_2	$\mathrm{ARCH}_{\mathrm{LM}}(16)$	11.890	$\chi^2(16)$	0.752
	univariate u_3	$\mathrm{ARCH}_{\mathrm{LM}}(16)$	17.420	$\chi^2(16)$	0.359
	univariate u_4	$\mathrm{ARCH}_{\mathrm{LM}}(16)$	18.701	$\chi^2(16)$	0.285

2.5 | VECM Projections for the Kannisto Model

In Section 1.9.4, we studied the projections of the KAN, KAN:2 and KAN:3 models using a random walk with drift which is considered as the standard approach in GAPC modelling. Our major criticism regarding the random walk approach was in the first place, the high degree of uncertainty of the resulting projections and in the second place, an inconsistent trend in the improvements in the remaining life expectancy compared to the historical development.

The objective of the following is to compare the random walk projection of Section 1.9.4 with the projections of the previously estimated VECMs of Sections 2.4.1 to 2.4.3. We begin by presenting how the central forecast and prediction intervals are obtained for VECM/VAR models. Following Lütkepohl (2007), the optimal h-step forecast at origin t, denoted by $\boldsymbol{\kappa}_t(h)$, of a VAR(p) driven by a white noise process u_t with covariance matrix Σ_u is given by

$$\boldsymbol{\kappa}_t(h) = A_1\boldsymbol{\kappa}_t(h-1) + \cdots + A_p\boldsymbol{\kappa}_t(h-p), \tag{2.58}$$

where $\boldsymbol{\kappa}(j) := \boldsymbol{\kappa}_{t+j}$ for $j \leq 0$. The forecast error can be expressed as

$$\boldsymbol{\kappa}_{t+h} - \boldsymbol{\kappa}_t(h) = u_{t+h} + \Phi_1 u_{t+h-1} + \cdots + \Phi_{h-1}u_{t+1},$$

where the matrices Φ_i for $i = 1,\ldots,h-1$ are recursively defined by

$$\Phi_i = \sum_{j=1}^{i} \Phi_{i-j}A_j, \quad i = 1,2,\ldots,$$

with $\Phi_0 = \mathbb{1}_K$ and $A_j = 0$ for $j > p$. The mean-squared error of an h-step forecast is then given by

$$\boldsymbol{\Sigma}_{\boldsymbol{\kappa}}(h) := \mathrm{MSE}[\boldsymbol{\kappa}_t(h)] = \mathbb{E}\left[(\boldsymbol{\kappa}_{t+h} - \boldsymbol{\kappa}_t(h))\,(\boldsymbol{\kappa}_{t+h} - \boldsymbol{\kappa}_t(h))'\right] = \sum_{j=0}^{h-1} \Phi_j\Sigma_u\Phi_j'.$$

Similar to Section 1.9.4, we can specify the $(1-\alpha)100\%$ forecast prediction interval (P.I.) of the i-th component for h periods ahead of the origin t by

$$\left[\kappa_{i,t}(h) - z_{(\alpha/2)}\sigma_i(h), \kappa_{i,t}(h) + z_{(\alpha/2)}\sigma_i(h)\right], \tag{2.59}$$

where $\sigma_i(h)$ is the square root of the i-th diagonal element of $\boldsymbol{\Sigma}_{\boldsymbol{\kappa}}(h)$ and $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution, see Lütkepohl (2007).

2.5.1 | VECM Projections for the KAN Model

In Figure 2.8, we illustrate the projections of the VECM for the KAN model as given eq. (2.50). The projections are obtained by using eqs. (2.58) and (2.59) and replacing the coefficient matrices A_1,A_2,A_3 and Σ_u by their estimates as provided in eqs. (2.51) and (2.52). The two upper panels of Figures 2.8(a) and 2.8(b) show the projected KAN components and their 95% prediction intervals. For comparison, the plots also display the random walk (RW) projections as obtained in Section 1.9.4. By this comparison, we can clearly see that the central projections of the VECM have different slopes and also narrower prediction intervals. The different slopes of both projections have direct implications on the resulting mortality structures. The lower projected $\kappa_t^{(1)}$ values of the VECM lead to lower mortality rates across all ages, while the lower $\kappa_t^{(2)}$ values of the RW imply lower mortality rates, in particular, for higher ages. For the interpretation of the KAN coefficients, see the discussion of Section 1.9.3 and the example illustrated in Figure 1.35.

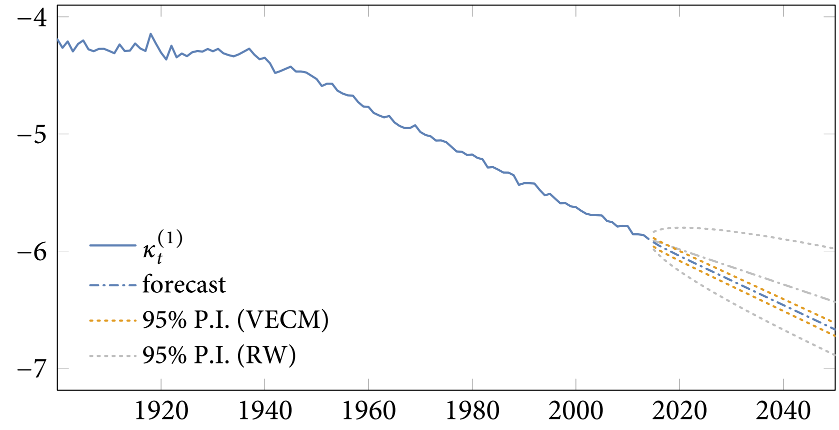
Figure 2.8(c) shows the influence of the KAN projection on the remaining life expectancy of the reference population of Swedish females aged 60. The life expectancy, projected by the VECM, is overall higher compared to the RW projection. By the VECM, we have an overall increase of the remaining life expectancy of about 3.6 years until the year 2050, rising from 25.85 to 29.44 years, which leads to an average improvement of about 36.5 days per year. This value is much closer to the historical improvements of, on average, 38.1 days per year during the past 36 years. Recall that the RW projections showed a lower average improvement of only 29.17 days per year. Another noticeable difference is the width of the prediction intervals. As criticized before in Section 1.9.4, the level of uncertainty of the future life expectancy, as given by the RW projection, would lead to non-competitive prices of mortality related claims. The width of the 95% prediction interval of the VECM at the period 2050 is with 2.33 years clearly smaller and also more realistic than the P.I width of 10.5 years of the RW projection.

Figure 2.9 shows the VECM projected hazard rates at 2050 with the corresponding 95% prediction intervals. For comparison, the plot includes the historical KAN hazard rates of the years 1910, 1970, and 2010. As the forecast for 2050 shows, there is a continuing trend of decreasing rates for the entire age range. Figure 2.10 illustrates the historical survival functions together with the forecast for 2050. Here, we can observe the continuing rectangularization of the survival function. That means, that the improvements of the life expectancy do not primarily come due to an increase of the highest attainable age but rather by a decrease of the mortality rates, in particular, at lower ages.

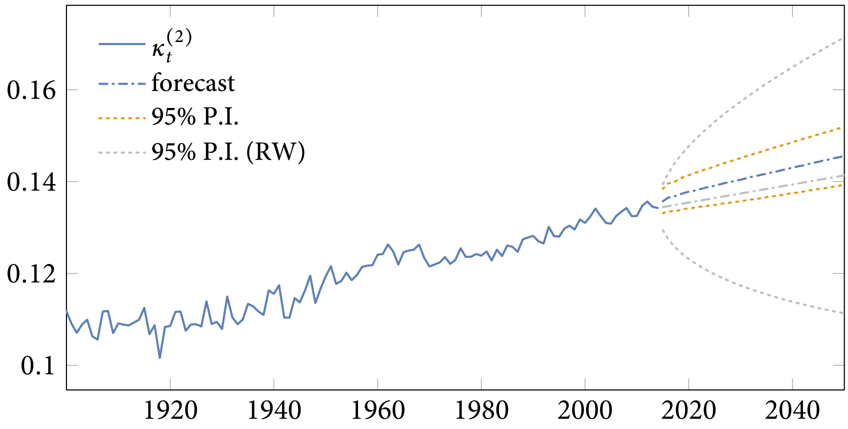
2.5.2 | VECM Projections for the KAN:2 Model

The projections for the VECM of the KAN:2 model, as given in eq. (2.53), are illustrated in Figure 2.11. Similar observations to the previous case can be made here. The central projections, obtained by the VECM, differ from those of the RW. The VECM implies higher improvements of the mortality rates across all ages. This can be determined by the lower $\kappa_t^{(1)}$ values of the VECM, as displayed in Figure 2.11(a). On the other side, higher $\kappa_t^{(3)}$ values of the VECM, compared to the RW projection, imply lower mortality improvement rates for very high ages. Furthermore, the prediction intervals of the VECM projection are noticeably smaller than those of the RW.

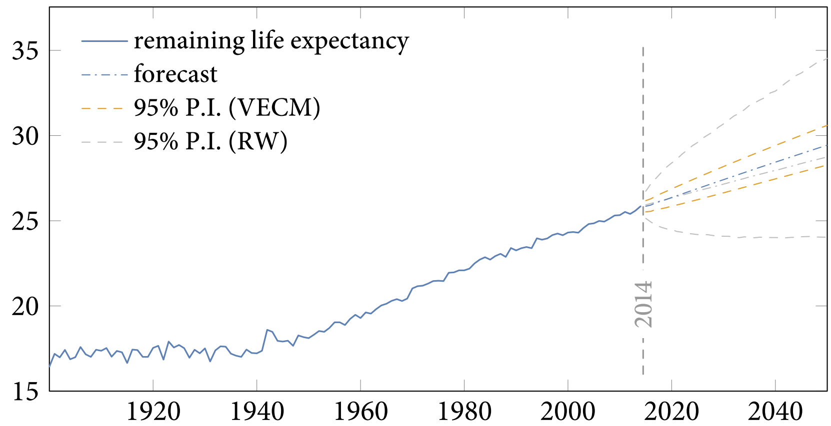
Figure 2.12 illustrates the KAN:2 projections of the remaining life expectancy of the Swedish female population aged 60. For comparison, we also include the projections by the RW (grey) and the VECM of the KAN model (green). Regarding the remaining life expectancy, the projections of the KAN and KAN:2 model are quantitatively very similar. The KAN:2 model leads to slightly higher values, where the remaining life expectancy increases to 29.51 years until the projection horizon at 2050. The average improvements of 38.1 days per year of the 36 periods ahead forecast coincides exactly with the historical average of the past 36 years. The prediction intervals of KAN:2 projections are also slightly narrower than those of the KAN model. Despite the similarities, the age-dependent



(a) Trajectories and projections of $\kappa_t^{(1)}$ under a VECM.



(b) Trajectories and projections of $\kappa_t^{(2)}$ under a VECM.



(c) Historical and projected remaining life expectancy for Swedish females aged 60.

Figure 2.8: VECM projections of the KAN model coefficient. Dashed lines represent the central forecasts and the dotted lines show the 95% prediction intervals.

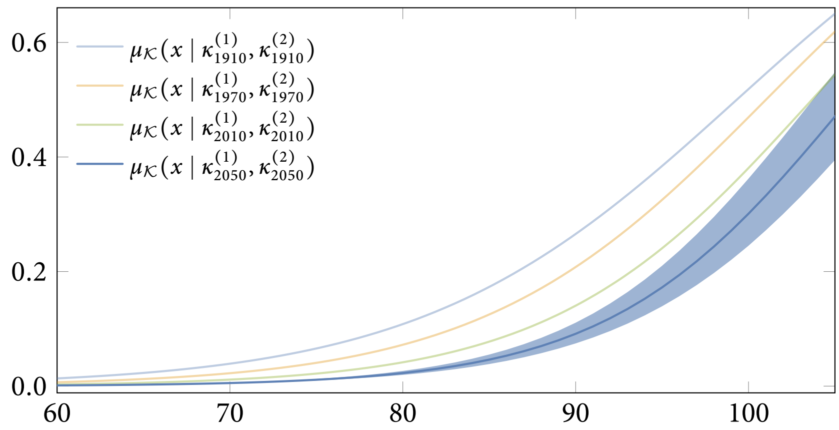


Figure 2.9: Projected hazard function at 2050 with 95% prediction intervals (Swedish females).

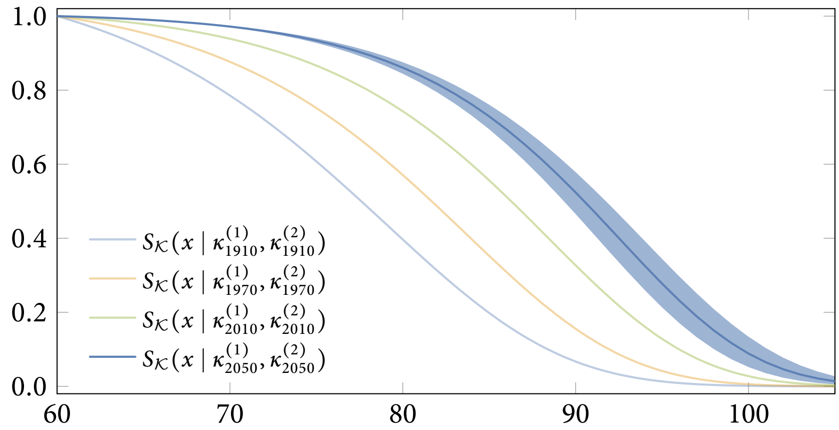


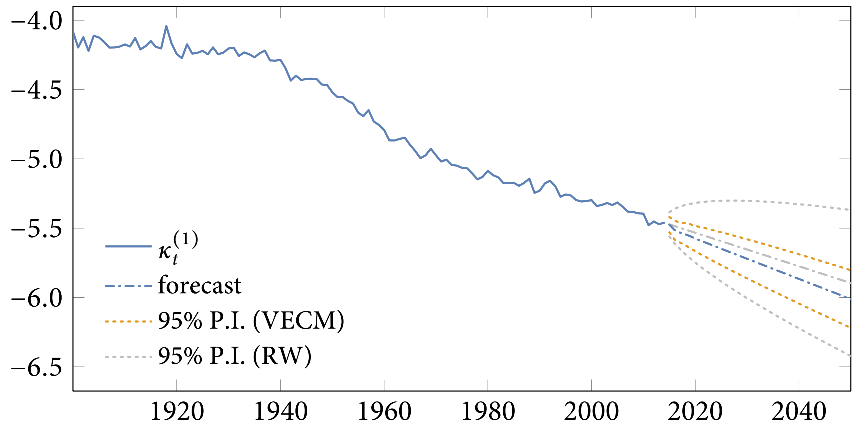
Figure 2.10: Projected survival function at 2050 with 95% prediction intervals (Swedish females).

mortality improvements of these two models differ significantly, as we illustrate in Figures 2.15 and 2.16.

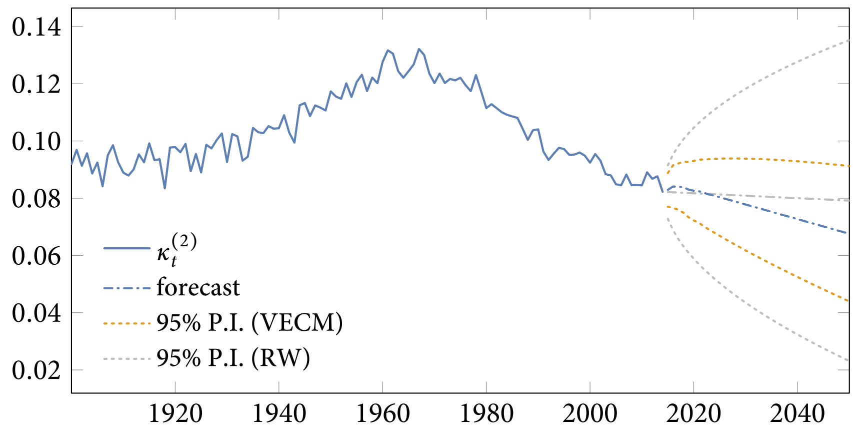
2.5.3 | VECM Projections for the KAN:3 Model

Figure 2.13 illustrates the projections of the KAN:3 coefficients by the VECM of eq. (2.55). As for both previous cases, these projections are obtained using eqs. (2.58) and (2.59) and substituting the estimates from eqs. (2.56) and (2.57) for the coefficient matrices A_1,A_2,A_3 and Σ_u. The comparison to the RW projection shows different slopes of the central forecasts as well as narrower prediction intervals as for the lower dimensional models KAN and KAN:2. The projected remaining life expectancy is quantitatively very similar to the KAN:2 model, as shown in Figure 2.12. The same applies to age-dependent mortality improvements, as will be illustrated in Figure 2.17.

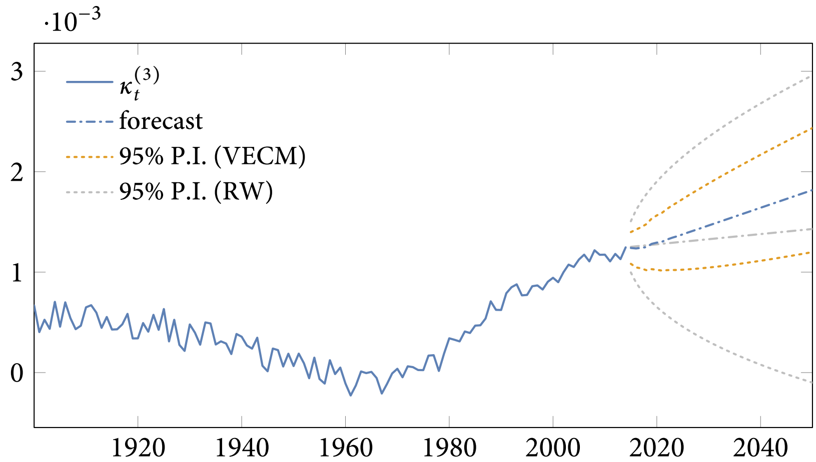
Next, we want to highlight a fundamental difference between the RW and VECM projection. Since the individual components of the Kannisto model are non-stationary, the RW is fitted to the first-order difference $\Delta\boldsymbol{\kappa}_t$, see eq. (1.90). However, by differencing variables, one potentially loses the long-run relations between those variables. Thus, it is highly recommended to first check



(a) Trajectories and projections of $\kappa_t^{(1)}$ under a VECM.



(b) Trajectories and projections of $\kappa_t^{(2)}$ under a VECM.



(c) Trajectories and projections of $\kappa_t^{(3)}$ under a VECM.

Figure 2.11: VECM projections of the KAN:2 model coefficient with 95% prediction intervals.

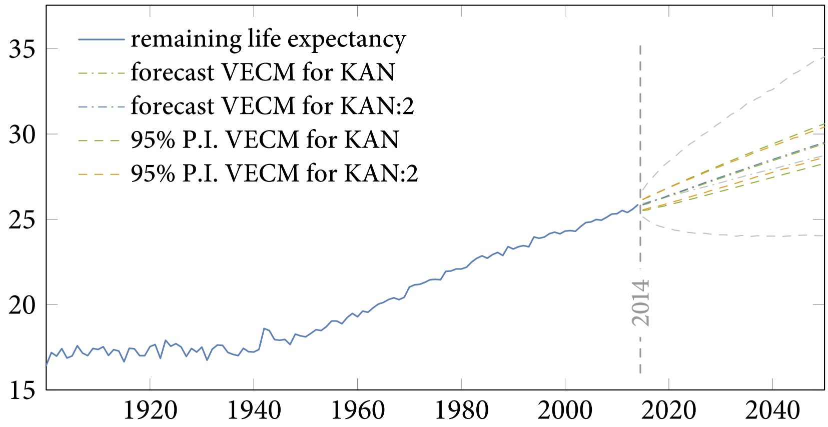


Figure 2.12: Historical and projected remaining life expectancy for Swedish females aged 60. Comparison of the VECM for KAN versus VECM for KAN:2.

the presence of cointegration relations before trying to stationarize the corresponding time series through differencing. If cointegration relations exist, then the individual components are not moving completely independent but rather follow a long-run equilibrium relation. This has a beneficial impact on the prediction intervals obtained by a VECM. Notice, from Figure 2.13, that prediction intervals of the individual coefficients increase over time. However, the prediction intervals of the process $\boldsymbol{\beta}'\boldsymbol{\kappa}_t + \xi_{co}t$ remain bounded since this process is stationary. The forecast of the equilibrium relation is illustrated in Figure 2.14. Notice that at the beginning of the projection period, the time series reverts towards their equilibrium state. Furthermore, we observe stable prediction intervals which match well to the historical progression of the time series.

2.5.4 | Mortality Improvements

In Figure 2.12, we showed the improvements of the remaining life expectancy obtained by the previously considered models. While this is a proper representation for the aggregated improvements, it contains no information about the age-related changes of the mortality rates. On the other hand, Figure 2.9 showed the absolute age-related changes of the mortality rates, but since the rates differ by two orders of magnitude on the involved range of ages, it is not a suitable representation to cover the changes for lower and higher ages simultaneously. A more appropriate representation can be achieved by considering the relative changes of the hazard rates, the so-called *mortality improvements*. In a time discrete setting, mortality improvements, denoted by $j(t,x)$, are defined as

$$j(t,x) := -\frac{\mu(t,x) - \mu(t-1,x)}{\mu(t-1,x)}, \tag{2.60}$$

where $\mu(t,x)$ are the hazard rates at time t for the age x. From eq. (2.60) it is clear that the mortality improvements are defined in terms of relative changes of the hazard rates from period $t-1$ to period t. Positive mortality improvements express decreasing hazard rates, while negative mortality improvements represent increasing hazard rates. For continuous time models, the definition of

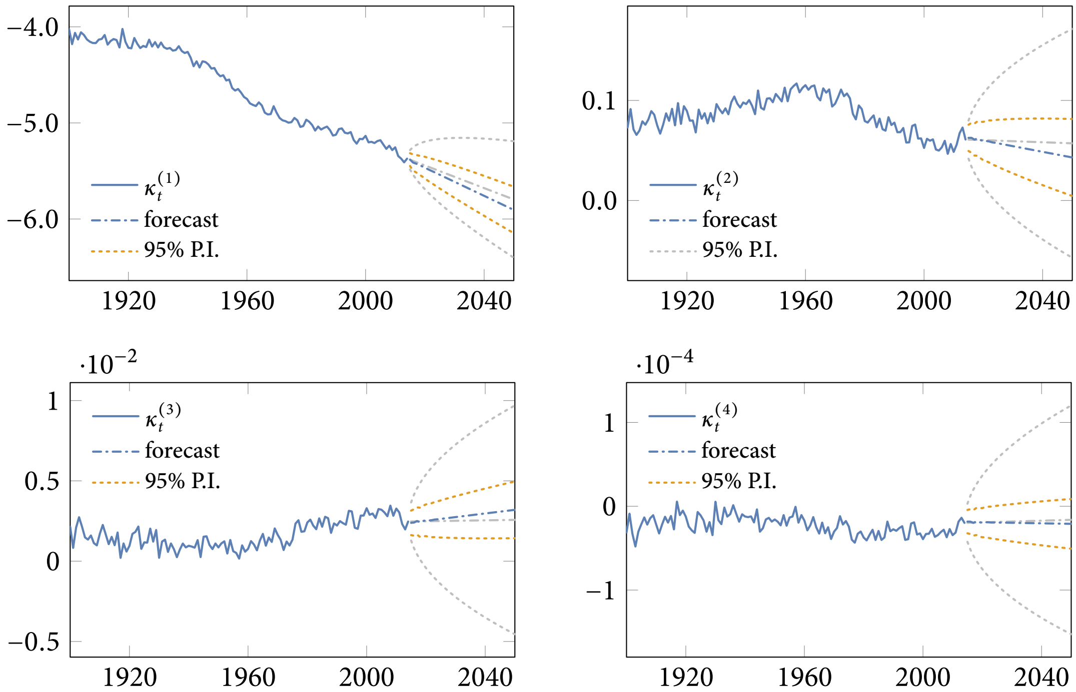


Figure 2.13: VECM projections of the KAN:3 coefficients. Dashed lines represent the central forecasts and dotted lines show the 95% prediction intervals.

mortality improvements is

$$j(t,x) := -\frac{d}{dt} \ln \mu(t,x). \tag{2.61}$$

This definition reveals a conceptional similarity to the relation of hazard rates and their survival functions. While mortality improvements are defined as (infinitesimal) relative changes of the hazard rates in time, hazard functions are defined as (infinitesimal) relative changes in age, as will be discussed in Section 3.2.1.

The objective of the following is to highlight the differences of the projection obtained by the VECMs for the KAN, KAN:2 and KAN:3 models. Figures 2.15 to 2.17 illustrate the mortality improvements of those three models. Note that the high fluctuations at the beginning of the projection occur due to the high fluctuation of the raw mortality rates, which are commonly greater than 10%. The historical mortality improvements, as shown in Figures 1.7 to 1.10, were obtained by the B/P-spline smoothing method. By comparing the forecasts of the KAN model, as displayed in Figure 2.15, with those of KAN:2 in Figure 2.16, we observe that the KAN model shows higher improvement rates of about 2% for the ages around 60, while the KAN:2 model projects improvement rates of only about 1.5%. On the other hand, the improvement rates of the KAN:2 model are higher compared to the KAN model for the ages between 70 and 90, reaching a maximum of 1.75% for at the ages around 75. Above the age of 90, the KAN model shows again higher improvement rates of about 1%

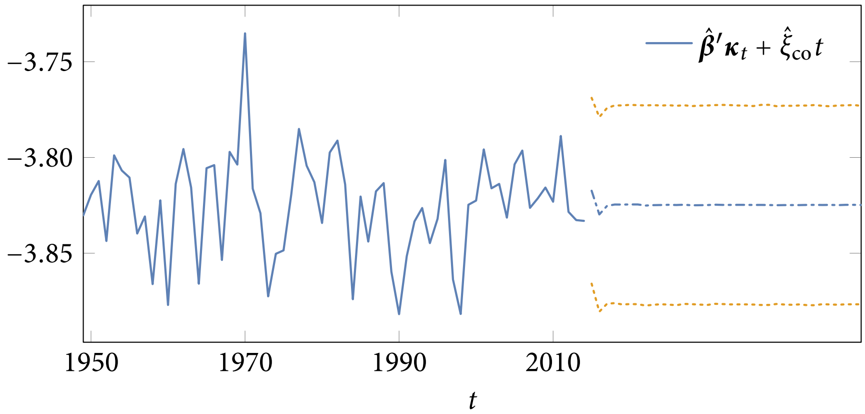


Figure 2.14: Estimated equilibrium relation $\hat{\beta}'\kappa_t + \hat{\xi}_{co}t$ of the KAN:3 model together with VECM projections and surrounding 95% prediction intervals.

for the age 95, compared to 0.75% of the KAN:2 model. Despite the deviations of the improvement rates, both models yield almost the same remaining life expectancy, as shown in Figure 2.12. The difference of the KAN:2 and KAN:3 model is only marginal. The KAN:3 model shows slightly higher improvements until the age of 75 and slightly lower values above. All projections have in common that the improvement rates go to zero as the age increases. This observation also provides an explanation for the rectangularization effect of the survival function since the historical, as well as the projected improvement rates, tend to be higher for lower ages.

2.5.5 | Projections of the Life Expectancy for Real and Synthetic Cohorts

In Section 1.9, we defined the Kannisto models as members of the GAPC mortality models family, by proposing specific logistic-type hazard rates (see eqs. (1.85) and (1.86)). This represents the static part of the mortality model, where for every fixed period the parametric hazard rates capture age-related mortality effects. In Section 2.4, we studied the dynamics of our model, where the evolution of the Kannisto parameters are modelled by discrete time series, such as VECMs. It is important to point out that calculations of survival probabilities, based on the hazard rates at fixed periods, would lead to survival probabilities of synthetic cohorts rather than real cohorts. In Section 1.2.1, we defined a synthetic cohort (T,x) as a group of individuals with age x in T, who throughout their life, experience the age-specific mortality rates $\mu(T,x+i)$ for $i=1,\ldots,i_{\max}$. This was referred to as the vertical arrangement in Table 1.1. Note that a real cohort $(T,x)^*$ is a group of individuals who, throughout their life, experience the age-specific mortality rates $\mu(T+i,x+i)$ for $i=1,\ldots,i_{\max}$. This case was referred to as the diagonal arrangement in Table 1.1.

Synthetic cohorts serve as an auxiliary construction, which is used to obtain a modelling approach for real cohorts. Note that the reason why the parametric hazard rates were fit to periodic mortalities (vertical arrangement) rather than to mortalities of real cohorts (diagonal arrangement), is that for recent real cohorts only few data are available. Thus, it is infeasible to fit a non-linear curve if only a few data points are available. However, a forecast from a dynamic model for the periodic hazard rates can be used to obtain projections of survival probabilities for real cohorts.

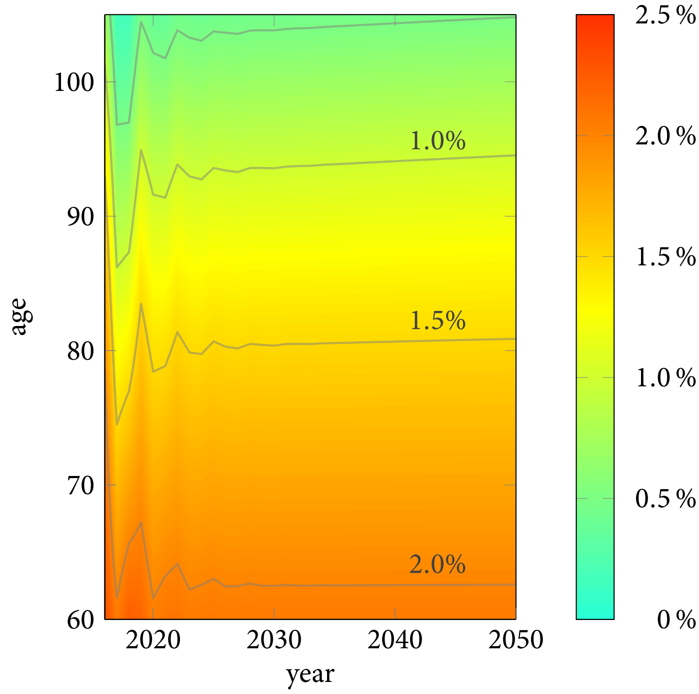


Figure 2.15: KAN projected mortality improvements of the Swedish female population.

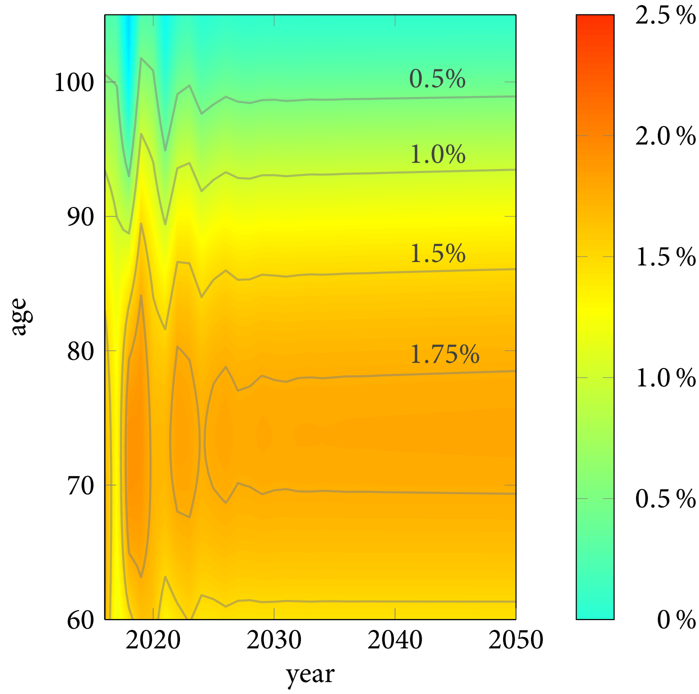


Figure 2.16: KAN:2 projected mortality improvements of the Swedish female population.

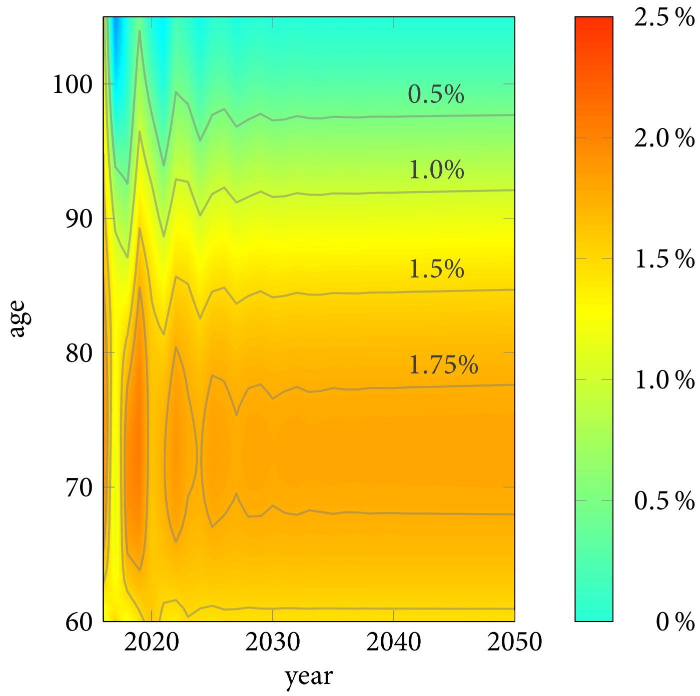


Figure 2.17: KAN:3 projected mortality improvements of the Swedish female population.

Figure 2.18 displays the projected survival functions for the Swedish female population of the synthetic cohort $(2015{,}60)$ and the real cohorts $(2015{,}60)^*$, $(2030{,}60)^*$ and $(2050{,}60)^*$. The life expectancy of the synthetic cohort $(2015{,}60)$ is 86.47. This calculation is based on the information available at 2015 and involves the mortality rates $\mu(2015{,}60),\ldots,\mu(2015{,}110)$. The life expectancy of the real cohort $(2015{,}60)^*$ is 88.54 and is based on the projected mortality rates $\mu(2015{,}60),\ldots,\mu(2065{,}110)$. The difference of more than two years implies that one would significantly underestimate the life expectancy of the real cohort by only using the periodic data of 2015 without including the corresponding mortality improvements. The projected life expectancies of the other two displayed real cohorts $(2030{,}60)^*$ and $(2050{,}60)^*$ are 90.22 and 92.31 years.

Another example for the improvements of the life expectancy of real cohorts, based on the VECM forecast, is shown in Figure 2.19, which illustrates the historical and the forecasted values for the ages 60, 70, 80, and 90. The solid lines represent the life expectancies purely based on available data, while the dotted lines show the life expectancies obtained by available and projected data. For instance, if we take the cohort $(1990{,}60)^*$, then the calculation of the life expectancy involves the already observed mortalities $\mu(1990{,}60),\ldots,\mu(2014{,}84)$ as well as the projected mortalities $\mu(2015{,}85),\ldots,\mu(2040{,}110)$. From Figure 2.19, we can observe that the life expectancies progress with different slopes. This can be attributed to the higher mortality improvements for lower ages. Notice also that the gaps between $e^*_{T,60}$, $e^*_{T,60}$, $e^*_{T,80}$ and $e^*_{T,90}$ continue to decrease due to the rectangularization effect of the involved survival functions.

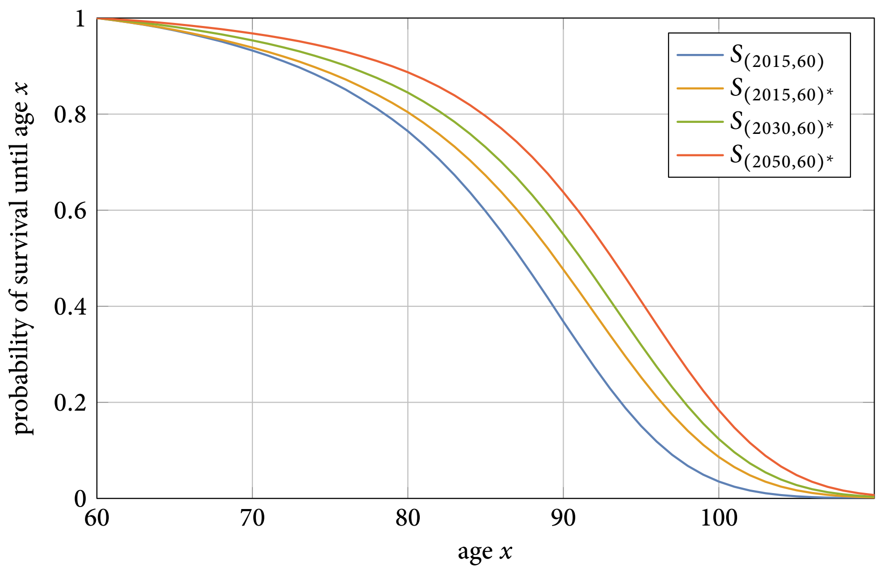


Figure 2.18: Projected survival functions of the synthetic cohort $(2015{,}60)$ and real cohorts $(2015{,}60)^*$, $(2030{,}60)^*$ and $(2050{,}60)^*$.

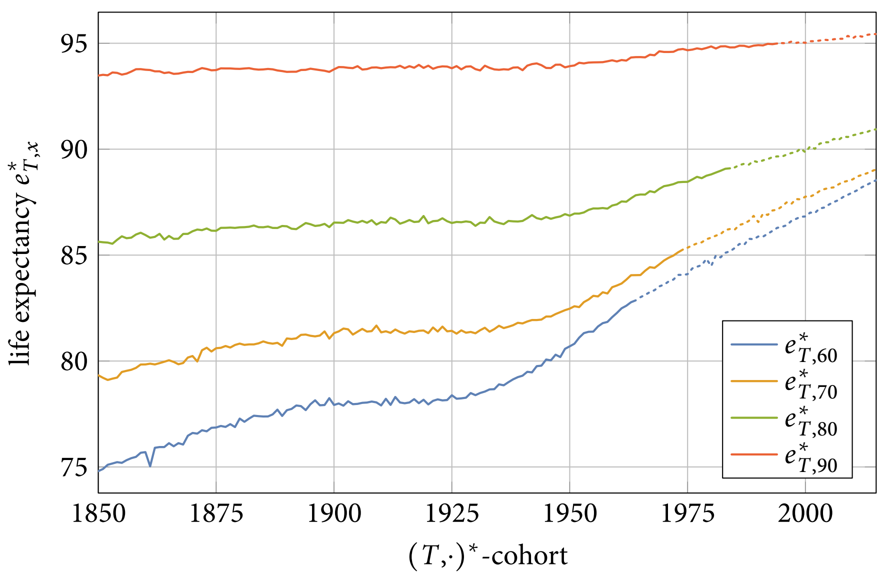


Figure 2.19: Real cohort life expectancies $e^*_{T,x}$ at age $x \in \{60, 70, 80, 90\}$ for Swedish female $(T,x)^*$-cohorts. This corresponds to the mean attained age of an x-survivor.

2.6 | Conclusion

This chapter presents the dynamic part of our proposed stochastic mortality model, where the evolution of the Kannisto parameters are modelled by a VECM. We recall that the Kannisto model is a multivariate parametric model for the age-related mortality structure (see Section 1.9.1). The Kannisto parameters can be understood as influencing factors of particular age groups on the mortality. For these factors, we use a VECM, a general modelling approach for the dynamics of a system of variables, which is, in particular, capable to capture long-run equilibrium relations between the individual components. By applying the common specification procedure for VECMs, we first demonstrate that cointegration relations exist for the time series of the models KAN, KAN:2 and KAN:3, and moreover that the estimated VECMs provide a good representation of the underlying DGPs. For all Kannisto time series, the preferred VECM uses one cointegration relation, which implies that the individual components do not move independently but rather have common stochastic trends.

The result that several stochastic factors, which drive the mortality rates, have long-run equilibrium relations has a strong impact on strategies managing longevity risk. On the one hand, there are oversimplified models with only one stochastic factor, like the Lee-Carter model, which implicitly assumes that changes in the mortality at one age can be perfectly hedged with changes at a different age. On the other hand, there are mortality models with multiple stochastic factors. The downside is that additional degrees of freedom could have a negative impact on the forecast mean squared error which leads to higher forecast uncertainties. However, if cointegration relations between these stochastic factors exist, then a VECM can be highly beneficial for the forecast performance.

By comparing the VECM projections to those obtained by the standard approach using a random walk, we demonstrate a better forecast performance of the VECM. Furthermore, the obtained central projections, as well as the prediction intervals, are more consistent with the historical experience. By using the VECM projections, we obtain substantially lower forecast MSEs. This is highly important for the life insurance industry since the forecast uncertainty is part of the risk which has to be priced for mortality-linked products. Finally, it is important to point out, that the VECM methodology is not limited to the Kannisto model and could also be successfully applied to other GAPC models with multiple stochastic factors.

Chapter 3 | Characterization of the Kannisto and the Extended Exponential Distribution

Contents

3.1 | Introduction

The objective of the following chapter is to characterize the Kannisto distribution which is determined by a logistic-type hazard rate function as used for the KAN model in Section 1.9. In the following, the age x is not treated as a discrete integer valued variable as in Chapter 1, but as a continuous variable. In analogy to probability densities or cumulative distribution functions, hazard rate functions are equivalent representatives of distributions. By characterizing the instantaneous risk of failure associated with certain age, hazard rate functions are the most preferred representatives of distributions in the context of survival analysis.

As noted in Section 1.9, the logistic-type hazard rate function, which we call the Kannisto hazard rate was first proposed by the demographer Väinö Kannisto who studied mortality rates of high ages, see Kannisto (1992) and Thatcher, Kannisto and Vaupel (1998). The approach by Kannisto was similar to those of his predecessors, who only aimed to identify a parametric curve which minimizes the Euclidean distance for the observed data. Commonly this process initially included a logarithmic or logit transformation of the data, followed by applying a (weighted) linear model to estimate the model parameters. Apart from this modelling approach, Kannisto did not study the properties of the continuous distribution induced by the logistic-type hazard rate function he proposed.

Unlike other life distributions, the Kannisto distribution is widely uncharacterized. Some results may be found in, e.g., Marshall and Olkin (2007) and Missov (2013). In the following, we will prove some properties of the Kannisto distribution and provide an extensive characterization. The chapter is outlined as follows. In Section 3.2, we initially review some standard concepts of survival analysis. Subsequently, we reveal some connections of the Kannisto distribution to other well-known distributions in Section 3.3. Some of these connections are based on the fact that the Kannisto distribution is a special case of the so-called extended exponential distribution which is analysed by Marshall and Olkin (2007), a standard reference on parametric and non-parametric life distributions. In Section 3.5, we provide an extensive characterization of both, the Kannisto and the extended exponential distribution, covering topics such as mean residual life function, moment generating function, central moments, order statistics, maximum, and minimum domain of attraction, Fisher information matrix, and Kullback-Leibler divergence.

3.2 | Survival Analysis

There exists a vast literature on mortality modelling based on parametric-type mortality rates. These models aim to describe the pattern of the instantaneous age-specific failure rate, also known as the hazard rate or in actuarial literature as the force of mortality. Traditional actuarial models can be found in Moivre (1725), Gompertz (1825), Makeham (1860), Beard (1961), Heligman and Pollard (1980) and Kannisto (1992). Carriere (1992), and Marshall and Olkin (2007) also cover other hazard rate models such as Weibull, Inverse-Gompertz, and Inverse-Weibull. For a comprehensive review of age-specific models for human populations, we refer to Gavrilov and Gavrilova (1991).

This section provides some background material and notations which are needed in the subsequent parts. The terminology covered here is standard in survival analysis and is mostly based on Marshall and Olkin (2007). We will begin by introducing hazard rate functions and other common representatives of lifetime distributions. We then proceed with basic concepts of survival analysis, including residual life distribution, competing risks and subsequently show how many familiar distributions are based on elementary hazard rate functions.

3.2.1 | Lifetime Distributions and their Representatives

In this section, we describe some basic concepts of univariate survival analysis. We introduce the notation and cover some results related to mortality modelling.

The *lifetime* of an individual of some population is represented by a continuous non-negative random variable X. Generally, lifetimes or survival times describe the duration between entering and escaping from particular states. In non-life related applications, the escaping is often called *failure*, while in life-related applications it stands for the *death* of an individual. Unless stated otherwise, we will only consider lifetime distributions with an unbounded support, i.e., all of the lifetime representative functions will be defined over the interval $[0,\infty)$.

In the following part, we summarize the so-called *lifetime representatives* which describe the distribution of a random variable. The following representatives will be discussed below (cf., e.g., Rinne, 2014):

- the *cumulative distribution function*, denoted by $F(\cdot)$, abbreviated by (CDF),
- the *survival function* denoted by $S(\cdot)$, also known as *reliability function*,
- the *probability density* denoted by $f(\cdot)$, abbreviated by (PDF),
- the *hazard rate function*, denoted by $h(\cdot)$ or $\mu(\cdot)$, also known as *instantaneous failure rate*, *force of mortality* or *mortality curve*,

- the *cumulative hazard rate* denoted by $H(\cdot)$,
- the *mean residual life function* denoted by $v(\cdot)$, where $v(x)$ is also called the mean future life of an x-survivor.

The representatives define the lifetime distributions and can be obtained from each other when they exist, see Marshall and Olkin (2007). In Section 3.2.7, we illustrate how popular distributions naturally arise from elementary hazard rate curves.

Definition 3.2.1. The function F defined for $x \in [0,\infty)$ by

$$F(x) := \mathbb{P}[X \leq x] \tag{3.1}$$

is called *distribution function* or *cumulative distribution function* of the lifetime random variable X.

The distribution function gives the probability of failure/death up to time or age x. To distinguish distribution functions associated to different lifetime variables, we use a subscript, such as F_X, to indicate the corresponding random variable. Any function satisfies the properties:

(a) $F(0) = 0$,

(b) $\lim_{x\to\infty} F(x) = 1$,

(c) $F(x_b) \geq F(x_a) \quad$ for all $x_b \geq x_a$,

(d) F is continuous,

is a distribution function of some continuous lifetime variable.

Definition 3.2.2. The function S defined for $x \in [0,\infty)$ by

$$S(x) := \mathbb{P}[X > x] \tag{3.2}$$

is called the *survival function* or sometimes the *reliability function*.

The survival function gives the probability of an individual surviving to time x or exceeding the age x. From eq. (3.1) and eq. (3.2) it is clear, that the relationship between the distribution function and the survival function is given by

$$S(x) = 1 - F(x).$$

A survival function S is monotone decreasing over $[0,\infty)$, is continuous and it satisfies $S(0) = 1$ and $\lim_{x\to\infty} S(x) = 0$. Furthermore, since we consider unbounded lifetimes, we have $S(x) > 0$ for all $x > 0$.

Definition 3.2.3. If f is a measurable non-negative function such that for all $x \in [0,\infty)$

$$F(x) = \int_0^x f(y)\,dy,$$

then f is called a *probability density function* or simply *density function* of F.

When a density exists then the random variable X is called *continuous* or *absolutely continuous* since the measure induced by X is absolutely continuous to the Lebesgue measure. Note that the density function is not unique since it can be changed on Lebesgue zero sets and remain a density function of X. The relations between a density function f and the survival function S of some random variable are given by

$$S(x) = \int_x^\infty f(y)\,dy$$

and

$$f(x) = -\frac{d}{dx}S(x).$$

Definition 3.2.4. A distribution with survival function S is defined to be *(right)-heavy-tailed* if and only if

$$\lim_{x\to\infty} S(x)\,e^{\lambda x} = \infty \quad \text{for all } \lambda > 0.$$

Otherwise, a distribution with survival function S is defined to be *light-tailed* if and only if

$$\lim_{x\to\infty} S(x)\,e^{\lambda x} < \infty \quad \text{for some } \lambda > 0.$$

Remark 3.2.5. Note that by the exponential Chebyshev inequality, the distribution is light-tailed if and only if the survival function is dominated by an exponential function, i.e., for some $c_1 > 0$ and $c_2 > 0$, we have $S(x) \le c_1 e^{-c_2 x}$ for all x.

Another important representative of a lifetime distribution, especially in mortality modelling, is the *hazard rate function* h also referred to as the *force of mortality* or *age-specific death rate* and is usually denoted by μ.

Definition 3.2.6. If F is an absolutely continuous lifetime distribution function with density f then the function h defined on the interval $[0,\infty)$ by

$$h(x) := \begin{cases} \frac{f(x)}{S(x)} & \text{if } S(x) > 0 \\ \infty & \text{if } S(x) = 0 \end{cases} \tag{3.3}$$

is called a *hazard rate* of F or X.

The hazard rate function characterizes the instantaneous risk of failure associated with certain age. Note that the hazard rate can be expressed as

$$h(x) = \lim_{\varepsilon\downarrow 0} \frac{\mathbb{P}\left[x \le X < x+\varepsilon \mid X \ge x\right]}{\varepsilon}.$$

Therefore, we have approximately

$$h(x)\,\varepsilon \approx \mathbb{P}\left[x \le X < x+\varepsilon \mid X \ge x\right],$$

i.e., for a small increment in time ε, the product $h(x)\,\varepsilon$ is the approximate probability of failure in the interval $[x,x+\varepsilon)$, given survival to x. This interpretation makes the hazard rate presumably the most preferred lifetime representative in theoretical considerations as well as in actuarial applications. The hazard rate reveals and characterizes the process of ageing more intuitively than other lifetime representatives and therefore lifetime distributions are often classified by the properties of their hazard rate function. These classifications distinguish between monotone increasing or decreasing hazard rates describing *wearout/ageing* or *wearin/improving* effects. Another important class of lifetime distributions are those with non-monotone bathtub-shaped hazard rates, which can be observed in life tables, where the infant mortality rates initially decreasing, see Figures 1.3 and 1.4.

Proposition 3.2.7. A function h is a hazard function of some non-negative random variable if and only if:

(a) $h(x) \ge 0, \quad \forall x > 0,$

(b) $\int_0^\infty h(y)dy = \infty,$

(c) $\exists x > 0 : \int_0^x h(y)dy < \infty,$

(d) $\int_0^x h(y)dy = \infty \Rightarrow h(z) = \infty, \quad \forall z > x.$

Unless stated otherwise, we will only consider non-bounded lifetimes, such that

$$\int_0^x h(y)dy < \infty, \text{ for all } x > 0,$$

thus h is a hazard function of some non-negative and non-bounded random variable if and only if properties (a) and (b) are satisfied.

Proof. First, suppose that h is a hazard function, then (a) follows since $f(x) \ge 0$ and $S(x) \ge 0$, thus

$$h(x) = \frac{f(x)}{S(x)} \ge 0.$$

(b) holds since $S(x) = e^{-\int_0^x h(y)\,dy}$ and $\lim_{x\to\infty} S(x) = 0$. Next, suppose that a function h fulfills (a) and (b). Define the function f as

$$f(x) := h(x)e^{-\int_0^x h(y)\,dy}.$$

To show that f is a probability density function of a non-negative random variable, we have to show the properties

(i) $f(x) \ge 0, \quad \forall x \ge 0,$

(ii) $\int_0^\infty f(x)\,dx = 1.$

The property (i) holds, since $h(x) \geq 0$ for all $x > 0$. With $S(x) = e^{-\int_0^x h(y)\,dy}$, we obtain

$$\begin{aligned}\int_0^\infty f(x)dx &= \int_0^\infty \left(h(x)e^{-\int_0^x h(y)\,dy}\right)dx \\ &= -\int_0^\infty \left(\frac{d}{dx}e^{-\int_0^x h(y)\,dy}\right)dx \\ &= -\int_0^\infty S'(x)\,dx \\ &= -\left(\underbrace{\lim_{x\to\infty} S(x)}_{=0} - S(0)\right) = 1.\end{aligned}$$

□

Definition 3.2.8. The function H defined on the interval $[0,\infty)$ by

$$H(x) := -\ln S(x) \tag{3.4}$$

where S is the survival function of X is called *cumulative hazard rate* of the lifetime X.

From eq. (3.3) and eq. (3.4) and the usual assumption for lifetime distributions, i.e., $F(0) = 0$, we see that

$$S(x) = e^{-H(x)} = e^{-\int_0^x h(y)\,dy}. \tag{3.5}$$

Furthermore, from eq. (3.5), we can see that when F is an absolutely continuous distribution then

$$h(x) = \frac{d}{dx}H(x) = -\frac{d}{dx}\ln S(x) = -\frac{\frac{d}{dx}S(x)}{S(x)}. \tag{3.6}$$

The cumulative hazard rate satisfies the following conditions:

(a) $H(0) = 0$,

(b) $\lim_{x\to\infty} H(x) = \infty$,

(c) $H(x)$ is non-decreasing.

Remark 3.2.9. From the right term of eq. (3.6), we see that the hazard rate is the negative of the so-called *logarithmic derivative* of the survival function. Let f be a real-valued function without roots, the logarithmic derivative $L(f)$ of f is defined by

$$L(f) = \frac{f'}{f}.$$

Intuitively it is clear that the logarithmic derivative of f denotes the relative infinitesimal change, which is the absolute infinitesimal change of f, namely f' scaled by the reciprocal of the function f, i.e., by $1/f$. Thus, the hazard rate h can be expressed as the logarithmic derivative of the survival function S by

$$h(x) = -\frac{S'(x)}{S(x)} = -L\big(S(x)\big).$$

The logarithmic derivative does also appear in the definition of mortality improvements (see eq. (2.61)), a quantity which describes the change of the hazard rate over time. Similar to a hazard rate, which is the negative logarithmic derivative of the survival function with respect to age, we introduced the mortality improvements as the negative logarithmic derivative of the hazard rate with respect to time.

Proposition 3.2.10. A distribution is light-tailed if and only if the cumulative hazard function $H(\cdot) = -\ln S(\cdot)$ satisfies

$$\lim_{x\to\infty} \frac{H(x)}{x} \neq 0.$$

Proof. First, suppose the distribution is light-tailed. By the exponential Chebyshev inequality, we have some $c_1, c_2 > 0$ such

$$S(x) = e^{-H(x)} \leq c_1 e^{-c_2 x} \quad \text{for all } x > 0.$$

This implies that $\lim_{x\to\infty} H(x)/x \geq c_2 \neq 0$. Suppose now that $\lim_{x\to\infty} H(x)/x > 0$, then there exist x_0 and $c > 0$ such $H(x) \leq cx$ for all $x \geq x_0$. Thus, we have

$$S(x) \leq e^{-cx} \quad \text{for all } x \geq x_0.$$

This implies that the distribution is light-tailed. □

3.2.2 | Expectation and Higher Central Moments

Clearly, the expectation of the random lifetime X corresponds to the individual life expectancy. Next, we recall how the expectation and higher raw moments of continuous lifetime random variables can be expressed in terms of its survival function.

Definition 3.2.11 (Expectation). Let X be a random variable with distribution function F, such that the integral $\int_{-\infty}^{\infty} |x| \, dF(x) < \infty$, then the *expectation* $\mathbb{E}[X]$ of X is defined by

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \, dF(x). \tag{3.7}$$

For non-negative random variables with survival function S, the expectation can also be expressed by

$$\mathbb{E}[X] = \int_{0}^{\infty} S(x) \, dx.$$

This relation follows from eq. (3.7) by using Fubini's theorem. This form is often more convenient for actuarial purposes. If the random variable corresponds to a lifetime, then the expectation is also called *life expectancy*. In the context of the thesis, the expectation is denoted by v. The *n-th moment* of a non-negative random variable X with distribution function F is defined as

$$v_n = \mathbb{E}[X^n] = \int_{0}^{\infty} x^n \, dF(x). \tag{3.8}$$

An alternative expression for the n-th moment of the non-negative random variable X in terms of the survival function is given by

$$v_n = \mathbb{E}[X^n] = n \int_{0}^{\infty} x^{n-1} S(x) \, dx.$$

This expression can be derived from eq. (3.8) using integration by parts.

The following hazard rate based criteria for the existence or non-existence of moments was originally given by Barlow, Marshall and Proschan (1963).

Proposition 3.2.12 (Marshall and Olkin, 2007). Let h be a hazard rate of a distribution on $[0,\infty)$, $n > 0$, and let v_n denote the n-th moment. If the inequality

$$n < \liminf_{x \to \infty} x \, h(x)$$

holds, then $v_n < \infty$. On the other hand, if

$$n > \limsup_{x \to \infty} x \, h(x)$$

holds, then $v_n = \infty$.

Proof. For a proof of this result, see Marshall and Olkin (2007, Proposition 20.B.6). □

A direct consequence of Proposition 3.2.12 is that for every distribution with non-decreasing hazard function all moments exist. On the other hand, if this criterion is applied to a distribution with a decreasing hazard rate, such as the Pareto distribution (see, eq. (3.22)), we obtain

$$\liminf_{x\to\infty} x\, h_{\mathcal{P}_{\mathrm{II}}}(x \mid k,\alpha,\mu) = \liminf_{x\to\infty} x \frac{\alpha}{k-\mu+x} = \alpha.$$

Thus, we have $\nu_n < \infty$ if $n < \alpha$.

3.2.3 | Residual Life Distribution

In survival analysis, one is often interested in the distribution of the remaining lifetime for individuals who accomplished to survive until a particular time. The relation of the remaining lifetime and the original lifetime is defined in the following.

Definition 3.2.13. Let S be the survival function of a lifetime distribution with an unbounded support X such that $S(0) = 1$ and $S(x) > 0$ for all $x > 0$. The *residual life distribution* S^t of S at t is defined by

$$S^t(x) = \frac{S(x+t)}{S(t)}, \quad x \geq 0. \tag{3.9}$$

It is clear from the above definition that the residual life distribution S^t is a conditional distribution of the remaining lifetime since we can express eq. (3.9)

$$S^t(x) = \frac{S(x+t)}{S(t)} = \mathbb{P}\left[X > x+t \mid X > t\right].$$

The corresponding conditional random variable $X - t \mid X > t$ with the survival function S^t is called the *residual lifetime of a t-survivor*. Suppose that X has a density f, then we obtain the density and the hazard rate of the residual lifetime distribution of a t-survivor by

$$f^t(x) = \frac{f(x+t)}{S(t)}, \quad \text{for } x \geq 0$$

and

$$h^t(x) = \frac{f(x+t)}{S(x+t)} = h(x+t), \quad \text{for } x \geq 0. \tag{3.10}$$

From the interpretation of S^t being a conditional distribution, it is clear that the hazard rate h^t associated to the remaining lifetime of a t-survivor is simply the original hazard rate h evaluated at the shifted argument $x + t$.

The logistic-type hazard rate functions are of particular interest of the thesis. The following proposition classifies the limiting residual life distribution of distributions where the hazard rate functions have finite positive limits.

Proposition 3.2.14 (Marshall and Olkin, 2007)**.** Let h be a hazard function and S the corresponding survival function of a lifetime variable. Suppose that the property

$$\lim_{x\to\infty} h(x) = \lambda, \quad \text{for some } \lambda \in \mathbb{R}_+ \tag{3.11}$$

holds, then the remaining lifetime with survival function S^t converges in distribution to an exponential distribution with parameter λ as $t \to \infty$.

Proof. By using the relation of eq. (3.10) and the assumption eq. (3.11), we conclude

$$S^t(x) = e^{-\int_0^x h^t(y)\,dy} = e^{-\int_0^x h(y+t)\,dy} = e^{-\int_t^{x+t} h(y)\,dy} \xrightarrow{t\to\infty} e^{-\lambda x}.$$

□

3.2.4 | Mean Residual Life Function

The next definition provides another distribution representative, which is mainly used in the area of survival analysis.

Definition 3.2.15. The *mean residual life function* $v(t)$ is the mean of the residual life distribution S^t as a function of t. Under the same assumption as for Definition 3.2.13, the mean residual life function is defined by

$$v(x) := \int_0^\infty \frac{S(x+t)}{S(t)}\,dx.$$

The mean residual life function $v(t)$ has the following properties:

(a) $v(t) \geq 0 \quad$ for all $t \geq 0$,

(b) $v(0) = \mathbb{E}[X] = v$,

(c) $S(t) = \frac{v(0)}{v(t)} e^{-\int_0^t \frac{1}{v(z)}\,dz}$.

The first two properties are clear from the definition. Property (c) shows how the survival function can be reconstructed from the mean residual life function. For a more detailed discussion and the proof of (c), see Cox (1962).

3.2.5 | Mixture Distributions

Survival data of humans might have some properties which can be better understood if the population is assumed to be non-homogeneous. Separation of gender, ethnic groups or various lifestyle characteristics is helpful to explain survival properties of the combined population as a mixture of individuals. The following definition provides mixture representations of inhomogeneous populations.

Definition 3.2.16 (Marshall and Olkin, 2007)**.** Let $\mathcal{F} = \{F(\cdot \mid \theta),\ \theta \in \Theta\}$ be a family of distributions and G a distribution on $\Theta \subset \mathbb{R}^d$, i.e., a distribution of the parameter θ. Then,

$$F(x) = \int_{\theta\in\Theta} F(x \mid \theta)\,dG(\theta) \tag{3.12}$$

is called *the mixture of F with respect to G* or *compound distribution* of F and G. $F(x|\theta)$ is known as the *kernel* and G is the *mixing (or compounding) distribution*.

In eq. (3.12), $\theta > 0$ is considered to be an unobservable random variable with distribution function G. A widely studied class of models is based on the assumption that the hazard function $\lambda(x)$ of a lifetime X has the form $\lambda_0(x) \times \theta$. These models are known as *frailty* models. $\lambda_0(x)$ is called the baseline hazard rate and θ represents an individual failure factor which scales the baseline hazard rate. Mixture models allow describing heterogeneity of populations. In Proposition 3.3.7, we will show a connection of the Kannisto and the Gompertz distributions, where the Kannisto distribution is obtained as a continuous mixture using a Gompertz kernel.

3.2.6 | Competing Risks

The lifetime can also be considered to be a composite if we decide to distinguish between different causes of failure. Models which refer to a cause-specific ending of life, such as different diseases or accidents, are also called *competing risks* models. In this framework, the failure is associated with the time of the first event of distinct causes of failures, where the single risk components are typically not observable, but only their minimum. Under these models, the lifetime is considered to be the *time-until-first-event*.

The following Proposition 3.2.17 shows that for independent competing risks the hazard function of the time-until-first-event is the sum of the single hazard rates of the involved risk components. Using this approach one can design complex age-specific failure profiles by combining risk components with decreasing and increasing hazard rates. Possible applications are bathtub-shaped hazard rates as observed for infants, where the hazard rate first decreases during the first years and eventually starts to increase, see Figures 1.3 and 1.4 for full age range empirical hazard rate profiles.

Proposition 3.2.17. Let $X_1, \ldots, X_n$ be independent (continuous) random variables with hazard functions $h_1, \ldots, h_n$. Then, the lifetime Y on n competing risks given as $Y := \min(X_1, \ldots, X_n)$ has the hazard function $h_Y = \sum_{i=1}^n h_i$.

Proof. From the independence of the random variables $X_1, \ldots, X_n$ follows that the survival function S_Y has the form

$$\begin{aligned} S_Y(y) &= \mathbb{P}\left[\min(X_1, \ldots, X_n) > y\right] \\ &= \mathbb{P}\left[X_1 > y, \ldots, X_n > y\right] \\ &= \prod_{i=1}^n \mathbb{P}\left[X_i > y\right] = \prod_{i=1}^n S_i(y), \end{aligned}$$

where $S_i(y) = \mathbb{P}\left[X_i > y\right]$ is the survival function of X_i. Since $S_i(y) = \exp\left(-\int_0^y h_i(x)dx\right)$, we have

$$S_Y(y) = \prod_{i=1}^n \exp\left(-\int_0^y h_i(x)dx\right) = \exp\left(-\int_0^y \sum_{i=1}^n h_i(x)dx\right) = \exp\left(-\int_0^y h_Y(x)dx\right).$$

□

3.2.7 | Distributions induced by Elementary Hazard Rate Functions

In the following section, we recall some continuous parametric distributions used in lifetime modelling from the perspective of hazard rates. We will define different types of hazard functions and derive the resulting distributions. This approach will show that many of the well-known distributions arise by a natural choice of hazard function. See, e.g., Rinne (2014) for an extended discussion on that topic.

For the clarification of categorical parameter types of the following distributions, we use the definition provided by Marshall and Olkin (2007).

Definition 3.2.18 (Scale, Frailty, and Tilt Parameters, (Marshall and Olkin, 2007)). A parametric family $S(\cdot \mid \beta)$ with $\beta > 0$ of the form $S(x \mid \beta) = S(\beta x \mid 1)$ is said to be a scale parameter family and β is called a scale parameter. A parameter θ is called a *frailty parameter* if $S(\cdot \mid \theta)$ is defined in terms of $S(\cdot)$ by the expression

$$S(x \mid \theta) = S(x)^{\theta}, \quad \theta > 0.$$

A parameter γ is called a *tilt parameter* if $S(\cdot \mid \gamma)$ is defined in terms of $S(\cdot)$ according to the expression

$$S(x \mid \gamma) = \frac{\gamma S(x)}{1 - (1-\gamma)S(x)}, \quad \gamma > 0. \tag{3.13}$$

A distribution with a tilt parameter is alternatively called a *proportional odds family*, since eq. (3.13) is equivalent to

$$\frac{F(x \mid \gamma)}{S(x \mid \gamma)} = \frac{1}{\gamma}\frac{F(x)}{S(x)}.$$

This section provides a foundation for Section 3.3, where we show many connections between popular distributions obtained by transformations, truncations, continuous mixtures, and as limiting distributions. In the following, we repeatedly start with an elementary type parametric hazard function and derive further representatives such as survival or density functions.

(a) A *constant* hazard rate

$$h(x \mid \lambda) = \lambda \mathbb{1}_{x \geq 0}, \quad \lambda > 0,$$

yields directly to

$$\begin{aligned} S(x \mid \lambda) &= e^{-\lambda x}\mathbb{1}_{x \geq 0}, \\ f(x \mid \lambda) &= \lambda e^{-\lambda x}\mathbb{1}_{x \geq 0}. \end{aligned} \tag{3.14}$$

This gives the exponential distribution $\mathrm{Exp}(\lambda)$.

(b) A *linear hazard rate* of the form

$$h(x \mid \lambda,\sigma) = (\lambda + \sigma x)\,\mathbb{1}_{x\geq 0}, \qquad \lambda,\sigma > 0, \tag{3.15}$$

leads to

$$S(x \mid \lambda,\sigma) = e^{-\lambda x - \frac{\sigma}{2}x^2}\,\mathbb{1}_{x\geq 0},$$
$$f(x \mid \lambda,\sigma) = (\lambda + \sigma x)e^{-\lambda x - \frac{\sigma^2}{2}x^2}\,\mathbb{1}_{x\geq 0}.$$

For $\lambda = 0$ this is the Rayleigh distribution with the scale parameter σ. From Proposition 3.2.17 follows that a competing risks lifetime with factors $\operatorname{Exp}(\lambda)$ and $\operatorname{Rayleigh}(\sigma)$ has a linear hazard rate as given in eq. (3.15).

(c) Starting with a *power hazard rate* given by

$$h(x \mid \alpha,\beta) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\mathbb{1}_{x\geq 0}, \qquad \alpha,\beta > 0,$$

we obtain

$$S(x \mid \alpha,\beta) = e^{-\left(\frac{x}{\beta}\right)^{\alpha}}\,\mathbb{1}_{x\geq 0},$$
$$f(x \mid \alpha,\beta) = \frac{\alpha}{\beta}e^{-\left(\frac{x}{\beta}\right)^{\alpha}}\left(\frac{x}{\beta}\right)^{\alpha-1}\mathbb{1}_{x\geq 0}.$$

This is the Weibull distribution, denoted by $\mathcal{WB}(\alpha,\beta)$, with shape parameter α and scale parameter β. The exponential distribution $\operatorname{Exp}(\lambda)$ and $\operatorname{Rayleigh}(\sigma)$ are special cases of the $\mathcal{WB}(\alpha,\beta)$ distribution, namely $\mathcal{WB}(1,{}^1\!/\!_\lambda)$ and $\mathcal{WB}(2,\sqrt{{}^2\!/\!_\sigma})$, respectively.

(d) For an *exponential hazard rate* of the form

$$h(x \mid \xi,\kappa) = \kappa\xi e^{\kappa x}\,\mathbb{1}_{x\geq 0}, \qquad \xi,\kappa > 0,$$

we obtain

$$S(x \mid \xi,\kappa) = e^{\xi(1-e^{\kappa x})}\,\mathbb{1}_{x\geq 0},$$
$$f(x \mid \xi,\kappa) = \kappa\xi e^{\xi(1-e^{\kappa x})+\kappa x}\,\mathbb{1}_{x\geq 0}.$$

This distribution with scale parameter κ and frailty parameter ξ is called *Gompertz* due to the parametric form of the hazard rate proposed by Gompertz (1825) and will be denoted by $\mathcal{G}(\xi,\kappa)$. The publication of Gompertz is widely considered as the first systematic attempt of age-specific modelling of human mortality rates. By studying mortality tables, Gompertz discovered that from the age of 30 onwards mortality rates tend to increase exponentially. A generalization was given by Makeham (1867, 1890) as a four parameter distribution, the so-called Gompertz-Makeham distribution $\mathcal{GM}(\xi,\kappa,\theta,\alpha)$ with the hazard rate

$$h(x \mid \kappa,\xi,\theta,\sigma) = \theta\kappa\xi + 2\kappa^2\xi\sigma x + \kappa\xi e^{\kappa x}\,\mathbb{1}_{x\geq 0}.$$

By Proposition 3.2.17, we see that the extended distribution $\mathcal{GM}(\xi,\kappa,{}^{\lambda}/(\xi\kappa),{}^{\sigma}/2\kappa^2\xi)$ with the hazard rate

$$h(x \mid \lambda,\sigma,\xi,\kappa) = \lambda + \sigma x + \kappa\xi e^{\kappa x}$$

results through a composition of an exponential, a Rayleigh and a restricted Gompertz distribution, i.e.,

$$\mathcal{GM}\left(\xi,\kappa,\frac{\lambda}{\xi\kappa},\frac{\sigma}{2\kappa^2\xi}\right) \sim \min(\mathrm{Exp}(\lambda),\ \mathrm{Rayleigh}(\sigma), \mathcal{GM}(\xi,\kappa)).$$

This shows the building blocks of the extended Gompertz-Makeham distribution that contains three risk factors, namely: an age-independent factor, a factor where the risk grows linear in age and one with an exponential growth.

(e) Kannisto (1992) proposed a logistic-type hazard rate

$$h(x \mid \alpha,\beta) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}\mathbb{1}_{x\geq 0}, \quad \alpha\in\mathbb{R},\beta>0 \tag{3.16}$$

which better tracks the observed mortality for higher ages, see Thatcher, Kannisto and Vaupel (1998). The authors studied the mortalities of several industrialized countries and showed that logistic-type hazard rates provide a better fit to historical data, since the observations show a sub-exponential growth for higher ages. From the hazard rate in eq. (3.16), we obtain the survival and the density as follows:

$$\begin{aligned} S(x \mid \alpha,\beta) &= (1+e^{\alpha})^{\frac{1}{\beta}}\left(1+e^{\alpha+\beta x}\right)^{-\frac{1}{\beta}}\mathbb{1}_{x\geq 0}, \\ f(x \mid \alpha,\beta) &= (1+e^{\alpha})^{\frac{1}{\beta}}\,e^{\alpha+\beta x}\left(1+e^{\alpha+\beta x}\right)^{-\frac{1+\beta}{\beta}}\mathbb{1}_{x\geq 0}. \end{aligned} \tag{3.17}$$

The distribution characterized by the hazard rate of eq. (3.16) will be called the *Kannisto distribution* and denoted by $\mathcal{K}(\alpha,\beta)$. Characteristic properties of the Kannisto distribution will be discussed in Section 3.5.

(f) Perks (1932) proposed a hazard rate of the form

$$h(x \mid \lambda,A,B,D,K) = \frac{A+Be^{\lambda x}}{1+De^{\lambda x}+Ke^{-\lambda x}}\mathbb{1}_{x\geq 0}.$$

This general form contains many of the hazard rate types introduced above, e.g., for:

- $B = D = K = 0$, we obtain a constant hazard rate, i.e., the distribution $\mathrm{Exp}(A)$,

- $A = D = K = 0$ this is the Gompertz distribution $\mathcal{G}(\lambda, {}^{B}\!/\!{}_{\lambda})$,
- $D = K = 0$, this results in the Gompertz-Makeham distribution $\mathcal{GM}(\lambda, {}^{B}\!/\!{}_{\lambda}, {}^{A}\!/\!{}_{B})$
- $A = K = 0$ and $B = D = e^{\alpha}$, is the Kannisto distribution $\mathcal{K}(\alpha, \lambda)$.

Marshall and Olkin (2007) showed that the survival function of the Perks type hazard rate is

$$S(x \mid \lambda, \alpha, \beta, \xi, \theta) = \left(\frac{\alpha}{e^{\lambda x} - (1-\alpha)}\right)^{\xi} \left(\frac{\beta}{e^{\lambda x} - (1-\beta)}\right)^{\theta} \tag{3.18}$$
$$\alpha, \beta > 0,\ \xi + \theta \geq 0,\ x \geq 0,$$

with the parametrization

$$\alpha = \frac{1}{2}\left(1 + \sqrt{1 - 4KD}\right) D + 1, \qquad \beta = \frac{1}{2}\left(1 - \sqrt{1 - 4KD}\right) D + 1,$$
$$\xi = \frac{B\left(1 - 2AD + \sqrt{1 - 4DK}\right)}{2D\sqrt{1 - 4DK}}, \qquad \theta = \frac{B\left(-1 + 2AD + \sqrt{1 - 4DK}\right)}{2D\sqrt{1 - 4DK}}.$$

For $\xi, \theta > 0$ eq. (3.18) is a product of two survival functions of the extended exponential distribution which will be defined below in eq. (3.20) below. The distribution with the Perks type hazard rate turns out to be distributed like the minimum of two extended independent exponential distributions. This is a consequence of Proposition 3.2.17 that shows that the hazard rate of the minimum of two random variables is the sum of the hazard rates. Since the hazard rates of the extended exponential family are decreasing if the tilt parameter is less than 1 ($\alpha, \beta < 1$ in eq. (3.18)), one can construct bathtub-shaped hazard rates from the Perks distribution. These types of hazard rates are useful to model initially decreasing hazard rates as in the case of infant mortality.

(g) An extended logistic-type hazard rate of the form

$$h(x \mid \gamma, \theta, \beta) = \frac{\beta\theta e^{\beta x}}{\gamma + e^{\beta x} - 1} \mathbb{1}_{x \geq 1} \qquad \gamma \geq 0,\ \theta > 0,\ \beta > 0, \tag{3.19}$$

generalizes the Kannisto hazard function of eq. (3.16). Both functions are logistic-type and can be obtained as solutions of the first-order non-linear Bernoulli differential equation

$$h'(x) = \kappa\, h(x)(c_{\max} - h(x)),$$

where $c_{\max}$ is the limiting value of h and $\kappa \in \mathbb{R}$. The distribution characterized by the hazard rate of eq. (3.19) will be called the *extended exponential distribution* and denoted by $\mathcal{E}(\gamma, \theta, \beta)$. This distribution received its name since it can be derived as a tilt and frailty parametric extension of the exponential distribution, see Marshall and Olkin (2007) and Proposition 3.3.1 for the precise definition of those extensions types. The survival and density functions of

$\mathcal{E}(\gamma,\theta,\beta)$ are

$$S(x \mid \gamma,\theta,\beta) = \left(\frac{\gamma}{e^{\beta x} - (1-\gamma)}\right)^{\theta} \mathbb{1}_{x\geq 0}, \tag{3.20}$$

$$f(x \mid \gamma,\theta,\beta) = \beta\theta\gamma^{\theta} e^{\beta x}(\gamma + e^{\beta x} - 1)^{-(\theta+1)}\mathbb{1}_{x\geq 0}. \tag{3.21}$$

(h) A *reciprocal hazard rate* function of the form

$$h(x \mid k,\alpha,\mu) = \frac{\alpha}{k-\mu+x}\mathbb{1}_{x\geq\mu} \quad k>0,\ \alpha>0,\ \mu\in\mathbb{R} \tag{3.22}$$

leads to

$$S(x \mid k,\alpha,\mu) = \left(\frac{x-\mu}{k}+1\right)^{-\alpha}\mathbb{1}_{x\geq\mu},$$
$$f(x \mid k,\alpha,\mu) = \alpha k^{\alpha}(k-\mu+x)^{-\alpha-1}\mathbb{1}_{x\geq\mu}.$$

This hazard rate characterizes the *Pareto type II* distribution and is denoted by $\mathcal{P}_{\mathrm{II}}(k,\alpha,\mu)$. For $\mu = 0$, we will abbreviate the notation and denote the distribution by $\mathcal{P}_{\mathrm{II}}(k,\alpha)$. This special case is also known as the *Lomax* distribution. In Proposition 3.3.4, we will give a connection of the Pareto type II distribution to the Kannisto as well as the extended exponential distribution.

(i) Next, we recall another well-known distribution which is not a lifetime distribution since it takes values in $\mathbb{R}$. The *logistic distribution*, denoted by $\mathcal{L}(\mu,s)$ is the distribution characterized by a logistic-type distribution function and should not be mistaken with $\mathcal{K}(\alpha,\beta)$ and $\mathcal{E}(\gamma,\theta,\beta)$ which are characterized by a logistic-type hazard rate function. The CDF of $\mathcal{L}(\mu,s)$ is given by

$$F(x \mid \mu,s) = \frac{1}{e^{-\frac{x-\mu}{s}}+1}, \quad \text{for } x\in\mathbb{R},\ \mu\in\mathbb{R},\ s>0$$

and the corresponding hazard, survival, and density functions are

$$\begin{aligned} h(x \mid \mu,s) &= \frac{1}{s\left(e^{\frac{\mu-x}{s}}+1\right)}, \\ S(x \mid \mu,s) &= \frac{1}{e^{\frac{x-\mu}{s}}+1}, \\ f(x \mid \mu,s) &= \frac{e^{\frac{\mu+x}{s}}}{s\left(e^{\frac{\mu}{s}}+e^{\frac{x}{s}}\right)^2}. \end{aligned} \tag{3.23}$$

The connection to the lifetime distributions $\mathcal{K}(\alpha,\beta)$ and $\mathcal{E}(\gamma,\theta,\beta)$ will be given in Proposition 3.3.10.

The next result shows that a distribution with a polynomial hazard function of order n can be decomposed into a competing risks model with of n independent Weibull distributed factors.

Proposition 3.2.19. A non-negative random variable Y with a polynomial hazard function given by

$$h_Y(x \mid a_0, \ldots, a_n) = \sum_{i=0}^{n} a_i x^i \mathbb{1}_{x \geq 0}, \qquad \text{with } a_i > 0 \quad \forall i \in \{0, \ldots n\},$$

can be composed of independent Weibull distributed competing risks $X_0, \ldots, X_n$, i.e.,

$$Y \sim \min(X_0, \ldots, X_n)$$

with $X_i \sim \mathcal{WB}\left(i + 1, \left({}^{i+1}/_{a_i}\right)^{1/i+1}\right)$.

Proof. We start with the cumulative hazard function H_Y, which is obtained by

$$\begin{aligned}
H_Y(x \mid a_0, \ldots, a_n) &= \int_0^x h_Y(y \mid a_0, \ldots, a_n)\, dy \\
&= \int_0^x \sum_{i=0}^{n} a_i y^i\, dy \\
&= \sum_{i=1}^{n} \frac{1}{i} a_i x^{i+1}.
\end{aligned}$$

The cumulative hazard function $H_Y(\cdot)$ is related to the survival function by $S_Y(\cdot) = \exp\{-H(\cdot)\}$. Using that, we obtain

$$\begin{aligned}
S_Y(x \mid a_0, \ldots, a_n) &= \exp\left(-H(x \mid a_0, \ldots, a_n)\right) \\
&= \exp\left(-\sum_{i=0}^{n} \frac{1}{i} a_i x^{i+1}\right) \\
&= \prod_{i=0}^{n} \exp\left(-\frac{1}{i} a_i x^{i+1}\right) \\
&= \prod_{i=0}^{n} S_{X_i}(x \mid a_i),
\end{aligned}$$

where the factors

$$S_{X_i}(x \mid a_i) = \exp\left(-\frac{1}{i} a_i x^{i+1}\right)$$

are survival functions of Weibull distributions $X_i \sim \mathcal{WB}\left(i + 1, \left({}^{i+1}/_{a_i}\right)^{1/i+1}\right)$. Hence, the survival function of $\min(X_0, \ldots, X_n)$ factorizes in the case of independent random variables, and we conclude $Y \sim \min(X_0, \ldots, X_n)$. □

3.3 | Connections of the Kannisto Distribution to Other Parametric Distributions

In the following part, we study the connections of the previously introduced logistic hazard rate distributions, namely the Kannisto and the extended exponential distribution, to other well-known distributions. We will show how the extended exponential distribution generalizes other distributions and demonstrate existing relations obtained by transformations, truncations, continuous mixtures, and by limiting distributions.

3.3.1 | Connection of the Extended Exponential Distribution and the Kannisto Distribution

First of all, one can easily verify that the extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$ is indeed a parametric extension of the exponential distribution $\mathrm{Exp}(\lambda)$. Choosing $\gamma = 1$ and $\theta = 1$, we see that the survival function of $\mathcal{E}(1,{}^{\lambda}/_{\beta},\beta)$

$$S_{\mathcal{E}}(x \mid 1,1,\beta) = e^{-\beta x} = S_{\mathrm{Exp}}(x \mid \beta)$$

coincides with the survival function of the exponential distribution $\mathrm{Exp}(\beta)$, see eq. (3.14). The next proposition shows that the Kannisto distribution is a special case of the extended exponential distribution.

Proposition 3.3.1. The extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$ is a generalization of the Kannisto distribution $\mathcal{K}(\alpha,\beta)$.

Proof. The extended exponential distribution has the survival function

$$S_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = \left(\frac{\gamma}{e^{\beta x} - (1-\gamma)}\right)^{\theta},$$

see the definition in eq. (3.20). For $\gamma = 1 + e^{-\alpha}$ and $\theta = {}^{1}/_{\beta}$, we obtain the survival function of the $\mathcal{K}(\alpha,\beta)$ distribution as given in eq. (3.17), i.e.,

$$S_{\mathcal{E}}(x \mid 1+e^{-\alpha},{}^{1}/_{\beta},\beta) = \left(\frac{1+e^{-\alpha}}{e^{-\alpha}+e^{\beta x}}\right)^{\frac{1}{\beta}} = \left(\frac{1+e^{\alpha+\beta x}}{1+e^{\alpha}}\right)^{-\frac{1}{\beta}} = S_{\mathcal{K}}(x \mid \alpha,\beta).$$

□

While $\mathcal{E}(\gamma,\theta,\beta)$ generalizes the exponential distribution as well as the Kannisto distribution, the exponential distribution cannot be obtained from the Kannisto distribution by a particular parameter choice. However, as next proposition shows, the exponential distribution results as a limiting distribution.

Proposition 3.3.2. For the Kannisto distribution $\mathcal{K}(\alpha,\beta)$, we have

$$\mathcal{K}(\alpha,\beta) \xrightarrow[\alpha\to\infty]{\mathcal{D}} \mathrm{Exp}(1), \tag{3.24}$$

$$\mathcal{K}(\alpha,\beta) \xrightarrow[\beta\to\infty]{\mathcal{D}} \mathrm{Exp}(1), \tag{3.25}$$

$$\mathcal{K}(\lambda/1-\lambda,\beta) \xrightarrow[\beta\to 0]{\mathcal{D}} \operatorname{Exp}(\lambda) \quad \text{for} \quad 0<\lambda<1. \tag{3.26}$$

Proof. For the convergence in distribution, we need to show that the survival function of the Kannisto distribution converges pointwise, for each $x \geq 0$, to the survival function of the exponential distribution. Thus, for eq. (3.24), we have

$$\lim_{\alpha\to\infty} S_{\mathcal{K}}(x \mid \alpha,\beta) = \lim_{\alpha\to\infty} \left(\frac{1+e^{\alpha+\beta x}}{1+e^{\alpha}}\right)^{-\frac{1}{\beta}} = e^{-x} = S_{\operatorname{Exp}}(x \mid 1).$$

To prove eq. (3.25), we apply the L'Hôpital's rule on eq. (3.27)

$$\begin{aligned}\lim_{\beta\to\infty} S_{\mathcal{K}}(x \mid \alpha,\beta) &= \lim_{\beta\to\infty} \left(\frac{1+e^{\alpha+\beta x}}{1+e^{\alpha}}\right)^{-\frac{1}{\beta}} \\ &= \lim_{\beta\to\infty} e^{-\frac{\ln\left(\frac{1+e^{\alpha+x\beta}}{1+e^{\alpha}}\right)}{\beta}} \\ &= \lim_{\beta\to\infty} e^{-\frac{e^{\alpha+x\beta}x}{1+e^{\alpha+x\beta}}} \\ &= e^{-x} = S_{\operatorname{Exp}}(x \mid 1).\end{aligned} \tag{3.27}$$

Equation (3.26) can be obtained by using the L'Hôpital's rule for eq. (3.28)

$$\begin{aligned}\lim_{\beta\to 0} S_{\mathcal{K}}(x \mid \lambda/1-\lambda,\beta) &= \lim_{\beta\to 0} \left(\lambda\left(e^{\beta x}-1\right)+1\right)^{-\frac{1}{\beta}} \\ &= \lim_{\beta\to 0} e^{-\frac{1}{\beta}\ln\left(\lambda\left(e^{\beta x}-1\right)+1\right)} \\ &= \lim_{\beta\to 0} e^{-\frac{\lambda x e^{\beta x}}{\lambda\left(e^{\beta x}-1\right)+1}} \\ &= e^{-\lambda x} = S_{\operatorname{Exp}}(x \mid \lambda).\end{aligned} \tag{3.28}$$

□

An overview of the relations between the exponential, Kannisto, and the extended exponential distribution is displayed in Figure 3.1. The next result shows that the Kannisto distribution is not stable under scaling, i.e., multiplying a Kannisto random variable with a positive constant does not result in a Kannisto distribution. The result shows also that the extended exponential distribution is stable under scaling.

Proposition 3.3.3. Let $X \sim \mathcal{K}(\alpha,\beta)$ be Kannisto distributed and $c \in \mathbb{R}_+$, then cX follows the extended exponential distribution $\mathcal{E}(1+e^{-\alpha},1/\beta,\beta/c)$.

Proof. As observed in Proposition 3.3.1 the Kannisto distribution $\mathcal{K}(\alpha,\beta)$ is a special case of extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$, where the frailty parameter θ is set to the inverse of the scale

$$
\begin{tikzcd}
\mathcal{E}(\gamma,\theta,\beta) \arrow[rr, "\gamma = 1 + e^{-\alpha},\, \theta = \frac{1}{\beta}"] \arrow[d, "\mathcal{E}(1,1,\lambda)"'] & & \mathcal{K}(\alpha,\beta) \arrow[dll, "\lim\limits_{\beta\to 0} \mathcal{K}(\ln(\lambda/(1-\lambda)),\beta),\ \lambda\in(0,1)"'] \arrow[d, "\lim\limits_{\alpha\to\infty} \mathcal{K}(\alpha,\beta),\ \lim\limits_{\beta\to\infty} \mathcal{K}(\alpha,\beta)"] \\
\operatorname{Exp}(\lambda) \arrow[rr, "\lambda = 1"] & & \operatorname{Exp}(1)
\end{tikzcd}
$$

Figure 3.1: Relationships between the Kannisto and the extended exponential distribution.

parameter β, and the tilt parameter set to $\gamma = 1 + e^{-\alpha}$. The proof can be obtained directly by computation of the transformed survival function. This leads to

$$S_{c\cdot X}(x) = \mathbb{P}\left[cX > x\right] = S_{\mathcal{E}}({}^{x}/_{c} \mid 1 + e^{-\alpha}, {}^{1}/_{\beta}, \beta) = \left(\frac{1 + e^{\alpha + \frac{\beta}{c}x}}{1 + e^{\alpha}}\right)^{-1/\beta} = S_{\mathcal{E}}(1 + e^{-\alpha}, {}^{1}/_{\beta}, {}^{\beta}/_{c}).$$

□

3.3.2 | Connection of the Extended Exponential Distribution and the Pareto Type II Distribution

The following proposition exhibits the connection of the Kannisto distribution and the Pareto type II distribution. The connection is interesting since the Pareto distribution is a heavy-tailed distribution with various applications especially in non-life actuarial science, while the Kannisto distribution arises in life actuarial science by observing a logistic-type growth of age-specific mortalities for high ages. Due to its popularity, the Pareto distribution is implemented in almost all statistical software packages. Thus, using the transformation of the following proposition, sampling from the non-standard Kannisto or extended exponential distribution can be implemented efficiently.

Proposition 3.3.4. Let X be a Pareto type II distribution with survival function

$$S_{\mathcal{P}_{II}}(x \mid \gamma,\theta) = (1 + {}^{x}/_{\gamma})^{-\theta}\, \mathbb{1}_{x\geq 0}, \quad \gamma,\theta > 0,$$

then $Y := g(X)$, with $g(X) = {}^{1}/_{\beta} \ln(X + 1)$ is $\mathcal{E}(\gamma,\theta,\beta)$ distributed. Furthermore, for $X \sim \mathcal{P}_{II}(1 + e^{-\alpha}, {}^{1}/_{\beta})$, we obtain the Kannisto $\mathcal{K}(\alpha,\beta)$ distribution through the transformation g.

Proof. The map $g : \mathbb{R}_{+} \to \mathbb{R}_{+}$ is bijective with $g^{-1}(y) = e^{\beta y} - 1$. The survival function of Y satisfies

$$\begin{aligned}
S_Y(y) &= \mathbb{P}\left[Y > y\right], \\
&= \mathbb{P}\left[g(X) > y\right], \\
&= \mathbb{P}\left[X > g^{-1}(y)\right], \\
&= \mathbb{P}\left[X > e^{\beta y} - 1\right], \\
&= S_{\mathcal{P}_{II}}\left[e^{\beta y} - 1\right],
\end{aligned}$$

$$= \left[\frac{\gamma}{e^{\beta x} - (1-\gamma)} \right]^{\theta},$$
$$= S_{\mathcal{E}}(y \mid \gamma, \theta, \beta),$$

which completes the proof of the first part. The second result for the Kannisto distribution follows directly from Proposition 3.3.1. □

Note that by using the inverse transformation of Proposition 3.3.4, $g^{-1}(Y) = e^{Y/\theta} - 1$ with $Y \sim \mathcal{K}\left(-\ln(\gamma - 1), \frac{1}{\theta}\right)$, we obtain the Pareto distribution $\mathcal{P}_{II}(\gamma,\theta)$ from the Kannisto distribution.

Alternatively, the extended exponential distribution or the Kannisto distribution can be sampled from the uniform distribution as covered in the following result.

Proposition 3.3.5. Let $U \sim \mathcal{U}(0,1)$ be uniformly distributed on the interval $(0,1)$ then

$$X := \frac{1}{\beta} \ln\left(1 - \gamma + \gamma(1-U)^{-\frac{1}{\theta}}\right) \tag{3.29}$$

is $\mathcal{E}(\gamma,\theta,\beta)$ distributed and

$$Y := \frac{1}{\beta} \ln\left((1 + e^{-\alpha})(1-U)^{-\beta} - e^{-\alpha}\right) \tag{3.30}$$

is $\mathcal{K}(\alpha,\beta)$ distributed.

Proof. This is a direct consequence of the inverse sampling theorem. The cumulative distribution function of the extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$ is given by

$$F_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = 1 - \left(\frac{\gamma}{e^{\beta x} - (1-\gamma)} \right)^{\theta} \mathbb{1}_{x \geq 0}. \tag{3.31}$$

For a strictly increasing distribution function F, *inverse* F^{-1} *of* F is defined by

$$F^{-1}(p) = \sup\{x : F(x) \leq p\}, \quad 0 < p < 1.$$

From eq. (3.31), we obtain

$$F_{\mathcal{E}}^{-1}(p) = \frac{1}{\beta} \ln\left(1 - \gamma + \gamma(1-p)^{-\frac{1}{\theta}}\right).$$

Equation (3.30) follows from eq. (3.29) by Proposition 3.3.1. □

3.3.3 | Connection of the Extended Exponential Distribution and the Gompertz Distribution

The next proposition shows how the Gompertz distribution, which is characterized by exponential increasing hazard rates can be obtained as a limiting distribution of the extended exponential distribution. This connection without a rigorous proof can be found in Marshall and Olkin (2007).

Proposition 3.3.6. The Gompertz distribution is a limiting distribution of the extended exponential distribution. More precisely, we have

$$\mathcal{E}(\gamma,\gamma\xi,\beta) \xrightarrow[\gamma\to\infty]{\mathcal{D}} \mathcal{G}(\xi,\beta)$$

Proof. This result is obtained by setting $\theta = \gamma\xi$ in the survival function of $\mathcal{E}(\gamma,\theta,\beta)$ and taking the limit $\gamma \to \infty$ for fixed ξ and β, i.e.,

$$\begin{aligned}
\lim_{\gamma\to\infty} S_{\mathcal{E}}(x \mid \gamma,\gamma\xi,\beta) &= \lim_{\gamma\to\infty} \left(\frac{\gamma}{e^{\beta x} - (1-\gamma)}\right)^{\gamma\xi} \mathbb{1}_{x\geq 0} \\
&= \lim_{\gamma\to\infty} e^{\gamma\xi \ln\left(\frac{\gamma}{e^{x\beta}+\gamma-1}\right)} \mathbb{1}_{x\geq 0} \\
&= \lim_{\gamma\to\infty} e^{\gamma\xi\left(\frac{\gamma}{e^{x\beta}+\gamma-1}-1\right)} \mathbb{1}_{x\geq 0} &(3.32)\\
&= e^{\xi(1-e^{\beta x})} \mathbb{1}_{x\geq 0} &(3.33)\\
&= S_{\mathcal{G}}(x \mid \xi,\beta),
\end{aligned}$$

where eqs. (3.32) and (3.33) follow by L'Hôpital's rule. □

The following proposition shows how the Kannisto distribution arises as a mixture of Gompertz distributions. This means that a proper mixture of individuals with inhomogeneous exponential hazard rates can result in a population with a bounded logistic hazard rate.

Proposition 3.3.7. The Kannisto distribution can be obtained as a mixture of Gompertz distributions by regarding the frailty parameter as a random variable following a gamma distribution. More specifically, the Kannisto distribution $\mathcal{K}(\alpha,\beta)$ results as a continuous mixture of a Gompertz distribution $\mathcal{G}(\xi,\beta)$, where the frailty parameter ξ is considered to be $\Gamma(1/(1+e^{-\alpha}),1/\beta)$ distributed.

Proof. Let $S_{\mathcal{G}}$ be the survival function of the Gompertz distribution, with

$$S_{\mathcal{G}}(x \mid \xi,\beta) = \exp\left(-\xi(e^{\beta x}-1)\right) \mathbb{1}_{x\geq 0}, \quad \xi,\beta > 0.$$

Let ξ be a $\Gamma(1/\gamma,\theta)$-distributed random variable with the distribution function $F_\Gamma(\xi \mid 1/\gamma,\theta)$ and density

$$f_\Gamma(\xi \mid 1/\gamma,\theta) = \frac{\gamma^\theta e^{-\gamma\xi}\xi^{\theta-1}}{\Gamma(\theta)}, \xi \geq 0,$$

where the $1/\gamma$ is the scale and θ the shape parameter. Using Definition 3.2.16, we now consider a continuous mixture for ξ in $\mathcal{G}(\xi,\beta)$ with respect to $\Gamma(1/\gamma,\theta)$. The survival function of the mixture

S_{Mix} has the form

$$\begin{aligned} S_{\text{Mix}}(x \mid \gamma,\theta,\beta) &= \int_0^\infty S_{\mathcal{G}}(x \mid \xi,\beta)\, dF_{\Gamma}(\xi \mid {}^1\!/\!_\gamma, \theta) \\ &= \int_0^\infty \exp\left(-\xi(e^{\beta x} - 1)\right) \frac{\gamma e^{-\gamma\xi}(\gamma\xi)^{\theta-1}}{\Gamma(\theta)}\, d\xi \\ &= \int_0^\infty \frac{(\gamma\xi)^\theta e^{-\xi(e^{\beta x}+\gamma-1)}}{\xi\Gamma(\theta)} d\xi \\ &= \left(\frac{\gamma}{e^{\beta x} - (1-\gamma)}\right)^\theta. \end{aligned} \tag{3.34}$$

The mixing distribution turns out to be the extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$. As shown in Proposition 3.3.1 the Kannisto distribution appears as a special case for $\gamma = 1 + e^{-\alpha}$ and $\theta = {}^1\!/\!_\beta$. Note, the connection of a Gompertz mixture and the extended exponential distribution was originally obtained by Marshall and Olkin (2007). □

Remark 3.3.8. Proposition 3.3.7 shows that the Kannisto distribution can be derived by continuous mixing of Gompertz distributions with respect to a gamma distribution. Observe that despite taking a mixture of Gompertz distributions with exponentially increasing hazard rates of the form

$$h_{\mathcal{G}}(x \mid \xi,\beta) := \xi\beta e^{\beta x}, \qquad x \geq 0, \quad \xi,\beta > 0, \tag{3.35}$$

the result is a Kannisto distribution with logistic-type hazard rates which are bounded by

$$\lim_{x\to\infty} h_{\mathcal{K}}(x \mid \alpha,\beta) = 1.$$

Thus, modelling lifetimes of populations by a Kannisto distribution allows a twofold interpretation of the composition of the underlying population. The first one is the *homogeneous population* interpretation, which treats the lifetimes of individuals as i.i.d. Kannisto distributed random variables. Since the entire group is considered to have the identical lifetime distribution, we indeed have a homogeneous group. On the other hand the *heterogeneous population* interpretation arises, when we do not assume the population to be composed of individuals with i.i.d. distributed lifetimes. This assumption is less restrictive and more realistic than the former one since groups of individual of the same cohort show indeed different lifestyles, diets or diseases which influence individual lifetimes. The differences can be modelled by the frailty parameter and its distribution. In the presented case, each member of the group has an individual Gompertz distributed remaining lifetime where the parameter ξ in eq. (3.35) is randomly chosen from a specified gamma distribution. Taking the above mixture leads to Kannisto distributed lifetimes of that population.

Remark 3.3.9. The mixture as given in eq. (3.34) can also be interpreted as a *Gamma-Gompertz* frailty model. Frailty models are popular in mortality analysis of heterogeneous population groups, see, e.g., Yashin (2004) and Vaupel and Yashin (1985). They are characterized in terms of hazard rates with a multiplicative decomposition into a frailty parameter and a baseline hazard rate h_0. In the case above, the hazard rate of the Gompertz distribution is

$$h_{\mathcal{G}}(x \mid \xi,\beta) = \xi\beta e^{\beta x} = \xi h_0(x \mid \beta),$$

where ξ is called the frailty parameter and is treated as non-negative random variable following, e.g., a gamma distribution such as in Proposition 3.3.7.

3.3.4 | Connection of the Extended Exponential Distribution and the Logistic Distribution

The following result shows that the extended exponential distribution and the Kannisto distribution correspond to zero-truncated and frailty extended logistic distribution.

Proposition 3.3.10. The extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$ can be obtained as a truncation of a frailty extended logistic distribution $\mathcal{L}(\mu,s)$ with frailty θ, $\mu = {}^{\ln(\gamma-1)}/_{\beta}$ and $s = {}^1/_\beta$.

Proof. Let $X \sim \mathcal{L}(\mu,s)$ be a logistic distributed random variable with survival function (see eq. (3.23))

$$S_{\mathcal{L}}(x \mid \mu,s) = \frac{1}{e^{\frac{x-\mu}{s}} + 1}, \quad \text{for } x \in \mathbb{R}, \mu \in \mathbb{R}, s > 0.$$

A frailty extension (see Definition 3.2.18) of $\mathcal{L}(\mu,s)$ is a distribution with survival function

$$S_{\mathcal{L}_\theta}(x \mid \mu,s,\theta) = \left(\frac{1}{e^{\frac{x-\mu}{s}} + 1}\right)^{\theta}, \quad \text{for } x \in \mathbb{R}, \mu \in \mathbb{R}, s > 0, \theta > 0.$$

Truncation of $S_{\mathcal{L}_\theta}(x \mid {}^{\ln(\gamma-1)}/_{\beta}, {}^1/_\beta, \theta)$ at 0 leads to a distribution with survival function

$$\frac{S_{\mathcal{L}_\theta}\left(x \mid \frac{\ln(\gamma-1)}{\beta}, \frac{1}{\beta}, \theta\right)}{S_{\mathcal{L}_\theta}\left(0 \mid \frac{\ln(\gamma-1)}{\beta}, \frac{1}{\beta}, \theta\right)} = \frac{\left(\frac{\gamma-1}{\gamma+e^{\beta x}-1}\right)^{\theta}}{\left(\frac{\gamma-1}{\gamma}\right)^{\theta}} = \left(\frac{\gamma}{\gamma + e^{\beta x} - 1}\right)^{\theta}.$$

The right-hand side reveals the survival function of the distribution $\mathcal{E}(\gamma,\theta,\beta)$ as defined in eq. (3.20). Note that by Proposition 3.3.1, a truncation at 0 with frailty $\theta = \frac{1}{\beta}$ and $\gamma = 1 + e^{-\alpha}$ leads to the Kannisto distribution $\mathcal{K}(\alpha,\beta)$. □

Figure 3.2 displays an overview of the connections revealed by Propositions 3.3.1, 3.3.4, 3.3.6, 3.3.7 and 3.3.10.

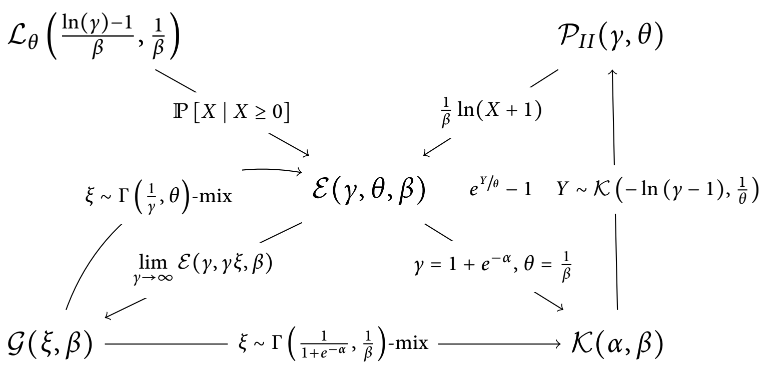


Figure 3.2: Relationships between the Gompertz, Pareto, Logistic, Kannisto, and the extended exponential distribution.

3.4 | Generalized Hypergeometric Functions

In Section 3.5, we will proceed with the characterization of the extended exponential and the Kannisto distribution. As will be shown there, many characteristics of interest such as the central moments, the Fisher information or the Kullback-Leibler divergence do not have closed form representations but can be expressed in terms of generalized hypergeometric functions. In preparation for the upcoming sections, we give a formal definition and add some remarks on generalized hypergeometric functions. A standard reference on the following topic is the book by Slater (1966).

The *generalized hypergeometric function* ${}_pF_q$ is defined by the following series expansion

$$ {}_pF_q[a_1,\dots,a_p;b_1,\dots,b_q;z] = \sum_{n=0}^{\infty} \frac{(a_1)_n \dots (a_p)_n}{(b_1)_n \dots (b_q)_n} \frac{z^n}{n!}; \quad p,q \in \mathbb{N}_0, \tag{3.36} $$

where $a_1,\dots,a_p$ and $b_1,\dots,b_q$ are complex numbers, and

$$ (a_i)_n = \frac{\Gamma(a_i+n)}{\Gamma(a_i)} = a_i(a_i+1)(a_i+2)\cdots(a_i+n-1), \quad (a_i)_0 = 1 $$

is a *Pochhammer symbol* or also called *ascending factorial*. The generalized hypergeometric function is invariant under permutations of the first p parameters $a_1,\dots,a_p$ and invariant under permutations of the last q parameters $b_1,\dots,b_q$. It is also clear from the definition that if anyone of the first p parameters, say a_i, coincides with one of the last q parameters, say b_j, then ${}_pF_q$ reduces to

$$ {}_{p-1}F_{q-1}[a_1,\dots,a_{i-1},a_{i+1},\dots,a_p;b_1,\dots,b_{j-1},b_{j+1},\dots,b_q;z]. $$

Using the gamma function, eq. (3.36) can be expressed as

$$ {}_pF_q\left[\begin{matrix} a_1,\dots,a_p \\ b_1,\dots,b_q \end{matrix};z\right] := \sum_{k=0}^{\infty} \prod_{i=1}^{p} \frac{\Gamma(k+a_i)}{\Gamma(a_i)} \prod_{j=1}^{q} \frac{\Gamma(b_j)}{\Gamma(k+b_j)} \frac{z^k}{k!} \quad p,q \in \mathbb{N}_0. \tag{3.37} $$

Note, the left-hand side of the above equation is a common alternative notation for generalized hypergeometric functions. Hypergeometric functions have been studied for more than two centuries by influencing mathematicians such as Euler, Gauss, and Riemann, see, e.g., Cattani (2006) for historical notes. For $p = 2$ and $q = 1$ the series eq. (3.37) is known as the *Gauss hypergeometric series*. It is well-known (Bailey, 1935; Slater, 1966) that for $p \le q$ the series converges for $|z| < \infty$. For $p = q + 1$ convergence occurs for $|z| < 1$ and when $p > q + 1$ the series diverges for all $z \neq 0$. For $|z| \ge 1$ many generalized hypergeometric functions ${}_pF_q$ can be analytically continued. For instance, Nørlund (1955) shows that for $p = q + 1$ any hypergeometric series in powers of z can be transformed into a series in powers of $\frac{z}{(z-1)}$ such that the convergence holds for $\mathrm{Re}(z) < \frac{1}{2}$. In particular, analytic continuation provided by Nørlund (1955) is given by the transformation

$$ {}_nF_{n-1}\left[\begin{matrix} a_1, a_2, \ldots, a_n \\ b_1, b_2, \ldots, b_{n-1}\end{matrix}; z\right] = (1-z)^{-a_1} \sum_{m=0}^{\infty} \frac{(a_1)_m}{m!} \, {}_nF_{n-1}\left[\begin{matrix} -m, a_2, \ldots, a_n \\ b_1, b_2, \ldots, b_{n-1}\end{matrix}; 1\right] \left(\frac{z}{z-1}\right)^m . \tag{3.38}$$

The series on the right-hand side converges for $z < \frac{1}{2}$. For $n = 2$, the relation given in eq. (3.38) reduces to the so-called Pfaff's transformations

$$ {}_2F_1\left[\begin{matrix} a, b \\ c \end{matrix}; z\right] = (1-z)^{-a} \, {}_2F_1\left[\begin{matrix} a, c-b \\ c \end{matrix}; \frac{z}{z-1}\right] = (1-z)^{-b} \, {}_2F_1\left[\begin{matrix} c-a, b \\ c \end{matrix}; \frac{z}{z-1}\right]. \tag{3.39}$$

The transformation of eq. (3.39) will be used in Sections 3.5.5 and 3.5.8. Efficient numerical methods for evaluation of generalized hypergeometric functions can be found in, e.g., Forrey (1997) and Pearson (2009).

3.5 | Characteristics and Properties of the Kannisto and the Extended Exponential Distribution

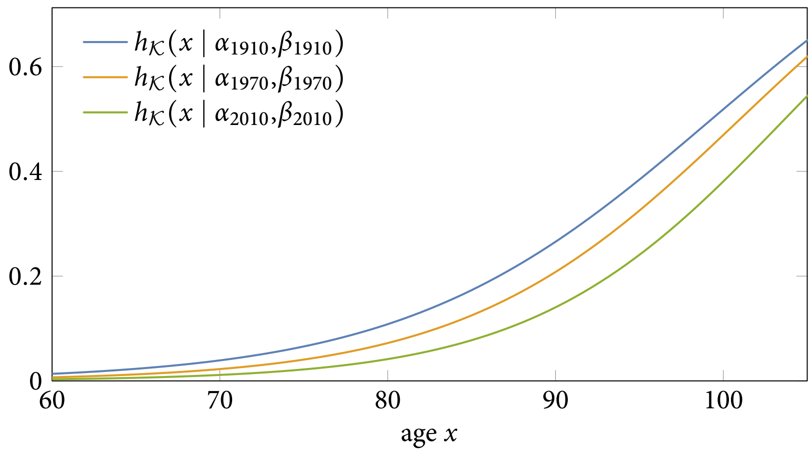
In the following section, we provide a characterization of the Kannisto and the extended exponential distribution and study their properties. Both distributions are, as defined in Section 3.2.7, specified by logistic hazard rate functions with two or three degrees of freedom. The representatives of both distributions are summarized in Tables 3.1 and 3.2. See also Figure 3.3 for example graphs of the Kannisto hazard rates, survival, and density function for some estimated parameters for the KAN model in Section 1.9.2.

3.5.1 | Basic Properties

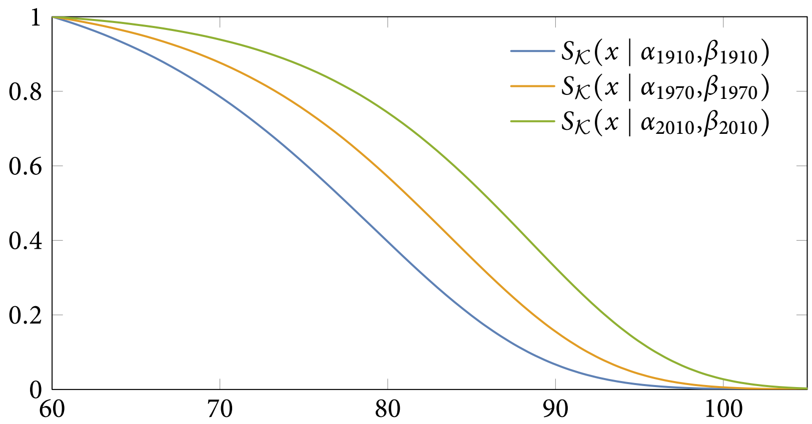
Initially, we deduce some basic properties of the Kannisto and the extended exponential distribution including some measures of central tendency, tail behaviour and characteristic of the hazard rate functions.

Unimodality

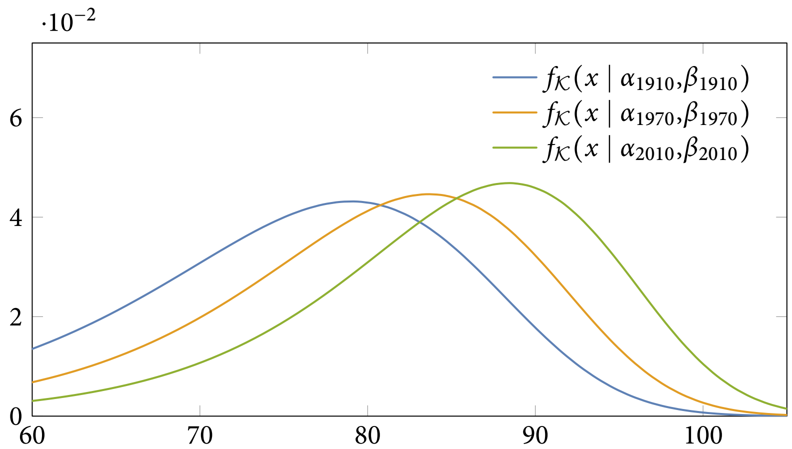
A distribution is called unimodal if the cumulative distribution function $F(x)$ is convex for $x < x^{\text{mode}}$ and concave for $x > x^{\text{mode}}$, see Marshall and Olkin (2007). By differentiating the density of $\mathcal{E}(\gamma, \theta, \beta)$



(a) Hazard rate function $h_{\mathcal{K}}(x \mid \alpha, \beta)$.



(b) Survival function $S_{\mathcal{K}}(x \mid \alpha, \beta)$.



(c) Probability density function $f_{\mathcal{K}}(x \mid \alpha, \beta)$.

Figure 3.3: Hazard, survival, and density function of the Kannisto distribution for the parameters of the Swedish female population at the years $t \in \{1910, 1970, 2010\}$.

Table 3.1: Representatives of the extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$.

ext. exponential distribution $X \sim \mathcal{E}(\gamma,\theta,\beta)$ with $\gamma > 1$, $\theta > 0$, $\beta > 0$	
PDF	$f_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = \gamma^{\theta}\beta\theta e^{\beta x}\left(\gamma + e^{\beta x} - 1\right)^{-(\theta+1)} \mathbb{1}_{x\geq 0}$
CDF	$F_{\mathcal{K}}(x \mid \alpha,\beta) = 1 - \left(\frac{\gamma}{e^{\beta x}-(1-\gamma)}\right)^{\theta} \mathbb{1}_{x\geq 0}$
survival function	$S_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = \left(\frac{\gamma}{e^{\beta x}-(1-\gamma)}\right)^{\theta} \mathbb{1}_{x\geq 0}$
hazard rate	$h_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = \frac{\beta\theta e^{\beta x}}{e^{\beta x}-(1-\gamma)} \mathbb{1}_{x\geq 0}$

Table 3.2: Representatives of the Kannisto distribution $\mathcal{K}(\alpha,\beta)$.

Kannisto distribution $X \sim \mathcal{K}(\alpha,\beta)$ with $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}_+$	
PDF	$f_{\mathcal{K}}(x \mid \alpha,\beta) = (1+e^{\alpha})^{\frac{1}{\beta}} e^{\alpha+\beta x}\left(1+e^{\alpha+\beta x}\right)^{-\frac{1+\beta}{\beta}} \mathbb{1}_{x\geq 0}$
CDF	$F_{\mathcal{K}}(x \mid \alpha,\beta) = 1 - (1+e^{\alpha})^{\frac{1}{\beta}}\left(1+e^{\alpha+\beta x}\right)^{-\frac{1}{\beta}} \mathbb{1}_{x\geq 0}$
survival function	$S_{\mathcal{K}}(x \mid \alpha,\beta) = (1+e^{\alpha})^{\frac{1}{\beta}}\left(1+e^{\alpha+\beta x}\right)^{-\frac{1}{\beta}} \mathbb{1}_{x\geq 0}$
hazard rate	$h_{\mathcal{K}}(x \mid \alpha,\beta) = e^{\alpha+\beta x}\left(1+e^{\alpha+\beta x}\right)^{-1} \mathbb{1}_{x\geq 0}$

one can verify that this density is unimodal with the mode at

$$x_{\mathcal{E}}^{\text{mode}} = \max\left\{0, \frac{1}{\beta}\ln\left(\frac{\gamma-1}{\theta}\right)\right\}. \tag{3.40}$$

The maximal value of the density obtained at the mode is given by

$$f_{\mathcal{E}}(x_{\mathcal{E}}^{\text{mode}}) = \begin{cases} \frac{\beta\theta}{\gamma}, & \text{if } x_{\mathcal{E}}^{\text{mode}} = 0 \\ \beta\left(\frac{\theta}{\theta+1}\right)^{\theta+1}\left(\frac{\gamma}{\gamma-1}\right)^{\theta}, & \text{if } x_{\mathcal{E}}^{\text{mode}} \neq 0. \end{cases} \tag{3.41}$$

From eqs. (3.40) and (3.41) and by Proposition 3.3.1, we see that the mode of the Kannisto distribution $\mathcal{K}(\alpha,\beta)$ is given by

$$x_{\mathcal{K}}^{\text{mode}} = \max\left\{0, \frac{1}{\beta}(\ln\beta - \alpha)\right\}$$

For the Kannisto density, the maximal value obtained at the mode is then given by

$$f_{\mathcal{K}}(x_{\mathcal{K}}^{\text{mode}}) = \begin{cases} \frac{1}{e^{-\alpha}+1}, & \text{if } x_{\mathcal{K}}^{\text{mode}} = 0 \\ \frac{\beta\left(\frac{e^{\alpha}+1}{\beta+1}\right)^{\frac{1}{\beta}}}{\beta+1}, & \text{if } x_{\mathcal{K}}^{\text{mode}} \neq 0. \end{cases}$$

Quantiles

The quantile functions of $\mathcal{E}(\gamma,\theta,\beta)$ and $\mathcal{K}(\alpha,\beta)$ are given by (cf., Proposition 3.3.5)

$$Q_{\mathcal{E}}(p \mid \gamma,\theta,\beta) = \frac{1}{\beta}\ln\left(1-\gamma+\gamma(1-p)^{-1/\theta}\right) \tag{3.42}$$

and

$$Q_{\mathcal{K}}(p \mid \alpha,\beta) = \frac{1}{\beta}\ln\left((e^{-\alpha}+1)\,(1-p)^{-\beta}-e^{-\alpha}\right).$$

From eq. (3.42), we obtain

$$\text{med}_{\mathcal{E}} = Q_{\mathcal{E}}\left(\frac{1}{2}\right) = \frac{1}{\beta}\ln\left(1-\gamma+\gamma 2^{1/\theta}\right),$$

as the median of $\mathcal{E}(\gamma,\theta,\beta)$ and

$$\text{med}_{\mathcal{K}} = Q_{\mathcal{K}}\left(\frac{1}{2}\right) = \frac{\ln\left(2^{\beta}\,(e^{\alpha}+1)-1\right)}{\beta} - \frac{\alpha}{\beta}$$

as the median of the distribution $\mathcal{K}(\alpha,\beta)$. While the mode and the median, which can be considered as measures of central tendency, are given in closed forms, this is not the case anymore for the expectation or higher moments, as we will see later.

Cumulative Hazard Rates

The cumulative hazard rates of the distributions $\mathcal{E}(\gamma,\theta,\beta)$ and $\mathcal{K}(\alpha,\beta)$ (see Definition 3.2.8) are given by

$$H_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = \int_0^x h_{\mathcal{E}}(x \mid \gamma,\theta,\beta)dy = \theta\ln\left(\frac{\gamma-1+e^{\beta x}}{\gamma}\right) \tag{3.43}$$

and

$$H_{\mathcal{K}}(x \mid \alpha,\beta) = \int_0^x h_{\mathcal{K}}(y \mid \alpha,\beta)dy = \frac{1}{\beta}\ln\left(\frac{1+e^{\alpha+\beta x}}{1+e^{\alpha}}\right).$$

By using the cumulative hazard rates, we derive the tail behaviour of both distributions in the next proposition.

Proposition 3.5.1. The Kannisto and the extended exponential distribution are light-tailed.

Proof. By Proposition 3.2.10, in order to prove that a distribution is light-tailed, we need to show

$$\lim_{x\to\infty}\frac{H(x)}{x} \neq 0,$$

Thus, by using eq. (3.43), we have

$$\lim_{x\to\infty}\frac{H_{\mathcal{E}}(x\mid\gamma,\theta,\beta)}{x}=\lim_{x\to\infty}\frac{\theta\ln\left(\frac{\gamma-1+e^{\beta x}}{\gamma}\right)}{x}=\lim_{x\to\infty}\frac{\beta\theta e^{\beta x}}{\gamma-1+e^{\beta x}}=\beta\theta\neq 0.$$

For the Kannisto distribution, we have by Proposition 3.3.1 $\lim_{x\to\infty} {}^{H_{\mathcal{K}}(x\,\mid\,\alpha,\beta)}/_{x}=1$. □

Initial and Limiting Values of the Hazard Rate Functions

The starting values of the $\mathcal{E}(\gamma,\theta,\beta)$ and $\mathcal{K}(\alpha,\beta)$ hazard rates are determined by

$$h_{\mathcal{E}}(0\mid\gamma,\theta,\beta)=\frac{\beta\theta}{\gamma}$$

and

$$h_{\mathcal{K}}(0\mid\alpha,\beta)=\frac{1}{1+e^{-\alpha}}.$$

The turning points of the logistic growth of the hazard rates are located at $x={}^{\ln(\gamma-1)}/_{\beta}$ for $\mathcal{E}(\gamma,\theta,\beta)$ and at $x=-{}^{\alpha}/_{\beta}$ for the $\mathcal{K}(\alpha,\beta)$ distribution. For the limiting hazard values, we have

$$\lim_{x\to\infty}h_{\mathcal{E}}(x\mid\gamma,\theta,\beta)=\beta\theta$$

and

$$\lim_{x\to\infty}h_{\mathcal{K}}(x\mid\alpha,\beta)=1.$$

While the Kannisto hazard rate is bounded by 1, the additional degree of freedom of the extended exponential distribution influences the upper limit of the hazard rate function.

3.5.2 | Residual Life Distribution

According to Definition 3.2.13 the survival function $S_{\mathcal{E}}^t$ of the residual life distribution of a t-survivor is given by

$$S_{\mathcal{E}}^t=\frac{S_{\mathcal{E}}(x+t\mid\gamma,\theta,\beta)}{S_{\mathcal{E}}(x\mid\gamma,\theta,\beta)}=\left(\frac{\gamma+e^{\beta(t+x)}-1}{\gamma+e^{\beta t}-1}\right)^{\theta}.$$

The next proposition states that the residual life distributions of $\mathcal{E}(x\mid\gamma,\theta,\beta)$ and $\mathcal{K}(x\mid\alpha,\beta)$ converge to exponential distributions.

Proposition 3.5.2. The residual life distribution of $\mathcal{E}(x\mid\gamma,\theta,\beta)$ converges in distribution to the exponential distribution $\mathrm{Exp}(\beta\theta)$.

Proof. This is a direct consequence of Proposition 3.2.14, where for any hazard rate h with a finite positive limit, $\lim_{x\to\infty}h(x)=\lambda$, the residual life distribution converges to an exponential distribution with parameter λ for $t \to \infty$. Alternatively, it can be obtained by direct calculation

$$\lim_{t\to\infty} S_{\mathcal{E}}^t = \lim_{t\to\infty} \frac{S_{\mathcal{E}}(x+t \mid \gamma,\theta,\beta)}{S_{\mathcal{E}}(x \mid \gamma,\theta,\beta)} = \lim_{t\to\infty} \left(\frac{\gamma + e^{\beta(t+x)} - 1}{\gamma + e^{\beta t} - 1} \right)^{\theta} = e^{-\beta\theta x}.$$

By Proposition 3.3.1, we have for $\mathcal{K}(\alpha,\beta)$ that the limiting distribution of a t-survivor converges to the distribution Exp (1) as $t \to \infty$. □

3.5.3 | Mean Residual Life Function

By Definition 3.2.15 the mean residual life function $v(x)$ is the remaining life expectancy of an x-survivor. In the following two results, we derive the mean residual life functions of the Kannisto and the extended exponential distribution.

Proposition 3.5.3. The mean residual life function $v_{\mathcal{E}}(x \mid \gamma,\theta,\beta)$ of an x-survivor with an extended exponential distributed lifetime $X \sim \mathcal{E}(\gamma,\theta,\beta)$ is given by

$$v_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = \frac{e^{-\beta\theta x} \left(e^{\beta x} + \gamma - 1\right)^{\theta}}{\beta\theta} \, {}_2F_1\left[\theta, \theta; \theta+1; e^{-\beta x}(1-\gamma)\right] \tag{3.44}$$

or equivalently by

$$v_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = \frac{1}{\beta} \left(\frac{\gamma - 1 + e^{\beta x}}{\gamma - 1} \right)^{\theta} B_{\frac{\gamma-1}{\gamma+e^{\beta x}-1}}(\theta,0), \tag{3.45}$$

where $B_z(a,b)$ is the incomplete beta function defined as

$$B_z(a,b) = \int_0^z t^{a-1}(1-t)^{b-1}\, dt. \tag{3.46}$$

Proof. The mean residual life function at age x is defined as

$$\begin{aligned} v_{\mathcal{E}}(x \mid \gamma,\theta,\beta) &= \int_0^\infty \frac{S_{\mathcal{E}}(x+t \mid \gamma,\theta,\beta)}{S_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}\, dt \\ &= \frac{1}{S_{\mathcal{E}}(x \mid \gamma,\theta,\beta)} \int_x^\infty S_{\mathcal{E}}(z \mid \gamma,\theta,\beta)\, dz. \end{aligned} \tag{3.47}$$

The factor outside the integral is the reciprocal of the survival function given as

$$\frac{1}{S_{\mathcal{E}}(x \mid \gamma,\theta,\beta)} = \left(\frac{\gamma}{\gamma + e^{\beta x} - 1} \right)^{-\theta}. \tag{3.48}$$

Let I be the integral term of eq. (3.47). The proof is obtained by direct computation using the integral representation of the incomplete beta function $B_z(a,b)$ and the connection to the hypergeometric function ${}_2F_1$ as will be given in eq. (3.54). We have

$$\begin{aligned}
I &= \int_x^\infty S_{\mathcal{E}}(z \mid \gamma,\theta,\beta)\, dz \\
&= \int_x^\infty \left(\frac{\gamma}{\gamma + e^{\beta z} - 1}\right)^\theta dz \\
&= \frac{\gamma^\theta}{\beta} \int_0^{(e^{\beta x}-(1-\gamma))^{-1}} \frac{s^{\theta-1}}{1-(\gamma-1)s} ds && (3.49) \\
&= \frac{\gamma^\theta}{\beta(\gamma-1)^\theta} \int_0^{\frac{\gamma-1}{\gamma-1+e^{\beta x}}} t^{\theta-1}(1-t)^{-1} dt && (3.50) \\
&= \frac{\gamma^\theta}{\beta(\gamma-1)^\theta} B_{\frac{\gamma-1}{\gamma-1+e^{x\beta}}}(\theta,0) && (3.51) \\
&= \frac{\gamma^\theta}{\beta\theta(\gamma-1+e^{x\beta})^\theta}\, {}_2F_1\left[\theta,1;\theta+1,\frac{\gamma-1}{\gamma-1+e^{x\beta}}\right] && (3.52) \\
&= \frac{\gamma^\theta}{\beta\theta} e^{-\beta\theta x}\, {}_2F_1\left[\theta,\theta;\theta+1;e^{-x\beta}(1-\gamma)\right]. && (3.53)
\end{aligned}$$

Equation (3.49) is obtained by substitution

$$z := \frac{1}{\beta} \ln\left(\frac{1+s-\gamma s}{s}\right),$$

and eq. (3.50) by substitution

$$s := \frac{t}{\gamma-1}.$$

The integraleq. (3.50) corresponds to the incomplete beta function. To derive eq. (3.52) and eq. (3.53), we use the identities

$$B_z(a,b) = \frac{z^a}{a} {}_2F_1[a,1-b;a+1,z] \tag{3.54}$$

and

$${}_2F_1[a,b;c;z] = (1-z)^{-a} {}_2F_1\left[a,c-b;c;\frac{z}{z-1}\right]$$

provided by Olver (2010). Multiplying eq. (3.53) or eq. (3.51) by the factor given in eq. (3.48) completes the proof. □

Corollary 3.5.4. The mean residual life function $v_{\mathcal{K}}(x \mid \alpha,\beta)$ of an x-survivor with a Kannisto distributed lifetime $X \sim \mathcal{K}(\alpha,\beta)$ is given by

$$\nu_{\mathcal{K}}(x \mid \alpha,\beta) = e^{-x}\left(e^{-\alpha} + e^{\beta x}\right)^{\frac{1}{\beta}}_2 F_1\left[\frac{1}{\beta},\frac{1}{\beta};1+\frac{1}{\beta};-e^{-(\alpha+x\beta)}\right]$$

or alternatively by

$$\nu_{\mathcal{K}}(x \mid \alpha,\beta) = \frac{1}{\beta}\left(1+e^{\alpha}\right)^{1/\beta} B_{\frac{1}{1+e^{\alpha+x\beta}}}\left(\frac{1}{\beta},0\right).$$

Proof. The result can be obtained directly from Propositions 3.3.1 and 3.5.3. □

An illustration of the Kannisto mean residual life function for some reference periods is provided in Figure 3.4. Note that for the mean residual lifetime of the extended exponential distribution, we have

$$\lim_{x\to\infty} \nu_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = \frac{1}{\beta\theta}.$$

Thus, the life expectancy of an x-survivor has a lower strictly positive bound $^1/_{\beta\theta}$. This means, regardless how high the attained age x is, the remaining life expectancy does not drop below that particular value. That specific property of the extended exponential or the Kannisto distribution differs from other popular life distributions such as the Weibull or Gompertz distribution. For

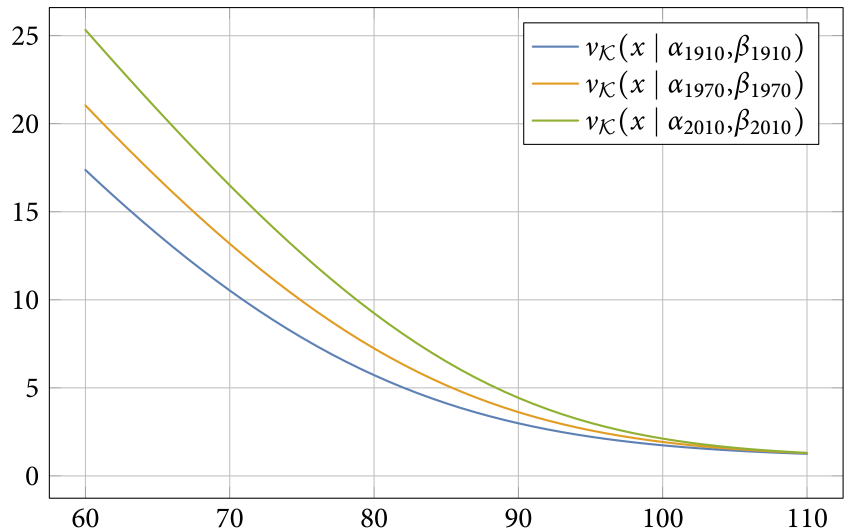


Figure 3.4: Post-age-60 mean residual life function $\nu_{\mathcal{K}}(x \mid \alpha_t,\beta_t)$ of the Kannisto distribution for Swedish females, for the years $t \in \{1910,1970,2010\}$.

instance, the mean residual life function of the Gompertz distribution is obtained by

$$\begin{aligned}
v_{\mathcal{G}}(x \mid \xi,\kappa) &= \int_0^\infty \frac{S_{\mathcal{G}}(x+t \mid \xi,\kappa)}{S_{\mathcal{G}}(x \mid \xi,\kappa)}\,dt \\
&= \int_0^\infty \frac{e^{\xi\left(1-e^{(t+x)\kappa}\right)}}{e^{\xi\left(1-e^{x\kappa}\right)}}\,dt \\
&= \int_0^\infty e^{-\xi e^{\kappa x}\left(e^{\kappa t}-1\right)}\,dt \\
&= -\frac{1}{\kappa} e^{\xi e^{\kappa x}}\,\mathrm{Ei}\left(-e^{x\kappa}\xi\right), \qquad (3.55)
\end{aligned}$$

where the exponential integral $\mathrm{Ei}(x)$ is defined for $x \in \mathbb{R}_{\neq 0}$ as

$$\mathrm{Ei}(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t}\,dt.$$

Taking the limit $x \to \infty$ in eq. (3.55) leads to

$$\lim_{x\to\infty} v_{\mathcal{G}}(x \mid \xi,\kappa) = 0.$$

Using Proposition 3.5.3 and the fact that evaluating the mean residual life at zero leads to the expectation of the corresponding lifetime (see page 164), we can deduce analytic expressions for the expectation of the $\mathcal{E}(\gamma,\theta,\beta)$ and $\mathcal{K}(\alpha,\beta)$ distribution.

Corollary 3.5.5. The expectation of the extended exponential distribution $X \sim \mathcal{E}(\gamma,\theta,\beta)$ is given by

$$\begin{aligned}
\mathbb{E}[X] &= \frac{\gamma^\theta}{\beta\theta}\,{}_2F_1[\theta,\theta;\theta+1;1-\gamma] \qquad (3.56)\\
&= \frac{\gamma^\theta}{\beta(\gamma-1)^\theta} B_{1-\frac{1}{\gamma}}(\theta,0). \qquad (3.57)
\end{aligned}$$

Proof. The mean residual life function evaluated at zero gives the expectation of the corresponding distribution, since $S_{\mathcal{E}}(0 \mid \gamma,\theta,\beta) = 1$, we have

$$v_{\mathcal{E}}(0 \mid \gamma,\theta,\beta) = \frac{1}{S_{\mathcal{E}}(0 \mid \gamma,\theta,\beta)} \int_0^\infty S_{\mathcal{E}}(z \mid \gamma,\theta,\beta)\,dz = \mathbb{E}[X].$$

Thus, setting $x = 0$ in eqs. (3.44) and (3.45) leads to eqs. (3.56) and (3.57). □

Corollary 3.5.6. The expectation of the Kannisto distribution $X \sim \mathcal{K}(\alpha,\beta)$ is given by

$$\begin{aligned}\mathbb{E}[X] &= (1+e^{-\alpha})^{\frac{1}{\beta}}\ {}_2F_1\left[\frac{1}{\beta},\frac{1}{\beta};1+\frac{1}{\beta};-e^{-\alpha}\right] \\ &= \frac{1}{\beta}(1+e^{\alpha})^{\frac{1}{\beta}}\, B_{\frac{1}{1+e^{\alpha}}}\left(\frac{1}{\beta},0\right).\end{aligned}$$

Proof. This follows directly from Corollary 3.5.4 using the fact that $v(0) = \mathbb{E}[X]$. □

3.5.4 | Moment Generating Function of the Kannisto and the Extended Exponential Distribution

The moment generating function is an alternative representative of a probability distribution of a real-valued random variable. The following result provides the *moment generating function* (mgf) of an extended exponential distribution.

Proposition 3.5.7. The moment generating function of the extended exponential distribution $X \sim \mathcal{E}(\gamma,\theta,\beta)$ is given by

$$\mathrm{mgf}_{\mathcal{E}}(s \mid \gamma,\theta,\beta) = \theta\gamma^{\theta}(\gamma-1)^{\frac{s}{\beta}-\theta} B_{1-\frac{1}{\gamma}}\left(\theta-\frac{s}{\beta},1+\frac{s}{\beta}\right), \quad \text{for } s < \beta\theta, \tag{3.58}$$

where $B_z(a,b)$ is the incomplete beta function as defined in eq. (3.46)

Proof. Recall from eq. (3.21) that the density $f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)$ of $X \sim \mathcal{E}(\gamma,\theta,\beta)$ is given by

$$f_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = \frac{\beta\theta}{\gamma} e^{\beta x}\left(\frac{\gamma}{\gamma+e^{\beta x}-1}\right)^{\theta+1}.$$

We proof eq. (3.58) by direct calculation. We have

$$\begin{aligned}\mathrm{mgf}_X(s \mid \gamma,\theta,\beta) &= \mathbb{E}[e^{sX}] \\ &= \frac{\beta\theta}{\gamma}\int_0^{\infty} e^{sx}\, e^{\beta x}\left(\frac{\gamma}{\gamma+e^{\beta x}-1}\right)^{\theta+1} dx \qquad (3.59) \\ &= \theta\int_0^1 z^{\theta-1}\left(\gamma\left(\frac{1}{z}-1\right)+1\right)^{s/\beta} dz \qquad (3.60) \\ &= \theta\gamma^{\theta}(\gamma-1)^{\frac{s}{\beta}-\theta}\int_0^{1-\frac{1}{\gamma}} t^{\theta-1}\left(\frac{1}{t}-1\right)^{s/\beta} dt \qquad (3.61) \\ &= \theta\gamma^{\theta}(\gamma-1)^{\frac{s}{\beta}-\theta}\int_0^{1-\frac{1}{\gamma}} t^{\theta-\frac{s}{\beta}-1}(1-t)^{\frac{s}{\beta}} dt \qquad (3.62) \\ &= \theta\gamma^{\theta}(\gamma-1)^{\frac{s}{\beta}-\theta} B_{1-\frac{1}{\gamma}}\left(\theta-\frac{s}{\beta},1+\frac{s}{\beta}\right), \quad \text{for } s < \beta\theta.\end{aligned}$$

To obtain eq. (3.60), we use the substitution $x := {}^{1}\!/\!_{\beta} \ln\left(\frac{\gamma-\gamma z+z}{z}\right)$ and for eq. (3.61) the substitution $z := t\frac{\gamma}{\gamma-1}$. The integral expression in eq. (3.62) is the incomplete beta function as given in eq. (3.46). □

By using the hypergeometric representation of the incomplete beta function, i.e.,

$$B_x(a,b) = \frac{x^a(1-x)^b}{a}\, {}_2F_1\left[a+b,1;a+1;x\right], \tag{3.63}$$

(see, e.g., Olver, 2010) we can represent the moment generating function of the extended exponential distribution in terms of a hypergeometric function. Using eq. (3.63), we obtain

$$\mathrm{mgf}_X(s \mid \gamma,\theta,\beta) = \frac{\beta\theta}{\gamma(\beta\theta - s)}\, {}_2F_1\left[1,\theta+1;\theta+1-\frac{s}{\beta};1-\frac{1}{\gamma}\right], \quad \text{for } s < \beta\theta. \tag{3.64}$$

Corollary 3.5.8. The moment generating function of the Kannisto distribution $X \sim \mathcal{K}(\alpha,\beta)$ is given by

$$\mathrm{mgf}_{\mathcal{K}}(s \mid \alpha,\beta) = \mathbb{E}[e^{sX}] = \frac{1}{\beta}e^{-\frac{\alpha}{\beta}}\left(1+e^{-\alpha}\right)^{\frac{s-1}{\beta}} B_{\frac{1}{1+e^{\alpha}}}\left(\frac{1-s}{\beta},1+\frac{s}{\beta}\right), \quad \text{for } s < 1. \tag{3.65}$$

Proof. We use again the fact that the Kannisto distribution is a special case of the extended exponential distribution as shown in Proposition 3.3.1 and substitute $\gamma = 1 + e^{-\alpha}$ and $\theta = {}^{1}\!/\!_{\beta}$ into eq. (3.58). □

The representation of eq. (3.65) in terms of a hypergeometric function follows by substituting $\theta = {}^{1}\!/\!_{\beta}$ and $\gamma = 1 + e^{-\alpha}$ into eq. (3.64). Thus, we have

$$\mathrm{mgf}_{\mathcal{K}}(s \mid \alpha,\beta) = \frac{1}{(1+e^{-\alpha})(1-s)}\, {}_2F_1\left[1,1+\frac{1}{\beta};1+\frac{1}{\beta}-\frac{s}{\beta};\frac{1}{1+e^{\alpha}}\right], \quad \text{for } s < 1.$$

The next result is not directly related to the moment generating function of the extended exponential distribution. However, it turns out that the life expectancy of a competing risks lifetime, as specified in the following proposition, involves an integral which is of the same type as in eq. (3.59).

Corollary 3.5.9. Let $Z := \min(X,Y)$ be a competing risks lifetime of two independent risk factors, where the first factor is $X \sim \mathcal{E}(\gamma,\theta,\beta)$ distributed and the second follows an exponential distribution, i.e., $Y \sim \mathrm{Exp}(\lambda)$. Then, the expectation Z is given by

$$\mathbb{E}[Z] = \frac{1}{\beta\theta+\lambda}\, {}_2F_1\left[1,\theta;1+\theta+\frac{\lambda}{\beta};1-\frac{1}{\gamma}\right].$$

Proof. Since Z is a competing risks lifetime variable, it has the hazard rate (see Proposition 3.2.17)

$$h_Z(x \mid \gamma,\theta,\beta,\lambda) = h_{\mathcal{E}} + h_{\mathrm{Exp}} = \frac{\beta\theta e^{\beta x}}{e^{\beta x}-(1-\gamma)} + \lambda,$$

where $h_{\mathcal{E}}$ is the hazard rate of $\mathcal{E}(\gamma,\theta,\beta)$ and h_{Exp} the hazard rate of an exponential distribution with parameter λ. The survival function S_Z is given by the product of the individual survival functions, i.e.,

$$S_Z(x \mid \gamma,\theta,\beta,\lambda) = e^{-\int_0^x h_Z(z|\gamma,\theta,\beta,\lambda)\,dz} = S_{\mathcal{E}}(x \mid \gamma,\theta,\beta)\, S_{\text{Exp}}(x \mid \lambda) = \left(\frac{\gamma}{e^{\beta x} - (1-\gamma)}\right)^{\theta} e^{-\lambda x}.$$

The expectation of Z is given by the integral $\int_0^\infty S_{\mathcal{E}}(x \mid \gamma,\theta,\beta)\, S_{\text{Exp}}(x \mid \lambda)\,dx$. Comparing the corresponding integral with the integral involved in the computation of eq. (3.59), we observe that $\mathbb{E}[Z]$ can be written in terms of the moment generating function of the extended exponential distribution as derived in Proposition 3.5.7. Using that, we obtain

$$\begin{aligned}
\mathbb{E}[Z] &= \int_0^\infty S_{\mathcal{E}}(x \mid \gamma,\theta,\beta)\, S_{\text{Exp}}(x \mid \lambda) dx \\
&= \int_0^\infty \left(\frac{\gamma}{e^{\beta x} - (1-\gamma)}\right)^{\theta} e^{-\lambda x}\,dx \\
&= \frac{\gamma}{\beta\theta}\,\text{mgf}_{\mathcal{E}}(-(\beta+\lambda) \mid \gamma,\theta-1,\beta) \\
&= \frac{1}{\beta\theta+\lambda}\,{}_2F_1\left[1,\theta;1+\theta+\frac{\lambda}{\beta};1-\frac{1}{\gamma}\right].
\end{aligned}$$

□

3.5.5 | Moments of the Kannisto and the Extended Exponential Distribution

In the following section, we will derive a concrete analytic expression of the n-th moment for the extended exponential as well as for the Kannisto distribution. Since the extended exponential distribution has a non-decreasing hazard function we know that by Proposition 3.2.12 all moments exist, i.e., for $X \sim \mathcal{E}(\gamma,\theta,\beta)$, we have $\mathbb{E}[X^n] < \infty$. The central moments could also be obtained by the moment generating function as given in Proposition 3.5.7. However, that would involve the differentiation of the incomplete beta function in both arguments, cf., eq. (3.58). The following result provides an expression of the moments of $\mathcal{E}(\gamma,\theta,\beta)$ in terms of generalized hypergeometric functions.

Theorem 3.5.10. The n-th moment of the extended exponential random variable $X \sim \mathcal{E}(\gamma,\theta,\beta)$ with survival function

$$S_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = \left(\frac{\gamma}{e^{\beta x} - (1-\gamma)}\right)^{\theta} \qquad \text{for } x \geq 0,\ \beta,\theta > 0, \gamma \geq 1,$$

is given by

$$\mathbb{E}[X^n] = \frac{n!\,\gamma^\theta}{\beta^n\theta^n}\,{}_{n+1}F_n\left[\underbrace{\theta,\ldots,\theta}_{n+1};\underbrace{1+\theta,\ldots,1+\theta}_{n};1-\gamma\right].$$

Proof. We start with the series representation of generalized hypergeometric function. Further transformations will be given below.

$$
\begin{aligned}
{}_{n+1}F_n\left[\underbrace{\theta,\ldots,\theta}_{n+1};\underbrace{1+\theta,\ldots,1+\theta}_{n};1-\gamma\right] &= \sum_{k=0}^{\infty}\frac{\overbrace{(\theta)_k\times\cdots\times(\theta)_k}^{n+1}}{\underbrace{(1+\theta)_k\times\ldots\times(1+\theta)_k}_{n}}\frac{(1-\gamma)^k}{k!}\\
&= \sum_{k=0}^{\infty}\frac{\Gamma(\theta+k)^{n+1}}{\Gamma(\theta)^{n+1}}\frac{\Gamma(1+\theta)^n}{\Gamma(k+1+\theta)^n}\frac{(1-\gamma)^k}{k!} && (3.66)\\
&= \sum_{k=0}^{\infty}\frac{\theta^n}{(\theta+k)^n}(\theta)_k\frac{(1-\gamma)^k}{k!} && (3.67)\\
&= \sum_{k=0}^{\infty}\theta^n(\theta)_k\frac{(1-\gamma)^k}{k!}\frac{1}{\Gamma(n)}\int_0^{\infty} z^{n-1}e^{-kz}e^{-\theta z}\,dz && (3.68)\\
&= \frac{\theta^n}{\Gamma(n)}\int_0^{\infty} z^{n-1}e^{-\theta z}\left(\sum_{k=0}^{\infty}(\theta)_k\frac{(1-\gamma)^k e^{-kz}}{k!}\right)dz && (3.69)\\
&= \frac{\theta^n}{\Gamma(n)}\int_0^{\infty} z^{n-1}e^{-\theta z}\left(\sum_{k=0}^{\infty}(\theta)_k\frac{(e^{-z}(1-\gamma))^k}{k!}\right)dz\\
&= \frac{\theta^n}{\Gamma(n)}\int_0^{\infty} z^{n-1}e^{-\theta z}\,{}_1F_0\left[\theta;;e^{-z}(1-\gamma)\right]dz && (3.70)\\
&= \frac{\theta^n}{\Gamma(n)}\int_0^{\infty} z^{n-1}e^{-\theta z}\left(1-e^{-z}(1-\gamma)\right)^{-\theta}dz && (3.71)\\
&= \frac{\theta^n}{(n-1)!}\int_0^{\infty} z^{n-1}\left(\frac{e^{-z}}{1-e^{-z}(1-\gamma)}\right)^{\theta}dz\\
&= \frac{\theta^n}{(n-1)!}\int_0^{\infty} z^{n-1}\left(\frac{1}{e^{z}-(1-\gamma)}\right)^{\theta}dz\\
&= \frac{\theta^n\beta^n}{(n-1)!}\int_0^{\infty} x^{n-1}\left(\frac{1}{e^{\beta x}-(1-\gamma)}\right)^{\theta}dx\\
&= \frac{\theta^n\beta^n}{\gamma^{\theta}(n-1)!}\int_0^{\infty} x^{n-1}\left(\frac{\gamma}{e^{\beta x}-(1-\gamma)}\right)^{\theta}dx && (3.72)\\
&= \frac{\theta^n\beta^n}{\gamma^{\theta}n!}\int_0^{\infty} n\,x^{n-1}S_{\mathcal{E}}(x\mid\gamma,\theta,\beta)\,dx\\
&= \frac{\theta^n\beta^n}{\gamma^{\theta}n!}\mathbb{E}[X^n].
\end{aligned}
$$

Equation (3.66) follows directly from the definition of the Pochhammer symbols as rising factorials given by

$$(\theta)_n = \theta(\theta+1)\cdots(\theta+n) = \frac{\Gamma(\theta+n)}{\Gamma(\theta)}.$$

Equation (3.67) follows from the identity

$$\frac{\theta^n}{(\theta+k)^n}(\theta)_k = \frac{\Gamma(\theta+k)^{n+1}}{\Gamma(\theta)^{n+1}}\frac{\Gamma(1+\theta)^n}{\Gamma(k+1+\theta)^n},$$

hence the gamma function satisfies the functional equation $\Gamma(\theta+1) = \theta\,\Gamma(\theta)$. Equation (3.68) is derived by the Laplace transformation where the term $\Gamma(n)(\theta+k)^{-n}$ is replaced by the Laplace transform of y^{n-1} at $\theta+k$, i.e.,

$$\frac{\Gamma(n)}{(\theta+k)^n} = \int_0^\infty y^{n-1}e^{-(\theta+k)y}\,dy.$$

Equation (3.69) is derived by interchanging the order summation and integration using Fubini's theorem, considering the sum as integration with respect to a counting measure on $\mathbb{N}_0$. Equation (3.70) is obtained by recognizing the series representation of a hypergeometric function ${}_1F_0$, i.e.,

$${}_1F_0\left[\theta;;e^{-z}(1-y)\right] = \sum_{k=0}^{\infty}(\theta)_k\frac{(e^{-z}(1-y))^k}{k!},$$

which is simplified in eq. (3.71) using the identity

$${}_1F_0[a;;z] = (1-z)^a,$$

which is stated in Olver (2010). Rearranging terms and the substitution $z = \beta x$ yields eq. (3.72). From eq. (3.72), we see that scaling the generalized hypergeometric function with the factor $\frac{n!\,y^\theta}{\beta^n\theta^n}$ results in the desired expression

$$\mathbb{E}[X^n] = \frac{n!\,y^\theta}{\beta^n\theta^n}\;{}_{n+1}F_n\left[\underbrace{\theta,\ldots,\theta}_{n+1};\underbrace{1+\theta,\ldots,1+\theta}_{n};1-y\right].$$

□

Corollary 3.5.11. The n-th moment of the Kannisto distribution $\mathcal{K}(\alpha,\beta)$ is given by

$$\mathbb{E}[X^n] = n!\,(1+e^{-\alpha})^{\frac{1}{\beta}}\ {}_{n+1}F_n\left[\underbrace{\frac{1}{\beta},\ldots,\frac{1}{\beta}}_{n+1};\underbrace{1+\frac{1}{\beta},\ldots,1+\frac{1}{\beta}}_{n};-e^{-\alpha}\right]. \tag{3.73}$$

Proof. The result is a consequence of Theorem 3.5.10 and the fact that the Kannisto distribution is a special case of the extended exponential distribution as shown in Proposition 3.3.1. □

For $n = 1$ in eq. (3.73), we obtain the expectation of a $\mathcal{K}(\alpha,\beta)$ distributed lifetime X as

$$\mathbb{E}[X] = (1+e^{-\alpha})^{\frac{1}{\beta}}\ {}_2F_1\left[\frac{1}{\beta},\frac{1}{\beta};1+\frac{1}{\beta};-e^{-\alpha}\right]. \tag{3.74}$$

This result corresponds to the expression derived from the mean residual life function as in Corollary 3.5.4. By using Pfaff's transformation formula for hypergeometric functions

$${}_2F_1\,[a,b;c;z] = (1-z)^{-b}{}_2F_1\left[c-a,b;c;\frac{z}{z-1}\right], \tag{3.75}$$

we can obtain

$$\mathbb{E}[X] = {}_2F_1\left[1,\frac{1}{\beta};1+\frac{1}{\beta};\frac{1}{1+e^{\alpha}}\right]$$

as an alternative representation of eq. (3.74).

Let $v_{\mathcal{K}}(\alpha,\beta)$ be the expectation of $\mathcal{K}(\alpha,\beta)$, then $v_{\mathcal{K}}(\alpha,\beta)$ is decreasing for an increasing α, since

$$\frac{1}{1+e^{\alpha_1}} \le \frac{1}{1+e^{\alpha_2}} \quad \text{for } \alpha_1 \ge \alpha_2,$$

and it is also decreasing for increasing β, since

$$\frac{({}^1/_\beta)_n}{(1+{}^1/_\beta)_n} = \frac{1}{\beta n+1}$$

is decreasing for increasing β and any fixed $n \in \mathbb{N}_0$.

Proposition 3.5.12. The variance of the Kannisto distribution is given by

$$\mathbb{V}(X) = C(\alpha,\beta)\left(2\,{}_3F_2\left[\frac{1}{\beta},\frac{1}{\beta},\frac{1}{\beta};1+\frac{1}{\beta},1+\frac{1}{\beta};-e^{-\alpha}\right] - C(\alpha,\beta)\,{}_2F_1\left[\frac{1}{\beta},\frac{1}{\beta};1+\frac{1}{\beta};-e^{-\alpha}\right]^2\right)$$

with

$$C(\alpha,\beta) := (1+e^{-\alpha})^{\frac{1}{\beta}}\,.$$

Proof. Use eq. (3.73) to substitute in $\mathbb{V}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. □

Next, we will list some asymptotic properties of the $\mathcal{K}(\alpha,\beta)$ and $\mathcal{E}(\gamma,\theta,\beta)$ moments which will require the following definition.

Definition 3.5.13 (Asymptotic equivalence). Let f and g be real valued functions. We say that f and g are *asymptotically equal* (at ∞), if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

Asymptotic equivalence will be denoted by $f(x) \sim g(x)$ and means that the growth of $f(x)$ and $g(x)$ is of the same type when x gets large.

Remark 3.5.14 (Approximation and Asymptotic Behaviour of the Kannisto Expectation). The expectation function $\nu_{\mathcal{K}}(\alpha,\beta)$ has the following asymptotic behaviour

$$\lim_{\alpha\to\infty} \nu_{\mathcal{K}}(\alpha,\beta) = \lim_{\beta\to\infty} \nu_{\mathcal{K}}(\alpha,\beta) = 1, \tag{3.76}$$

$$\lim_{\beta\to 0} \nu_{\mathcal{K}}(\alpha,\beta) = 1 + e^{-\alpha}, \tag{3.77}$$

$$\lim_{\alpha\to-\infty} \frac{\nu_{\mathcal{K}}(\alpha,\beta)}{\alpha} = -\frac{1}{\beta}.$$

The limits of eqs. (3.76) and (3.77) can also be established from the limiting distributions, which is the exponential distribution with parameter 1 in the first case and the exponential distribution with parameter $(1 + e^{-\alpha})^{-1}$ in the second case, cf., Figure 3.1 for an overview of the Kannisto limiting distributions. Taylor series expansion of $\nu_{\mathcal{K}}(\alpha,\beta)$ for α at $-\infty$, which is essentially an expansion of

$$\nu_{\mathcal{K}}\left(\ln\left(\frac{1-z}{z}\right),\beta\right) = {}_2F_1\left[1,\frac{1}{\beta};1+\frac{1}{\beta};z\right]$$

at $z = 1$ with $z := \frac{1}{1+e^{\alpha}}$ shows that for $e^{\alpha} \approx 0$ the expectation can be approximated by

$$\mathbb{E}[X] = {}_2F_1\left[1,\frac{1}{\beta};1+\frac{1}{\beta};\frac{1}{1+e^{\alpha}}\right] \approx -\frac{\alpha}{\beta} - \frac{\psi(1)+\psi\left({}^1\!/\!_\beta\right)}{\beta},$$

where ψ denotes the polygamma function, defined by

$$\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)},$$

and $\psi(1) = -\tilde{\gamma} \approx -0.57721$ is also known as the Euler-Mascheroni constant. Comparing this linearization with regard to the parameter α, we observe a relative error of about 2.5% in the parameter region obtained for real life data of the past century, see Table 1.15. Note that the approximation tends to improve since the observed α values decrease over time.

Remark 3.5.15 (Asymptotic Properties of the $\mathcal{E}(\gamma,\theta,\beta)$ n-th moments).
In the following part, we list some asymptotic properties and special cases for the central moments of the distribution $\mathcal{E}(\gamma,\theta,\beta)$.

(a) For $X \sim \mathcal{E}(\gamma,\theta,\beta)$, we have

$$\lim_{\gamma\to\infty} \frac{\mathbb{E}[X^n]}{(-1)^n\beta^{-n}\ln(\gamma-1)^n} = 1,$$

i.e., for $\gamma \to \infty$ the n-th moment is asymptotically equivalent to the function

$$\lambda(\gamma \mid \beta,n) := (-1)^n\beta^{-n}\ln(\gamma-1)^n.$$

(b) For $\gamma \to 1$, the extended exponential distribution $X \sim \mathcal{E}(\gamma,\theta,\beta)$ converges to the exponential distribution $\mathrm{Exp}(\beta\theta)$, hence we have

$$\lim_{\gamma\to 1}\mathbb{E}[X^n] = \frac{n!}{\beta^n\theta^n}.$$

(c) Since

$${}_{n+1}F_n\left[\theta,\ldots,\theta;1+\theta,\ldots,1+\theta;1-\gamma\right]\Big|_{\theta=0} = 1,$$

we have

$$\lim_{\theta\to 0}\frac{\mathbb{E}[X^n]}{\frac{n!\,\gamma^\theta}{\beta^n\theta^n}} = 1.$$

(d) For $\theta = 1$, we have the special case where only a tilt extension of the exponential distribution is considered. The generalized hypergeometric function ${}_{n+1}F_n[\theta,\ldots,\theta;1+\theta,\ldots,1+\theta]$ reduces to the polylogarithm function. This can be deduced by using the series representation of the generalized hypergeometric function and by simplifying the integer valued Pochhammer functions. Thus,

$$\begin{aligned}
{}_{n+1}F_n\left[\begin{matrix}1,\ldots,1\\2,\ldots,2\end{matrix};1-\gamma\right] &= \sum_{m=0}^{\infty}\frac{\prod_{i=1}^{n+1}(1)_m}{\prod_{j=1}^{n}(2)_m}\frac{(1-\gamma)^m}{m!}\\
&= \sum_{m=0}^{\infty}\frac{(1-\gamma)^m}{(1+m)^n}\\
&= \frac{\mathrm{Li}_n(1-\gamma)}{1-\gamma}.
\end{aligned}$$

The last expression contains the polylogarithm function defined by

$$\mathrm{Li}_n(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^n}.$$

Thus, for $X \sim \mathcal{E}(\gamma,1,\beta)$, we conclude

$$\mathbb{E}[X^n] = \frac{n!\,\gamma}{\beta^n(1-\gamma)}\,\mathrm{Li}_n(1-\gamma).$$

This can be further generalized for arbitrary $\theta \in \mathbb{N}$. For θ being an integer, we have the following reduction of the generalized hypergeometric function

$$_{n+1}F_n\left[\begin{matrix}\theta,\ldots,\theta \\ 1+\theta,\ldots,1+\theta\end{matrix};1-\gamma\right] = \frac{\theta^n \sum_{i=1}^{\theta} S(\theta,i)\,\mathrm{Li}_{n+1-i}(1-\gamma)}{(\theta-1)!(1-\gamma)^\theta}, \tag{3.78}$$

where $S(\theta,i)$ denotes the Stirling number of the first kind, defined as the number of permutation of θ elements with exactly i cycles multiplied by the factor $(-1)^{\theta-i}$, see Olver (2010). Using the reduction formula of eq. (3.78), we obtain

$$\mathbb{E}[X^n] = \frac{\left(\frac{\gamma}{1-\gamma}\right)^\theta}{\Gamma(\theta)\beta^n}\sum_{i=1}^{\theta} S(\theta,i)\,\mathrm{Li}_{n+1-i}(1-\gamma).$$

This special case expression is useful since θ is a frailty parameter of the $\mathcal{E}(\gamma,\theta,\beta)$ distribution and thus, we can consider the survival function of $\mathcal{E}(\gamma,\theta,\beta)$ as the θ-fold product of tilt extended exponential distribution survival functions which is indeed a competing risks model with θ independent risk factors following the distribution $\mathcal{E}(\gamma,1,\beta)$.

(e) For $X \sim \mathcal{E}(\gamma,\theta,\beta)$, we have the following asymptotic behaviour of the n-th moment

$$\mathbb{E}[X^n] \sim \frac{n!\gamma^\theta}{\beta^n\theta^n} \quad \text{as } n \to \infty,$$

since

$$_{n+1}F_n[\theta,\ldots,\theta;1+\theta,\ldots,1+\theta;1-\gamma] \overset{n\to\infty}{\longrightarrow} 1 \quad \text{for } \theta > 0, \gamma > 1.$$

Thus, we obtain the asymptotic moment ratios as

$$\frac{\mathbb{E}[X^n]}{\mathbb{E}[X^{n-1}]} \sim \frac{n}{\beta\theta} \quad \text{as } n \to \infty.$$

This asymptotic ratio property of the $\mathcal{E}(\gamma,\theta,\beta)$ moments is inherited from the underlying exponential distribution, where for $Y \sim \mathrm{Exp}(\lambda)$, we have

$$\frac{\mathbb{E}[Y^n]}{\mathbb{E}[Y^{n-1}]} = \frac{\lambda^{-n} n!}{\lambda^{-(n-1)}(n-1)!} = \frac{n}{\lambda} \quad \forall n \in \mathbb{N}.$$

(f) For $X \sim \mathcal{E}(\gamma,\theta,\beta)$ and $n \in \mathbb{N}$, we have

$$n!\left(\frac{1}{\beta\theta}\right)^n < \mathbb{E}[X^n] < n!\left(\frac{\gamma}{\beta\theta}\right)^n. \tag{3.79}$$

The inequality holds, since

$$\frac{\beta\theta}{\gamma} < h_{\mathcal{E}}(x \mid \gamma,\theta,\beta) < \beta\theta \quad \text{for all } x > 0.$$

Both inequalities of eq. (3.79) arise from the n-th moments of the exponential distributions $\mathrm{Exp}(\gamma/\beta\theta)$ and $\mathrm{Exp}(1/\beta\theta)$.

(g) For $X \sim \mathcal{E}(\gamma,\theta,\beta)$ and $n \in \mathbb{N}$, we have

$$\frac{n!}{(\beta\theta)^{n-1}}\mathbb{E}[X] \le \frac{n}{\beta\theta}\mathbb{E}[X^{n-1}] \le \mathbb{E}[X^n] \le n!\mathbb{E}[X]^n \tag{3.80}$$

and for the coefficient of variation

$$\frac{\sqrt{\mathbb{V}[X]}}{\mathbb{E}[X]} \le 1. \tag{3.81}$$

The last inequity of eq. (3.80) holds for the class of increasing hazard rate distributions, see Rinne (2014) which directly implies eq. (3.81). The first inequality is obtained by an iterative application of the second inequality, which follows from the fact

$$1 \le \frac{{}_{n+1}F_n\left[\theta,\ldots,\theta;1+\theta,\ldots,1+\theta;1-\gamma\right]}{{}_{n}F_{n-1}\left[\theta,\ldots,\theta;1+\theta,\ldots,1+\theta;1-\gamma\right]} \quad \text{for } \gamma \ge 1.$$

Remark 3.5.16 (Approximate Expressions for Expectation, Variance, Skewness, and Kurtosis). Taylor series expansion of the $\mathcal{E}(\gamma,\theta,\beta)$ moments at $\gamma = \infty$ lead to the following approximations:

$$\begin{aligned}
\mathbb{E}[X] &= \frac{1}{\beta}\left(\ln(\gamma) - \psi^{(0)}(\theta) + \psi^{(0)}(1)\right) + \mathcal{O}\left(\gamma^{-1}\right),\\
\mathbb{V}[X] &= \frac{\pi^2}{\beta^2} + \frac{\psi^{(1)}(\theta)}{\beta^2} + \mathcal{O}\left(\gamma^{-1}\right),\\
\mathrm{skew}[X] &= \frac{6\sqrt{6}\left(\psi^{(2)}(1) - \psi^{(2)}(\theta)\right)}{\left(6\psi^{(1)}(\theta) + \pi^2\right)^{3/2}} + \mathcal{O}\left(\gamma^{-1}\right),
\end{aligned}$$

$$\operatorname{kurt}[X] = \frac{9\left(60\psi^{(1)}(\theta)^2 + 20\pi^2\psi^{(1)}(\theta) + 20\psi^{(3)}(\theta) + 3\pi^4\right)}{5\left(6\psi^{(1)}(\theta) + \pi^2\right)^2} + \mathcal{O}\left(\gamma^{-1}\right).$$

The function $\psi^{(m)}$ denotes the polygamma function of order m defined as the $(m+1)$-th derivative of the logarithm of the gamma function.

$$\psi^{(m)}(z) := \frac{\partial^m}{\partial z^m}\psi(z) = \frac{\partial^{m+1}}{\partial z^{m+1}}\ln\Gamma(z).$$

The next higher Taylor expansion yields the following approximations:

$$\mathbb{E}[X] = \frac{1}{\beta\gamma}\left(\psi^{(0)}(1)(\gamma+\theta) + (\gamma+\theta)\ln(\gamma) - (\gamma+\theta)\psi^{(0)}(\theta) + \theta - 1\right) + \mathcal{O}\left(\gamma^{-2}\right), \tag{3.82}$$

$$\begin{aligned}\mathbb{V}[X] = \frac{1}{6\beta^2\gamma}\Big(&-6\theta\ln(\gamma)(2\psi^{(0)}(1) + \ln(\gamma) + 2) - 6(\psi^{(0)}(1)(\psi^{(0)}(1) + 2) + 2)\theta + \pi^2(\gamma+\theta)\\ &+ 12\theta\psi^{(0)}(\theta)(\psi^{(0)}(1) + \ln(\gamma) + 1) + 6(\gamma+\theta)\psi^{(1)}(\theta) - 6\theta\psi^{(0)}(\theta)^2\Big) + \mathcal{O}\left(\gamma^{-2}\right).\end{aligned} \tag{3.83}$$

We observe a very high accuracy of the approximations in eqs. (3.82) and (3.83). On average, the relative deviation, in the region of estimated Kannisto parameters, is about 0.1% for the expectation approximation, and about −1.8% for the variance approximation in eq. (3.83).

3.5.6 | Order Statistics

In the next result, we derive the probability density function of the r-th order statistic from an n sample of the extended exponential distribution. In actuarial applications, the following result can be used to obtain statements about the occurrence of the r-th death in a group of n individuals. For example, the probability that the first death takes place within one year or the expected age at death of the last survivor, which is useful in actuarial pension calculations.

Theorem 3.5.17. Let $X_{r:n}$ by the r-th order statistic from an i.i.d. sequence $X_1,\ldots,X_n$, with $X_i \sim \mathcal{E}(\gamma,\theta,\beta)$. The probability density function of the r-th order statistic is given by,

$$f_{\mathcal{E}_{r:n}}(x \mid \gamma,\theta,\beta) = \frac{n!}{(r-1)!(n-r)!}\sum_{i=0}^{r-1}\frac{(-1)^i\binom{r-1}{i}}{(n-r+i+1)}f_{\mathcal{E}}(x \mid \gamma, \theta(n-r+i+1),\beta). \tag{3.84}$$

Proof. In general, the probability density function of the r-th order statistic of X is given by (Arnold, Balakrishnan and Nagaraja, 1992)

$$f_{X_{r:n}}(x) = \frac{n!}{(r-1)!(n-r)!}f_X(x)[F_X(x)]^{r-1}[1-F_X(x)]^{n-r}.$$

Thus, for $X \sim \mathcal{E}(\gamma,\theta,\beta)$, we have

$$f_{X_{r:n}}(x \mid \gamma,\theta,\beta) = \frac{n!}{(r-1)!(n-r)!}f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)[F_{\mathcal{E}}(x \mid \gamma,\theta,\beta)]^{r-1}[1-F_{\mathcal{E}}(x \mid \gamma,\theta,\beta)]^{n-r}$$

$$= \frac{n!}{(r-1)!(n-r)!} h_{\mathcal{E}}(x \mid \gamma,\theta,\beta)[F_{\mathcal{E}}(x \mid \gamma,\theta,\beta)]^{r-1}[1 - F_{\mathcal{E}}(x \mid \gamma,\theta,\beta)]^{n-r+1} \tag{3.85}$$

$$= \frac{n!}{(r-1)!(n-r)!} \sum_{i=0}^{r-1} \binom{r-1}{i} (-1)^i h_{\mathcal{E}}(x \mid \gamma,\theta,\beta)[1 - F_{\mathcal{E}}(x \mid \gamma,\theta,\beta)]^{n-r+i+1} \tag{3.86}$$

$$= \frac{n!}{(r-1)!(n-r)!} \sum_{i=0}^{r-1} \binom{r-1}{i} (-1)^i h_{\mathcal{E}}(x \mid \gamma,\theta,\beta) S_{\mathcal{E}}(x \mid \gamma,\theta,\beta)^{n-r+i+1}$$

$$= \frac{n!}{(r-1)!(n-r)!} \sum_{i=0}^{r-1} \binom{r-1}{i} (-1)^i \frac{h_{\mathcal{E}}(x \mid \gamma,\theta(n-r+i+1),\beta)}{n-r+i+1} \times S_{\mathcal{E}}(x \mid \gamma,\theta(n-r+i+1),\beta) \tag{3.87}$$

$$= \frac{n!}{(r-1)!(n-r)!} \sum_{i=0}^{r-1} \frac{(-1)^i \binom{r-1}{i}}{(n-r+i+1)} f_{\mathcal{E}}(x \mid \gamma, \theta(n-r+i+1),\beta),$$

where for eq. (3.85), we used the relation

$$f_{\mathcal{E}} = h_{\mathcal{E}} S_{\mathcal{E}} = h_{\mathcal{E}}[1 - F_{\mathcal{E}}], \tag{3.88}$$

and eq. (3.86) is obtained by the binomial expansion, i.e.,

$$\begin{aligned} F_{\mathcal{E}}(x \mid \gamma,\theta,\beta)^{r-1} &= [1 - (1 - F_{\mathcal{E}}(x \mid \gamma,\theta,\beta))]^{r-1} \\ &= \sum_{i=0}^{r-1} \binom{r-1}{i} (-1)^i [1 - F_{\mathcal{E}}(x \mid \gamma,\theta,\beta)]^i. \end{aligned}$$

To obtain eq. (3.87), we use the fact that $\mathcal{E}(\gamma,\theta,\beta)$ belongs to the family of frailty distributions. Thus, we have

$$S_{\mathcal{E}}(x \mid \gamma,\theta,\beta)^{n-r+i+1} = S_{\mathcal{E}}(x \mid \gamma,\theta(n-r+i+1),\beta)$$

and

$$h_{\mathcal{E}}(\gamma,\theta,\beta) = \frac{\beta\theta e^{\beta x}}{\gamma + e^{\beta x} - 1} = \frac{1}{(n-r+i+1)} h_{\mathcal{E}}(\gamma,\theta(n-r+i+1),\beta).$$

The proof is completed by reusing the relation of eq. (3.88). □

Note that the distribution density of the r-th order statistic given in eq. (3.84) has the form of a discrete mixture of extended exponential distributions with different frailty parameters $\theta(n-r+i+1)$ weighted by $w_i(r,n)$, with

$$w_i(r,n) = \frac{(-1)^i \binom{r-1}{i} n!}{(n-r+i+1)(r-1)!(n-r)!}, \quad \text{for } 0 \le r \le n \text{ and } 0 \le i \le r-1. \tag{3.89}$$

Corollary 3.5.18. The first-order statistic $X_{1:n}$ of an i.i.d. sample $X_1,\ldots,X_n$ from the extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$ follows the distribution

$$X_{1:n} \sim \mathcal{E}(\gamma,n\theta,\beta).$$

Proof. The proof follows directly from Theorem 3.5.17 by setting $r = 1$ or by using the relation

$$\begin{aligned} F_{X_{1:n}} &= 1 - [1 - F_{\mathcal{E}}(x \mid \gamma,\theta,\beta)]^n \\ &= 1 - S_{\mathcal{E}}(x \mid \gamma,n\theta,\beta) \\ &= F_{\mathcal{E}}(x \mid \gamma,n\theta,\beta), \end{aligned} \tag{3.90}$$

where eq. (3.90) follows from the fact that $\mathcal{E}(\gamma,\theta,\beta)$ belongs to the frailty family. □

Notice here that the first-order statistic remains extended exponentially distributed. This property is also shared by the exponential distribution, where for i.i.d. $X_i \sim \text{Exp}(\lambda)$ the following holds $X_{1:n} = \min(X_1,\ldots,X_n) \sim \text{Exp}(n\lambda)$. Furthermore, since θ is the frailty parameter, we can conclude by Proposition 3.2.17 that the distribution of the first-order statistic corresponds to a competing risks lifetime with n individual $\mathcal{E}(\gamma,\theta,\beta)$ factors.

Corollary 3.5.19. The m-th moment of the r-th order statistic $X_{r:n}$ of the extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$ is given by

$$\begin{aligned} \mathbb{E}[X_{r:n}^m] = \sum_{i=0}^{r-1} w_i(r,n) &\frac{m!\,\gamma^{\theta(n-r+i+1)}}{\beta^m(\theta(n-r+i+1))^m} \\ &{}_{m+1}F_m\left[\begin{matrix}\theta(n-r+i+1),\ldots,\theta(n-r+i+1) \\ 1+\theta(n-r+i+1)\ldots,1+\theta(n-r+i+1)\end{matrix};1-\gamma\right], \end{aligned}$$

where $w_i(r,n)$ are the weights as given in eq. (3.89). In particular, we have

$$\mathbb{E}[X_{1:n}^m] = \frac{m!\,\gamma^{\theta n}}{\beta^m(\theta n)^m}\,{}_{m+1}F_m\left[\begin{matrix}\theta n,\ldots,\theta n \\ 1+\theta n\ldots,1+\theta n\end{matrix};1-\gamma\right],$$

and

$$\mathbb{E}[X_{n:n}^m] = \sum_{i=0}^{n-1} \frac{(-1)^i n\binom{n-1}{i}}{i+1} \frac{m!\,\gamma^{\theta(i+1)}}{\beta^m(\theta(i+1))^m}\,{}_{m+1}F_m\left[\begin{matrix}\theta(i+1),\ldots,\theta(i+1) \\ 1+\theta(i+1)\ldots,1+\theta(i+1)\end{matrix};1-\gamma\right].$$

Proof. This result is obtained by combining Theorems 3.5.10 and 3.5.17. □

3.5.7 | Maximum and Minimum Domain of Attraction

As defined in Section 3.2.1, the lifetime of individuals is considered to be an unbounded random variable, i.e., $F(x) < 1$ for all $x < \infty$. If one is interested in the distribution of the maximum lifespan of n individuals, i.e., the distribution of the n-th order statistic, one obviously has,

$$F_{X_{n:n}}(x) = F_X^n(x) \overset{n\to\infty}{\longrightarrow} 0,$$

for all x, and thus

$$X_{n:n} \xrightarrow[n\to\infty]{\mathbb{P}} \infty.$$

The n-th order statistic $X_{n:n}$ has to be standardized to accomplish a non-degenerate limit behaviour.

Definition 3.5.20 (Minimum and Maximum Domain of Attraction, (Marshall and Olkin, 2007))**.** Let $X_1,X_2,\ldots,$ be a sequence of independent and identically distributed random variables with distribution function F, and for $n = 1,2,\ldots$ let

$$U_n = \min\{X_1,\ldots,X_n\} \qquad V_n = \max\{X_1,\ldots,X_n\}$$

be the minima and maxima of a sample of length n. If there exist sequences a_n and b_n such that the normalization $(U_n - b_n)/a_n$ converges to a non-degenerate distribution G, i.e.,

$$\frac{U_n - b_n}{a_n} \xrightarrow[n\to\infty]{\mathcal{D}} G,$$

then the distribution F is said to belong to the *minimum domain of attraction* of G. Respectively, if there exist sequences a_n and b_n, such that $(V_n - b_n)/a_n$ converges in distribution to a random variable with distribution H,

$$\frac{V_n - b_n}{a_n} \xrightarrow[n\to\infty]{\mathcal{D}} H,$$

then the sample distribution F is said to belong to the *maximum domain of attraction* of H.

A fundamental result, known as the Fisher-Tippett-Gnedenko theorem, states that there are only three types of limiting distributions, which are the Gumbel, Fréchet, and Weibull distributions, also known as *extreme value distributions*. Partial results were first found by Fréchet (1927) and Fisher and Tippett (1928) and later in full generality by Gnedenko (1943).

In the following results, we will prove that the extended exponential distribution and the Gompertz distribution belong to the Gumbel maximum domain of attraction and to the Weibull minimum domain of attraction.

Theorem 3.5.21 (Maximum Domain of Attraction of $\mathcal{E}(\gamma,\theta,\beta)$)**.** Let $X_1,\ldots,X_n$ be an i.i.d. sequence of the extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$ and $F_{\text{Gumbel}}(x \mid 0,1) = e^{-e^{-x}}$ the CDF of the Gumbel distribution $\text{Gum}(0,1)$. Let the normalizing constants be defined as

$$a_n = \frac{1}{\theta\beta} \quad \text{and} \quad b_n = \frac{\ln(\gamma)}{\beta} + \frac{\ln(n)}{\beta\theta},$$

then, we have

$$\frac{\max\{X_1,\ldots,X_n\} - b_n}{a_n} \xrightarrow[n\to\infty]{\mathcal{D}} \text{Gum}(0,1),$$

i.e.,

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{\max\{X_1,\ldots,X_n\} - b_n}{a_n} \leq x\right] = F_{\text{Gumbel}}(x \mid 0,1),$$

for all $x \geq 0$. As a consequence, the extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$ belongs to the Gumbel maximum domain of attraction.

Proof. The proof is obtained directly. For all $x \geq 0$ we have

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left[\frac{\max\{X_1,\ldots,X_n\} - b_n}{a_n} \leq x\right] &= \lim_{n\to\infty} \mathbb{P}\left(\frac{X_{n:n} - b_n}{a_n} < x\right) \\
&= \lim_{n\to\infty} F_{X_{n:n}}(a_n x + b_n) \\
&= \lim_{x\to\infty} F_{\mathcal{E}}^n(a_n x + b_n) \\
&= \lim_{x\to\infty} (1 - S_{\mathcal{E}}(a_n x + b_n))^n \\
&= \lim_{x\to\infty} \left(1 - \left(\frac{\gamma}{\gamma + \gamma(ne^x)^{\frac{1}{\theta}} - 1}\right)^{\theta}\right)^n \\
&= \lim_{x\to\infty} \left(1 - \frac{e^{-x}}{n}\right)^n \qquad (3.91) \\
&= e^{-e^{-x}} = F_{\text{Gumbel}}(x \mid 0,1).
\end{aligned}$$

Equation (3.91) is obtained by observing

$$\left(\frac{\gamma}{\gamma + \gamma(ne^x)^{\frac{1}{\theta}} - 1}\right)^{\theta} \sim \frac{e^{-x}}{n}.$$

□

Note that the above theorem states, that for large $n \gg 1$ the distribution of the n-th order statistic is approximately Gumbel distributed, i.e.,

$$X_{n:n} \overset{\text{asym.}}{\sim} \text{Gum}\left(\frac{\ln(\gamma)}{\beta} + \frac{\ln(n)}{\beta\theta}, \frac{1}{\beta\theta}\right) =: Y, \qquad (3.92)$$

with expectation

$$\mathbb{E}[Y] \simeq \frac{\tilde{\gamma}}{\beta\theta} + \frac{\ln(\gamma)}{\beta} + \frac{\ln(n)}{\beta\theta},$$

where $\tilde{\gamma} \approx 0.577$ is the Euler-Mascheroni constant. Note that the expectation is logarithmically increasing in the sample size n. However, the variance of the asymptotical distribution, given by

$$\mathbb{V}[Y] \simeq \frac{\pi^2}{6\beta^2\theta^2},$$

does not depend on the sample size.

Theorem 3.5.22 (Minimum Domain of Attraction of $\mathcal{E}(\gamma,\theta,\beta)$)**.** Let $X_1,\ldots,X_n$ be an i.i.d. sequence of the extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$ and $F_{\text{Weibull}}(x \mid 1,1) = 1 - e^{-x}$ the CDF of the Weibull distribution $\mathcal{W}(1,1)$. Let the normalizing constants be defined as

$$a_n = \frac{\gamma}{\beta\theta n} \quad \text{and} \quad b_n = 0, \tag{3.93}$$

then, we have

$$\frac{\min\{X_1,\ldots,X_n\} - b_n}{a_n} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{W}(1,1).$$

Therefore, the extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$ belongs to the Weibull minimum domain of attraction.

Proof. The convergence of the normalized sample minimum in distribution to the Weibull distribution $\mathcal{W}(1,1)$ holds, since for all $x \geq 0$ we have

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left[\frac{\min\{X_1,\ldots,X_n\} - b_n}{a_n} \leq x\right] &= \lim_{n\to\infty} \mathbb{P}\left(\frac{X_{1:n} - b_n}{a_n} < x\right) \\
&= \lim_{n\to\infty} F_{X_{1:n}}(a_n x + b_n) \\
&= \lim_{x\to\infty} 1 - (1 - F_{\mathcal{E}}(a_n x + b_n))^n \\
&= \lim_{x\to\infty} 1 - S_{\mathcal{E}}^n(a_n x + b_n) \\
&= \lim_{x\to\infty} 1 - \left(\frac{\gamma}{\gamma + e^{\frac{\gamma x}{\theta n}} - 1}\right)^{\theta n} \\
&= 1 - e^{-x} \qquad (3.94) \\
&= F_{\text{Weibull}}(x \mid 1,1).
\end{aligned}$$

Equation (3.94) holds, since

$$\left(\frac{\gamma}{\gamma + e^{\frac{\gamma x}{\theta n}} - 1}\right)^{\theta n} \sim e^{-x}.$$

□

Thus, for a large sample size n, we have approximately

$$X_{1:n} \overset{\text{asym.}}{\sim} \mathcal{W}\left(1,\frac{\gamma}{\beta\theta n}\right) =: Z,$$

with expectation and variance given by

$$\mathbb{E}[Z] \simeq \frac{\gamma}{\beta\theta n} \quad \text{and} \quad \mathbb{V}[Z] \simeq \left(\frac{\gamma}{\beta\theta n}\right)^2.$$

Next, we derive the normalizing sequences for maximum and minimum domain of attraction of the Gompertz distribution in order to compare their asymptotic behaviour with the extended exponential and the Kannisto distribution.

Theorem 3.5.23 (Maximum Domain of Attraction of $\mathcal{G}(\xi,\kappa)$)**.** Let $X_1,\ldots,X_n$ be an i.i.d. sequence of the Gompertz distribution $\mathcal{G}(\xi,\kappa)$ and $F_{\text{Gumbel}}(x \mid 0,1) = e^{-e^{-x}}$ the CDF of the Gumbel distribution $\text{Gum}(0,1)$. Suppose the normalizing constants are given as

$$a_n = \frac{1}{\kappa\ln(n)} \quad \text{and} \quad b_n = \frac{\xi}{\kappa\ln(n)} + \frac{1}{\kappa}\ln\left(\frac{1}{\xi}\ln(n)\right),$$

then, we have

$$\frac{\max\{X_1,\ldots,X_n\} - b_n}{a_n} \xrightarrow[n\to\infty]{\mathcal{D}} \text{Gum}(0,1)$$

Proof. The proof is obtained directly. For all $x \geq 0$ we have

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left[\frac{\max\{X_1,\ldots,X_n\} - b_n}{a_n} \leq x\right] &= \lim_{n\to\infty} \mathbb{P}\left(\frac{X_{n:n} - b_n}{a_n} < x\right) \\
&= \lim_{n\to\infty} F_{X_{n:n}}(a_n x + b_n) \\
&= \lim_{x\to\infty} F_{\mathcal{G}}^n(a_n x + b_n) \\
&= \lim_{x\to\infty} (1 - S_{\mathcal{G}}(a_n x + b_n))^n \\
&= \lim_{n\to\infty} \left(1 - e^{\xi} n^{-e^{\frac{\xi+x}{\ln(n)}}}\right)^n \\
&= \lim_{n\to\infty} \left(1 - \frac{e^{-x}}{n}\right)^n \qquad (3.95) \\
&= e^{-e^{-x}} = F_{\text{Gum}}(x \mid 0,1).
\end{aligned}$$

Equation (3.95) holds, since

$$n^{-e^{\frac{x+\xi}{\ln(n)}}} \sim \frac{e^{-(x+\xi)}}{n}.$$

□

Theorem 3.5.24 (Minimum Domain of Attraction of $\mathcal{G}(\xi,\kappa)$)**.** Let $X_1,\ldots,X_n$ be an i.i.d. sequence of the Gompertz distribution $\mathcal{G}(\xi,\kappa)$ and $F_{\text{Weibull}}(x \mid 1,1) = 1 - e^{-x}$ the CDF of the Weibull distribution $\mathcal{W}$. Suppose the normalizing constants are given as

$$a_n = \frac{1}{\kappa\xi n} \quad \text{and} \quad b_n = 0, \tag{3.96}$$

then, we have

$$\frac{\min\{X_1,\ldots,X_n\} - b_n}{a_n} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{W}(1,1).$$

Therefore, the Gompertz distribution $\mathcal{G}(\xi,\kappa)$ belongs to the Weibull minimum domain of attraction.

Proof. The convergence of the normalized sample minimum in distribution to the Weibull distribution $\mathcal{W}(1,1)$ holds, since:

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left[\frac{\min\{X_1,\ldots,X_n\} - b_n}{a_n} \le x\right] &= \lim_{n\to\infty} \mathbb{P}\left(\frac{X_{1:n} - b_n}{a_n} < x\right) \\
&= \lim_{n\to\infty} F_{X_{1:n}}(a_n x + b_n) \\
&= \lim_{x\to\infty} 1 - (1 - F_{\mathcal{G}}(a_n x + b_n))^n \\
&= \lim_{x\to\infty} 1 - S_{\mathcal{G}}^n(a_n x + b_n) \\
&= 1 - \left(e^{\xi - \xi e^{\frac{x}{n\xi}}}\right)^n \\
&= 1 - e^{-x} \qquad (3.97) \\
&= F_{\text{Weibull}}(x \mid 1,1).
\end{aligned}$$

Equation (3.97) holds, since

$$e^{\xi - \xi e^{\frac{x}{n\xi}}} \sim 1 - \frac{x}{n},$$

which follows from a Taylor expansion of the transformed variable $s := \frac{1}{n}$ at $s = 0$. □

The results of Theorems 3.5.22 and 3.5.24 reveal a close connection between the Gompertz and the extended exponential distribution at the lower tail. The results show not only that both lifetime distributions belong to the Weibull minimum domain of attraction, but they also share the same growth behaviour of the normalizing sequences, see eqs. (3.93) and (3.96). For instance, with

$$\kappa = \frac{\beta(\gamma-1)}{\gamma} \quad \text{and} \quad \xi = \frac{\theta}{\gamma - 1},$$

both hazard rate functions $h_{\mathcal{G}}(x \mid \xi,\kappa)$ and $h_{\mathcal{E}}(x \mid \gamma,\theta,\beta)$ coincide at $x = 0$, which leads to the same normalizing sequences as given in eq. (3.93). In contrast to that, Theorems 3.5.21 and 3.5.23 reveal fundamental differences between both distributions at the right tail. They are both in the Gumbel maximum domain of attraction, but they show different growth behaviour of the normalizing sequences. In particular, we have the shifting sequence

$$b_n^{\mathcal{E}} = \frac{\ln(\gamma)}{\beta} + \frac{\ln(n)}{\beta\theta} \sim \frac{\ln(n)}{\beta\theta} \in \mathcal{O}(\ln n)$$

for the extended exponential distribution and

$$b_n^{\mathcal{G}} = \frac{\xi}{\kappa \ln(n)} + \frac{1}{\kappa}\ln\left(\frac{1}{\xi}\ln(n)\right) \sim \frac{1}{\kappa}\ln(\ln(n)) \in \mathcal{O}(\ln\ln n),$$

for the Gompertz distribution, respectively. Note that $b_n^{\mathcal{G}} \in o(b_n^{\mathcal{E}})$, i.e., $b_n^{\mathcal{G}}$ is asymptotically negligible compared to $b_n^{\mathcal{E}}$.

For Kannisto distributed lifetimes, i.e., $X_i \sim \mathcal{K}(\alpha,\beta)$ for $i = 1,\dots,n$, we conclude from eq. (3.92) that the expected maximal lifespan can be approximated by

$$\mathbb{E}[X_{n:n}] \overset{\text{asym.}}{\simeq} \tilde{\gamma} + \frac{1}{\beta}\ln(1+e^{-\alpha}) + \ln(n) \overset{\alpha \ll 0}{\approx} \tilde{\gamma} - \frac{\alpha}{\beta} + \ln(n). \tag{3.98}$$

Evaluating the last term of eq. (3.98), at the parameter estimates $(\alpha_{2011},\beta_{2011}) = (-5.894, 0.138)$ for the reference population at 2011 and using the population size $n := E^c_{2011,60} = 6 \times 10^4$, yields an expected maximal lifespan of 105.69 years, which appears to be a plausible approximation.

3.5.8 | Fisher Information Matrix for the Kannisto Distribution and the Extended Exponential Distribution

The Fisher information matrix plays an important role in statistics such as for the derivation of the asymptotic covariance matrix of maximum likelihood estimators. To determine the exact form of the Fisher information matrix for the Kannisto and the extended exponential distribution, we will need some integral representations of particular generalized hypergeometric functions, which are provided in the following Lemmas 3.5.25 and 3.5.26.

Lemma 3.5.25. Let $k > 1$, $\theta > 0$ and $n \in \mathbb{N}_0$. Then, we obtain the following integral representation.

$${}_{n+1}F_n\left[\begin{matrix} k-1+\theta,\dots,k-1+\theta \\ k+\theta,\dots k+\theta \end{matrix} ;(1-\gamma)\right] = \frac{(k-1+\theta)^{n+1}}{n!}\int_{\gamma}^{\infty} x^{-(k+\theta)}\ln^n(1-\gamma+x)\,dx.$$

Proof. We start with the series representation of the generalized geometric function ${}_{n+1}F_n$ where all upper and all the lower coefficients coincide. Further transformations are described below.

$${}_{n+1}F_n\left[\begin{matrix} k-1+\theta,\dots,k-1+\theta \\ k+\theta,\dots k+\theta \end{matrix} ;(1-\gamma)\right]$$

$$
\begin{aligned}
&= \sum_{m=0}^{\infty} \frac{(k-1+\theta)^n}{(k-1+\theta+m)^n} (k-1+\theta)_m \frac{(1-\gamma)^m}{m!} \\
&= \frac{(k-1+\theta)^n}{n!} \sum_{m=0}^{\infty} (k-1+\theta)_m \int_0^{\infty} z^n (k-1+\theta+m) e^{-(k-1+\theta+m)z} dz \qquad (3.99)\\
&= \frac{(k-1+\theta)^n}{n!} \int_0^{\infty} z^n e^{-z(k-1+\theta)} \sum_{m=1}^{\infty} (k-1+\theta)_m (k-1+\theta+m) \frac{((1-\gamma)e^{-z})^m}{m!} dz \\
&= \frac{(k-1+\theta)^n}{n!} \int_0^{\infty} z^n e^{-z(k-1+\theta)} (k-1+\theta)\, {}_1F_0[k+\theta;;(1-\gamma)e^{-z}]\, dz \qquad (3.100)\\
&= \frac{(k-1+\theta)^{n+1}}{n!} \int_0^{\infty} z^n e^{-z(\theta+k-1)} \left(1-(1-\gamma)e^{-z}\right)^{-(k+\theta)} dz \\
&= \frac{(k-1+\theta)^{n+1}}{n!} \int_{\gamma}^{\infty} x^{-(k+\theta)} \ln^n(1-\gamma+x)\, dx \qquad (3.101)
\end{aligned}
$$

For eq. (3.99), we write $(k-1+\theta+m)^{-n}$ as an integral using a Laplace transform, i.e.,

$$
(k-1+\theta+m)^{-n} = \frac{k-1+\theta+m}{\Gamma(n+1)} \int_0^{\infty} z^n e^{-(k-1+\theta+m)z}\, dz.
$$

To obtain eq. (3.100), we use the identity

$$
\sum_{m=0}^{\infty} (a)_m (a+m) \frac{z^m}{m!} = a\, {}_1F_0[a+1;;z] = a(1-z)^{-(a+1)},
$$

since

$$
(a+m)(a)_m = (a+m)\frac{\Gamma(a+m)}{\Gamma(a)} = \frac{\Gamma(a+1+m)}{\Gamma(a)} = \frac{a\Gamma(a+1+m)}{\Gamma(a+1)} = a(a+1)_m.
$$

The last transformation in eq. (3.101) is obtained by the substitution $z = \ln(1-\gamma+x)$. □

Lemma 3.5.26. Let $\theta > 0$, $k > 2$, and $n \in \mathbb{N}$. Then, the following integral representation holds:

$$
\begin{aligned}
&{}_{n+1}F_n\left[\begin{matrix}\theta+k-2,\ldots,\theta+k-2,\theta+k \\ \theta+k-1,\ldots,\theta+k-1\end{matrix};1-\gamma\right] \\
&\quad = \frac{(\theta+k-2)^n}{(n-1)!} \int_{\gamma}^{\infty} x^{-(\theta+k)} (1-\gamma+x) \ln^{n-1}(1-\gamma+x)\, dx.
\end{aligned}
$$

Proof. The proof is obtained by the same methods as in Lemma 3.5.25. We start with the series representation of the generalized geometric function ${}_{n+1}F_n$. Further transformations are described below.

$$
\begin{aligned}
{}_{n+1}F_n&\left[\begin{matrix}\theta+k-2,\dots,\theta+k-2,\theta+k\\ \theta+k-1,\dots,\theta+k-1\end{matrix};1-\gamma\right]\\
&=\sum_{m=0}^{\infty}\frac{(\theta+k-2)_m^n(\theta+k)_m}{(\theta+k-1)_m^n}\frac{(1-\gamma)^m}{m!}\\
&=\sum_{m=0}^{\infty}\frac{(\theta+k-2)^n\Gamma(k+m+\theta)}{(\theta+k-2+m)^n\Gamma(k+\theta)}\frac{(1-\gamma)^m}{m!}\\
&=(\theta+k-2)^n\sum_{m=0}^{\infty}\frac{(\theta+k)_m}{(\theta+k-2+m)^n}\frac{(1-\gamma)^m}{m!} && (3.102)\\
&=\frac{(\theta+k-2)^n}{\Gamma(n)}\sum_{m=0}^{\infty}(\theta+k)_m\frac{(1-\gamma)^m}{m!}\int_0^{\infty}z^{n-1}e^{-(\theta+k-2+m)z}\,dz && (3.103)\\
&=\frac{(\theta+k-2)^n}{\Gamma(n)}\int_0^{\infty}z^{n-1}e^{-(\theta+k-2)}\,{}_1F_0(\theta+k;;(1-\gamma)e^{-z})\,dz\\
&=\frac{(\theta+k-2)^n}{\Gamma(n)}\int_0^{\infty}z^{n-1}e^{-z(\theta+k-2)}\left(1-(1-\gamma)e^{-z}\right)^{-(\theta+k)}\,dz\\
&=\frac{(\theta+k-2)^n}{(n-1)!}\int_{\gamma}^{\infty}x^{-(\theta+k)}(1-\gamma+x)\ln^{n-1}(1-\gamma+x)\,dx. && (3.104)
\end{aligned}
$$

To obtain eq. (3.102), we rewrite the term $(\theta+k-2+m)^{-n}$ as an integral using a Laplace transform, i.e.,

$$(\theta+k-2+m)^{-n}=\frac{1}{\Gamma(n)}\int_0^{\infty}z^{n-1}e^{-(\theta+k-2+m)z}\,dz.$$

Equation (3.103) follows from Fubini's theorem by changing the order of integration and by replacing the series with the hypergeometric function ${}_1F_0$, i.e.,

$${}_1F_0[\theta+k;;(1-\gamma)e^{-z}]=\sum_{m=0}^{\infty}(\theta+k)_m\frac{((1-\gamma)e^{-z})^m}{m!}.$$

Finally, substitution $z=\ln(1-\gamma+x)$ gives eq. (3.104). □

In the following part, we determine the Fisher information matrix for the Kannisto distribution. For a random variable X with probability density $f(\cdot\mid\boldsymbol{\theta})$, where $\boldsymbol{\theta}=(\theta_1\dots,\theta_n)$, the Fisher information matrix $\mathbb{I}(\boldsymbol{\theta})$ is an $n\times n$ symmetric matrix with elements given by

$$\mathbb{I}_{i,j}(\boldsymbol{\theta})=\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial\ln f(X\mid\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial\ln f(X\mid\boldsymbol{\theta})}{\partial\theta_j}\right].$$

If the density has continuous second partial derivatives $\frac{\partial^2 f(\cdot|\theta_1,\ldots,\theta_n)}{\partial_{\theta_i}\partial_{\theta_j}}$ for all i and j, then $\mathbb{I}_{i,j}(\boldsymbol{\theta})$ can be expressed by

$$\mathbb{I}_{i,j}(\boldsymbol{\theta}) = -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f(X \mid \Theta)}{\partial\theta_i\partial\theta_j}\right]. \tag{3.105}$$

For the Kannisto distribution $\mathcal{K}(\alpha,\beta)$ with density

$$f_{\mathcal{K}}(x \mid \alpha,\beta) = (1+e^{\alpha})^{\frac{1}{\beta}} e^{\alpha+\beta x}\left(1+e^{\alpha+\beta x}\right)^{-\frac{1+\beta}{\beta}} \mathbb{1}_{x\geq 0}$$

all second partial derivatives exist and are continuous, such that we will use the expression given in eq. (3.105) for the computation of the matrix coefficients.

For the following part, we will change the Kannisto parametrization by defining $\gamma := 1+e^{-\alpha}$ and $\theta := {}^{1}/_{\beta}$, such that by that choice, we have

$$\mathcal{K}(\alpha,\beta) = \mathcal{K}(\ln({}^{1}/_{\gamma-1}), {}^{1}/_{\theta}),$$

where $\gamma > 1$ and $\theta > 0$. With that parametrization, the logarithmic density is given by

$$\ln(f(x \mid \gamma,\theta)) = \theta\ln(\gamma) - (\theta+1)\ln\left(\gamma + e^{x/\theta} - 1\right) + \frac{x}{\theta}.$$

The second partial derivatives required for the Fisher information matrix are:

$$\begin{aligned}
\frac{\partial^2 \ln f_{\mathcal{K}}(x \mid \gamma,\theta)}{\partial\gamma\partial\gamma} &= \frac{\theta+1}{\left(\gamma+e^{x/\theta}-1\right)^2} - \frac{\theta}{\gamma^2},\\
\frac{\partial^2 \ln f_{\mathcal{K}}(x \mid \gamma,\theta)}{\partial\gamma\partial\theta} &= \frac{1}{\gamma} - \frac{(\theta+1)xe^{x/\theta}}{\theta^2\left(\gamma+e^{x/\theta}-1\right)^2} - \frac{1}{\gamma+e^{x/\theta}-1},\\
\frac{\partial^2 \ln f_{\mathcal{K}}(x \mid \gamma,\theta)}{\partial\theta\partial\theta} &= \frac{x(\gamma-1)\left(-2\theta+2e^{x/\theta}\theta+2\gamma\theta\right)}{\left(-1+e^{x/\theta}+\gamma\right)^2\theta^4} + \frac{x^2(\gamma-1)\left(-e^{x/\theta}-e^{x/\theta}\theta\right)}{\left(-1+e^{x/\theta}+\gamma\right)^2\theta^4}.
\end{aligned}$$

Theorem 3.5.27 (Fisher Information Matrix of the Kannisto Distribution). The Fisher information matrix of the Kannisto distribution $\mathcal{K}(\ln({}^{1}/_{\gamma-1}), {}^{1}/_{\theta})$ is given by

$$\mathbb{I}(\gamma,\theta) = \begin{pmatrix} \mathbb{I}_{\gamma,\gamma} & \mathbb{I}_{\gamma,\theta} \\ \mathbb{I}_{\theta,\gamma} & \mathbb{I}_{\theta,\theta} \end{pmatrix},$$

with the following coefficients:

$$\mathbb{I}_{\gamma,\gamma}(\gamma,\theta) = \frac{\theta}{\gamma^2(\theta+2)} \tag{3.106}$$

$$\mathbb{I}_{\gamma,\theta}(\gamma,\theta) = \mathbb{I}_{\theta,\gamma}(\gamma,\theta) = -\frac{1}{\gamma(\theta+1)} + \frac{\gamma^{\theta+1}}{\gamma(\theta+1)}\, {}_3F_2[\theta+1,\theta+1,\theta+3;\theta+2,\theta+2;1-\gamma] \tag{3.107}$$

$$\mathbb{I}_{\theta,\theta}(\gamma,\theta) = \frac{2(\gamma-1)^2\gamma^\theta}{\theta(\theta+2)^3}\,{}_3F_2[\theta+2,\theta+2,\theta+2;\theta+3,\theta+3;1-\gamma]$$

Proof. We begin by proving eq. (3.106). The calculation of this term only requires standard integration techniques. We have,

$$\begin{aligned}
\mathbb{I}_{\gamma,\gamma}(\gamma,\theta) &= -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f_{\mathcal{K}}(x\mid\gamma,\theta)}{\partial\gamma\partial\gamma}\right] \\
&= \frac{\theta}{\gamma^2} - \int_0^\infty \frac{\theta+1}{\left(\gamma+e^{\beta x}-1\right)^2} f_{\mathcal{K}}(x\mid\gamma,\theta)\,dx \\
&= \frac{\theta}{\gamma^2} - \theta(\theta+1)\gamma^\theta \int_\gamma^\infty x^{-(\theta+3)}\,dx \\
&= \frac{\theta}{\gamma^2} - \frac{\theta(\theta+1)}{\gamma^2(\theta+2)} \\
&= \frac{\theta}{\gamma^2(\theta+2)}.
\end{aligned}$$

To show eq. (3.107), we express $\mathbb{I}_{\gamma,\theta}(\gamma,\theta)$ in terms of standard integrals and an integral of the type of Lemma 3.5.26. We obtain

$$\begin{aligned}
\mathbb{I}_{\gamma,\theta}(\gamma,\theta) &= -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f_{\mathcal{K}}(x\mid\gamma,\theta)}{\partial\gamma\partial\theta}\right] \\
&= -\frac{1}{\gamma} + \int_0^\infty \left(\frac{1}{-1+e^{x/\theta}+\gamma} + \frac{e^{x/\theta}x(1+\theta)}{\left(-1+e^{x/\theta}+\gamma\right)^2\theta^2}\right) f_{\mathcal{K}}(x\mid\gamma,\theta)\,dx \\
&= -\frac{1}{\gamma} + \theta\gamma^\theta \int_\gamma^\infty x^{-(\theta+2)}\,dx + (\theta+1)\gamma^\theta \int_\gamma^\infty x^{-(\theta+3)}(1-\gamma+x)\ln(1-\gamma+x)\,dx \qquad (3.108) \\
&= -\frac{1}{\gamma} + \frac{\theta}{\gamma(\theta+1)} + \frac{\gamma^\theta}{\theta+1}\,{}_3F_2[\theta+1,\theta+1,\theta+3;\theta+2,\theta+2;1-\gamma] \\
&= -\frac{1}{\gamma(\theta+1)} + \frac{\gamma^{\theta+1}}{\gamma(\theta+1)}\,{}_3F_2[\theta+1,\theta+1,\theta+3;\theta+2,\theta+2;1-\gamma],
\end{aligned}$$

where Lemma 3.5.26 is applied to eq. (3.108). The remaining element $\mathbb{I}_{\gamma,\theta}(\gamma,\theta)$ is obtained by a similar procedure.

$$\mathbb{I}_{\gamma,\theta}(\gamma,\theta) = -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f_{\mathcal{K}}(x\mid\gamma,\theta)}{\partial\theta\partial\theta}\right]$$

$$
\begin{aligned}
&= -\int_0^\infty \frac{x(\gamma-1)\left(-2\theta+2e^{x/\theta}\theta+2\gamma\theta\right)}{\left(-1+e^{x/\theta}+\gamma\right)^2\theta^4} f_{\mathcal{K}}(x\mid\gamma,\theta)\,dx \\
&\quad -\int_0^\infty \frac{x^2(\gamma-1)\left(-e^{x/\theta}-e^{x/\theta}\theta\right)}{\left(-1+e^{x/\theta}+\gamma\right)^2\theta^4} f_{\mathcal{K}}(x\mid\gamma,\theta)\,dx \\
&= -\frac{2(\gamma-1)\gamma^\theta}{\theta}\int_\gamma^\infty x^{-(\theta+2)}\ln(1-\gamma+z)\,dx \\
&\quad +\frac{(\gamma-1)\gamma^\theta(1+\theta)}{\theta}\int_\gamma^\infty x^{-(\theta+3)}(1+x-\gamma)\ln^2(1+x-\gamma)\,dx \qquad (3.109)\\
&= -\frac{2(\gamma-1)\gamma^\theta}{\theta(\theta+1)^2}\,{}_2F_1[\theta+1,\theta+1;\theta+2;1-\gamma] \\
&\quad +\frac{2(\gamma-1)\gamma^\theta}{\theta(\theta+1)^2}\,{}_4F_3[\theta+1,\theta+1,\theta+1,\theta+3;\theta+2,\theta+2,\theta+2;1-\gamma] \\
&= -\frac{2(\gamma-1)\gamma^\theta}{\theta(\theta+1)^2}\,{}_2F_1[\theta+1,\theta+1;\theta+2;1-\gamma] \\
&\quad +\frac{2(\gamma-1)\gamma^\theta}{\theta(\theta+1)^2}\,{}_4F_3[\theta+1,\theta+1,\theta+1,\theta+3;\theta+2,\theta+2,\theta+2;1-\gamma] \\
&= \frac{2(\gamma-1)\gamma^\theta}{\theta(\theta+1)^2}\left({}_4F_3\left[\begin{matrix}\theta+1,\theta+1,\theta+1,\theta+3\\ \theta+2,\theta+2,\theta+2\end{matrix};1-\gamma\right]-{}_2F_1\left[\begin{matrix}\theta+1,\theta+1\\ \theta+2\end{matrix};1-\gamma\right]\right) \\
&= \frac{2(\gamma-1)^2\gamma^\theta}{\theta(\theta+2)^3}\,{}_3F_2[\theta+2,\theta+2,\theta+2;\theta+3,\theta+3;1-\gamma],
\end{aligned}
$$

where for eq. (3.109) we use Lemmas 3.5.25 and 3.5.26. The last identity follows from rearranging the series coefficients of ${}_4F_3+{}_2F_1$ to a generalized hypergeometric ${}_3F_2$. More specifically, we use the identity

$$
\begin{aligned}
&{}_4F_3\left[\begin{matrix}\theta+1,\theta+1,\theta+1,\theta+3\\ \theta+2,\theta+2,\theta+2\end{matrix};1-\gamma\right]-{}_2F_1\left[\begin{matrix}\theta+1,\theta+1\\ \theta+2\end{matrix};1-\gamma\right] \\
&\quad= \sum_{m=0}^\infty \frac{-m(1+\theta)^2\Gamma(1+m+\theta)}{(1+m+\theta)^2(2+\theta)(1+\theta)\Gamma(1+\theta)}\frac{(1-\gamma)^m}{m!} \\
&\quad= \frac{\gamma-1}{(\theta+2)(\theta+1)}\sum_{m=0}^\infty \frac{(1+\theta)^2\Gamma(1+m+\theta)}{(1+m+\theta)^2\Gamma(1+\theta)}\frac{m(1-\gamma)^{m-1}}{m!} \\
&\quad= \frac{\gamma-1}{(\theta+2)(\theta+1)}\frac{d}{d(1-\gamma)}\,{}_3F_2\left[\begin{matrix}\theta+1,\theta+1,\theta+1\\ \theta+2,\theta+2\end{matrix};1-\gamma\right] \\
&\quad= \frac{\gamma-1}{(\theta+2)(\theta+1)}\frac{(\theta+1)^3}{(\theta+2)^2}\,{}_3F_2\left[\begin{matrix}\theta+2,\theta+2,\theta+2\\ \theta+3,\theta+3\end{matrix};1-\gamma\right] \qquad (3.110)
\end{aligned}
$$

$$= \frac{(\gamma-1)(\theta+1)^2}{(\theta+2)^3} {}_3F_2\left[\begin{matrix}\theta+2,\theta+2,\theta+2\\ \theta+3,\theta+3\end{matrix};1-\gamma\right].$$

Equation (3.110) is obtained by the differentiation rule

$$\frac{d^n}{dz^n}\,{}_pF_q\left[\begin{matrix}a_1,\ldots,a_p\\ b_1,\ldots,b_q\end{matrix};z\right] = \frac{(a_1)_n\cdots(a_p)_n}{(b_1)_n\cdots(b_q)_n}\,{}_pF_q\left[\begin{matrix}a_1+n,\ldots,a_p+n\\ b_1+n,\ldots,b_q+n\end{matrix};z\right],$$

see, e.g., Olver (2010). □

Retransformation to the original parametrization of the Kannisto distribution yields the following matrix coefficients:

$$\mathbb{I}_{\alpha,\alpha}(\alpha,\beta) = \frac{1}{(e^{-\alpha}+1)^2(2\beta+1)},$$

$$\mathbb{I}_{\alpha,\beta}(\alpha,\beta) = \mathbb{I}_{\beta,\alpha}(\alpha,\beta)$$

$$= -\frac{e^{\alpha}\beta}{e^{\alpha}\beta+e^{\alpha}+\beta+1} + \frac{\beta^3(e^{-\alpha}+1)^{\frac{1}{\beta}}}{(\beta+1)^3}\,{}_3F_2\left[\frac{1}{\beta}+1,\frac{1}{\beta}+1,\frac{1}{\beta}+3;\frac{1}{\beta}+2,\frac{1}{\beta}+2;-e^{-\alpha}\right],$$

$$\mathbb{I}_{\beta,\beta}(\alpha,\beta) = \frac{2e^{-2\alpha}\beta(e^{-\alpha}+1)^{\frac{1}{\beta}}}{\left(\frac{1}{\beta}+2\right)^3}\,{}_3F_2\left[\frac{1}{\beta}+2,\frac{1}{\beta}+2,\frac{1}{\beta}+2;\frac{1}{\beta}+3,\frac{1}{\beta}+3;-e^{-\alpha}\right].$$

Theorem 3.5.28 (Fisher Information Matrix of the Extended Exponential Distribution)**.** The Fisher information matrix of the extended exponential distribution $\mathcal{E}(\gamma,\theta,\beta)$

$$\mathbb{I}(\gamma,\theta,\beta) = \mathbb{I}_{i,j}(\Theta) = -\mathbb{E}_{\theta}\left[\frac{\partial^2 \ln f(X\mid\Theta)}{\partial\theta_i\partial\theta_j}\right] = \begin{pmatrix}\mathbb{I}_{\gamma,\gamma} & \mathbb{I}_{\gamma,\theta} & \mathbb{I}_{\gamma,\beta}\\ \mathbb{I}_{\theta,\gamma} & \mathbb{I}_{\theta,\theta} & \mathbb{I}_{\theta,\beta}\\ \mathbb{I}_{\beta,\gamma} & \mathbb{I}_{\beta,\theta} & \mathbb{I}_{\beta,\beta}\end{pmatrix}$$

has the following coefficients:

$$\mathbb{I}_{\gamma,\gamma} = \frac{\theta}{\gamma^2(\theta+2)} \tag{3.111}$$

$$\mathbb{I}_{\gamma,\theta} = \mathbb{I}_{\theta,\gamma} = -\frac{1}{\gamma(\theta+1)} \tag{3.112}$$

$$\mathbb{I}_{\gamma,\beta} = \mathbb{I}_{\beta,\gamma} = -\frac{\theta\gamma^{\theta}}{\beta(\theta+1)}\,{}_3F_2[\theta+1,\theta+1,\theta+3;\theta+2,\theta+2;1-\gamma] \tag{3.113}$$

$$\mathbb{I}_{\theta,\theta} = \frac{1}{\theta^2} \tag{3.114}$$

$$\mathbb{I}_{\theta,\beta} = \mathbb{I}_{\beta,\theta} = \frac{\gamma^{\theta}}{\beta\theta}\,{}_3F_2[\theta,\theta,\theta+2;\theta+1,\theta+1;1-\gamma] \tag{3.115}$$

$$\mathbb{I}_{\beta,\beta} = \frac{1}{\beta^2} + \frac{2(\gamma-1)\theta\gamma^{\theta}}{\beta^2(\theta+1)^2}\,{}_4F_3[\theta+1,\theta+1,\theta+1,\theta+3;\theta+2,\theta+2,\theta+2;1-\gamma] \tag{3.116}$$

Proof. The density of the extended exponential distribution is given by

$$f_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = \beta\theta\gamma^{\theta} e^{\beta x} \left(\gamma + e^{\beta x} - 1\right)^{-(\theta+1)} \quad \text{for } x \geq 0.$$

Thus, for the logarithmic density, we have

$$\ln(f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)) = \ln\left(\beta\theta\gamma^{\theta}\right) - (\theta+1)\ln\left(e^{\beta x} - (1-\gamma)\right) + \beta x.$$

The Fisher information matrix can be expressed as

$$\mathbb{I}_{i,j}(\Theta) = -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f(X \mid \Theta)}{\partial\theta_i \partial\theta_j}\right],$$

since the density $f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)$ has continuous second partial derivatives regards to all parameters. The second partial derivatives of the logarithmic density required for the Fisher information matrix are:

$$\begin{aligned}
\frac{\partial^2 \ln f(x \mid \gamma,\theta,\beta)}{\partial\gamma\partial\gamma} &= \frac{\theta+1}{\left(\gamma + e^{\beta x} - 1\right)^2} - \frac{\theta}{\gamma^2}, \\
\frac{\partial^2 \ln f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{\partial\gamma\partial\theta} &= \frac{1}{\gamma} - \frac{1}{\gamma + e^{\beta x} - 1}, \\
\frac{\partial^2 \ln f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{\partial\gamma\partial\beta} &= \frac{(\theta+1)xe^{\beta x}}{\left(\gamma + e^{\beta x} - 1\right)^2}, \\
\frac{\partial^2 \ln f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{\partial\theta\partial\theta} &= -\frac{1}{\theta^2}, \\
\frac{\partial^2 \ln f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{\partial\theta\partial\beta} &= -\frac{xe^{\beta x}}{\gamma + e^{\beta x} - 1}, \\
\frac{\partial^2 \ln f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{\partial\beta\partial\beta} &= -\frac{1}{\beta^2} - \frac{(\gamma-1)(\theta+1)x^2 e^{\beta x}}{\left(\gamma + e^{\beta x} - 1\right)^2}.
\end{aligned}$$

In the following, we proof by direct calculation the matrix coefficients as provided in eqs. (3.111) to (3.116). For standard integrals, such as for the coefficients $\mathbb{I}_{\gamma,\theta}$ and $\mathbb{I}_{\gamma,\gamma}$ we use

$$\int_{\gamma}^{\infty} x^{-(\theta+k)}\, dx = \frac{\gamma^{-(\theta+k-1)}}{\theta+k-1}, \quad \text{for } \gamma > 0, \theta + k > 1, \tag{3.117}$$

and for integrals involving generalized hypergeometric functions, we use Lemma 3.5.26 such as for $\mathbb{I}_{\gamma,\beta}$, $\mathbb{I}_{\theta,\beta}$ and $\mathbb{I}_{\beta,\beta}$. Equation (3.111) is derived by using eq. (3.117).

$$\mathbb{I}_{\gamma,\gamma}(\gamma,\theta,\beta) = -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{\partial\gamma\partial\gamma}\right]$$

$$
\begin{aligned}
&= \frac{\theta}{\gamma^2} - \int_0^\infty \frac{\theta+1}{\left(\gamma + e^{\beta x} - 1\right)^2} f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)\, dx \\
&= \frac{\theta}{\gamma^2} - \theta(\theta+1)\gamma^\theta \int_\gamma^\infty x^{-(\theta+3)}\, dx \\
&= \frac{\theta}{\gamma^2} - \frac{\theta(\theta+1)}{\gamma^2(\theta+2)} \\
&= \frac{\theta}{\gamma^2(\theta+2)}.
\end{aligned}
$$

Equation (3.112) can also be shown by using eq. (3.117).

$$
\begin{aligned}
\mathbb{I}_{\gamma,\theta}(\gamma,\theta,\beta) &= -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{\partial\gamma\partial\theta}\right] \\
&= -\frac{1}{\gamma} + \int_0^\infty \frac{1}{\gamma + e^{\beta x} - 1} f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)\, dx \\
&= -\frac{1}{\gamma} + \gamma^\theta\theta \int_\gamma^\infty x^{-(\theta+2)}\, dx \\
&= -\frac{1}{\gamma} + \frac{\theta}{\gamma(\theta+1)} \\
&= -\frac{1}{\gamma(\theta+1)}.
\end{aligned}
$$

Equation (3.113) is obtained by using Lemma 3.5.26 for eq. (3.118).

$$
\begin{aligned}
\mathbb{I}_{\gamma,\beta}(\gamma,\theta,\beta) &= -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{\partial\gamma\partial\beta}\right] \\
&= -(\theta+1)\int_0^\infty \frac{x e^{\beta x}}{\left(\gamma + e^{\beta x} - 1\right)^2} f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)\, dx \\
&= -(\theta+1)\beta\theta\gamma^\theta \int_0^\infty x e^{2\beta x}\left(\gamma + e^{\beta x} - 1\right)^{-\theta-3} dx \\
&= -\frac{(\theta+1)\theta\gamma^\theta}{\beta}\int_\gamma^\infty x^{-(\theta+3)}(1-\gamma+x)\ln(1-\gamma+x)\, dx \qquad (3.118) \\
&= -\frac{\theta\gamma^\theta}{\beta(\theta+1)}\, {}_3F_2[\theta+1,\theta+1,\theta+3;\theta+2,\theta+2;1-\gamma].
\end{aligned}
$$

The computation of eq. (3.114) is trivial since the corresponding second derivative is constant.

$$\begin{aligned}\mathbb{I}_{\theta,\theta}(\gamma,\theta,\beta) &= -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{\partial\theta\partial\theta}\right] \\ &= \frac{1}{\theta^2}\int_0^\infty f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)\,dx \\ &= \frac{1}{\theta^2}.\end{aligned}$$

To show eq. (3.115), we apply Lemma 3.5.26 to eq. (3.119)

$$\begin{aligned}\mathbb{I}_{\theta,\beta}(\gamma,\theta,\beta) &= -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{\partial\theta\partial\beta}\right] \\ &= \int_0^\infty \frac{xe^{\beta x}}{\gamma+e^{\beta x}-1} f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)\,dx \\ &= \beta\theta\gamma^\theta \int_0^\infty xe^{2\beta x}\left(\gamma+e^{\beta x}-1\right)^{-\theta-2}dx \\ &= \frac{\theta\gamma^\theta}{\beta}\int_\gamma^\infty x^{-(\theta+2)}(1-\gamma+x)\ln(1-\gamma+x)\,dx \qquad (3.119) \\ &= \frac{\gamma^\theta}{\beta\theta}\,{}_3F_2[\theta,\theta,\theta+2;\theta+1,\theta+1;1-\gamma].\end{aligned}$$

The remaining coefficient of eq. (3.116) is shown by using Lemma 3.5.26 for eq. (3.120)

$$\begin{aligned}\mathbb{I}_{\beta,\beta}(\gamma,\theta,\beta) &= -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{\partial\beta\partial\beta}\right] \\ &= \frac{1}{\beta^2}+(\gamma-1)(\theta+1)\int_0^\infty \frac{x^2e^{\beta x}}{\left(\gamma+e^{\beta x}-1\right)^2} f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)\,dx \\ &= \frac{1}{\beta^2}+(\gamma-1)(\theta+1)\beta\theta\gamma^\theta\int_0^\infty x^2e^{2\beta x}\left(\gamma+e^{\beta x}-1\right)^{-\theta-3}dx \\ &= \frac{1}{\beta^2}+\frac{(\gamma-1)(\theta+1)\theta\gamma^\theta}{\beta^2}\int_\gamma^\theta x^{-(\theta+3)}(1-\gamma+x)\ln^2(1-\gamma+x)\,dx \qquad (3.120) \\ &= \frac{1}{\beta^2}+\frac{2(\gamma-1)\theta\gamma^\theta}{\beta^2(\theta+1)^2}\,{}_4F_3[\theta+1,\theta+1,\theta+1,\theta+3;\theta+2,\theta+2,\theta+2;1-\gamma].\end{aligned}$$

□

3.5.9 | Kullback-Leibler Divergence

For continuous distributions, the Kullback-Leibler divergence $D_{\mathrm{KL}}(P \parallel Q)$ is defined as

$$D_{\mathrm{KL}}(P \parallel Q) = \int_{-\infty}^{\infty} f_P(x) \ln \frac{f_P(x)}{f_Q(x)} dx,$$

where $f_P(x)$ and $f_Q(x)$ denote the probability density functions of P and Q, respectively. The Kullback-Leibler divergence can be seen as a non-symmetrical distance from Q to P, or as a measure of the information loss when the measure Q is used to approximate P, see Burnham and Anderson (2002). The Kullback-Leibler divergence is non-negative and zero if and only if $P = Q$ almost everywhere.

In the following proposition, we derive the Kullback-Leibler divergence between the Kannisto distribution $\mathcal{K}(\alpha,\beta)$ and Gompertz distribution $\mathcal{G}(\beta,{}^{e^\alpha}/_\beta)$. The particular choice of the Gompertz parameters is made to obtain similar behaviour at the lower tail of the distribution, see the illustration Figure 3.5.

Proposition 3.5.29 (Kullback-Leibler Divergence between the Gompertz and the Kannisto Distribution)**.** The Kullback-Leibler divergence between the Gompertz distribution $\mathcal{G}(\beta,{}^{e^\alpha}/_\beta)$ and the Kannisto distribution $\mathcal{K}(\alpha,\beta)$ is given by

$$D_{\mathrm{KL}}\big(\mathcal{K}(\alpha,\beta) \parallel \mathcal{G}(\beta,{}^{e^\alpha}/_\beta)\big) = \frac{e^\alpha + \beta^2 + (1-\beta)\ln\left(e^\alpha + 1\right)}{\beta - 1}, \quad \text{for } 0 < \beta < 1. \tag{3.121}$$

Proof. The Kullback-Leibler divergence of two probability distributions P and Q with Lebesgue densities $f_P(x)$ respectively $f_Q(x)$ is defined as

$$D_{\mathrm{KL}}(P \parallel Q) = \int_{-\infty}^{\infty} f_P(x) \ln \frac{f_P(x)}{f_Q(x)} dx,$$

see, Kullback and Leibler (1951). With

$$f_{\mathcal{K}}(x \mid \alpha,\beta) = (1 + e^\alpha)^{\frac{1}{\beta}} e^{\alpha+\beta x} \left(1 + e^{\alpha+\beta x}\right)^{-\frac{1+\beta}{\beta}}, \quad \text{for } x \geq 0,$$

and

$$f_{\mathcal{G}}(x \mid \beta,{}^{e^\alpha}/_\beta) = e^{\alpha+\beta x + \frac{e^\alpha\left(1-e^{\beta x}\right)}{\beta}}, \quad \text{for } x \geq 0,$$

we have

$$\ln\left(\frac{f_{\mathcal{K}}(x \mid \alpha,\beta)}{f_{\mathcal{G}}(x \mid \beta,{}^{e^\alpha}/_\beta)}\right) = \frac{\ln\left(e^\alpha + 1\right) + e^\alpha\left(e^{\beta x} - 1\right) - (\beta + 1)\ln\left(e^{\alpha+\beta x} + 1\right)}{\beta}. \tag{3.122}$$

By substitution of eq. (3.122) in the definition of the Kullback-Leibler divergence, we obtain

$$
\begin{aligned}
D_{\mathrm{KL}}\big(\mathcal{K}(\alpha,\beta) \,\|\, \mathcal{G}(\beta,{}^{e^\alpha}/_\beta)\big) &= \int_0^\infty f_{\mathcal{K}}(x) \ln \frac{f_{\mathcal{K}}(x)}{f_{\mathcal{G}}(x)}\,dx \\
&= \frac{\ln(e^\alpha+1)}{\beta} \int_0^\infty f_{\mathcal{K}}(x)\,dx + \frac{e^\alpha}{\beta}\int_0^\infty \left(e^{\beta x}-1\right) f_{\mathcal{K}}(x)\,dx \\
&\quad - \frac{\beta-1}{\beta}\int_0^\infty \ln\left(e^{\alpha+\beta x}+1\right) f_{\mathcal{K}}(x)\,dx \\
&= \frac{\ln(e^\alpha+1)}{\beta} + \frac{e^\alpha}{\beta}\left(\frac{1+e^{-\alpha}\beta}{1-\beta}-1\right) \\
&\quad - \frac{\beta+1}{\beta}\left(\alpha+\beta+\ln\left(1+e^{-\alpha}\right)\right) \\
&= \frac{e^\alpha+\beta^2+(1-\beta)\ln(e^\alpha+1)}{\beta-1},
\end{aligned}
\tag{3.123}
$$

where the second integral in eq. (3.123) only exists for $\beta < 1$, since by Proposition 3.3.4

$$
\int_0^\infty e^{\beta x} f_{\mathcal{K}}(x)\,dx = \beta\,\mathbb{E}[X],
$$

for $X \sim \mathcal{P}_{II}(1+e^{-\alpha},\beta)$. □

Notice, that for the Kullback-Leibler divergence obtained in Proposition 3.5.29, we have the following limits:

$$
\begin{aligned}
\lim_{\alpha\to\infty} D_{\mathrm{KL}}\big(\mathcal{K}(\alpha,\beta) \,\|\, \mathcal{G}(\beta,{}^{e^\alpha}/_\beta)\big) &= \infty, \\
\lim_{\alpha\to-\infty} D_{\mathrm{KL}}\big(\mathcal{K}(\alpha,\beta) \,\|\, \mathcal{G}(\beta,{}^{e^\alpha}/_\beta)\big) &= \frac{\beta^2}{1-\beta}, \\
\lim_{\beta\to 0} D_{\mathrm{KL}}\big(\mathcal{K}(\alpha,\beta) \,\|\, \mathcal{G}(\beta,{}^{e^\alpha}/_\beta)\big) &= e^\alpha - \ln(e^\alpha+1)\,.
\end{aligned}
$$

We see that the Kullback-Leibler divergence as obtained eq. (3.121) is exponentially decreasing for α such that for $e^\alpha \approx 0$, that is for $\alpha \ll 0$, we have

$$
D_{\mathrm{KL}}\big(\mathcal{K}(\alpha,\beta) \,\|\, \mathcal{G}(\beta,{}^{e^\alpha}/_\beta)\big) \approx \frac{\beta^2}{1-\beta}.
$$

Proposition 3.5.30. Let $\mathcal{E}(\gamma,\theta,\beta)$ and $\mathcal{E}(\delta,\tau,\beta)$ be two extended exponential distributions with the same scale parameter β. The Kullback-Leibler divergence between $\mathcal{E}(\gamma,\theta,\beta)$ and $\mathcal{E}(\delta,\tau,\beta)$ is given

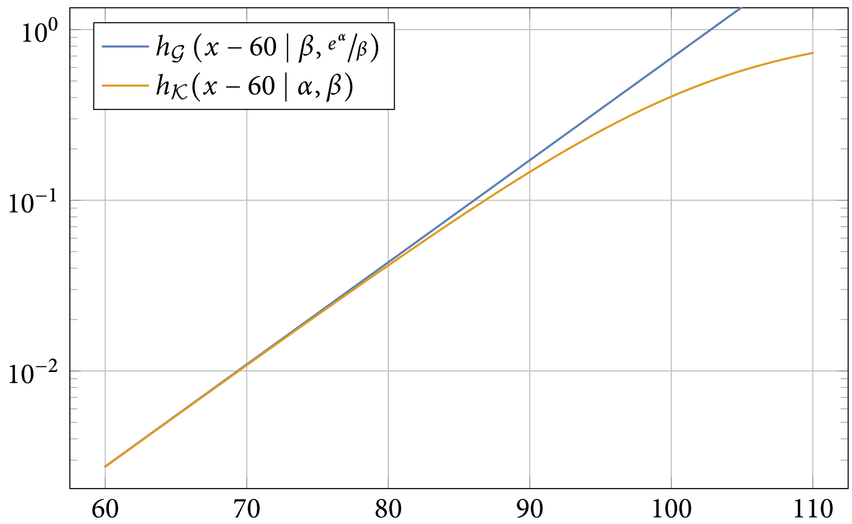


Figure 3.5: Comparison of the post-age-60 hazard rates $h_{\mathcal{K}}(x \mid \alpha_{2010}, \beta_{2010})$ of the Kannisto distribution with estimated parameters ($\alpha = -5.89, \beta = 0.138$) for the year 2010 of the Swedish female population and the Gompertz distribution evaluated for the same parameters. The hazard rates almost coincide for lower ages but diverge for higher ages.

by

$$D_{\mathrm{KL}}\big(\mathcal{E}(\gamma,\theta,\beta) \parallel \mathcal{E}(\delta,\tau,\beta)\big) = \ln\left(\frac{\theta\gamma^{\theta}}{\tau\delta^{\tau}}\right) - (1+\theta)\left(\ln(\gamma) + \frac{1}{\theta}\right) + (1+\tau)\left(\ln(\delta) + \frac{1}{\theta}\,{}_2F_1\left[1,\theta;\theta+1;1-\frac{\delta}{\gamma}\right]\right).$$

Proof. The Kullback-Leibler divergence for non-negative continuous distributions $\mathcal{E}(\gamma,\theta,\beta)$ and $\mathcal{E}(\delta,\tau,\beta)$ with corresponding densities $f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)$ and $f_{\mathcal{E}}(x \mid \delta,\tau,\beta)$ is defined as

$$D_{\mathrm{KL}}(\mathcal{E}(\gamma,\theta,\beta) \parallel \mathcal{E}(\delta,\tau,\beta)) = \int_0^{\infty} f_{\mathcal{E}}(x \mid \gamma,\theta,\beta) \ln \frac{f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{f_{\mathcal{E}}(x \mid \delta,\tau,\beta)}\, dx. \tag{3.124}$$

The density of $f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)$ is given by

$$f_{\mathcal{E}}(x \mid \gamma,\theta,\beta) = \beta\theta\gamma^{\theta} e^{\beta x}\left(e^{\beta x} - (1-\gamma)\right)^{-(\theta+1)},$$

and the logarithmic quotient of the corresponding densities can be simplified to

$$\ln \frac{f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)}{f_{\mathcal{E}}(x \mid \delta,\tau,\beta)} = \ln\left(\frac{\theta\gamma^{\theta}}{\tau\delta^{\tau}}\right) - (\theta+1)\ln\left(\gamma + e^{\beta x} - 1\right) + (\tau+1)\ln\left(\delta + e^{\beta x} - 1\right). \tag{3.125}$$

We begin evaluating eq. (3.124) by splitting the integration into three parts, namely for each summand of the right-hand side of eq. (3.125). Integration of the first term

$$I_1 = \ln\left(\frac{\theta\gamma^\theta}{\tau\delta^\tau}\right)$$

regarding the density is trivial since the term is constant. Integration of the second term I_2, with

$$I_2 = -(\theta+1)\ln\left(\gamma + e^{\beta x} - 1\right),$$

leads to

$$\int_0^\infty I_2\, f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)\, dx = -(\theta+1)\beta\theta\gamma^\theta \int_0^\infty \ln\left(\gamma + e^{\beta x} - 1\right) e^{\beta x} \left(\gamma + e^{\beta x} - 1\right)^{-(\theta+1)} dx.$$

Substituting $x = {}^1/_\beta \ln(1-\gamma+z)$ leads to

$$-(\theta+1)\beta\theta\gamma^\theta \int_\gamma^\infty \frac{1}{\beta} z^{-(\theta+1)} \ln(z)\, dz = -(\theta+1)\beta\theta\gamma^\theta \frac{\gamma^{-\theta}(\theta\ln(\gamma)+1)}{\beta\theta^2} = -(1+\theta)\left(\ln(\gamma) + \frac{1}{\theta}\right).$$

The integration of the last term

$$I_3 = (1+\tau)\ln\left(\delta + e^{\beta x} - 1\right)$$

leads to

$$\begin{aligned}
\int_0^\infty I_3\, f_{\mathcal{E}}(x \mid \gamma,\theta,\beta)\, dx &= (1+\tau)\beta\theta\gamma^\theta \int_0^\infty \ln\left(\delta + e^{\beta x} - 1\right) e^{\beta x} \left(\gamma + e^{\beta x} - 1\right)^{-(\theta+1)} dx \\
&= (1+\tau)\theta\gamma^\theta \int_\gamma^\infty z^{-(\theta+1)} \ln(\delta-\gamma+z)\, dz \qquad (3.126) \\
&= (1+\tau)\theta\gamma^\theta \int_\gamma^\infty z^{-(\theta+1)} \left(\ln(\delta) + \ln\left(1 - \frac{\gamma}{\delta} + \frac{z}{\delta}\right)\right) dz \\
&= (1+\tau)\theta\gamma^\theta \left(\ln(\delta) \int_\gamma^\infty z^{-(\theta+1)}\, dz + \int_\gamma^\infty z^{-(\theta+1)} \ln\left(1 - \frac{\gamma}{\delta} + \frac{z}{\delta}\right) dz\right) \\
&= (1+\tau)\theta\gamma^\theta \left(\frac{\ln(\delta)}{\theta\gamma^\theta} + \frac{1}{\delta^\theta} \int_{\frac{\gamma}{\delta}}^\infty z^{-(\theta+1)} \ln\left(1 - \frac{\gamma}{\delta} + z\right) dz\right) \\
&= (1+\tau)\theta\gamma^\theta \left(\frac{\ln(\delta)}{\theta\gamma^\theta} + \frac{1}{\delta^\theta\theta^2}\, {}_2F_1\left[\theta,\theta;\theta+1;1-\frac{\gamma}{\delta}\right]\right) \qquad (3.127)
\end{aligned}$$

$$= (1+\tau)\theta\gamma^{\theta}\left(\frac{\ln(\delta)}{\theta\gamma^{\theta}} + \frac{1}{\delta^{\theta}\theta^{2}}\frac{\delta^{\theta}}{\gamma^{\theta}}\,{}_2F_1\left[1,\theta;\theta+1;1-\frac{\delta}{\gamma}\right]\right) \tag{3.128}$$
$$= (1+\tau)\left(\ln(\delta) + \frac{1}{\theta}\,{}_2F_1\left[1,\theta;\theta+1;1-\frac{\delta}{\gamma}\right]\right),$$

where eq. (3.126) is obtained by the substitution $z = e^{\beta x} - (1-\gamma)$. Equation (3.127) follows from the integral representation given in Lemma 3.5.25. For eq. (3.128), we use Pfaff's transformation formula for hypergeometric functions to change the argument $1 - {}^{\gamma}/_{\delta}$ to $1 - {}^{\delta}/_{\gamma}$, by using the transformation of eq. (3.75). □

Corollary 3.5.31. The Kullback-Leibler divergence between the Kannisto distributions $\mathcal{K}(\alpha,\beta)$ and $\mathcal{K}(\lambda,\beta)$ is given by

$$D_{\mathrm{KL}}(\mathcal{K}(\alpha,\beta)\,\|\,\mathcal{K}(\lambda,\beta)) = (1+\beta)\left({}_2F_1\left[1,\frac{1}{\beta};1+\frac{1}{\beta};\frac{1-e^{\alpha-\lambda}}{1+e^{\alpha}}\right]-1\right) + \ln\left(\frac{1+e^{-\lambda}}{1+e^{-\alpha}}\right).$$

Proof. Corollary 3.5.31 follows directly from Proposition 3.5.30 for $\tau = \theta = \frac{1}{\beta}$, $\gamma = 1 + e^{-\alpha}$ and $\delta = 1 + e^{-\lambda}$, since by Proposition 3.3.1 the Kannisto distribution is a special case of the extended exponential distribution. □

3.6 | Conclusion

In this chapter, we provided an extensive characterization of the Kannisto and the extended exponential life distributions which are determined by logistic-type hazard rate functions. Logistic-type hazard rates have been originally studied by Kannisto (1992) and Thatcher, Kannisto and Vaupel (1998) and play an important role in mortality modelling of elderly populations. However, the corresponding distributions remained widely uncharacterized. Our contributions show how these distributions are connected to other well-known life and non-life distributions. These connections were obtained through transformations, continuous mixtures, truncations, and as limiting distributions. Furthermore, we derived analytic expressions for the mean residual life function, moment generating function, central moments, Fisher information matrix, and Kullback-Leibler divergence. The Kannisto distribution has been widely uncharacterized in terms of these quantities and the results provided here are the main contributions of this chapter. Moreover, we proved that the extended exponential distribution and the Kannisto distribution belong to the minimum domain of attraction of the Weibull distribution and to the maximum domain of attraction of the Gumbel distribution. We also provided the maximum and minimum domain of attraction for the Gompertz distribution and quantified how the Kannisto and the Gompertz distributions differ in terms of the population maximal lifespan.

In the academical literature as well as in practical applications, the Gompertz distribution plays a major role in mortality modelling of high ages. As we demonstrated, the Kannisto distribution can be obtained as a continuous mixture of Gompertz distributions. This allows a non-homogeneous interpretation of the underlying population. In the previous Section 1.7, we demonstrated that the non-canonical logit link function is often preferable over the canonical logarithmic link in the

Poisson GAPC setting. This implies that in many cases a logistic growth of the mortality rates at high ages does provide a more accurate representation. For the most elementary predictor function $\eta(t,x) = \kappa_t^{(0)} + \kappa_t^{(1)} x$ the choice between a logit or a logarithmic link equally complies to the choice between a Kannisto or a Gompertz distributed lifetime. By studying their connections and showing some important characteristics, we provide deeper insights to parametric hazard rate models for higher ages, which can be beneficially applied in actuarial science or life insurance industry.

Bibliography

Ahn, Sung K. and Gregory C. Reinsel (1990): 'Estimation for partially nonstationary multivariate autoregressive models'. *Journal of the American Statistical Association*, vol. **85**(411): pp. 813–823.

Akaike, Hirotugu (1973): 'Information theory and an extension of the maximum likelihood principle'. *Second International Symposium on Information Theory*. Ed. by Petrov, B. N. and F. Csaki. Budapest: Akadémiai Kiado: pp. 267–281.

– (1981): 'Likelihood of a model and information criteria'. *Journal of econometrics*, vol. **16**(1): pp. 3–14.

Alai, Daniel H. and Michael Sherris (2014): 'Rethinking age-period-cohort mortality trend models'. *Scandinavian Actuarial Journal*, vol. **2014**(3): pp. 208–227.

Arnold, Barry C., Narayanaswamy Balakrishnan and Haikady Navada Nagaraja (1992): *A first course in order statistics*. Vol. 54. Siam.

Aro, Helena and Teemu Pennanen (2011): 'A user-friendly approach to stochastic mortality modelling'. *European Actuarial Journal*, vol. **1**(2): pp. 151–167.

Bailey, Wilfrid Norman (1935): 'Generalized hypergeometric series'.

Ballotta, Laura and Steven Haberman (2003): 'Valuation of guaranteed annuity conversion options'. *Insurance: Mathematics and Economics*, vol. **33**(1): pp. 87–108.

– (2006): 'The fair valuation problem of guaranteed annuity options: The stochastic mortality environment case'. *Insurance: Mathematics and Economics*, vol. **38**(1): pp. 195–214.

Barlow, Richard E., Albert Waldron Marshall and Frank Proschan (1963): 'Properties of probability distributions with monotone hazard rate'. *The Annals of Mathematical Statistics*: pp. 375–389.

Bauer, Daniel, Fred Espen Benth and Rudiger Kiesel (2012): 'Modeling the forward surface of mortality'. *SIAM Journal on Financial Mathematics*, vol. **3**(1): pp. 639–666.

Beard, R.E. (1961): 'A theory of mortality based on actuarial, biological, and medical considerations'. *Proceedings of International Population Conference, New York*. Vol. 1: pp. 611–625.

Berkum, Frank van, Katrien Antonio and Michel Vellekoop (2014): 'The impact of multiple structural changes on mortality predictions'. *Scandinavian Actuarial Journal*: pp. 1–23.

Biffis, Enrico (2005): 'Affine processes for dynamic mortality and actuarial valuations'. *Insurance: mathematics and economics*, vol. **37**(3): pp. 443–468.

Biffis, Enrico and Pietro Millossovich (2006): 'A bidimensional approach to mortality risk'. *Decisions in Economics and Finance*, vol. **29**(2): pp. 71–94.

Booth, Heather (2006): 'Demographic forecasting: 1980 to 2005 in review'. *International Journal of Forecasting*, vol. **22**(3): pp. 547–581.

Booth, Heather and Leonie Tickle (2008): 'Mortality modelling and forecasting: A review of methods'. *Annals of actuarial science*, vol. **3**(1-2): pp. 3–43.

Börger, Matthias, Daniel Fleischer and Nikita Kuksin (2014): 'Modeling the mortality trend under modern solvency regimes'. *Astin Bulletin*, vol. **44**(01): pp. 1–38.

Breusch, Trevor Stanley (1978): 'Testing for autocorrelation in dynamic linear models'. *Australian Economic Papers*, vol. **17**(31): pp. 334–355.

Brockwell, Peter J. and Richard A. Davis (2013): *Time series: theory and methods*. Springer Science & Business Media.

Brouhns, Natacha, Michel Denuit and Jeroen K. Vermunt (2002): 'A Poisson log-bilinear regression approach to the construction of projected lifetables'. *Insurance: Mathematics and Economics*, vol. **31**(3): pp. 373–393.

Burnham, Kenneth P. and David R. Anderson (2002): *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Cairns, Andrew J.G., David Blake and Kevin Dowd (2006): 'A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration'. *Journal of Risk and Insurance*, vol. **73**(4): pp. 687–718.

– (2008): 'Modelling and management of mortality risk: a review'. *Scandinavian Actuarial Journal*, vol. **2008**(2-3): pp. 79–113.

Cairns, Andrew J.G., David Blake, Kevin Dowd, Guy D. Coughlan, David Epstein and Marwa Khalaf-Allah (2011): 'Mortality density forecasts: An analysis of six stochastic mortality models'. *Insurance: Mathematics and Economics*, vol. **48**(3): pp. 355–367.

Cairns, Andrew J.G., David Blake, Kevin Dowd, Guy D. Coughlan, David Epstein, Alen Ong and Igor Balevich (2009): 'A quantitative comparison of stochastic mortality models using data from England and Wales and the United States'. *North American Actuarial Journal*, vol. **13**(1): pp. 1–35.

Carriere, Jacques F. (1992): 'Parametric models for life tables'. *Transactions of the Society of Actuaries*, vol. **44**: pp. 77–99.

Cattani, EDUARDO (2006): 'Three lectures on hypergeometric functions'. *Notes for a course*.

Charnes, Abraham, E.L. Frome and Po-Lung Yu (1976): 'The equivalence of generalized least squares and maximum likelihood estimates in the exponential family'. *Journal of the American Statistical Association*, vol. **71**(353): pp. 169–171.

Clayton, D. and E. Schifflers (1987): 'Models for temporal variation in cancer rates. I: age–period and age–cohort models'. *Statistics in medicine*, vol. **6**(4): pp. 449–467.

Cox, David Roxbee (1962): *Renewal theory*. Vol. 58. Methuen.

Currie, Iain D. (2006): 'Smoothing and forecasting mortality rates with P-splines'. *URL http://www. macs. hw. ac. uk/˜ iain/research/talks/Mortality. pdf.*

– (2012): 'Forecasting with the age-period-cohort model?' *Proceedings of 27th International Workshop on Statistical Modelling*: pp. 87–92.

– (2016): 'On fitting generalized linear and non-linear models of mortality'. *Scandinavian Actuarial Journal*, vol. **2016**(4): pp. 356–383.

Currie, Iain D., Maria Durban and Paul H.C. Eilers (2004): 'Smoothing and forecasting mortality rates'. *Statistical modelling*, vol. **4**(4): pp. 279–298.

Dahl, Mikkel (2004): 'Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts'. *Insurance: mathematics and economics*, vol. **35**(1): pp. 113–136.

Dahl, Mikkel and Thomas Møller (2006): 'Valuation and hedging of life insurance liabilities with systematic mortality risk'. *Insurance: mathematics and economics*, vol. **39**(2): pp. 193–217.

Dickey, David A. and Wayne A. Fuller (1981): 'Likelihood ratio statistics for autoregressive time series with a unit root'. *Econometrica: Journal of the Econometric Society*: pp. 1057–1072.

Diderichsen, Finn and Johan Hallqvist (1997): 'Trends in occupational mortality among middle-aged men in Sweden 1961-1990.' *International journal of epidemiology*, vol. **26**(4): pp. 782–787.

Dobson, Annette J. and Adrian Barnett (2008): *An introduction to generalized linear models*. CRC press.

Dowd, Kevin, Andrew J.G. Cairns, David Blake, Guy D. Coughlan, David Epstein and Marwa Khalaf-Allah (2010a): 'Backtesting stochastic mortality models: an ex post evaluation of multiperiod-ahead density forecasts'. *North American Actuarial Journal*, vol. **14**(3): pp. 281–298.

– (2010b): 'Evaluating the goodness of fit of stochastic mortality models'. *Insurance: Mathematics and Economics*, vol. **47**(3): pp. 255–265.

Efron, Bradley and David V. Hinkley (1978): 'Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information'. *Biometrika*, vol. **65**(3): pp. 457–483.

Engle, Robert F. (1982): 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation'. *Econometrica: Journal of the Econometric Society*: pp. 987–1007.

Engle, Robert F. and Clive W. J. Granger (1987): 'Co-integration and error correction: representation, estimation, and testing'. *Econometrica: journal of the Econometric Society*: pp. 251–276.

Fahrmeir, Ludwig and Heinz Kaufmann (1985): 'Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models'. *The Annals of Statistics*: pp. 342–368.

Fisher, Ronald Aylmer (1935): 'Appendix to BLISS, CI: The case of zero survivors'. *Ann. Appl. Biol*, vol. **22**: pp. 164–165.

Fisher, Ronald Aylmer and Leonard Henry Caleb Tippett (1928): 'Limiting forms of the frequency distribution of the largest or smallest member of a sample'. *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 24. 02. Cambridge Univ Press: pp. 180–190.

Forrey, Robert C. (1997): 'Computing the hypergeometric function'. *Journal of computational physics*, vol. **137**(1): pp. 79–100.

Fréchet, Maurice (1927): 'Sur la loi de probabilité de l'écart maximum'. *Annales de la societe Polonaise de Mathematique*. Vol. 6: pp. 93–116.

Fuller, Wayne A. (1976): *Introduction to statistical time series*. John Wiley & Sons.

Gaille, Séverine and Michael Sherris (2011): 'Modelling mortality with common stochastic long-run trends'. *The Geneva Papers on Risk and Insurance Issues and Practice*, vol. **36**(4): pp. 595–621.

– (2012): *Causes-of-Death Mortality: What Can Be Learned from Cointegration*. Tech. rep. Working paper. Faculty of Business and Economics (HEC Lausanne), University of Lausanne.

Gatzert, Nadine and Hannah Wesker (2014): 'Mortality risk and its effect on shortfall and risk management in life insurance'. *Journal of Risk and Insurance*, vol. **81**(1): pp. 57–90.

Gavrilov, Leonid Anatolevich and Natalia Sergeevna Gavrilova (1991): 'The biology of life span: a quantitative approach.'

Gnedenko, Boris (1943): 'Sur la distribution limite du terme maximum d'une serie aleatoire'. *Annals of mathematics*: pp. 423–453.

Godfrey, Leslie G (1978): 'Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables'. *Econometrica: Journal of the Econometric Society*: pp. 1303–1310.

Gompertz, Benjamin (1825): 'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies'. *Philosophical transactions of the Royal Society of London*: pp. 513–583.

Granger, Clive W. J. (1981): 'Some properties of time series data and their use in econometric model specification'. *Journal of econometrics*, vol. **16**(1): pp. 121–130.

Haberman, Steven and Arthur Renshaw (2009): 'On age-period-cohort parametric mortality rate projections'. *Insurance: Mathematics and Economics*, vol. **45**(2): pp. 255–270.

– (2011): 'A comparative study of parametric mortality projection models'. *Insurance: Mathematics and Economics*, vol. **48**(1): pp. 35–55.

Halley, Edmund (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'. *Philosophical Transactions of the Royal Society London*, vol. **17**: pp. 596–610.

Hamilton, James Douglas (1994): *Time series analysis*. Vol. 2. Princeton university press Princeton.

Hannan, Edward J. and Barry G. Quinn (1979): 'The determination of the order of an autoregression'. *Journal of the Royal Statistical Society. Series B (Methodological)*: pp. 190–195.

Hardin, James W., Joseph W. Hilbe et al. (2012): 'Generalized Linear Models and Extensions'. *Stata Press books*.

Harrell, Frank (2015): *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

Hatzopoulos, Petros and Steven Haberman (2011): 'A dynamic parameterization modeling for the age–period–cohort mortality'. *Insurance: Mathematics and Economics*, vol. **49**(2): pp. 155–174.

Heligman, Larry and John H. Pollard (1980): 'The age pattern of mortality'. *Journal of the Institute of Actuaries*, vol. **107**(01): pp. 49–80.

Himes, Christine L., Samuel H. Preston and Gretchen A. Condran (1994): 'A relational model of mortality at older ages in low mortality countries'. *Population Studies*, vol. **48**(2): pp. 269–291.

Hobcraft, John, Jane Menken and Samuel H. Preston (1985): *Age, period, and cohort effects in demography: a review*. Springer.

Hunt, Andrew and David Blake (2014): 'On the structure and classification of mortality models'. *Work in Progress*.

Hunt, Andrew and Andrés M. Villegas (2015): 'Robustness and convergence in the Lee–Carter model with cohort effects'. *Insurance: Mathematics and Economics*, vol. **64**: pp. 186–202.

Jarque, Carlos M. and Anil K. Bera (1987): 'A test for normality of observations and regression residuals'. *International Statistical Review/Revue Internationale de Statistique*: pp. 163–172.

Johansen, Søren (1988): 'Statistical analysis of cointegration vectors'. *Journal of economic dynamics and control*, vol. **12**(2): pp. 231–254.

– (1995): *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

Johansen, Søren and Katarina Juselius (1990): 'Maximum likelihood estimation and inference on cointegration—with applications to the demand for money'. *Oxford Bulletin of Economics and statistics*, vol. **52**(2): pp. 169–210.

Juselius, Katarina (2006): *The cointegrated VAR model: methodology and applications*. Oxford University Press.

Kannisto, Väinö (1992): *Presentation at a workshop on old age mortality at Odense University*. Odense University, Odense, Denmark.

Knight, Keith (2000): *Mathematical statistics*. Chapman & Hall/CRC Press.

Kullback, Solomon and Richard A. Leibler (1951): 'On information and sufficiency'. *The annals of mathematical statistics*: pp. 79–86.

Kwiatkowski, Denis, Peter C.B. Phillips, Peter Schmidt and Yongcheol Shin (1992): 'Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?' *Journal of econometrics*, vol. **54**(1): pp. 159–178.

Lazar, Dorina and Michel Denuit (2009): 'A multivariate time series approach to projected life tables'. *Applied Stochastic Models in Business and Industry*, vol. **25**(6): pp. 806–823.

Lee, Ronald D. and Lawrence R. Carter (1992): 'Modeling and forecasting US mortality'. *Journal of the American statistical association*, vol. **87**(419): pp. 659–671.

Li, Hong and Yang Lu (2017): 'Coherent Forecasting of Mortality Rates: A Spatial-Temporal Approach'. *ASTIN Bulletin*, vol. **47**(2): pp. 563–600.

Liu, Xiaoming (2008): 'Stochastic Mortality Modelling'. PhD thesis. University of Toronto.

Lomnicki, Zbigniew A. (1961): 'Tests for departure from normality in the case of linear stochastic processes'. *Metrika*, vol. **4**(1): pp. 37–62.

Lovász, Enrico (2011): 'Analysis of Finnish and Swedish mortality data with stochastic mortality models'. *European Actuarial Journal*, vol. **1**(2): pp. 259–289.

Lütkepohl, Helmut (2007): *New introduction to multiple time series analysis*. Springer.

Lütkepohl, Helmut and Markus Krätzig (2004): *Applied time series econometrics*. Cambridge University Press.

Macdonald, Angus, Adrian Gallop, Keith Miller, Stephen Richards, Rajeev Shah and Richard Willets (2007): 'Stochastic projection methodologies: Lee–Carter model features, example results and implications'. *Continuous Mortality Investigation Bureau*, vol. **25**.

MacKinnon, James G., Alfred A. Haug and Leo Michelis (1999): 'Numerical distribution functions of likelihood ratio tests for cointegration'. *Journal of Applied Econometrics*, vol. **14**(5): pp. 563–577.

Makeham, William Matthew (1860): 'On the law of mortality and the construction of annuity tables'. *The Assurance Magazine, and Journal of the Institute of Actuaries*, vol. **8**(6): pp. 301–310.

– (1867): 'On the law of mortality'. *Journal of the Institute of Actuaries (1866)*: pp. 325–358.

– (1890): 'On the Further Development of Gompertz's Law (Concluded)'. *Journal of the Institute of Actuaries*, vol. **28**(4): pp. 316–332.

Marshall, Albert Waldron and Ingram Olkin (2007): *Life distributions: structure of nonparametric, semiparametric and parametric families*. Springer.

McCullagh, Peter and John A. Nelder (1989): *Generalized linear models*. Vol. 37. CRC press.

Milevsky, Moshe A. and S. David Promislow (2001): 'Mortality derivatives and the option to annuitise'. *Insurance: Mathematics and Economics*, vol. **29**(3): pp. 299–318.

Missov, Trifon I. (2013): 'Gamma-Gompertz life expectancy at birth'. *Demographic research*, vol. **28**: pp. 259–270.

MOIVRE, Abraham DE (1725): *Annuities Upon Lives: Or, the Valuation of Annuities Upon Any Number of Lives*. 3. WP et al.

MOKYR, Joel (2003): *The Oxford encyclopedia of economic history*. Oxford University Press.

NELDER, John A. and Robert William Maclagan WEDDERBURN (1972): 'Generalized Linear Models'. *Journal of the Royal Statistical Society. Series A (General)*, vol. **135**(3): pp. 370–384.

NORBERG, Ragnar (2010): 'Forward mortality and other vital rates—Are they the way forward?' *Insurance: Mathematics and Economics*, vol. **47**(2): pp. 105–112.

NØRLUND, Niels Erik (1955): 'Hypergeometric functions'. *Acta Mathematica*, vol. **94**(1): pp. 289–349.

OEPPEN, Jim, James W. VAUPEL et al. (2002): 'Broken limits to life expectancy'. *Science*, vol. **296**(5570): pp. 1029–1031.

O'HARE, Colin and Youwei LI (2012): 'Explaining young mortality'. *Insurance: Mathematics and Economics*, vol. **50**(1): pp. 12–25.

OLIVIERI, Annamaria (2001): 'Uncertainty in mortality projections: an actuarial perspective'. *Insurance: mathematics and economics*, vol. **29**(2): pp. 231–245.

OLVER, Frank W. J. (2010): *NIST handbook of mathematical functions*. Cambridge University Press.

OSTERWALD-LENUM, Michael (1992): 'A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics'. *Oxford bulletin of economics and statistics*, vol. **54**(3): pp. 461–472.

PEARSON, John William (2009): 'Computation of hypergeometric functions'. PhD thesis. University of Oxford.

PERKS, Wilfred (1932): 'On some experiments in the graduation of mortality statistics'. *Journal of the Institute of Actuaries (1886-1994)*, vol. **63**(1): pp. 12–57.

PFAFF, Bernhard (2008): *Analysis of integrated and cointegrated time series with R*. Springer.

PFAFF, Bernhard, Eric ZIVOT and Matthieu STIGLER (2016): *Unit root and cointegration tests for time series data.* R package version 1.3-0.

PIERCE, Donald A. and Daniel W. SCHAFER (1986): 'Residuals in generalized linear models'. *Journal of the American Statistical Association*, vol. **81**(396): pp. 977–986.

PITACCO, Ermanno (2016): *High age mortality and frailty. Some remarks and hints for actuarial modeling*. Tech. rep. CEPAR Working Paper 2016/19.

PITACCO, Ermanno, Michel DENUIT, Steven HABERMAN and Annamaria OLIVIERI (2009): *Modelling longevity dynamics for pensions and annuity business*. OUP Oxford.

PITACCO, Ermanno and Edit RROJI (2016): 'Assessing longevity risk in a portfolio of life annuities'.

PLAT, Richard (2009): 'On stochastic mortality modeling'. *Insurance: Mathematics and Economics*, vol. **45**(3): pp. 393–404.

POLLARD, John H., B. BENJAMIN and A.S. SOLIMAN (1987): 'Projection of age-specific mortality rates.' *Population Bulletin of the United Nations*: pp. 55–69.

R Core Team (2015): *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Renshaw, Arthur and Steven Haberman (2003): 'Lee–Carter mortality forecasting with age-specific enhancement'. *Insurance: Mathematics and Economics*, vol. **33**(2): pp. 255–272.

– (2006): 'A cohort-based extension to the Lee–Carter model for mortality reduction factors'. *Insurance: Mathematics and Economics*, vol. **38**(3): pp. 556–570.

Renshaw, Arthur, Steven Haberman and Petros Hatzopoulos (1996): 'The modelling of recent mortality trends in United Kingdom male assured lives'. *British Actuarial Journal*, vol. **2**(02): pp. 449–477.

Rinne, Horst (2014): *The Hazard Rate. Theory and inference (with supplementary MATLAB-Programs)*. Justus-Liebig-Universität.

Roberts, Richard (2012): 'Did anyone learn anything from the Equitable Life?' *Lessons and learning from the financial crises', Kings College*.

Rodríguez, Germán (2008): *Lecture notes on generalized linear models*.

Salhi, Yahia and Stéphane Loisel (2011): *Longevity basis risk modeling: A co-integration based approach*. Tech. rep. Tech. rep., ISFA, Université Lyon 1.

Schwarz, Gideon et al. (1978): 'Estimating the dimension of a model'. *The annals of statistics*, vol. **6**(2): pp. 461–464.

Sithole, Terry Z., Steven Haberman and Richard J. Verrall (2000): 'An investigation into parametric models for mortality projections, with applications to immediate annuitants' and life office pensioners' data'. *Insurance: Mathematics and Economics*, vol. **27**(3): pp. 285–312.

Slater, Lucy Joan (1966): *Generalized hypergeometric functions*. University Press Cambridge.

Tabeau, Ewa, Anneke van den Berg Jeths and Christopher Heathcote (2001): *Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective*. Vol. 9. Springer Science & Business Media.

Tappe, Stefan and Stefan Weber (2014): 'Stochastic mortality models: an infinite-dimensional approach'. *Finance and Stochastics*, vol. **18**(1): pp. 209–248.

Taubenberger, Jeffery K. and David M. Morens (2006): '1918 Influenza: the mother of all pandemics'. *Rev Biomed*, vol. **17**: pp. 69–79.

Thatcher, A. Roger, Väinö Kannisto, James W. Vaupel et al. (1998): *The force of mortality at ages 80 to 120*. Vol. 22. Odense University Press Odense.

Tuljapurkar, Shripad and Carl Boe (1998): 'Mortality change and forecasting: how much and how little do we know?' *North American Actuarial Journal*, vol. **2**(4): pp. 13–47.

Turner, Heather and David Firth (2015): *Generalized nonlinear models in R: An overview of the gnm package*. R package version 1.0-8.

VAUPEL, James W. and Anatoli I. YASHIN (1985): 'Heterogeneity's ruses: some surprising effects of selection on population dynamics'. *The American Statistician*, vol. **39**(3): pp. 176–185.

VILLEGAS, Andrés M., Vladimir K. KAISHEV and Pietro MILLOSSOVICH (2015): 'StMoMo: An R Package for Stochastic Mortality Modelling'. *7th Australasian Actuarial Education and Research Symposium*.

VILLEGAS, Andrés M., Pietro MILLOSSOVICH and Vladimir K. KAISHEV (2016): *StMoMo: An R Package for Stochastic Mortality Modelling*. R package version 0.3.1.

WASSERMAN, Larry (2013): *All of statistics: a concise course in statistical inference*. Springer Science & Business Media.

WILLETS, R.C. (2004): *The cohort effect: insights and explanations*. Cambridge Univ Press.

WILMOTH, John R. (1997): 'In search of limits'. *Between Zeus and the Salmon: the biodemography of longevity*, vol. **3**: pp. 38–64.

– (2000): 'Demography of longevity: past, present, and future trends'. *Experimental gerontology*, vol. **35**(9): pp. 1111–1129.

WILMOTH, John R., K. ANDREEV, D. JDANOV, Dana A. GLEI, C. BOE, M. BUBENHEIM, D. PHILIPOV, V. SHKOLNIKOV and P. VACHON (2007): 'Methods protocol for the human mortality database'. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://mortality. org [version 31/05/2007]*.

WONG-FUPUY, Carlos and Steven HABERMAN (2004): 'Projecting mortality trends: recent developments in the United Kingdom and the United States'. *North American Actuarial Journal*, vol. **8**(2): pp. 56–83.

YASHIN, Anatoli I. (2004): 'Semiparametric models in the studies of aging and longevity'. *Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life*. Springer: pp. 149–164.

List of Figures

List of Tables

Appendix A | Empirical Data & Model Estimates

Contents

A.1 | Characteristics of the Kannisto distribution

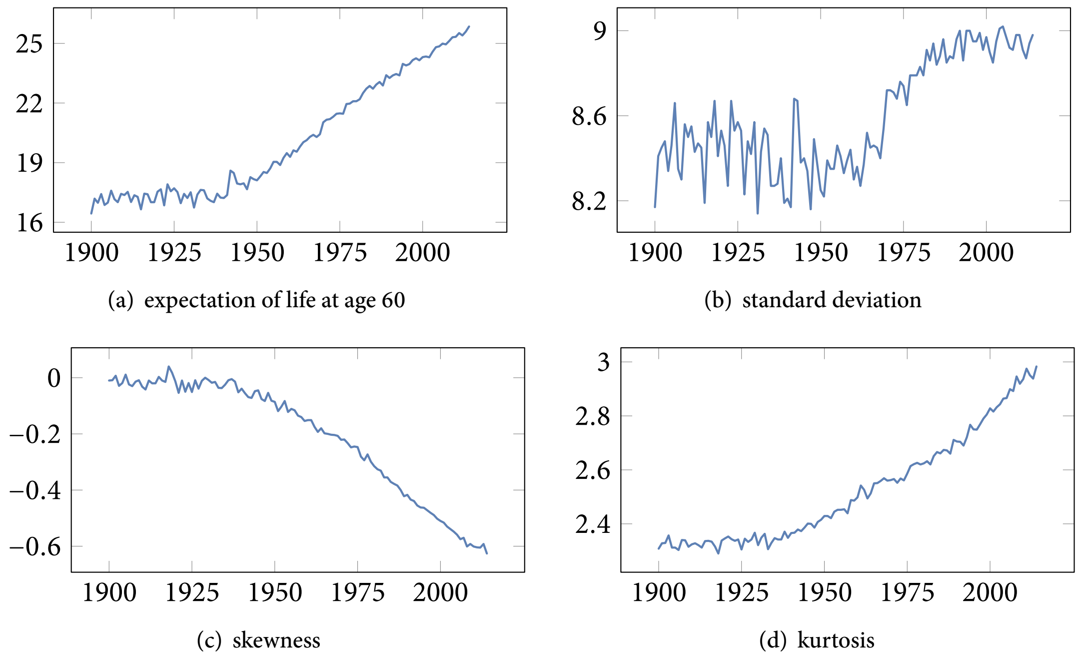


Figure A.1: Mean, standard deviation, skewness, and kurtosis of the $\mathcal{K}(\alpha_t,\beta_t)$ distribution at the estimated parameters α and β for the for Swedish females between 1900 and 2014.

A.2 | Mortality Improvements

A.2.1 | Sweden Improvements

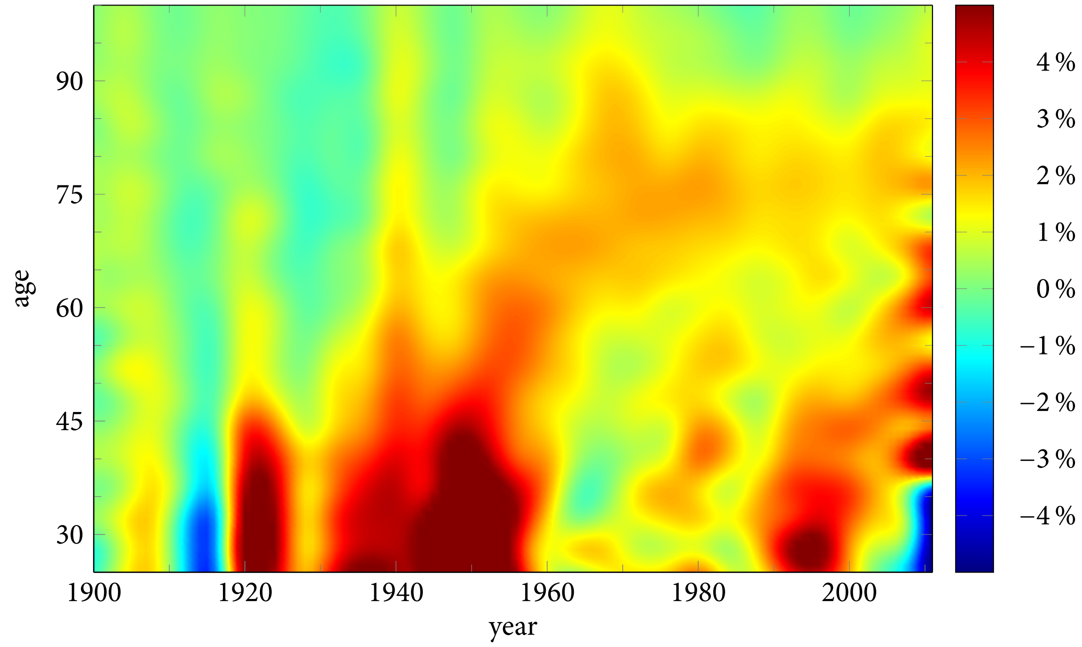


Figure A.2: Annual mortality improvements of Sweden (females).

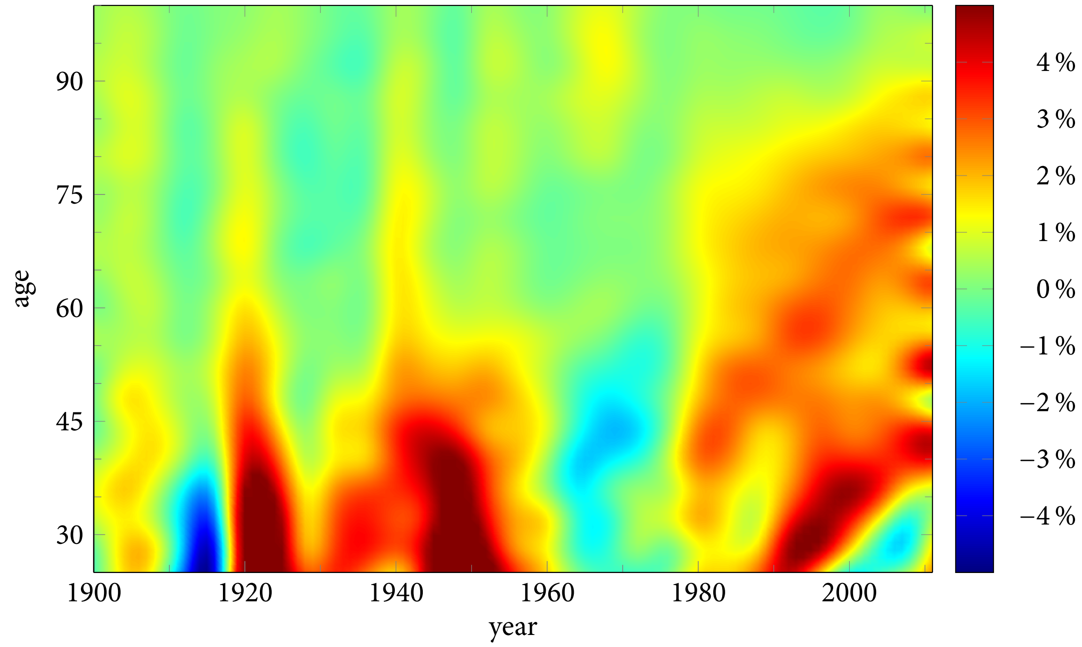


Figure A.3: Annual mortality improvements of Sweden (males).

A.2.2 | UK Improvements

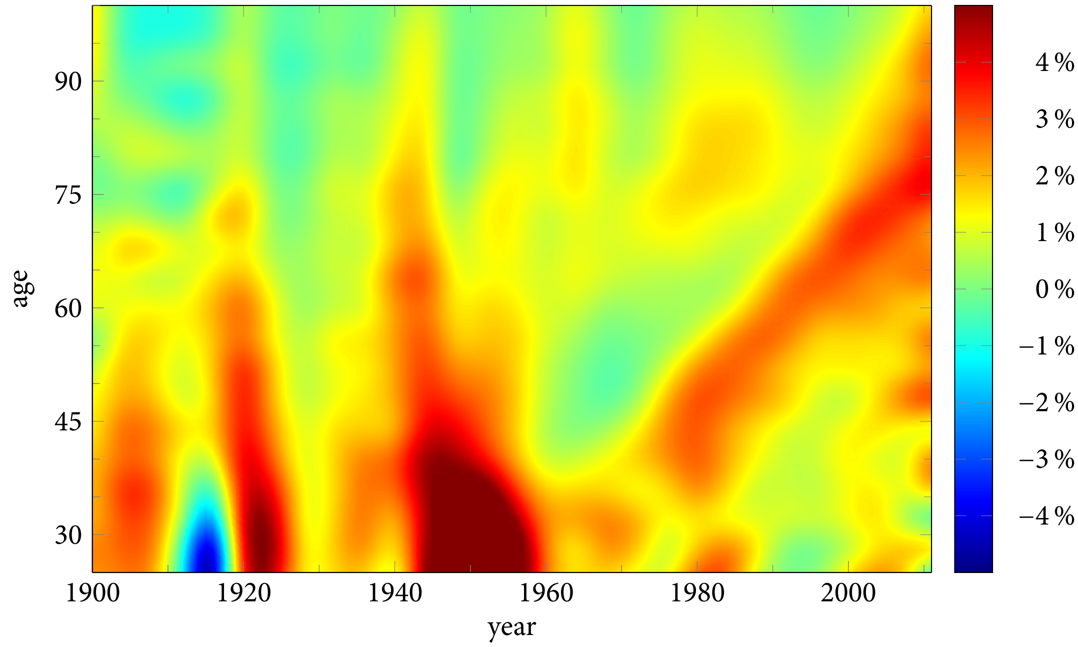


Figure A.4: Annual mortality improvements of UK-Wales (females).

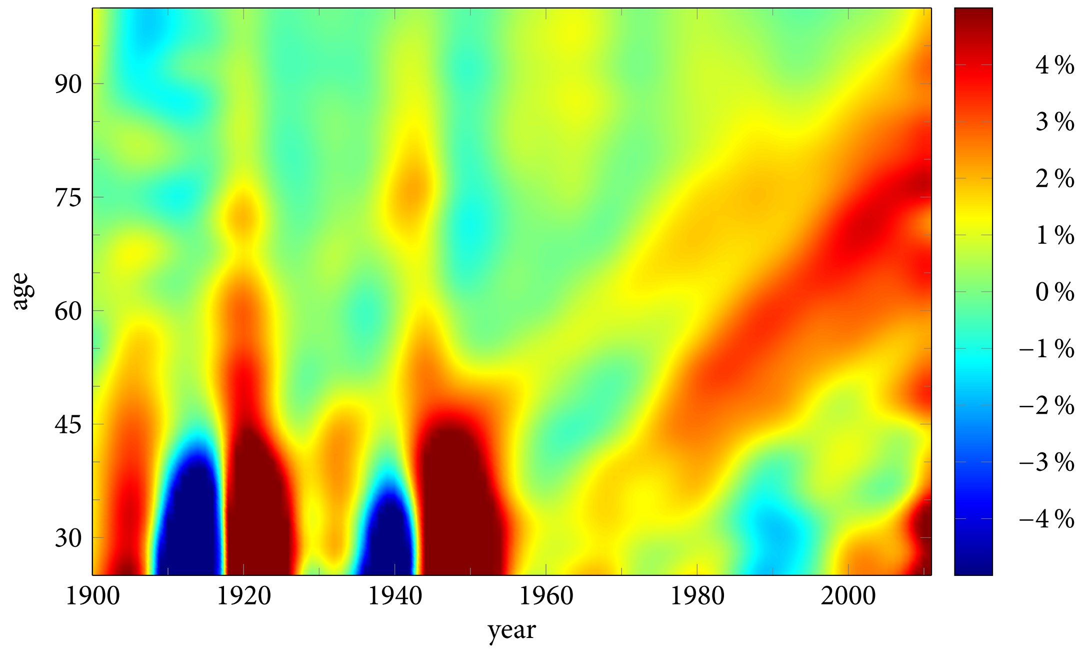


Figure A.5: Annual mortality improvements of UK-Wales (males).

A.2.3 | France Improvements

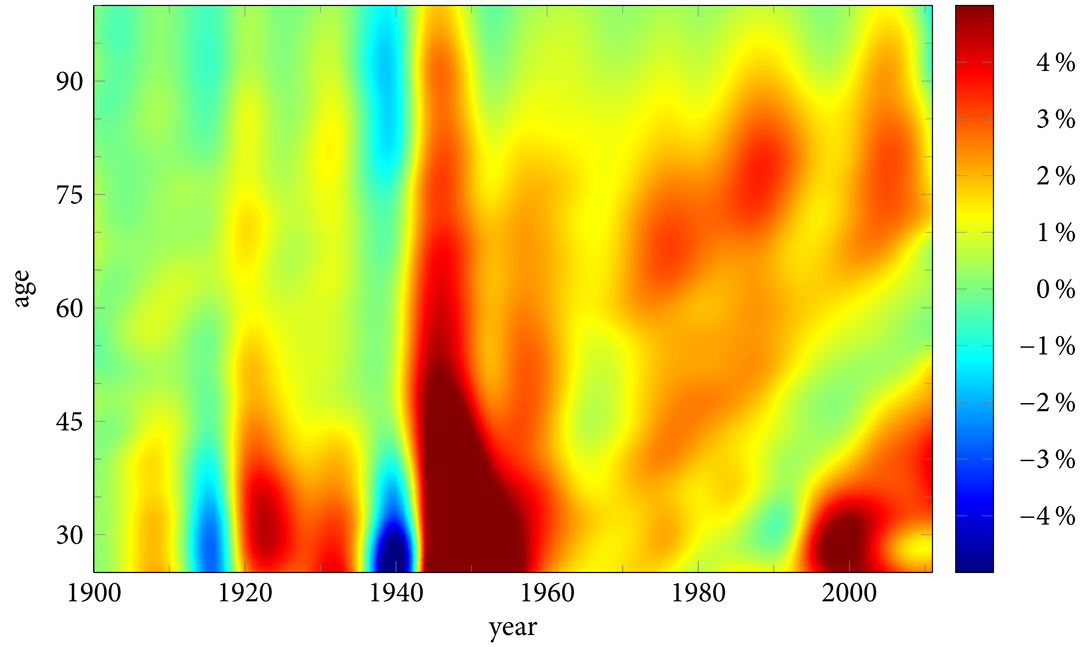


Figure A.6: Annual mortality improvements of France (females).

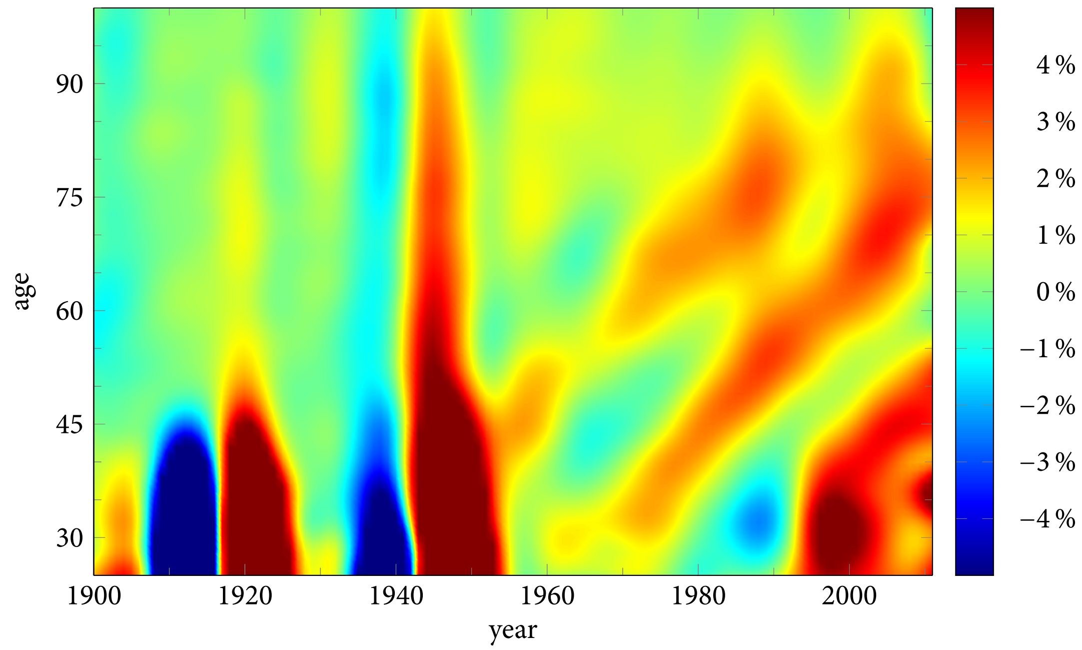


Figure A.7: Annual mortality improvements of France (males).

A.2.4 | Denmark Improvements

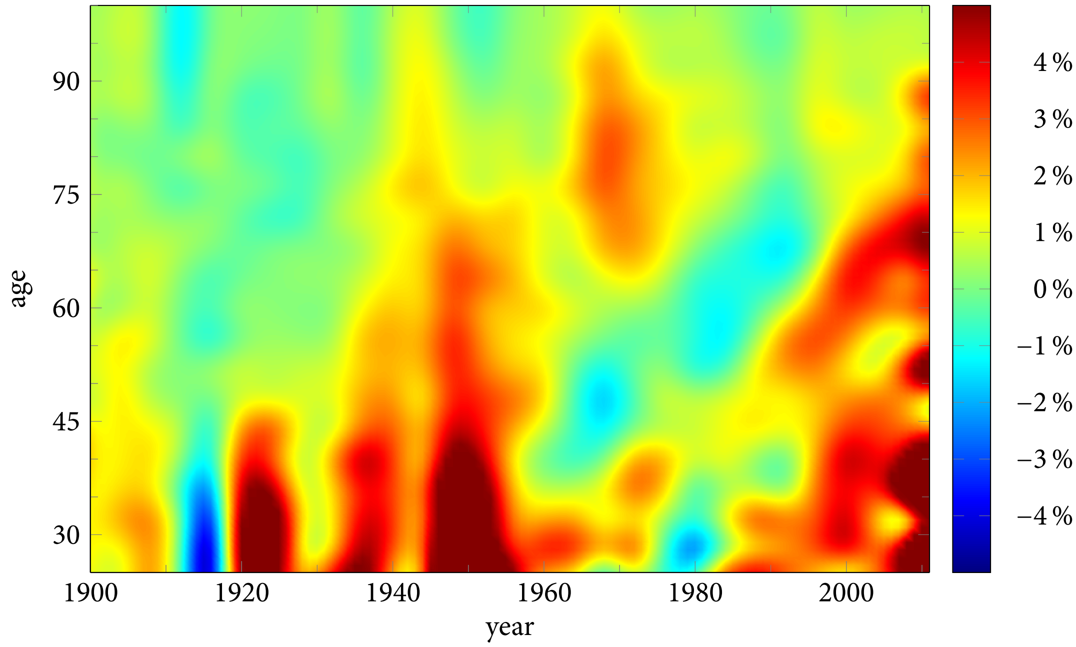


Figure A.8: Annual mortality improvements of Denmark (females).

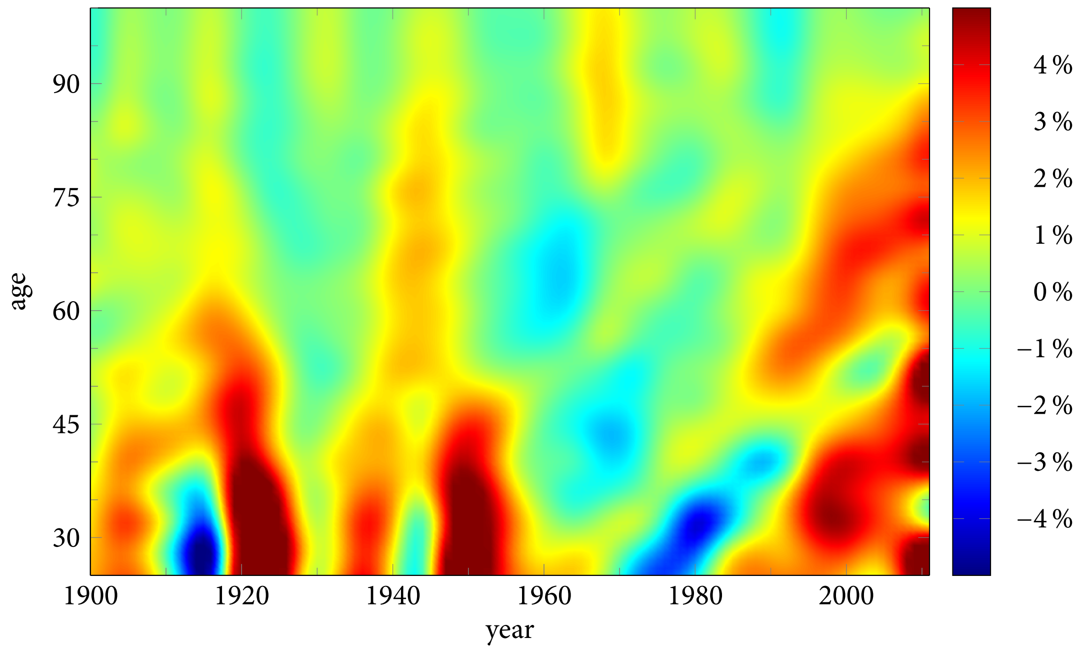


Figure A.9: Annual mortality improvements of Denmark (males).

A.2.5 | Switzerland Improvements

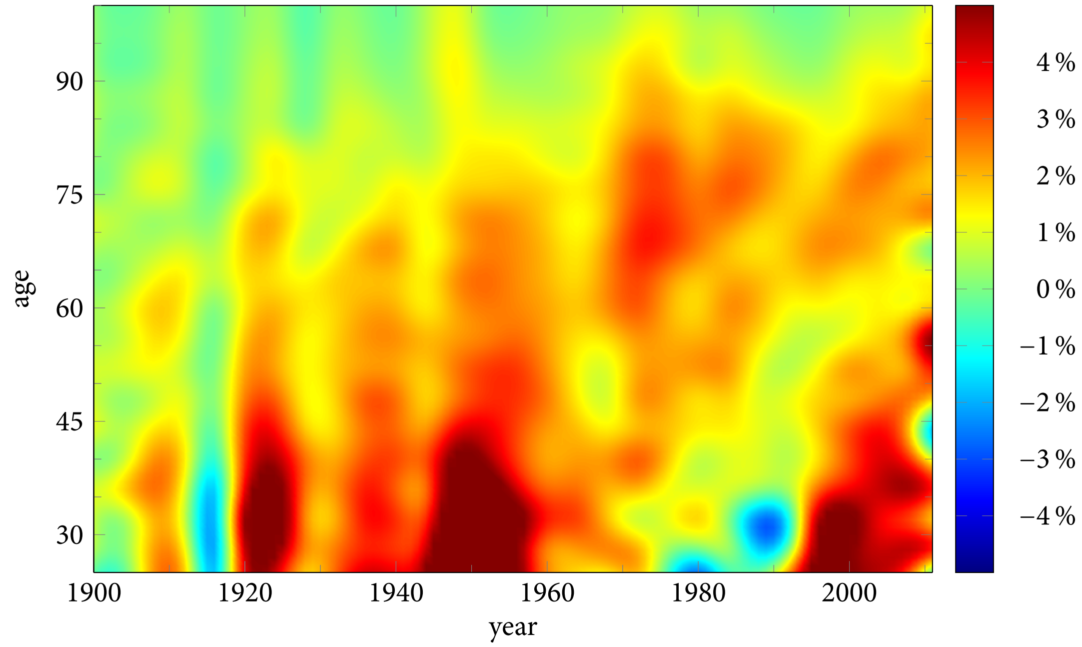


Figure A.10: Annual mortality improvements of Switzerland (females).

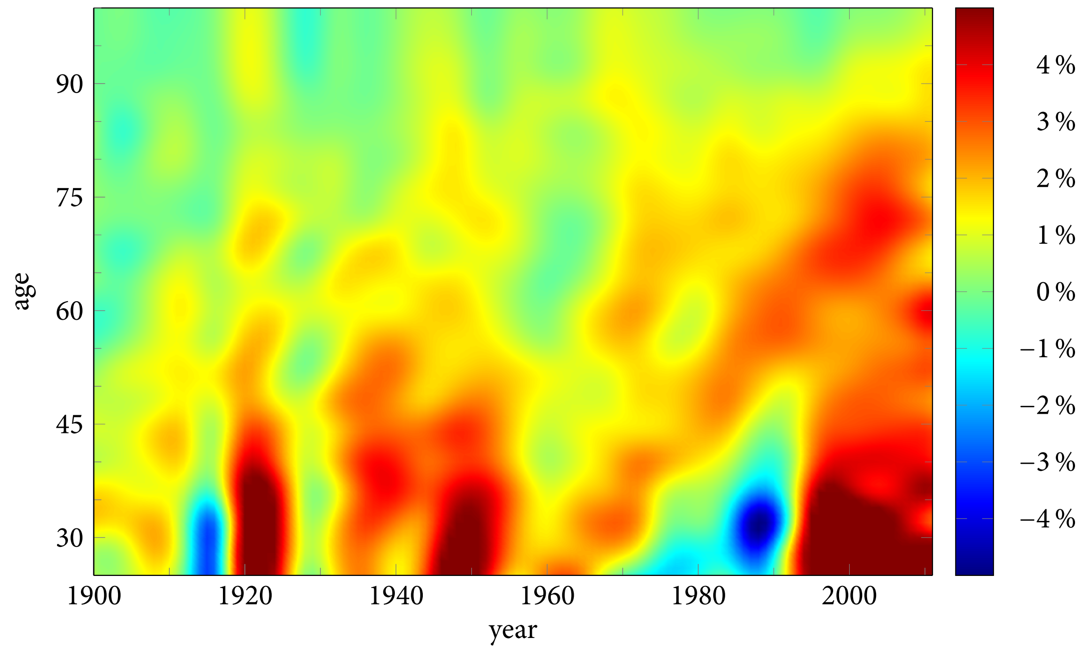


Figure A.11: Annual mortality improvements of Switzerland (males).

A.2.6 | Finland Improvements

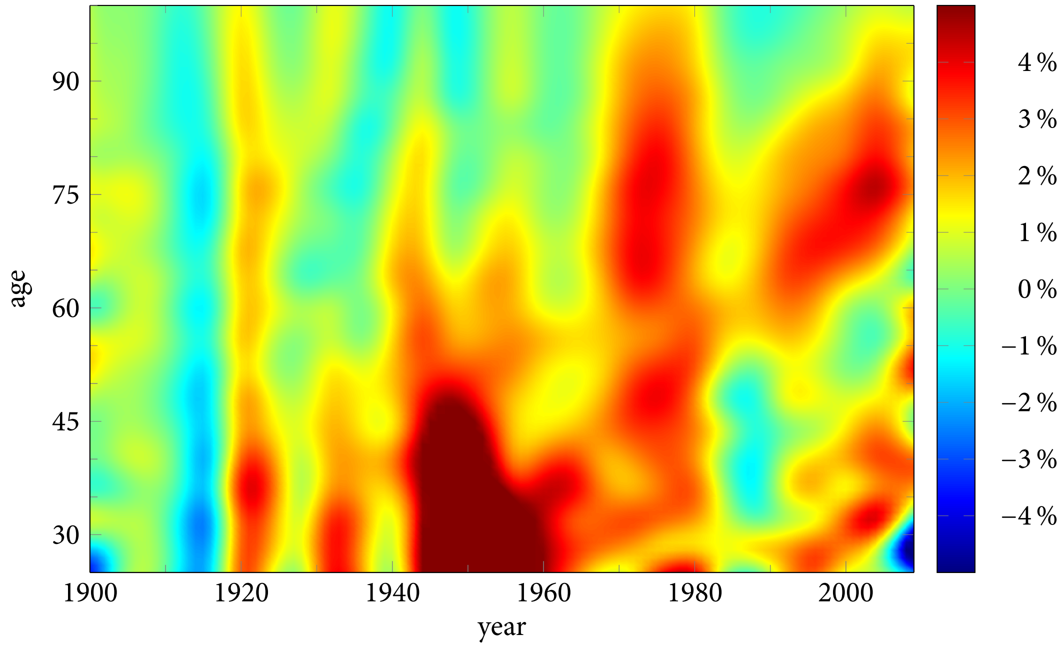


Figure A.12: Annual mortality improvements of Finland (females).

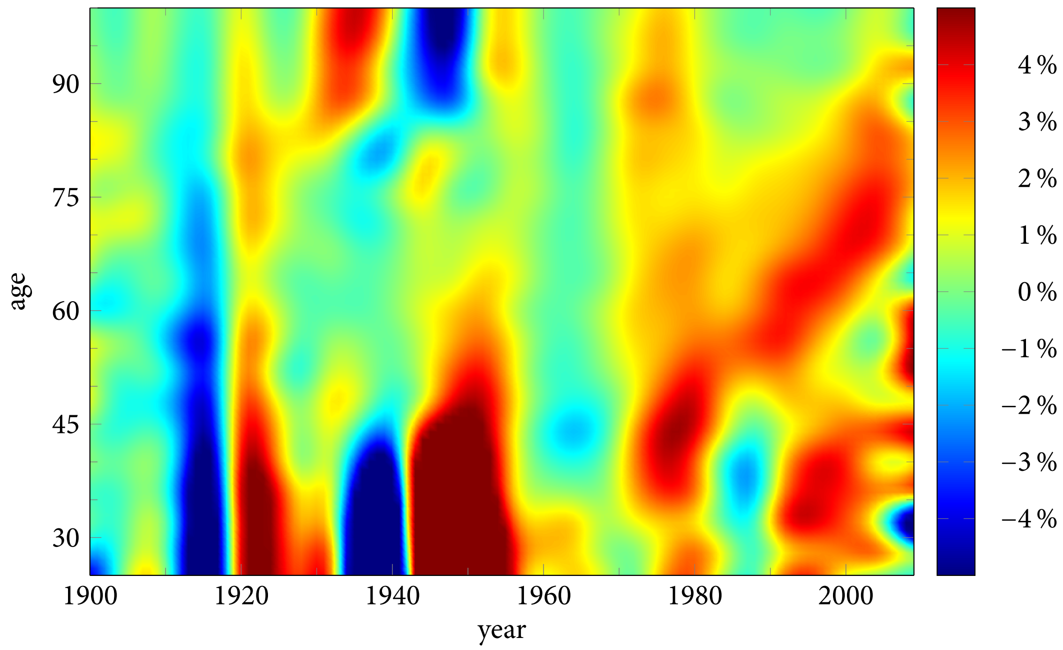


Figure A.13: Annual mortality improvements of Finland (males).

A.2.7 | Russia Improvements

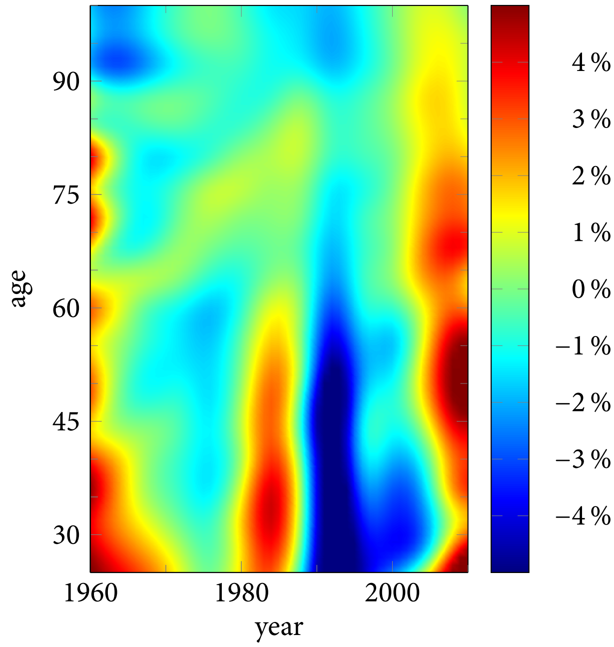


Figure A.14: Annual mortality improvements of Russia (females).

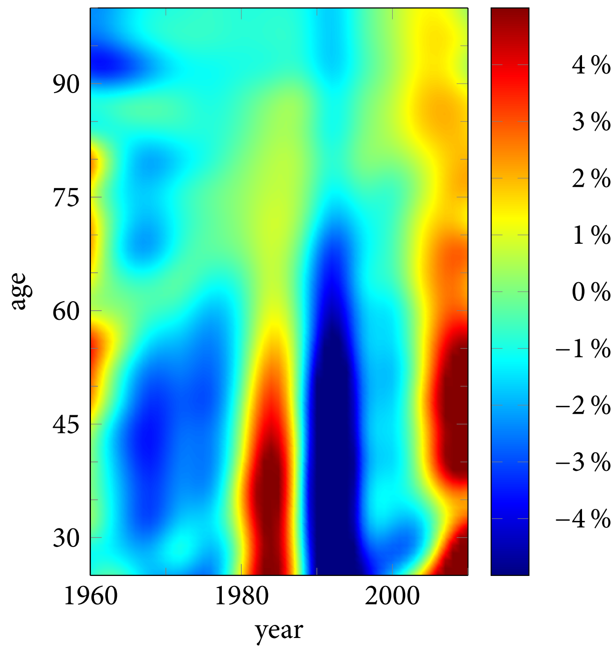
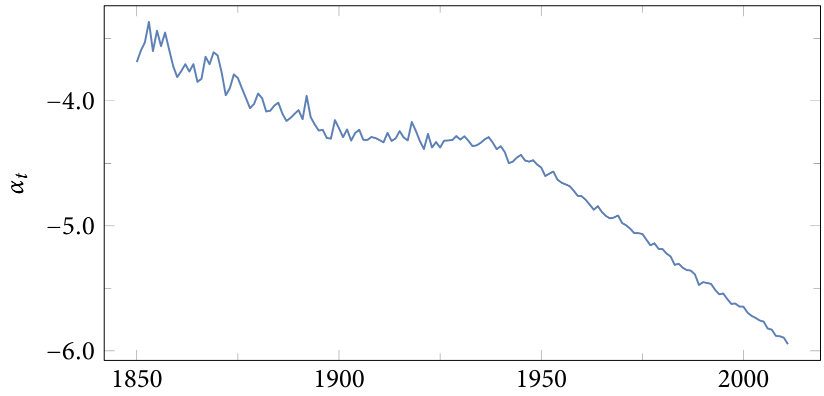


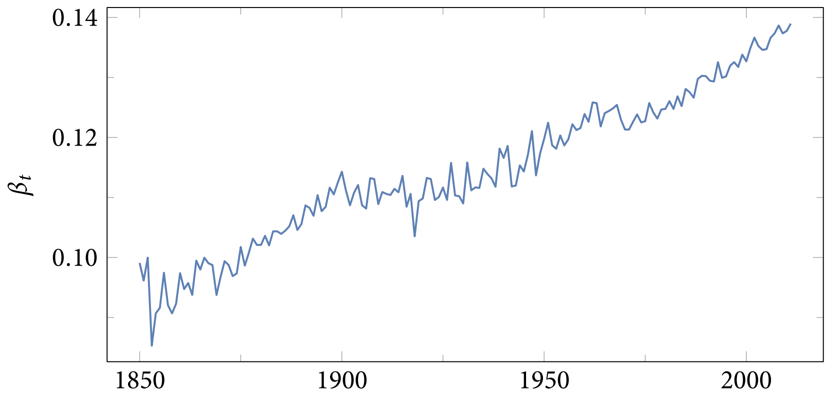
Figure A.15: Annual mortality improvements of Russia (males).

A.3 | Estimated Kannisto Time Series

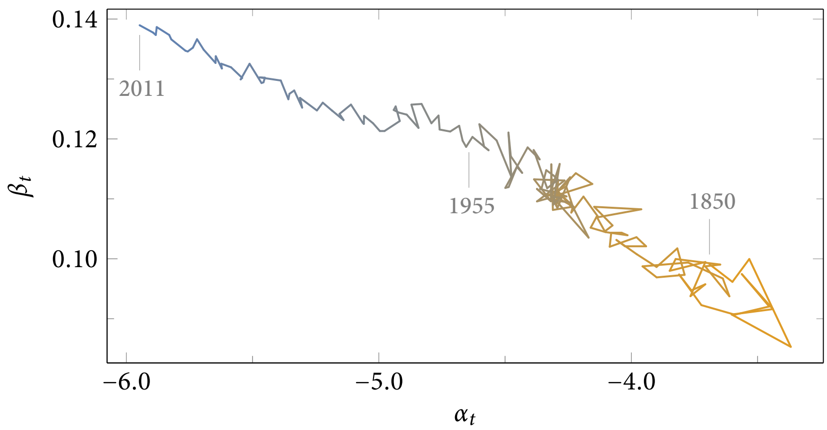
A.3.1 | Sweden



(a) Estimated parameter series α_t.

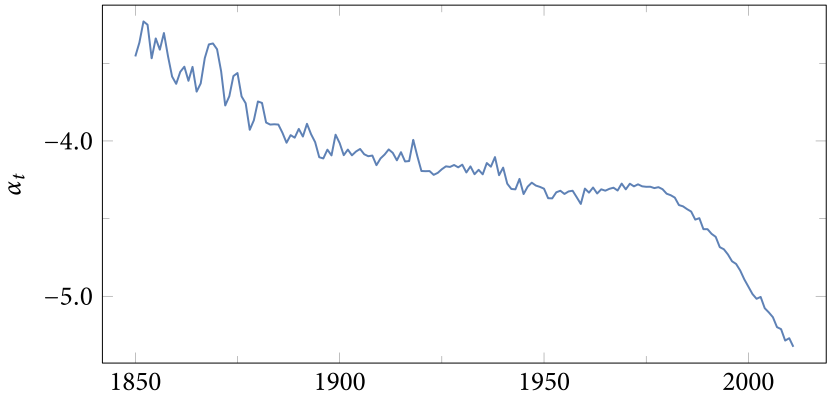


(b) Estimated parameter series β_t.

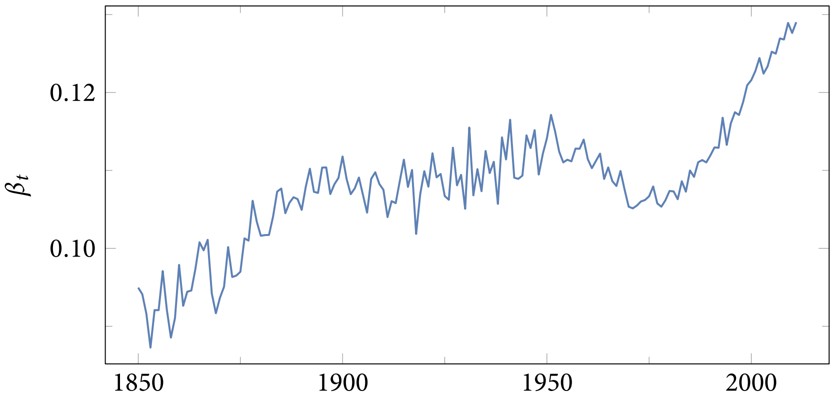


(c) Parametric plot (α_t,β_t).

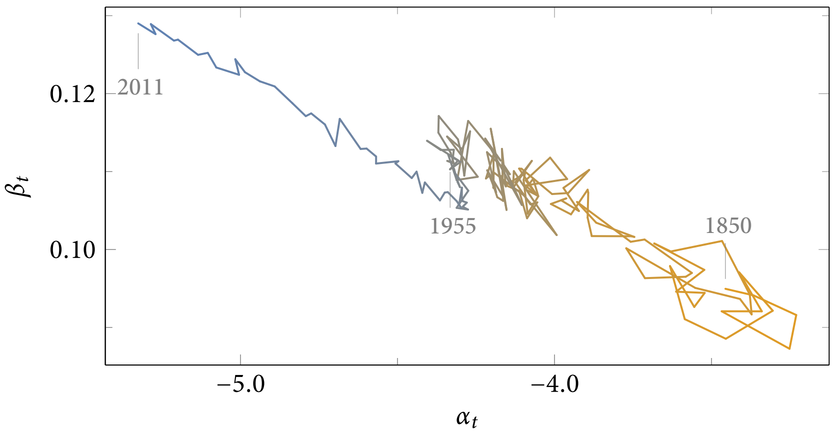
Figure A.16: KAN model estimates for Sweden (females).



(a) Estimated parameter series α_t.



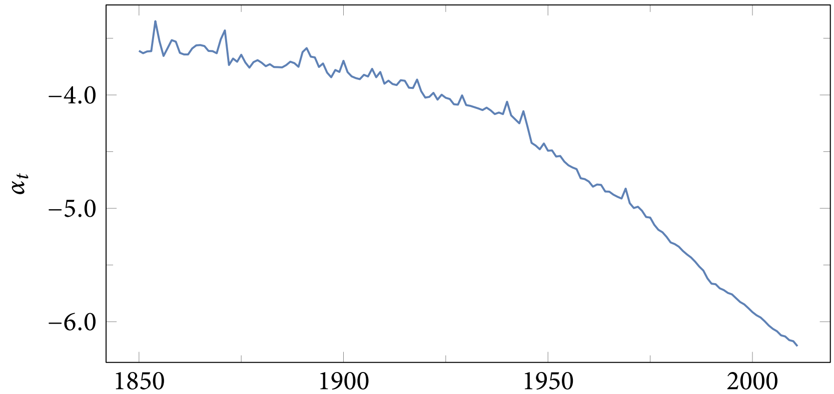
(b) Estimated parameter series β_t.



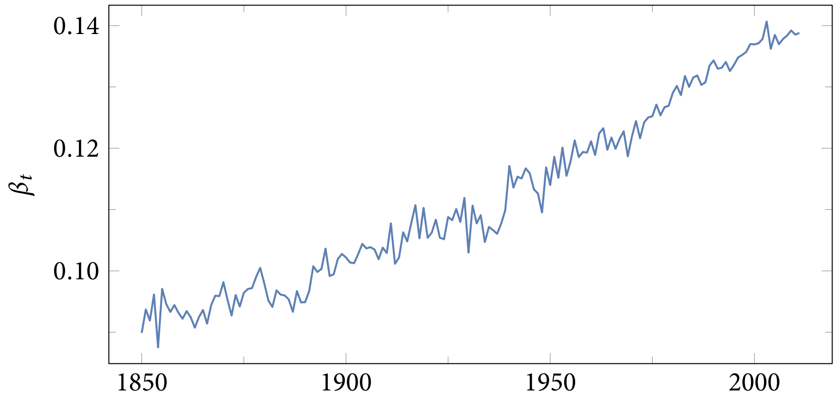
(c) Parametric plot (α_t,β_t).

Figure A.17: KAN model estimates for Sweden (males).

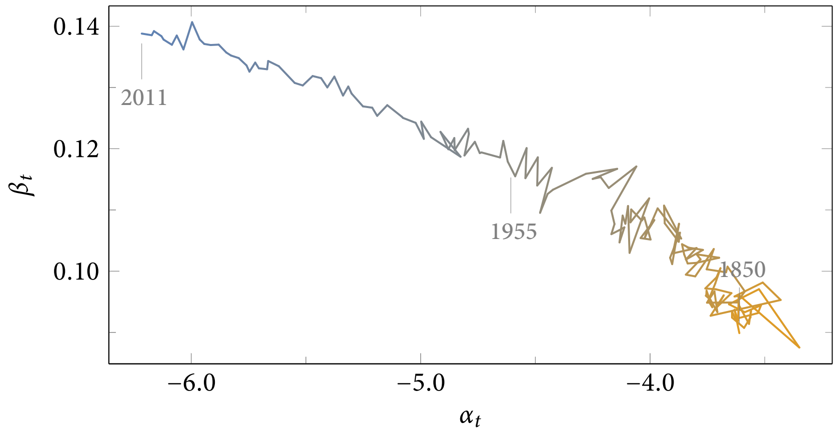
A.3.2 | France



(a) Estimated parameter series α_t.

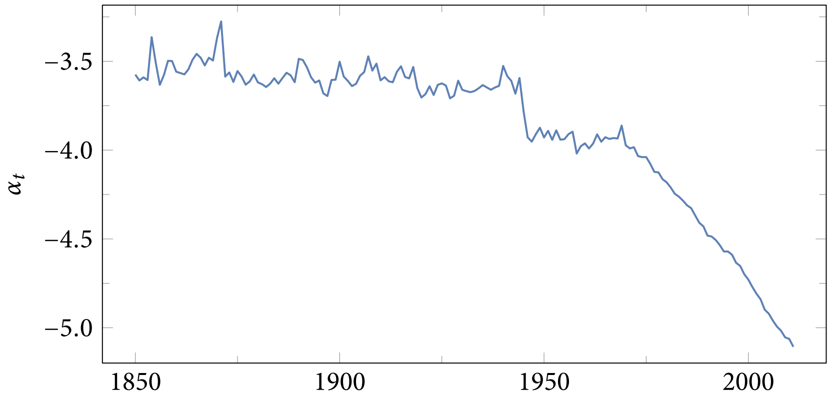


(b) Estimated parameter series β_t.

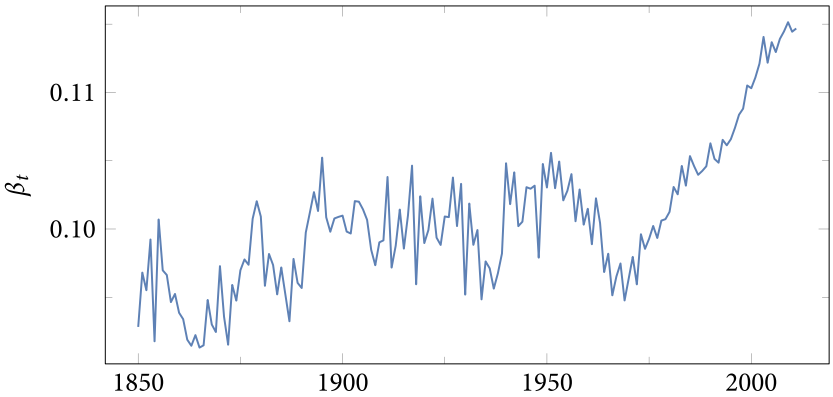


(c) Parametric plot (α_t, β_t).

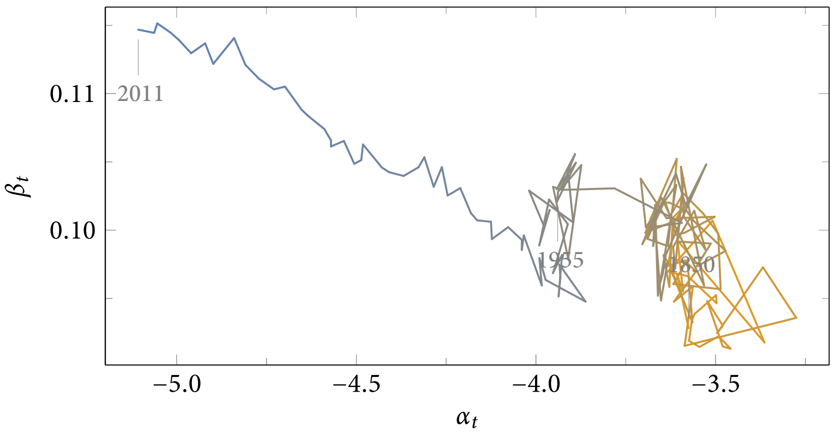
Figure A.18: KAN model estimates for France (females).



(a) Estimated parameter series α_t.



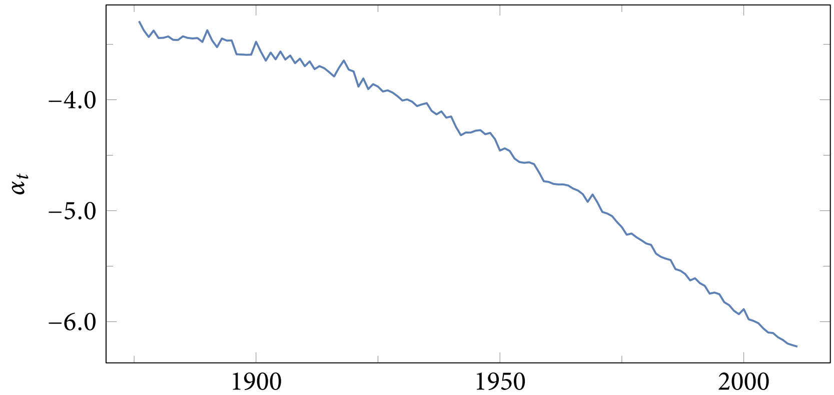
(b) Estimated parameter series β_t.



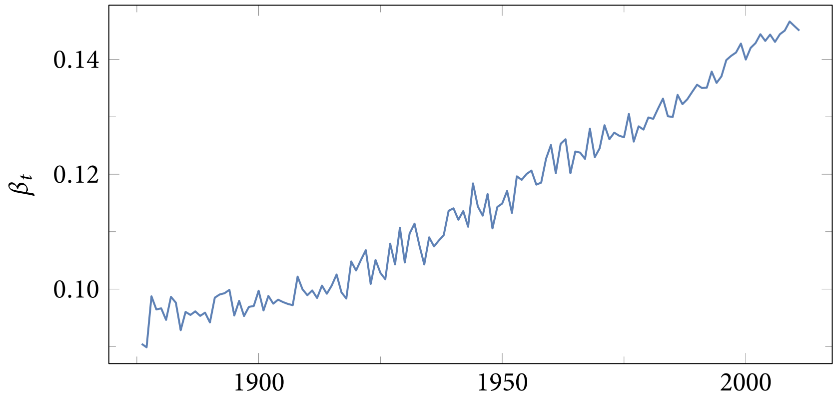
(c) Parametric plot (α_t,β_t).

Figure A.19: KAN model estimates for France (males).

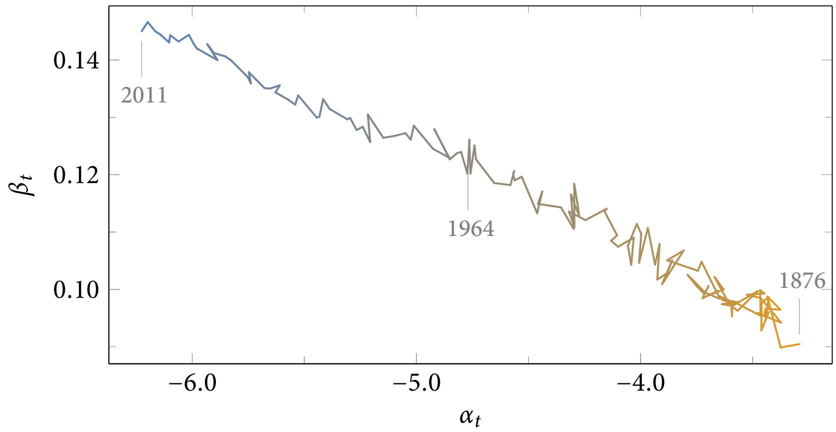
A.3.3 | Switzerland



(a) Estimated parameter series α_t.

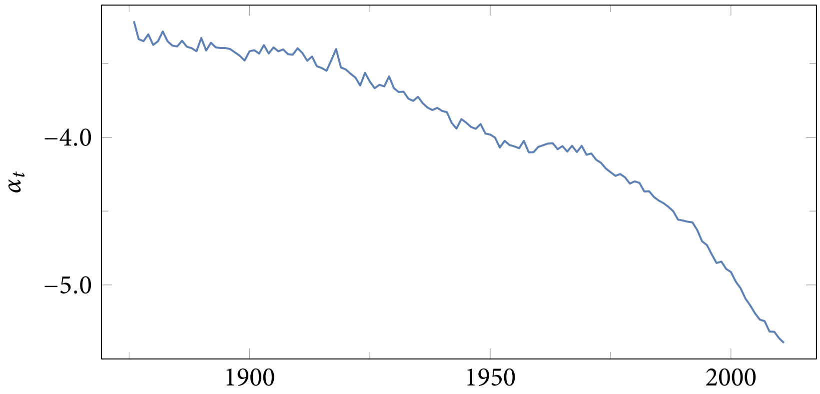


(b) Estimated parameter series β_t.

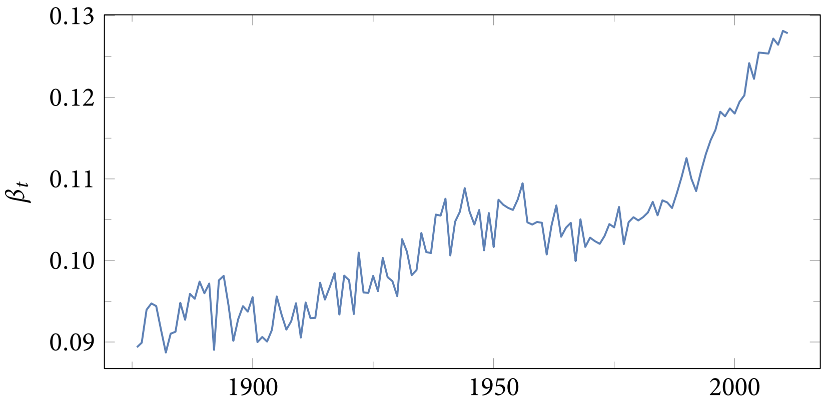


(c) Parametric plot (α_t, β_t).

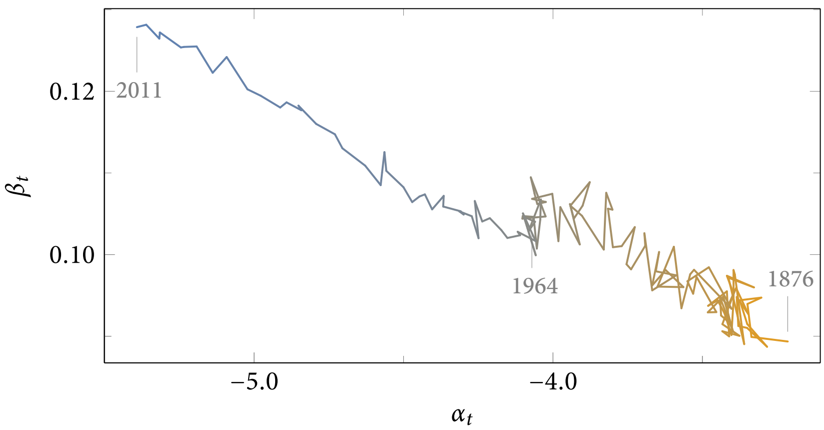
Figure A.20: KAN model estimates for Switzerland (females).



(a) Estimated parameter series α_t.



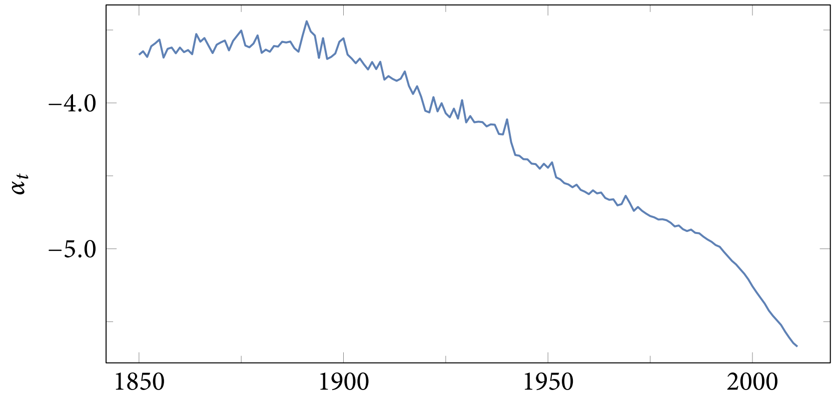
(b) Estimated parameter series β_t.



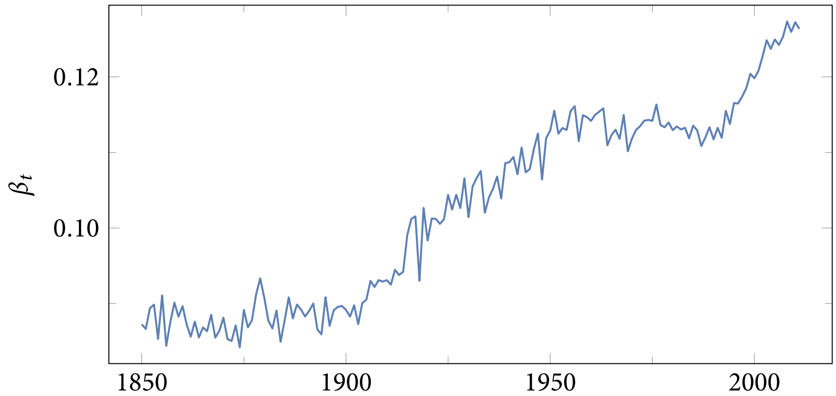
(c) Parametric plot (α_t,β_t).

Figure A.21: KAN model estimates for Switzerland (males).

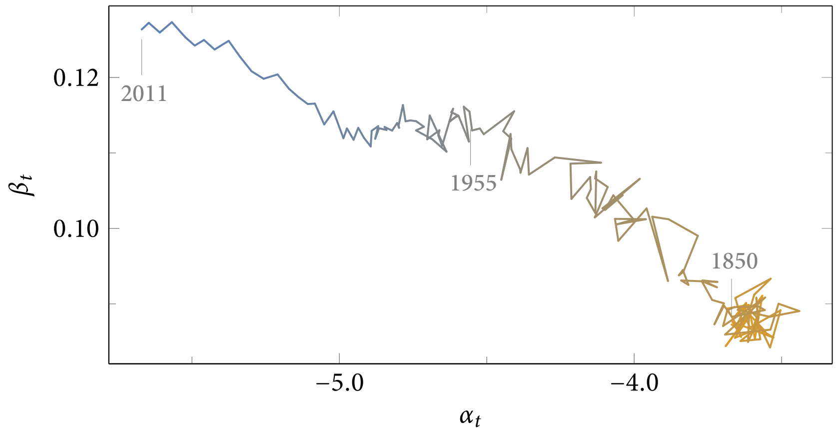
A.3.4 | UK-Wales



(a) Estimated parameter series α_t.

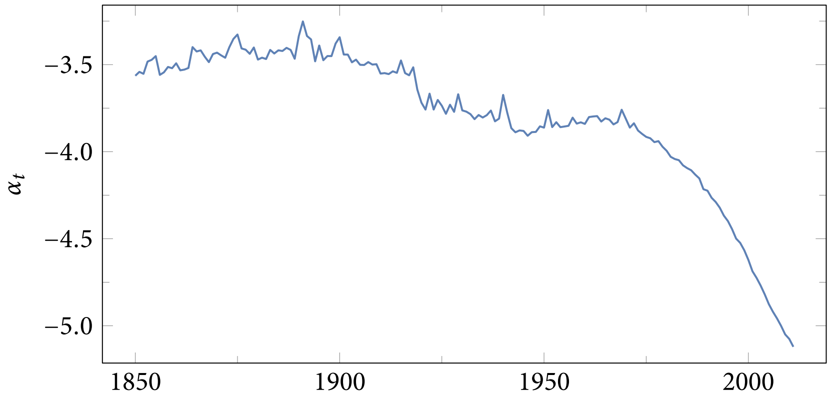


(b) Estimated parameter series β_t.

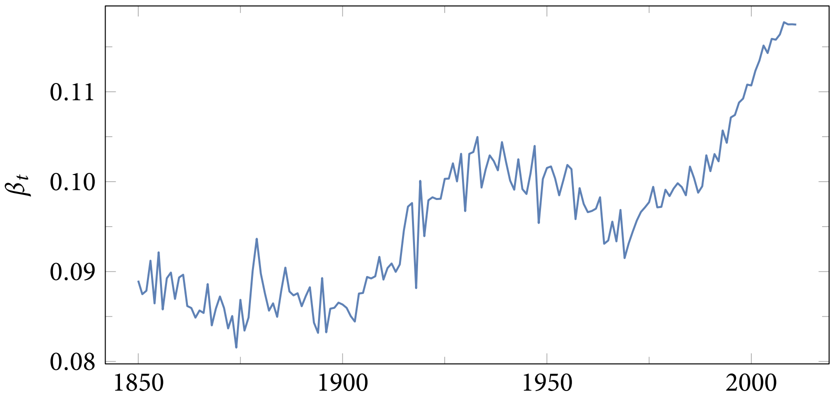


(c) Parametric plot (α_t,β_t).

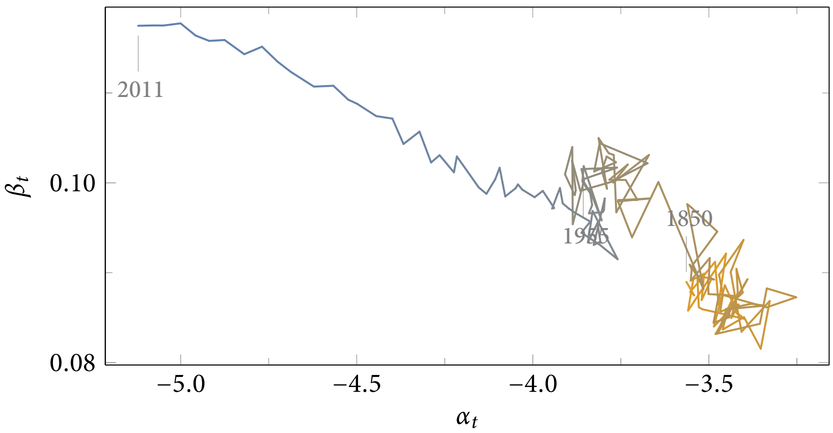
Figure A.22: KAN model estimates for UKWales (females).



(a) Estimated parameter series α_t.



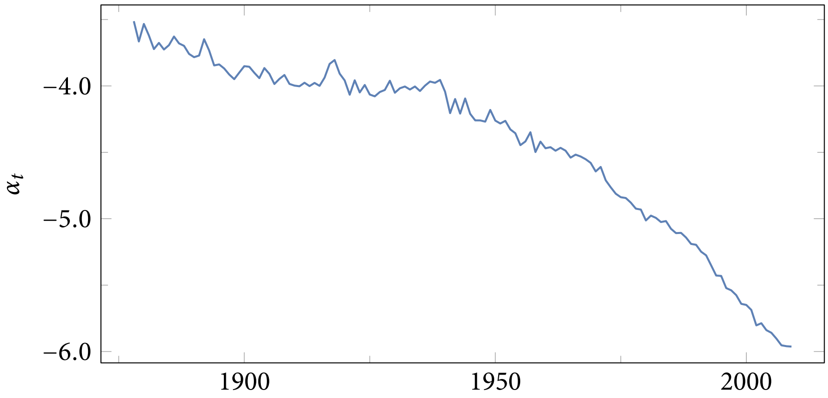
(b) Estimated parameter series β_t.



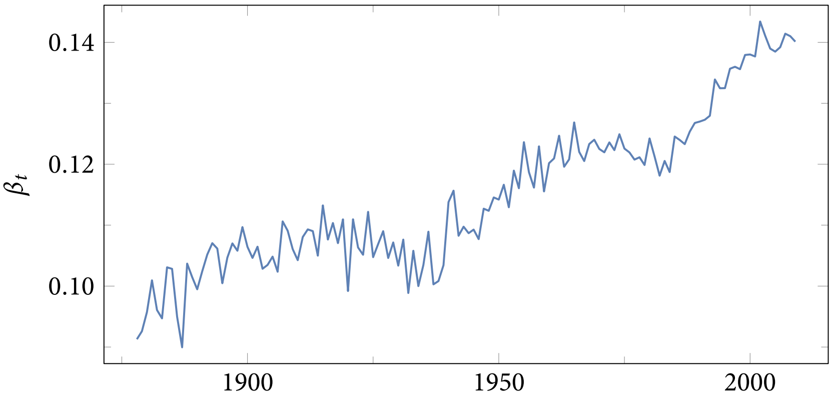
(c) Parametric plot (α_t,β_t).

Figure A.23: KAN model estimates for UKWales (males).

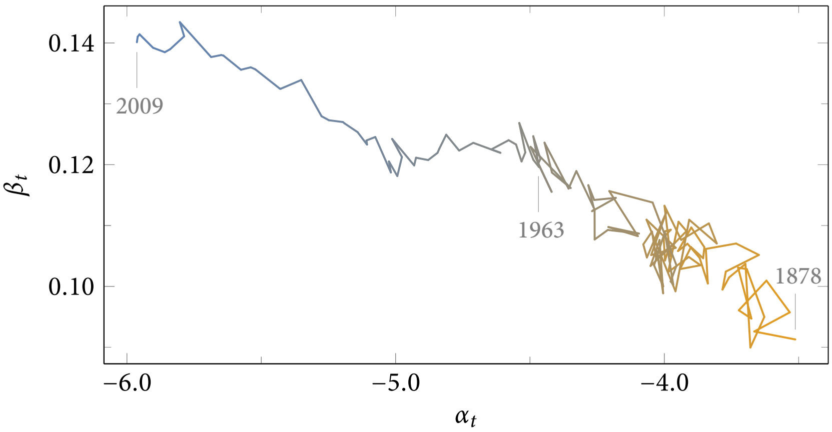
A.3.5 | Finland



(a) Estimated parameter series α_t.

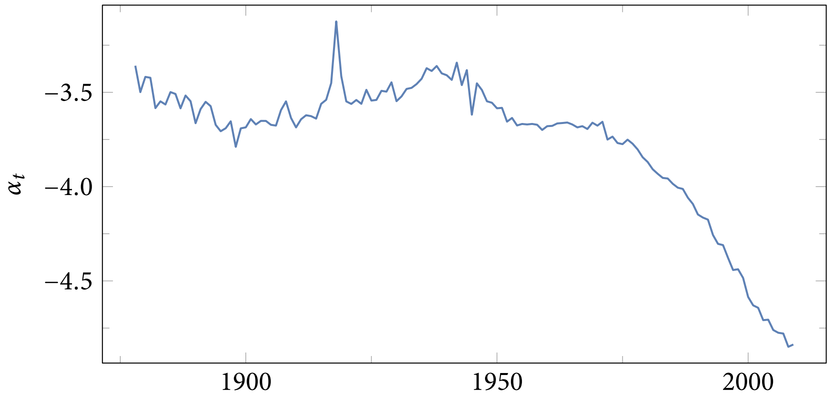


(b) Estimated parameter series β_t.

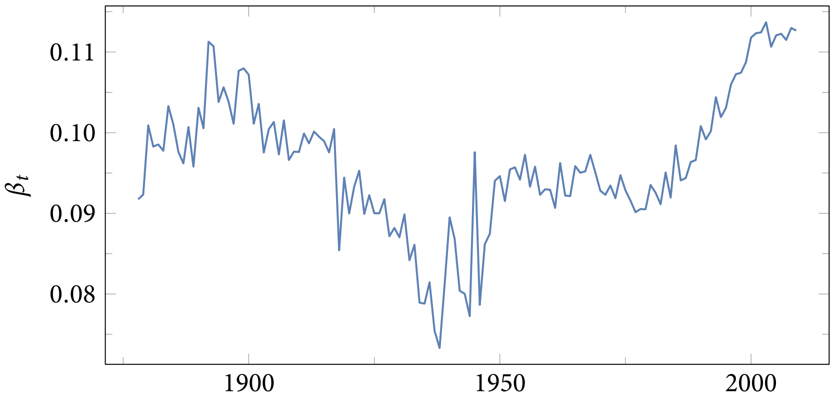


(c) Parametric plot (α_t,β_t).

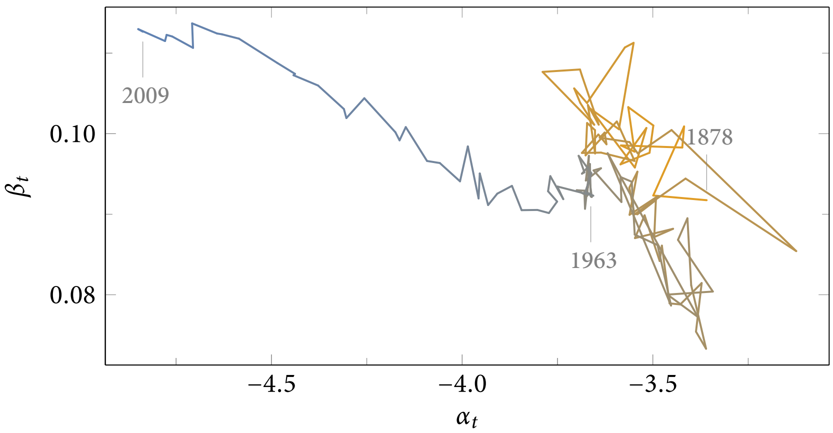
Figure A.24: KAN model estimates for Finnland (females).



(a) Estimated parameter series α_t.



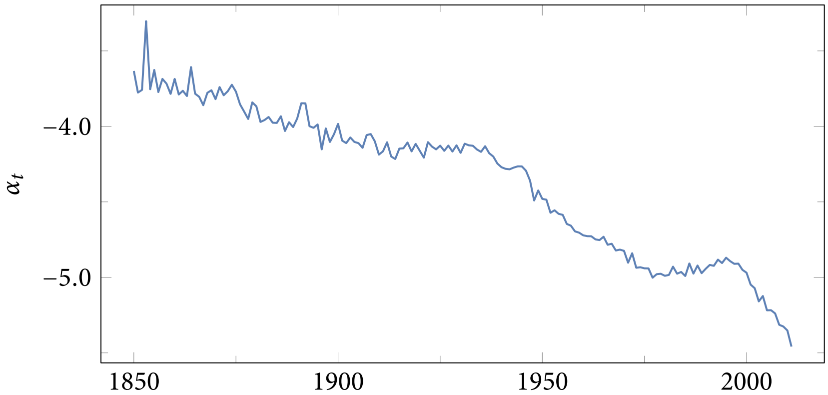
(b) Estimated parameter series β_t.



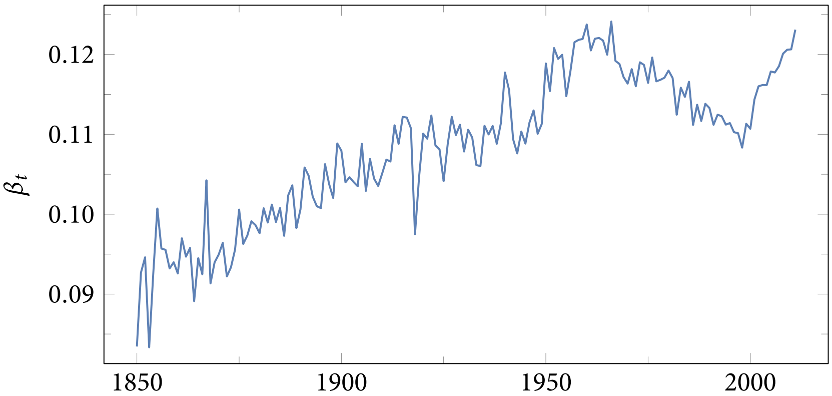
(c) Parametric plot (α_t,β_t).

Figure A.25: KAN model estimates for Finnland (males).

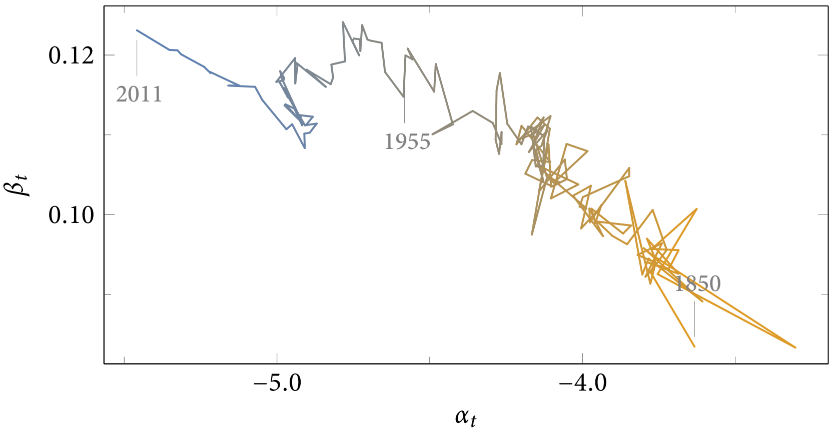
A.3.6 | Denmark



(a) Estimated parameter series α_t.

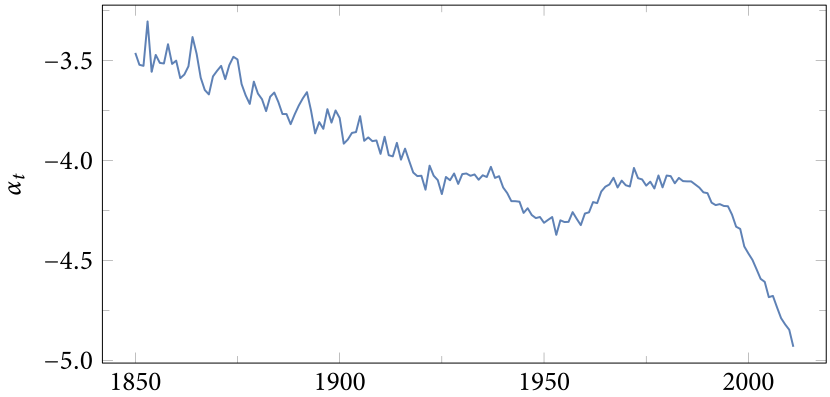


(b) Estimated parameter series β_t.

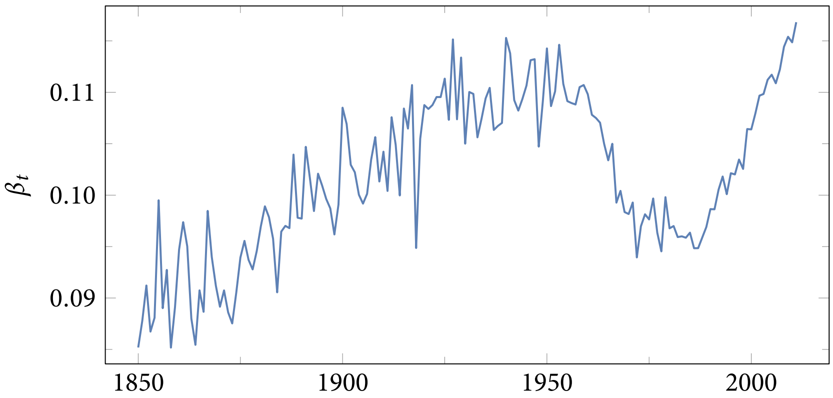


(c) Parametric plot (α_t, β_t).

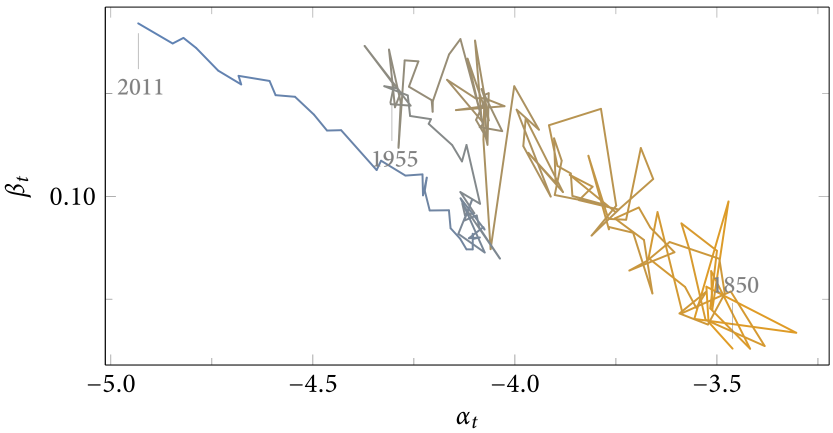
Figure A.26: KAN model estimates for Denmark (females).



(a) Estimated parameter series α_t.



(b) Estimated parameter series β_t.



(c) Parametric plot (α_t,β_t).

Figure A.27: KAN model estimates for Denmark (males).

Appendix B | GAPC Regression Tables

Contents

B.1 | Significance of Individual GAPC Parameters

B.1.1 | APC Model

Table B.1: Regression table of the APC model for Swedish females. Only 143 of 315 parameters (≈ 45%) are significant on the 5% level. p-value significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' . ' 0.1. ' ' 1.

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
α_{60}	−5.219	7.358×10^{-1}	−7.093	1.3×10^{-12}	***
α_{61}	−5.146	7.422×10^{-1}	−6.933	4.1×10^{-12}	***
α_{62}	−5.066	7.487×10^{-1}	−6.766	1.3×10^{-11}	***
α_{63}	−4.971	7.551×10^{-1}	−6.583	4.6×10^{-11}	***
α_{64}	−4.880	7.616×10^{-1}	−6.408	1.5×10^{-10}	***
α_{65}	−4.786	7.680×10^{-1}	−6.232	4.6×10^{-10}	***
α_{66}	−4.700	7.744×10^{-1}	−6.069	1.3×10^{-9}	***
α_{67}	−4.601	7.809×10^{-1}	−5.893	3.8×10^{-9}	***
α_{68}	−4.508	7.873×10^{-1}	−5.726	1.0×10^{-8}	***
α_{69}	−4.412	7.938×10^{-1}	−5.559	2.7×10^{-8}	***
α_{70}	−4.305	8.002×10^{-1}	−5.380	7.4×10^{-8}	***
α_{71}	−4.204	8.067×10^{-1}	−5.212	1.9×10^{-7}	***
α_{72}	−4.104	8.131×10^{-1}	−5.048	4.5×10^{-7}	***
α_{73}	−3.999	8.196×10^{-1}	−4.880	1.1×10^{-6}	***
α_{74}	−3.888	8.260×10^{-1}	−4.707	2.5×10^{-6}	***
α_{75}	−3.793	8.324×10^{-1}	−4.556	5.2×10^{-6}	***
α_{76}	−3.679	8.389×10^{-1}	−4.386	1.2×10^{-5}	***
continued …					

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
α_{77}	-3.571	8.453×10^{-1}	-4.225	2.4×10^{-5}	***
α_{78}	-3.474	8.518×10^{-1}	-4.078	4.5×10^{-5}	***
α_{79}	-3.365	8.582×10^{-1}	-3.921	8.8×10^{-5}	***
α_{80}	-3.263	8.647×10^{-1}	-3.774	1.6×10^{-4}	***
α_{81}	-3.156	8.711×10^{-1}	-3.623	2.9×10^{-4}	***
α_{82}	-3.053	8.776×10^{-1}	-3.478	5.0×10^{-4}	***
α_{83}	-2.950	8.840×10^{-1}	-3.337	8.5×10^{-4}	***
α_{84}	-2.847	8.905×10^{-1}	-3.197	1.4×10^{-3}	**
α_{85}	-2.742	8.969×10^{-1}	-3.057	2.2×10^{-3}	**
α_{86}	-2.640	9.034×10^{-1}	-2.923	3.5×10^{-3}	**
α_{87}	-2.540	9.098×10^{-1}	-2.792	5.2×10^{-3}	**
α_{88}	-2.441	9.163×10^{-1}	-2.664	7.7×10^{-3}	**
α_{89}	-2.341	9.227×10^{-1}	-2.537	1.1×10^{-2}	*
α_{90}	-2.249	9.292×10^{-1}	-2.421	1.5×10^{-2}	*
α_{91}	-2.153	9.356×10^{-1}	-2.301	2.1×10^{-2}	*
α_{92}	-2.056	9.421×10^{-1}	-2.183	2.9×10^{-2}	*
α_{93}	-1.967	9.485×10^{-1}	-2.074	3.8×10^{-2}	*
α_{94}	-1.884	9.550×10^{-1}	-1.972	4.9×10^{-2}	*
α_{95}	-1.795	9.614×10^{-1}	-1.867	6.2×10^{-2}	.
α_{96}	-1.713	9.679×10^{-1}	-1.769	7.7×10^{-2}	.
α_{97}	-1.637	9.743×10^{-1}	-1.680	9.3×10^{-2}	.
α_{98}	-1.570	9.808×10^{-1}	-1.600	1.1×10^{-1}	
α_{99}	-1.509	9.873×10^{-1}	-1.528	1.3×10^{-1}	
α_{100}	-1.421	9.938×10^{-1}	-1.430	1.5×10^{-1}	
α_{101}	-1.343	1.000	-1.343	1.8×10^{-1}	
α_{102}	-1.326	1.007	-1.317	1.9×10^{-1}	
α_{103}	-1.287	1.013	-1.270	2.0×10^{-1}	
α_{104}	-1.247	1.020	-1.222	2.2×10^{-1}	
α_{105}	-1.211	1.027	-1.179	2.4×10^{-1}	
α_{106}	-1.223	1.035	-1.181	2.4×10^{-1}	
κ_{1901}	-1.082×10^{-1}	1.219×10^{-2}	-8.870	$<2.0 \times 10^{-16}$	***
κ_{1902}	-9.004×10^{-2}	1.647×10^{-2}	-5.466	4.6×10^{-8}	***
κ_{1903}	-1.470×10^{-1}	2.195×10^{-2}	-6.697	2.1×10^{-11}	***
κ_{1904}	-7.846×10^{-2}	2.774×10^{-2}	-2.828	4.7×10^{-3}	**
κ_{1905}	-1.067×10^{-1}	3.386×10^{-2}	-3.152	1.6×10^{-3}	**
κ_{1906}	-1.889×10^{-1}	4.012×10^{-2}	-4.709	2.5×10^{-6}	***
κ_{1907}	-1.213×10^{-1}	4.635×10^{-2}	-2.616	8.9×10^{-3}	**
κ_{1908}	-1.047×10^{-1}	5.267×10^{-2}	-1.987	4.7×10^{-2}	*
κ_{1909}	-1.771×10^{-1}	5.905×10^{-2}	-2.999	2.7×10^{-3}	**
κ_{1910}	-1.668×10^{-1}	6.542×10^{-2}	-2.550	1.1×10^{-2}	*
κ_{1911}	-1.917×10^{-1}	7.182×10^{-2}	-2.669	7.6×10^{-3}	**
κ_{1912}	-1.300×10^{-1}	7.821×10^{-2}	-1.663	9.6×10^{-2}	.
κ_{1913}	-1.753×10^{-1}	8.464×10^{-2}	-2.071	3.8×10^{-2}	*
κ_{1914}	-1.658×10^{-1}	9.106×10^{-2}	-1.821	6.9×10^{-2}	.
κ_{1915}	-7.798×10^{-2}	9.747×10^{-2}	-8.000×10^{-1}	4.2×10^{-1}	
continued …					

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
κ_{1916}	-2.021×10^{-1}	1.039×10^{-1}	-1.945	5.2×10^{-2}	.
κ_{1917}	-1.959×10^{-1}	1.104×10^{-1}	-1.775	7.6×10^{-2}	.
κ_{1918}	-1.694×10^{-1}	1.168×10^{-1}	-1.450	1.5×10^{-1}	
κ_{1919}	-1.515×10^{-1}	1.232×10^{-1}	-1.229	2.2×10^{-1}	
κ_{1920}	-2.220×10^{-1}	1.297×10^{-1}	-1.711	8.7×10^{-2}	.
κ_{1921}	-2.328×10^{-1}	1.361×10^{-1}	-1.710	8.7×10^{-2}	.
κ_{1922}	-1.307×10^{-1}	1.426×10^{-1}	-9.160×10^{-1}	3.6×10^{-1}	
κ_{1923}	-2.818×10^{-1}	1.491×10^{-1}	-1.891	5.9×10^{-2}	.
κ_{1924}	-2.359×10^{-1}	1.555×10^{-1}	-1.517	1.3×10^{-1}	
κ_{1925}	-2.574×10^{-1}	1.620×10^{-1}	-1.589	1.1×10^{-1}	
κ_{1926}	-2.383×10^{-1}	1.684×10^{-1}	-1.415	1.6×10^{-1}	
κ_{1927}	-1.523×10^{-1}	1.748×10^{-1}	-8.710×10^{-1}	3.8×10^{-1}	
κ_{1928}	-2.294×10^{-1}	1.813×10^{-1}	-1.265	2.1×10^{-1}	
κ_{1929}	-2.048×10^{-1}	1.878×10^{-1}	-1.091	2.8×10^{-1}	
κ_{1930}	-2.457×10^{-1}	1.942×10^{-1}	-1.265	2.1×10^{-1}	
κ_{1931}	-1.280×10^{-1}	2.007×10^{-1}	-6.380×10^{-1}	5.2×10^{-1}	
κ_{1932}	-2.278×10^{-1}	2.071×10^{-1}	-1.100	2.7×10^{-1}	
κ_{1933}	-2.637×10^{-1}	2.136×10^{-1}	-1.235	2.2×10^{-1}	
κ_{1934}	-2.604×10^{-1}	2.201×10^{-1}	-1.183	2.4×10^{-1}	
κ_{1935}	-1.968×10^{-1}	2.265×10^{-1}	-8.690×10^{-1}	3.8×10^{-1}	
κ_{1936}	-1.840×10^{-1}	2.330×10^{-1}	-7.900×10^{-1}	4.3×10^{-1}	
κ_{1937}	-1.789×10^{-1}	2.394×10^{-1}	-7.470×10^{-1}	4.5×10^{-1}	
κ_{1938}	-2.369×10^{-1}	2.459×10^{-1}	-9.640×10^{-1}	3.4×10^{-1}	
κ_{1939}	-1.939×10^{-1}	2.524×10^{-1}	-7.690×10^{-1}	4.4×10^{-1}	
κ_{1940}	-1.933×10^{-1}	2.588×10^{-1}	-7.470×10^{-1}	4.6×10^{-1}	
κ_{1941}	-2.076×10^{-1}	2.653×10^{-1}	-7.820×10^{-1}	4.3×10^{-1}	
κ_{1942}	-3.843×10^{-1}	2.718×10^{-1}	-1.414	1.6×10^{-1}	
κ_{1943}	-3.687×10^{-1}	2.782×10^{-1}	-1.325	1.9×10^{-1}	
κ_{1944}	-2.881×10^{-1}	2.847×10^{-1}	-1.012	3.1×10^{-1}	
κ_{1945}	-2.831×10^{-1}	2.911×10^{-1}	-9.720×10^{-1}	3.3×10^{-1}	
κ_{1946}	-2.791×10^{-1}	2.976×10^{-1}	-9.380×10^{-1}	3.5×10^{-1}	
κ_{1947}	-2.290×10^{-1}	3.041×10^{-1}	-7.530×10^{-1}	4.5×10^{-1}	
κ_{1948}	-3.206×10^{-1}	3.105×10^{-1}	-1.032	3.0×10^{-1}	
κ_{1949}	-2.963×10^{-1}	3.170×10^{-1}	-9.350×10^{-1}	3.5×10^{-1}	
κ_{1950}	-2.775×10^{-1}	3.234×10^{-1}	-8.580×10^{-1}	3.9×10^{-1}	
κ_{1951}	-2.938×10^{-1}	3.299×10^{-1}	-8.910×10^{-1}	3.7×10^{-1}	
κ_{1952}	-3.278×10^{-1}	3.364×10^{-1}	-9.740×10^{-1}	3.3×10^{-1}	
κ_{1953}	-3.131×10^{-1}	3.428×10^{-1}	-9.130×10^{-1}	3.6×10^{-1}	
κ_{1954}	-3.323×10^{-1}	3.493×10^{-1}	-9.510×10^{-1}	3.4×10^{-1}	
κ_{1955}	-3.732×10^{-1}	3.558×10^{-1}	-1.049	2.9×10^{-1}	
κ_{1956}	-3.644×10^{-1}	3.622×10^{-1}	-1.006	3.1×10^{-1}	
κ_{1957}	-3.317×10^{-1}	3.687×10^{-1}	-9.000×10^{-1}	3.7×10^{-1}	
κ_{1958}	-3.699×10^{-1}	3.752×10^{-1}	-9.860×10^{-1}	3.2×10^{-1}	
κ_{1959}	-3.938×10^{-1}	3.816×10^{-1}	-1.032	3.0×10^{-1}	
κ_{1960}	-3.549×10^{-1}	3.881×10^{-1}	-9.150×10^{-1}	3.6×10^{-1}	

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
κ_{1961}	-3.900×10^{-1}	3.945×10^{-1}	-9.890×10^{-1}	3.2×10^{-1}	
κ_{1962}	-3.674×10^{-1}	4.010×10^{-1}	-9.160×10^{-1}	3.6×10^{-1}	
κ_{1963}	-3.960×10^{-1}	4.075×10^{-1}	-9.720×10^{-1}	3.3×10^{-1}	
κ_{1964}	-4.194×10^{-1}	4.139×10^{-1}	-1.013	3.1×10^{-1}	
κ_{1965}	-4.168×10^{-1}	4.204×10^{-1}	-9.910×10^{-1}	3.2×10^{-1}	
κ_{1966}	-4.277×10^{-1}	4.269×10^{-1}	-1.002	3.2×10^{-1}	
κ_{1967}	-4.307×10^{-1}	4.333×10^{-1}	-9.940×10^{-1}	3.2×10^{-1}	
κ_{1968}	-4.008×10^{-1}	4.398×10^{-1}	-9.110×10^{-1}	3.6×10^{-1}	
κ_{1969}	-4.140×10^{-1}	4.463×10^{-1}	-9.280×10^{-1}	3.5×10^{-1}	
κ_{1970}	-4.840×10^{-1}	4.527×10^{-1}	-1.069	2.9×10^{-1}	
κ_{1971}	-4.883×10^{-1}	4.592×10^{-1}	-1.063	2.9×10^{-1}	
κ_{1972}	-4.787×10^{-1}	4.657×10^{-1}	-1.028	3.0×10^{-1}	
κ_{1973}	-4.796×10^{-1}	4.721×10^{-1}	-1.016	3.1×10^{-1}	
κ_{1974}	-4.900×10^{-1}	4.786×10^{-1}	-1.024	3.1×10^{-1}	
κ_{1975}	-4.767×10^{-1}	4.850×10^{-1}	-9.830×10^{-1}	3.3×10^{-1}	
κ_{1976}	-4.559×10^{-1}	4.915×10^{-1}	-9.270×10^{-1}	3.5×10^{-1}	
κ_{1977}	-5.104×10^{-1}	4.980×10^{-1}	-1.025	3.1×10^{-1}	
κ_{1978}	-4.977×10^{-1}	5.044×10^{-1}	-9.870×10^{-1}	3.2×10^{-1}	
κ_{1979}	-4.975×10^{-1}	5.109×10^{-1}	-9.740×10^{-1}	3.3×10^{-1}	
κ_{1980}	-4.852×10^{-1}	5.174×10^{-1}	-9.380×10^{-1}	3.5×10^{-1}	
κ_{1981}	-4.804×10^{-1}	5.238×10^{-1}	-9.170×10^{-1}	3.6×10^{-1}	
κ_{1982}	-5.115×10^{-1}	5.303×10^{-1}	-9.640×10^{-1}	3.3×10^{-1}	
κ_{1983}	-5.174×10^{-1}	5.368×10^{-1}	-9.640×10^{-1}	3.4×10^{-1}	
κ_{1984}	-5.232×10^{-1}	5.432×10^{-1}	-9.630×10^{-1}	3.4×10^{-1}	
κ_{1985}	-4.843×10^{-1}	5.497×10^{-1}	-8.810×10^{-1}	3.8×10^{-1}	
κ_{1986}	-4.955×10^{-1}	5.562×10^{-1}	-8.910×10^{-1}	3.7×10^{-1}	
κ_{1987}	-4.994×10^{-1}	5.626×10^{-1}	-8.880×10^{-1}	3.7×10^{-1}	
κ_{1988}	-4.508×10^{-1}	5.691×10^{-1}	-7.920×10^{-1}	4.3×10^{-1}	
κ_{1989}	-5.002×10^{-1}	5.756×10^{-1}	-8.690×10^{-1}	3.8×10^{-1}	
κ_{1990}	-4.629×10^{-1}	5.820×10^{-1}	-7.950×10^{-1}	4.3×10^{-1}	
κ_{1991}	-4.669×10^{-1}	5.885×10^{-1}	-7.930×10^{-1}	4.3×10^{-1}	
κ_{1992}	-4.602×10^{-1}	5.950×10^{-1}	-7.740×10^{-1}	4.4×10^{-1}	
κ_{1993}	-4.194×10^{-1}	6.014×10^{-1}	-6.970×10^{-1}	4.9×10^{-1}	
κ_{1994}	-4.825×10^{-1}	6.079×10^{-1}	-7.940×10^{-1}	4.3×10^{-1}	
κ_{1995}	-4.554×10^{-1}	6.144×10^{-1}	-7.410×10^{-1}	4.6×10^{-1}	
κ_{1996}	-4.382×10^{-1}	6.208×10^{-1}	-7.060×10^{-1}	4.8×10^{-1}	
κ_{1997}	-4.408×10^{-1}	6.273×10^{-1}	-7.030×10^{-1}	4.8×10^{-1}	
κ_{1998}	-4.369×10^{-1}	6.337×10^{-1}	-6.890×10^{-1}	4.9×10^{-1}	
κ_{1999}	-3.962×10^{-1}	6.402×10^{-1}	-6.190×10^{-1}	5.4×10^{-1}	
κ_{2000}	-3.976×10^{-1}	6.467×10^{-1}	-6.150×10^{-1}	5.4×10^{-1}	
κ_{2001}	-3.764×10^{-1}	6.531×10^{-1}	-5.760×10^{-1}	5.6×10^{-1}	
κ_{2002}	-3.421×10^{-1}	6.596×10^{-1}	-5.190×10^{-1}	6.0×10^{-1}	
κ_{2003}	-3.632×10^{-1}	6.661×10^{-1}	-5.450×10^{-1}	5.9×10^{-1}	
κ_{2004}	-3.759×10^{-1}	6.725×10^{-1}	-5.590×10^{-1}	5.8×10^{-1}	
κ_{2005}	-3.603×10^{-1}	6.790×10^{-1}	-5.310×10^{-1}	6.0×10^{-1}	

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
κ_{2006}	-3.465×10^{-1}	6.855×10^{-1}	-5.060×10^{-1}	6.1×10^{-1}	
κ_{2007}	-3.150×10^{-1}	6.919×10^{-1}	-4.550×10^{-1}	6.5×10^{-1}	
κ_{2008}	-3.101×10^{-1}	6.984×10^{-1}	-4.440×10^{-1}	6.6×10^{-1}	
κ_{2009}	-3.203×10^{-1}	7.049×10^{-1}	-4.540×10^{-1}	6.5×10^{-1}	
κ_{2010}	-3.013×10^{-1}	7.113×10^{-1}	-4.240×10^{-1}	6.7×10^{-1}	
κ_{2011}	-2.962×10^{-1}	7.178×10^{-1}	-4.130×10^{-1}	6.8×10^{-1}	
κ_{2012}	-2.540×10^{-1}	7.243×10^{-1}	-3.510×10^{-1}	7.3×10^{-1}	
κ_{2013}	-2.616×10^{-1}	7.307×10^{-1}	-3.580×10^{-1}	7.2×10^{-1}	
κ_{2014}	-2.769×10^{-1}	7.372×10^{-1}	-3.760×10^{-1}	7.1×10^{-1}	
γ_{1800}	1.001	1.112	9.000×10^{-1}	3.7×10^{-1}	
γ_{1801}	1.159	1.042	1.113	2.7×10^{-1}	
γ_{1802}	1.501	1.012	1.483	1.4×10^{-1}	
γ_{1803}	1.093	9.893×10^{-1}	1.105	2.7×10^{-1}	
γ_{1804}	1.034	9.763×10^{-1}	1.060	2.9×10^{-1}	
γ_{1805}	1.098	9.680×10^{-1}	1.134	2.6×10^{-1}	
γ_{1806}	1.042	9.593×10^{-1}	1.086	2.8×10^{-1}	
γ_{1807}	9.671×10^{-1}	9.511×10^{-1}	1.017	3.1×10^{-1}	
γ_{1808}	1.037	9.436×10^{-1}	1.099	2.7×10^{-1}	
γ_{1809}	1.089	9.369×10^{-1}	1.163	2.4×10^{-1}	
γ_{1810}	1.071	9.301×10^{-1}	1.151	2.5×10^{-1}	
γ_{1811}	1.101	9.233×10^{-1}	1.193	2.3×10^{-1}	
γ_{1812}	1.115	9.167×10^{-1}	1.216	2.2×10^{-1}	
γ_{1813}	1.150	9.102×10^{-1}	1.264	2.1×10^{-1}	
γ_{1814}	1.137	9.037×10^{-1}	1.259	2.1×10^{-1}	
γ_{1815}	1.168	8.971×10^{-1}	1.302	1.9×10^{-1}	
γ_{1816}	1.196	8.906×10^{-1}	1.343	1.8×10^{-1}	
γ_{1817}	1.196	8.842×10^{-1}	1.353	1.8×10^{-1}	
γ_{1818}	1.170	8.777×10^{-1}	1.332	1.8×10^{-1}	
γ_{1819}	1.209	8.712×10^{-1}	1.388	1.7×10^{-1}	
γ_{1820}	1.176	8.648×10^{-1}	1.360	1.7×10^{-1}	
γ_{1821}	1.202	8.583×10^{-1}	1.401	1.6×10^{-1}	
γ_{1822}	1.206	8.519×10^{-1}	1.416	1.6×10^{-1}	
γ_{1823}	1.183	8.454×10^{-1}	1.399	1.6×10^{-1}	
γ_{1824}	1.207	8.390×10^{-1}	1.439	1.5×10^{-1}	
γ_{1825}	1.225	8.325×10^{-1}	1.472	1.4×10^{-1}	
γ_{1826}	1.228	8.261×10^{-1}	1.487	1.4×10^{-1}	
γ_{1827}	1.226	8.196×10^{-1}	1.496	1.3×10^{-1}	
γ_{1828}	1.195	8.132×10^{-1}	1.470	1.4×10^{-1}	
γ_{1829}	1.262	8.067×10^{-1}	1.565	1.2×10^{-1}	
γ_{1830}	1.231	8.003×10^{-1}	1.539	1.2×10^{-1}	
γ_{1831}	1.229	7.938×10^{-1}	1.549	1.2×10^{-1}	
γ_{1832}	1.203	7.874×10^{-1}	1.528	1.3×10^{-1}	
γ_{1833}	1.219	7.809×10^{-1}	1.561	1.2×10^{-1}	
γ_{1834}	1.261	7.745×10^{-1}	1.628	1.0×10^{-1}	
γ_{1835}	1.247	7.680×10^{-1}	1.623	1.0×10^{-1}	

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\|z\|)$	Signif. code
γ_{1836}	1.227	7.616×10^{-1}	1.612	1.1×10^{-1}	
γ_{1837}	1.258	7.552×10^{-1}	1.666	9.6×10^{-2}	.
γ_{1838}	1.257	7.487×10^{-1}	1.679	9.3×10^{-2}	.
γ_{1839}	1.211	7.423×10^{-1}	1.632	1.0×10^{-1}	
γ_{1840}	1.273	7.358×10^{-1}	1.730	8.4×10^{-2}	.
γ_{1841}	1.269	7.294×10^{-1}	1.739	8.2×10^{-2}	.
γ_{1842}	1.280	7.230×10^{-1}	1.770	7.7×10^{-2}	.
γ_{1843}	1.258	7.165×10^{-1}	1.756	7.9×10^{-2}	.
γ_{1844}	1.282	7.101×10^{-1}	1.805	7.1×10^{-2}	.
γ_{1845}	1.288	7.036×10^{-1}	1.831	6.7×10^{-2}	.
γ_{1846}	1.283	6.972×10^{-1}	1.841	6.6×10^{-2}	.
γ_{1847}	1.286	6.908×10^{-1}	1.861	6.3×10^{-2}	.
γ_{1848}	1.282	6.843×10^{-1}	1.873	6.1×10^{-2}	.
γ_{1849}	1.288	6.779×10^{-1}	1.900	5.7×10^{-2}	.
γ_{1850}	1.316	6.715×10^{-1}	1.960	5.0×10^{-2}	.
γ_{1851}	1.275	6.650×10^{-1}	1.917	5.5×10^{-2}	.
γ_{1852}	1.311	6.586×10^{-1}	1.991	4.6×10^{-2}	*
γ_{1853}	1.263	6.522×10^{-1}	1.936	5.3×10^{-2}	.
γ_{1854}	1.301	6.457×10^{-1}	2.015	4.4×10^{-2}	*
γ_{1855}	1.307	6.393×10^{-1}	2.045	4.1×10^{-2}	*
γ_{1856}	1.317	6.329×10^{-1}	2.081	3.7×10^{-2}	*
γ_{1857}	1.295	6.264×10^{-1}	2.067	3.9×10^{-2}	*
γ_{1858}	1.320	6.200×10^{-1}	2.129	3.3×10^{-2}	*
γ_{1859}	1.314	6.136×10^{-1}	2.141	3.2×10^{-2}	*
γ_{1860}	1.336	6.071×10^{-1}	2.200	2.8×10^{-2}	*
γ_{1861}	1.317	6.007×10^{-1}	2.193	2.8×10^{-2}	*
γ_{1862}	1.304	5.943×10^{-1}	2.194	2.8×10^{-2}	*
γ_{1863}	1.325	5.879×10^{-1}	2.255	2.4×10^{-2}	*
γ_{1864}	1.321	5.814×10^{-1}	2.272	2.3×10^{-2}	*
γ_{1865}	1.296	5.750×10^{-1}	2.254	2.4×10^{-2}	*
γ_{1866}	1.322	5.686×10^{-1}	2.325	2.0×10^{-2}	*
γ_{1867}	1.319	5.622×10^{-1}	2.346	1.9×10^{-2}	*
γ_{1868}	1.322	5.558×10^{-1}	2.378	1.7×10^{-2}	*
γ_{1869}	1.298	5.493×10^{-1}	2.364	1.8×10^{-2}	*
γ_{1870}	1.316	5.429×10^{-1}	2.423	1.5×10^{-2}	*
γ_{1871}	1.307	5.365×10^{-1}	2.436	1.5×10^{-2}	*
γ_{1872}	1.309	5.301×10^{-1}	2.469	1.4×10^{-2}	*
γ_{1873}	1.301	5.237×10^{-1}	2.485	1.3×10^{-2}	*
γ_{1874}	1.301	5.172×10^{-1}	2.515	1.2×10^{-2}	*
γ_{1875}	1.280	5.108×10^{-1}	2.506	1.2×10^{-2}	*
γ_{1876}	1.290	5.044×10^{-1}	2.558	1.1×10^{-2}	*
γ_{1877}	1.282	4.980×10^{-1}	2.575	1.0×10^{-2}	*
γ_{1878}	1.276	4.916×10^{-1}	2.596	9.4×10^{-3}	**
γ_{1879}	1.264	4.852×10^{-1}	2.605	9.2×10^{-3}	**
γ_{1880}	1.263	4.788×10^{-1}	2.638	8.3×10^{-3}	**
continued …					

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
γ_{1881}	1.248	4.724×10^{-1}	2.641	8.3×10^{-3}	**
γ_{1882}	1.259	4.660×10^{-1}	2.702	6.9×10^{-3}	**
γ_{1883}	1.227	4.596×10^{-1}	2.669	7.6×10^{-3}	**
γ_{1884}	1.221	4.532×10^{-1}	2.694	7.1×10^{-3}	**
γ_{1885}	1.200	4.468×10^{-1}	2.686	7.2×10^{-3}	**
γ_{1886}	1.209	4.404×10^{-1}	2.745	6.1×10^{-3}	**
γ_{1887}	1.184	4.340×10^{-1}	2.728	6.4×10^{-3}	**
γ_{1888}	1.203	4.276×10^{-1}	2.813	4.9×10^{-3}	**
γ_{1889}	1.167	4.212×10^{-1}	2.771	5.6×10^{-3}	**
γ_{1890}	1.159	4.148×10^{-1}	2.793	5.2×10^{-3}	**
γ_{1891}	1.150	4.084×10^{-1}	2.817	4.9×10^{-3}	**
γ_{1892}	1.129	4.021×10^{-1}	2.809	5.0×10^{-3}	**
γ_{1893}	1.109	3.957×10^{-1}	2.802	5.1×10^{-3}	**
γ_{1894}	1.108	3.893×10^{-1}	2.847	4.4×10^{-3}	**
γ_{1895}	1.110	3.829×10^{-1}	2.898	3.8×10^{-3}	**
γ_{1896}	1.085	3.766×10^{-1}	2.882	4.0×10^{-3}	**
γ_{1897}	1.068	3.702×10^{-1}	2.884	3.9×10^{-3}	**
γ_{1898}	1.073	3.638×10^{-1}	2.948	3.2×10^{-3}	**
γ_{1899}	1.048	3.575×10^{-1}	2.932	3.4×10^{-3}	**
γ_{1900}	1.029	3.511×10^{-1}	2.932	3.4×10^{-3}	**
γ_{1901}	1.010	3.448×10^{-1}	2.930	3.4×10^{-3}	**
γ_{1902}	9.928×10^{-1}	3.384×10^{-1}	2.933	3.4×10^{-3}	**
γ_{1903}	9.708×10^{-1}	3.321×10^{-1}	2.923	3.5×10^{-3}	**
γ_{1904}	9.595×10^{-1}	3.258×10^{-1}	2.945	3.2×10^{-3}	**
γ_{1905}	9.416×10^{-1}	3.194×10^{-1}	2.948	3.2×10^{-3}	**
γ_{1906}	9.262×10^{-1}	3.131×10^{-1}	2.958	3.1×10^{-3}	**
γ_{1907}	9.086×10^{-1}	3.068×10^{-1}	2.962	3.1×10^{-3}	**
γ_{1908}	9.008×10^{-1}	3.005×10^{-1}	2.998	2.7×10^{-3}	**
γ_{1909}	8.706×10^{-1}	2.942×10^{-1}	2.960	3.1×10^{-3}	**
γ_{1910}	8.639×10^{-1}	2.879×10^{-1}	3.001	2.7×10^{-3}	**
γ_{1911}	8.403×10^{-1}	2.816×10^{-1}	2.984	2.8×10^{-3}	**
γ_{1912}	8.257×10^{-1}	2.753×10^{-1}	2.999	2.7×10^{-3}	**
γ_{1913}	7.949×10^{-1}	2.690×10^{-1}	2.955	3.1×10^{-3}	**
γ_{1914}	7.863×10^{-1}	2.628×10^{-1}	2.992	2.8×10^{-3}	**
γ_{1915}	7.775×10^{-1}	2.565×10^{-1}	3.031	2.4×10^{-3}	**
γ_{1916}	7.427×10^{-1}	2.503×10^{-1}	2.967	3.0×10^{-3}	**
γ_{1917}	7.208×10^{-1}	2.441×10^{-1}	2.954	3.1×10^{-3}	**
γ_{1918}	7.018×10^{-1}	2.378×10^{-1}	2.951	3.2×10^{-3}	**
γ_{1919}	6.506×10^{-1}	2.316×10^{-1}	2.809	5.0×10^{-3}	**
γ_{1920}	6.858×10^{-1}	2.254×10^{-1}	3.042	2.4×10^{-3}	**
γ_{1921}	6.113×10^{-1}	2.193×10^{-1}	2.788	5.3×10^{-3}	**
γ_{1922}	6.183×10^{-1}	2.131×10^{-1}	2.901	3.7×10^{-3}	**
γ_{1923}	5.698×10^{-1}	2.070×10^{-1}	2.753	5.9×10^{-3}	**
γ_{1924}	5.476×10^{-1}	2.009×10^{-1}	2.726	6.4×10^{-3}	**
γ_{1925}	5.166×10^{-1}	1.948×10^{-1}	2.652	8.0×10^{-3}	**

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
γ_{1926}	4.699×10^{-1}	1.887×10^{-1}	2.490	1.3×10^{-2}	*
γ_{1927}	4.696×10^{-1}	1.827×10^{-1}	2.570	1.0×10^{-2}	*
γ_{1928}	4.145×10^{-1}	1.767×10^{-1}	2.346	1.9×10^{-2}	*
γ_{1929}	4.109×10^{-1}	1.707×10^{-1}	2.407	1.6×10^{-2}	*
γ_{1930}	3.797×10^{-1}	1.648×10^{-1}	2.304	2.1×10^{-2}	*
γ_{1931}	3.757×10^{-1}	1.589×10^{-1}	2.364	1.8×10^{-2}	*
γ_{1932}	3.490×10^{-1}	1.531×10^{-1}	2.280	2.3×10^{-2}	*
γ_{1933}	3.223×10^{-1}	1.473×10^{-1}	2.188	2.9×10^{-2}	*
γ_{1934}	2.975×10^{-1}	1.416×10^{-1}	2.100	3.6×10^{-2}	*
γ_{1935}	2.930×10^{-1}	1.360×10^{-1}	2.155	3.1×10^{-2}	*
γ_{1936}	2.915×10^{-1}	1.304×10^{-1}	2.235	2.5×10^{-2}	*
γ_{1937}	2.533×10^{-1}	1.250×10^{-1}	2.027	4.3×10^{-2}	*
γ_{1938}	2.411×10^{-1}	1.196×10^{-1}	2.017	4.4×10^{-2}	*
γ_{1939}	2.308×10^{-1}	1.143×10^{-1}	2.018	4.4×10^{-2}	*
γ_{1940}	2.530×10^{-1}	1.092×10^{-1}	2.317	2.1×10^{-2}	*
γ_{1941}	2.114×10^{-1}	1.043×10^{-1}	2.027	4.3×10^{-2}	*
γ_{1942}	2.526×10^{-1}	9.944×10^{-2}	2.541	1.1×10^{-2}	*
γ_{1943}	2.052×10^{-1}	9.484×10^{-2}	2.163	3.1×10^{-2}	*
γ_{1944}	2.278×10^{-1}	9.045×10^{-2}	2.519	1.2×10^{-2}	*
γ_{1945}	2.062×10^{-1}	8.641×10^{-2}	2.386	1.7×10^{-2}	*
γ_{1946}	1.895×10^{-1}	8.275×10^{-2}	2.290	2.2×10^{-2}	*
γ_{1947}	1.952×10^{-1}	7.949×10^{-2}	2.455	1.4×10^{-2}	*
γ_{1948}	1.540×10^{-1}	7.684×10^{-2}	2.004	4.5×10^{-2}	*
γ_{1949}	1.415×10^{-1}	7.481×10^{-2}	1.891	5.9×10^{-2}	.
γ_{1950}	1.256×10^{-1}	7.364×10^{-2}	1.705	8.8×10^{-2}	.
γ_{1951}	1.126×10^{-1}	7.357×10^{-2}	1.531	1.3×10^{-1}	
γ_{1952}	2.086×10^{-2}	7.555×10^{-2}	2.760×10^{-1}	7.8×10^{-1}	
γ_{1953}	1.922×10^{-2}	7.971×10^{-2}	2.410×10^{-1}	8.1×10^{-1}	

B.1.2 | PLAT Model

Table B.2: Regression table of the PLAT:2 model for Swedish females. Only 26 of 428 parameters (≈ 6%) are significant on the 5% level. p-value significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' . ' 0.1 ' ' 1.

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
α_{60}	−5.652	8.203×10^{-1}	−6.890	5.6×10^{-12}	***
α_{61}	−5.571	7.936×10^{-1}	−7.020	2.2×10^{-12}	***
α_{62}	−5.484	7.745×10^{-1}	−7.082	1.4×10^{-12}	***
α_{63}	−5.382	7.633×10^{-1}	−7.051	1.8×10^{-12}	***
α_{64}	−5.284	7.602×10^{-1}	−6.951	3.6×10^{-12}	***
α_{65}	−5.183	7.649×10^{-1}	−6.776	1.2×10^{-11}	***
α_{66}	−5.090	7.769×10^{-1}	−6.552	5.7×10^{-11}	***
α_{67}	−4.984	7.956×10^{-1}	−6.264	3.8×10^{-10}	***
α_{68}	−4.882	8.202×10^{-1}	−5.952	2.6×10^{-9}	***
α_{69}	−4.779	8.501×10^{-1}	−5.621	1.9×10^{-8}	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\|z\|)$	Signif. code
α_{70}	-4.664	8.843×10^{-1}	-5.273	1.3×10^{-7}	***
α_{71}	-4.554	9.223×10^{-1}	-4.937	7.9×10^{-7}	***
α_{72}	-4.445	9.632×10^{-1}	-4.615	3.9×10^{-6}	***
α_{73}	-4.331	1.007	-4.303	1.7×10^{-5}	***
α_{74}	-4.210	1.052	-4.002	6.3×10^{-5}	***
α_{75}	-4.105	1.099	-3.736	1.9×10^{-4}	***
α_{76}	-3.981	1.147	-3.472	5.2×10^{-4}	***
α_{77}	-3.863	1.196	-3.231	1.2×10^{-3}	**
α_{78}	-3.754	1.245	-3.016	2.6×10^{-3}	**
α_{79}	-3.634	1.295	-2.807	5.0×10^{-3}	**
α_{80}	-3.521	1.345	-2.619	8.8×10^{-3}	**
α_{81}	-3.403	1.395	-2.440	1.5×10^{-2}	*
α_{82}	-3.289	1.444	-2.277	2.3×10^{-2}	*
α_{83}	-3.175	1.494	-2.125	3.4×10^{-2}	*
α_{84}	-3.061	1.543	-1.984	4.7×10^{-2}	*
α_{85}	-2.946	1.592	-1.850	6.4×10^{-2}	.
α_{86}	-2.835	1.641	-1.728	8.4×10^{-2}	.
α_{87}	-2.726	1.689	-1.614	1.1×10^{-1}	
α_{88}	-2.618	1.736	-1.508	1.3×10^{-1}	
α_{89}	-2.512	1.783	-1.409	1.6×10^{-1}	
α_{90}	-2.414	1.829	-1.320	1.9×10^{-1}	
α_{91}	-2.313	1.874	-1.234	2.2×10^{-1}	
α_{92}	-2.213	1.919	-1.153	2.5×10^{-1}	
α_{93}	-2.123	1.963	-1.081	2.8×10^{-1}	
α_{94}	-2.038	2.006	-1.016	3.1×10^{-1}	
α_{95}	-1.951	2.049	-9.520×10^{-1}	3.4×10^{-1}	
α_{96}	-1.872	2.090	-8.950×10^{-1}	3.7×10^{-1}	
α_{97}	-1.801	2.131	-8.450×10^{-1}	4.0×10^{-1}	
α_{98}	-1.739	2.171	-8.010×10^{-1}	4.2×10^{-1}	
α_{99}	-1.685	2.210	-7.620×10^{-1}	4.5×10^{-1}	
α_{100}	-1.605	2.248	-7.140×10^{-1}	4.8×10^{-1}	
α_{101}	-1.538	2.285	-6.730×10^{-1}	5.0×10^{-1}	
α_{102}	-1.530	2.322	-6.590×10^{-1}	5.1×10^{-1}	
α_{103}	-1.507	2.357	-6.390×10^{-1}	5.2×10^{-1}	
α_{104}	-1.486	2.392	-6.210×10^{-1}	5.3×10^{-1}	
α_{105}	-1.475	2.426	-6.080×10^{-1}	5.4×10^{-1}	
α_{106}	-1.512	2.459	-6.150×10^{-1}	5.4×10^{-1}	
$\kappa^{(1)}_{1901}$	-1.385×10^{-1}	6.360×10^{-2}	-2.178	2.9×10^{-2}	*
$\kappa^{(1)}_{1902}$	-1.543×10^{-1}	1.237×10^{-1}	-1.248	2.1×10^{-1}	
$\kappa^{(1)}_{1903}$	-2.170×10^{-1}	1.833×10^{-1}	-1.184	2.4×10^{-1}	
$\kappa^{(1)}_{1904}$	-1.641×10^{-1}	2.420×10^{-1}	-6.780×10^{-1}	5.0×10^{-1}	
$\kappa^{(1)}_{1905}$	-2.327×10^{-1}	2.998×10^{-1}	-7.760×10^{-1}	4.4×10^{-1}	
$\kappa^{(1)}_{1906}$	-3.308×10^{-1}	3.565×10^{-1}	-9.280×10^{-1}	3.5×10^{-1}	
$\kappa^{(1)}_{1907}$	-2.506×10^{-1}	4.122×10^{-1}	-6.080×10^{-1}	5.4×10^{-1}	
$\kappa^{(1)}_{1908}$	-2.527×10^{-1}	4.669×10^{-1}	-5.410×10^{-1}	5.9×10^{-1}	

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\|z\|)$	Signif. code
$\kappa^{(1)}_{1909}$	-3.659×10^{-1}	5.206×10^{-1}	-7.030×10^{-1}	4.8×10^{-1}	
$\kappa^{(1)}_{1910}$	-3.590×10^{-1}	5.732×10^{-1}	-6.260×10^{-1}	5.3×10^{-1}	
$\kappa^{(1)}_{1911}$	-4.013×10^{-1}	6.248×10^{-1}	-6.420×10^{-1}	5.2×10^{-1}	
$\kappa^{(1)}_{1912}$	-3.602×10^{-1}	6.753×10^{-1}	-5.330×10^{-1}	5.9×10^{-1}	
$\kappa^{(1)}_{1913}$	-4.171×10^{-1}	7.249×10^{-1}	-5.750×10^{-1}	5.7×10^{-1}	
$\kappa^{(1)}_{1914}$	-4.189×10^{-1}	7.734×10^{-1}	-5.420×10^{-1}	5.9×10^{-1}	
$\kappa^{(1)}_{1915}$	-3.381×10^{-1}	8.208×10^{-1}	-4.120×10^{-1}	6.8×10^{-1}	
$\kappa^{(1)}_{1916}$	-5.005×10^{-1}	8.673×10^{-1}	-5.770×10^{-1}	5.6×10^{-1}	
$\kappa^{(1)}_{1917}$	-4.966×10^{-1}	9.127×10^{-1}	-5.440×10^{-1}	5.9×10^{-1}	
$\kappa^{(1)}_{1918}$	-5.263×10^{-1}	9.571×10^{-1}	-5.500×10^{-1}	5.8×10^{-1}	
$\kappa^{(1)}_{1919}$	-4.844×10^{-1}	1.000	-4.840×10^{-1}	6.3×10^{-1}	
$\kappa^{(1)}_{1920}$	-5.615×10^{-1}	1.043	-5.380×10^{-1}	5.9×10^{-1}	
$\kappa^{(1)}_{1921}$	-5.667×10^{-1}	1.084	-5.230×10^{-1}	6.0×10^{-1}	
$\kappa^{(1)}_{1922}$	-4.844×10^{-1}	1.124	-4.310×10^{-1}	6.7×10^{-1}	
$\kappa^{(1)}_{1923}$	-6.593×10^{-1}	1.163	-5.670×10^{-1}	5.7×10^{-1}	
$\kappa^{(1)}_{1924}$	-6.199×10^{-1}	1.202	-5.160×10^{-1}	6.1×10^{-1}	
$\kappa^{(1)}_{1925}$	-6.507×10^{-1}	1.239	-5.250×10^{-1}	6.0×10^{-1}	
$\kappa^{(1)}_{1926}$	-6.485×10^{-1}	1.275	-5.090×10^{-1}	6.1×10^{-1}	
$\kappa^{(1)}_{1927}$	-5.476×10^{-1}	1.310	-4.180×10^{-1}	6.8×10^{-1}	
$\kappa^{(1)}_{1928}$	-6.579×10^{-1}	1.344	-4.900×10^{-1}	6.2×10^{-1}	
$\kappa^{(1)}_{1929}$	-6.435×10^{-1}	1.377	-4.670×10^{-1}	6.4×10^{-1}	
$\kappa^{(1)}_{1930}$	-6.967×10^{-1}	1.409	-4.940×10^{-1}	6.2×10^{-1}	
$\kappa^{(1)}_{1931}$	-5.584×10^{-1}	1.440	-3.880×10^{-1}	7.0×10^{-1}	
$\kappa^{(1)}_{1932}$	-6.851×10^{-1}	1.470	-4.660×10^{-1}	6.4×10^{-1}	
$\kappa^{(1)}_{1933}$	-7.330×10^{-1}	1.499	-4.890×10^{-1}	6.2×10^{-1}	
$\kappa^{(1)}_{1934}$	-7.321×10^{-1}	1.527	-4.790×10^{-1}	6.3×10^{-1}	
$\kappa^{(1)}_{1935}$	-6.627×10^{-1}	1.554	-4.270×10^{-1}	6.7×10^{-1}	
$\kappa^{(1)}_{1936}$	-6.598×10^{-1}	1.580	-4.180×10^{-1}	6.8×10^{-1}	
$\kappa^{(1)}_{1937}$	-6.680×10^{-1}	1.604	-4.160×10^{-1}	6.8×10^{-1}	
$\kappa^{(1)}_{1938}$	-7.331×10^{-1}	1.628	-4.500×10^{-1}	6.5×10^{-1}	
$\kappa^{(1)}_{1939}$	-6.683×10^{-1}	1.651	-4.050×10^{-1}	6.9×10^{-1}	
$\kappa^{(1)}_{1940}$	-6.770×10^{-1}	1.673	-4.050×10^{-1}	6.9×10^{-1}	
$\kappa^{(1)}_{1941}$	-6.862×10^{-1}	1.693	-4.050×10^{-1}	6.9×10^{-1}	
$\kappa^{(1)}_{1942}$	-8.950×10^{-1}	1.713	-5.220×10^{-1}	6.0×10^{-1}	
$\kappa^{(1)}_{1943}$	-8.843×10^{-1}	1.732	-5.110×10^{-1}	6.1×10^{-1}	
$\kappa^{(1)}_{1944}$	-7.900×10^{-1}	1.749	-4.520×10^{-1}	6.5×10^{-1}	
$\kappa^{(1)}_{1945}$	-7.952×10^{-1}	1.766	-4.500×10^{-1}	6.5×10^{-1}	
$\kappa^{(1)}_{1946}$	-7.783×10^{-1}	1.782	-4.370×10^{-1}	6.6×10^{-1}	
$\kappa^{(1)}_{1947}$	-7.168×10^{-1}	1.796	-3.990×10^{-1}	6.9×10^{-1}	
$\kappa^{(1)}_{1948}$	-8.363×10^{-1}	1.810	-4.620×10^{-1}	6.4×10^{-1}	
$\kappa^{(1)}_{1949}$	-7.970×10^{-1}	1.822	-4.370×10^{-1}	6.6×10^{-1}	
$\kappa^{(1)}_{1950}$	-7.660×10^{-1}	1.834	-4.180×10^{-1}	6.8×10^{-1}	
$\kappa^{(1)}_{1951}$	-7.690×10^{-1}	1.844	-4.170×10^{-1}	6.8×10^{-1}	
$\kappa^{(1)}_{1952}$	-8.217×10^{-1}	1.854	-4.430×10^{-1}	6.6×10^{-1}	

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\|z\|)$	Signif. code
$\kappa^{(1)}_{1953}$	-8.040×10^{-1}	1.862	-4.320×10^{-1}	6.7×10^{-1}	
$\kappa^{(1)}_{1954}$	-8.096×10^{-1}	1.870	-4.330×10^{-1}	6.7×10^{-1}	
$\kappa^{(1)}_{1955}$	-8.548×10^{-1}	1.876	-4.560×10^{-1}	6.5×10^{-1}	
$\kappa^{(1)}_{1956}$	-8.376×10^{-1}	1.882	-4.450×10^{-1}	6.6×10^{-1}	
$\kappa^{(1)}_{1957}$	-7.916×10^{-1}	1.886	-4.200×10^{-1}	6.7×10^{-1}	
$\kappa^{(1)}_{1958}$	-8.224×10^{-1}	1.890	-4.350×10^{-1}	6.6×10^{-1}	
$\kappa^{(1)}_{1959}$	-8.379×10^{-1}	1.892	-4.430×10^{-1}	6.6×10^{-1}	
$\kappa^{(1)}_{1960}$	-7.849×10^{-1}	1.894	-4.140×10^{-1}	6.8×10^{-1}	
$\kappa^{(1)}_{1961}$	-8.108×10^{-1}	1.894	-4.280×10^{-1}	6.7×10^{-1}	
$\kappa^{(1)}_{1962}$	-7.708×10^{-1}	1.894	-4.070×10^{-1}	6.8×10^{-1}	
$\kappa^{(1)}_{1963}$	-7.952×10^{-1}	1.892	-4.200×10^{-1}	6.7×10^{-1}	
$\kappa^{(1)}_{1964}$	-8.215×10^{-1}	1.890	-4.350×10^{-1}	6.6×10^{-1}	
$\kappa^{(1)}_{1965}$	-7.950×10^{-1}	1.886	-4.210×10^{-1}	6.7×10^{-1}	
$\kappa^{(1)}_{1966}$	-7.914×10^{-1}	1.882	-4.210×10^{-1}	6.7×10^{-1}	
$\kappa^{(1)}_{1967}$	-7.819×10^{-1}	1.876	-4.170×10^{-1}	6.8×10^{-1}	
$\kappa^{(1)}_{1968}$	-7.337×10^{-1}	1.870	-3.920×10^{-1}	6.9×10^{-1}	
$\kappa^{(1)}_{1969}$	-7.444×10^{-1}	1.863	-4.000×10^{-1}	6.9×10^{-1}	
$\kappa^{(1)}_{1970}$	-8.022×10^{-1}	1.854	-4.330×10^{-1}	6.7×10^{-1}	
$\kappa^{(1)}_{1971}$	-7.874×10^{-1}	1.845	-4.270×10^{-1}	6.7×10^{-1}	
$\kappa^{(1)}_{1972}$	-7.571×10^{-1}	1.834	-4.130×10^{-1}	6.8×10^{-1}	
$\kappa^{(1)}_{1973}$	-7.350×10^{-1}	1.823	-4.030×10^{-1}	6.9×10^{-1}	
$\kappa^{(1)}_{1974}$	-7.311×10^{-1}	1.811	-4.040×10^{-1}	6.9×10^{-1}	
$\kappa^{(1)}_{1975}$	-6.940×10^{-1}	1.798	-3.860×10^{-1}	7.0×10^{-1}	
$\kappa^{(1)}_{1976}$	-6.437×10^{-1}	1.784	-3.610×10^{-1}	7.2×10^{-1}	
$\kappa^{(1)}_{1977}$	-6.810×10^{-1}	1.769	-3.850×10^{-1}	7.0×10^{-1}	
$\kappa^{(1)}_{1978}$	-6.462×10^{-1}	1.753	-3.690×10^{-1}	7.1×10^{-1}	
$\kappa^{(1)}_{1979}$	-6.207×10^{-1}	1.736	-3.580×10^{-1}	7.2×10^{-1}	
$\kappa^{(1)}_{1980}$	-5.856×10^{-1}	1.718	-3.410×10^{-1}	7.3×10^{-1}	
$\kappa^{(1)}_{1981}$	-5.549×10^{-1}	1.699	-3.270×10^{-1}	7.4×10^{-1}	
$\kappa^{(1)}_{1982}$	-5.660×10^{-1}	1.680	-3.370×10^{-1}	7.4×10^{-1}	
$\kappa^{(1)}_{1983}$	-5.406×10^{-1}	1.659	-3.260×10^{-1}	7.4×10^{-1}	
$\kappa^{(1)}_{1984}$	-5.257×10^{-1}	1.638	-3.210×10^{-1}	7.5×10^{-1}	
$\kappa^{(1)}_{1985}$	-4.575×10^{-1}	1.616	-2.830×10^{-1}	7.8×10^{-1}	
$\kappa^{(1)}_{1986}$	-4.452×10^{-1}	1.593	-2.800×10^{-1}	7.8×10^{-1}	
$\kappa^{(1)}_{1987}$	-4.265×10^{-1}	1.569	-2.720×10^{-1}	7.9×10^{-1}	
$\kappa^{(1)}_{1988}$	-3.493×10^{-1}	1.544	-2.260×10^{-1}	8.2×10^{-1}	
$\kappa^{(1)}_{1989}$	-3.740×10^{-1}	1.519	-2.460×10^{-1}	8.1×10^{-1}	
$\kappa^{(1)}_{1990}$	-3.136×10^{-1}	1.492	-2.100×10^{-1}	8.3×10^{-1}	
$\kappa^{(1)}_{1991}$	-2.958×10^{-1}	1.465	-2.020×10^{-1}	8.4×10^{-1}	
$\kappa^{(1)}_{1992}$	-2.672×10^{-1}	1.438	-1.860×10^{-1}	8.5×10^{-1}	
$\kappa^{(1)}_{1993}$	-2.015×10^{-1}	1.409	-1.430×10^{-1}	8.9×10^{-1}	
$\kappa^{(1)}_{1994}$	-2.448×10^{-1}	1.380	-1.770×10^{-1}	8.6×10^{-1}	
$\kappa^{(1)}_{1995}$	-1.975×10^{-1}	1.351	-1.460×10^{-1}	8.8×10^{-1}	
$\kappa^{(1)}_{1996}$	-1.599×10^{-1}	1.321	-1.210×10^{-1}	9.0×10^{-1}	

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(1)}_{1997}$	-1.439×10^{-1}	1.290	-1.120×10^{-1}	9.1×10^{-1}	
$\kappa^{(1)}_{1998}$	-1.226×10^{-1}	1.259	-9.700×10^{-2}	9.2×10^{-1}	
$\kappa^{(1)}_{1999}$	-6.553×10^{-2}	1.228	-5.300×10^{-2}	9.6×10^{-1}	
$\kappa^{(1)}_{2000}$	-5.169×10^{-2}	1.196	-4.300×10^{-2}	9.7×10^{-1}	
$\kappa^{(1)}_{2001}$	-1.664×10^{-2}	1.164	-1.400×10^{-2}	9.9×10^{-1}	
$\kappa^{(1)}_{2002}$	2.968×10^{-2}	1.132	2.600×10^{-2}	9.8×10^{-1}	
$\kappa^{(1)}_{2003}$	2.051×10^{-2}	1.100	1.900×10^{-2}	9.9×10^{-1}	
$\kappa^{(1)}_{2004}$	1.855×10^{-2}	1.068	1.700×10^{-2}	9.9×10^{-1}	
$\kappa^{(1)}_{2005}$	4.312×10^{-2}	1.036	4.200×10^{-2}	9.7×10^{-1}	
$\kappa^{(1)}_{2006}$	6.347×10^{-2}	1.005	6.300×10^{-2}	9.5×10^{-1}	
$\kappa^{(1)}_{2007}$	1.010×10^{-1}	9.746×10^{-1}	1.040×10^{-1}	9.2×10^{-1}	
$\kappa^{(1)}_{2008}$	1.107×10^{-1}	9.449×10^{-1}	1.170×10^{-1}	9.1×10^{-1}	
$\kappa^{(1)}_{2009}$	1.060×10^{-1}	9.164×10^{-1}	1.160×10^{-1}	9.1×10^{-1}	
$\kappa^{(1)}_{2010}$	1.285×10^{-1}	8.893×10^{-1}	1.440×10^{-1}	8.9×10^{-1}	
$\kappa^{(1)}_{2011}$	1.340×10^{-1}	8.639×10^{-1}	1.550×10^{-1}	8.8×10^{-1}	
$\kappa^{(1)}_{2012}$	1.771×10^{-1}	8.406×10^{-1}	2.110×10^{-1}	8.3×10^{-1}	
$\kappa^{(1)}_{2013}$	1.712×10^{-1}	8.197×10^{-1}	2.090×10^{-1}	8.3×10^{-1}	
$\kappa^{(1)}_{2014}$	1.557×10^{-1}	8.019×10^{-1}	1.940×10^{-1}	8.5×10^{-1}	
$\kappa^{(2)}_{1900}$	-1.082×10^{-2}	1.187×10^{-1}	-9.100×10^{-2}	9.3×10^{-1}	
$\kappa^{(2)}_{1901}$	-9.011×10^{-3}	1.177×10^{-1}	-7.700×10^{-2}	9.4×10^{-1}	
$\kappa^{(2)}_{1902}$	-6.681×10^{-3}	1.166×10^{-1}	-5.700×10^{-2}	9.5×10^{-1}	
$\kappa^{(2)}_{1903}$	-7.994×10^{-3}	1.156×10^{-1}	-6.900×10^{-2}	9.4×10^{-1}	
$\kappa^{(2)}_{1904}$	-7.985×10^{-3}	1.145×10^{-1}	-7.000×10^{-2}	9.4×10^{-1}	
$\kappa^{(2)}_{1905}$	-4.582×10^{-3}	1.135×10^{-1}	-4.000×10^{-2}	9.7×10^{-1}	
$\kappa^{(2)}_{1906}$	-4.404×10^{-3}	1.125×10^{-1}	-3.900×10^{-2}	9.7×10^{-1}	
$\kappa^{(2)}_{1907}$	-8.231×10^{-3}	1.114×10^{-1}	-7.400×10^{-2}	9.4×10^{-1}	
$\kappa^{(2)}_{1908}$	-7.618×10^{-3}	1.104×10^{-1}	-6.900×10^{-2}	9.4×10^{-1}	
$\kappa^{(2)}_{1909}$	-3.762×10^{-3}	1.093×10^{-1}	-3.400×10^{-2}	9.7×10^{-1}	
$\kappa^{(2)}_{1910}$	-5.207×10^{-3}	1.083×10^{-1}	-4.800×10^{-2}	9.6×10^{-1}	
$\kappa^{(2)}_{1911}$	-4.601×10^{-3}	1.073×10^{-1}	-4.300×10^{-2}	9.7×10^{-1}	
$\kappa^{(2)}_{1912}$	-3.456×10^{-3}	1.062×10^{-1}	-3.300×10^{-2}	9.7×10^{-1}	
$\kappa^{(2)}_{1913}$	-3.584×10^{-3}	1.052×10^{-1}	-3.400×10^{-2}	9.7×10^{-1}	
$\kappa^{(2)}_{1914}$	-3.692×10^{-3}	1.041×10^{-1}	-3.500×10^{-2}	9.7×10^{-1}	
$\kappa^{(2)}_{1915}$	-4.381×10^{-3}	1.031×10^{-1}	-4.300×10^{-2}	9.7×10^{-1}	
$\kappa^{(2)}_{1916}$	-3.474×10^{-4}	1.020×10^{-1}	-3.000×10^{-3}	10.0×10^{-1}	
$\kappa^{(2)}_{1917}$	-1.565×10^{-3}	1.010×10^{-1}	-1.500×10^{-2}	9.9×10^{-1}	
$\kappa^{(2)}_{1918}$	5.096×10^{-3}	9.996×10^{-2}	5.100×10^{-2}	9.6×10^{-1}	
$\kappa^{(2)}_{1919}$	2.521×10^{-4}	9.892×10^{-2}	3.000×10^{-3}	10.0×10^{-1}	
$\kappa^{(2)}_{1920}$	-1.563×10^{-4}	9.788×10^{-2}	-2.000×10^{-3}	10.0×10^{-1}	
$\kappa^{(2)}_{1921}$	-2.373×10^{-3}	9.684×10^{-2}	-2.500×10^{-2}	9.8×10^{-1}	
$\kappa^{(2)}_{1922}$	-6.484×10^{-4}	9.580×10^{-2}	-7.000×10^{-3}	9.9×10^{-1}	
$\kappa^{(2)}_{1923}$	1.703×10^{-3}	9.476×10^{-2}	1.800×10^{-2}	9.9×10^{-1}	
$\kappa^{(2)}_{1924}$	1.554×10^{-3}	9.371×10^{-2}	1.700×10^{-2}	9.9×10^{-1}	
$\kappa^{(2)}_{1925}$	1.864×10^{-3}	9.267×10^{-2}	2.000×10^{-2}	9.8×10^{-1}	

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z \rvert)$	Signif. code
$\kappa^{(2)}_{1926}$	3.389×10^{-3}	9.163×10^{-2}	3.700×10^{-2}	9.7×10^{-1}	
$\kappa^{(2)}_{1927}$	2.435×10^{-4}	9.059×10^{-2}	3.000×10^{-3}	10.0×10^{-1}	
$\kappa^{(2)}_{1928}$	4.330×10^{-3}	8.955×10^{-2}	4.800×10^{-2}	9.6×10^{-1}	
$\kappa^{(2)}_{1929}$	5.053×10^{-3}	8.851×10^{-2}	5.700×10^{-2}	9.5×10^{-1}	
$\kappa^{(2)}_{1930}$	6.155×10^{-3}	8.746×10^{-2}	7.000×10^{-2}	9.4×10^{-1}	
$\kappa^{(2)}_{1931}$	2.416×10^{-3}	8.642×10^{-2}	2.800×10^{-2}	9.8×10^{-1}	
$\kappa^{(2)}_{1932}$	5.854×10^{-3}	8.538×10^{-2}	6.900×10^{-2}	9.5×10^{-1}	
$\kappa^{(2)}_{1933}$	7.146×10^{-3}	8.434×10^{-2}	8.500×10^{-2}	9.3×10^{-1}	
$\kappa^{(2)}_{1934}$	7.086×10^{-3}	8.330×10^{-2}	8.500×10^{-2}	9.3×10^{-1}	
$\kappa^{(2)}_{1935}$	5.848×10^{-3}	8.226×10^{-2}	7.100×10^{-2}	9.4×10^{-1}	
$\kappa^{(2)}_{1936}$	7.046×10^{-3}	8.121×10^{-2}	8.700×10^{-2}	9.3×10^{-1}	
$\kappa^{(2)}_{1937}$	8.854×10^{-3}	8.017×10^{-2}	1.100×10^{-1}	9.1×10^{-1}	
$\kappa^{(2)}_{1938}$	9.808×10^{-3}	7.913×10^{-2}	1.240×10^{-1}	9.0×10^{-1}	
$\kappa^{(2)}_{1939}$	6.462×10^{-3}	7.809×10^{-2}	8.300×10^{-2}	9.3×10^{-1}	
$\kappa^{(2)}_{1940}$	7.926×10^{-3}	7.705×10^{-2}	1.030×10^{-1}	9.2×10^{-1}	
$\kappa^{(2)}_{1941}$	7.280×10^{-3}	7.601×10^{-2}	9.600×10^{-2}	9.2×10^{-1}	
$\kappa^{(2)}_{1942}$	1.236×10^{-2}	7.497×10^{-2}	1.650×10^{-1}	8.7×10^{-1}	
$\kappa^{(2)}_{1943}$	1.345×10^{-2}	7.392×10^{-2}	1.820×10^{-1}	8.6×10^{-1}	
$\kappa^{(2)}_{1944}$	1.178×10^{-2}	7.288×10^{-2}	1.620×10^{-1}	8.7×10^{-1}	
$\kappa^{(2)}_{1945}$	1.387×10^{-2}	7.184×10^{-2}	1.930×10^{-1}	8.5×10^{-1}	
$\kappa^{(2)}_{1946}$	1.247×10^{-2}	7.080×10^{-2}	1.760×10^{-1}	8.6×10^{-1}	
$\kappa^{(2)}_{1947}$	1.136×10^{-2}	6.976×10^{-2}	1.630×10^{-1}	8.7×10^{-1}	
$\kappa^{(2)}_{1948}$	1.648×10^{-2}	6.872×10^{-2}	2.400×10^{-1}	8.1×10^{-1}	
$\kappa^{(2)}_{1949}$	1.509×10^{-2}	6.767×10^{-2}	2.230×10^{-1}	8.2×10^{-1}	
$\kappa^{(2)}_{1950}$	1.418×10^{-2}	6.663×10^{-2}	2.130×10^{-1}	8.3×10^{-1}	
$\kappa^{(2)}_{1951}$	1.319×10^{-2}	6.559×10^{-2}	2.010×10^{-1}	8.4×10^{-1}	
$\kappa^{(2)}_{1952}$	1.733×10^{-2}	6.455×10^{-2}	2.680×10^{-1}	7.9×10^{-1}	
$\kappa^{(2)}_{1953}$	1.817×10^{-2}	6.351×10^{-2}	2.860×10^{-1}	7.7×10^{-1}	
$\kappa^{(2)}_{1954}$	1.749×10^{-2}	6.247×10^{-2}	2.800×10^{-1}	7.8×10^{-1}	
$\kappa^{(2)}_{1955}$	1.972×10^{-2}	6.142×10^{-2}	3.210×10^{-1}	7.5×10^{-1}	
$\kappa^{(2)}_{1956}$	2.009×10^{-2}	6.038×10^{-2}	3.330×10^{-1}	7.4×10^{-1}	
$\kappa^{(2)}_{1957}$	1.977×10^{-2}	5.934×10^{-2}	3.330×10^{-1}	7.4×10^{-1}	
$\kappa^{(2)}_{1958}$	2.050×10^{-2}	5.830×10^{-2}	3.520×10^{-1}	7.3×10^{-1}	
$\kappa^{(2)}_{1959}$	2.120×10^{-2}	5.726×10^{-2}	3.700×10^{-1}	7.1×10^{-1}	
$\kappa^{(2)}_{1960}$	2.109×10^{-2}	5.622×10^{-2}	3.750×10^{-1}	7.1×10^{-1}	
$\kappa^{(2)}_{1961}$	2.194×10^{-2}	5.517×10^{-2}	3.980×10^{-1}	6.9×10^{-1}	
$\kappa^{(2)}_{1962}$	2.150×10^{-2}	5.413×10^{-2}	3.970×10^{-1}	6.9×10^{-1}	
$\kappa^{(2)}_{1963}$	2.350×10^{-2}	5.309×10^{-2}	4.430×10^{-1}	6.6×10^{-1}	
$\kappa^{(2)}_{1964}$	2.693×10^{-2}	5.205×10^{-2}	5.170×10^{-1}	6.0×10^{-1}	
$\kappa^{(2)}_{1965}$	2.579×10^{-2}	5.101×10^{-2}	5.060×10^{-1}	6.1×10^{-1}	
$\kappa^{(2)}_{1966}$	2.640×10^{-2}	4.997×10^{-2}	5.280×10^{-1}	6.0×10^{-1}	
$\kappa^{(2)}_{1967}$	2.755×10^{-2}	4.893×10^{-2}	5.630×10^{-1}	5.7×10^{-1}	
$\kappa^{(2)}_{1968}$	2.772×10^{-2}	4.788×10^{-2}	5.790×10^{-1}	5.6×10^{-1}	
$\kappa^{(2)}_{1969}$	3.111×10^{-2}	4.684×10^{-2}	6.640×10^{-1}	5.1×10^{-1}	

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(2)}_{1970}$	3.284×10^{-2}	4.580×10^{-2}	7.170×10^{-1}	4.7×10^{-1}	
$\kappa^{(2)}_{1971}$	3.336×10^{-2}	4.476×10^{-2}	7.450×10^{-1}	4.6×10^{-1}	
$\kappa^{(2)}_{1972}$	3.366×10^{-2}	4.372×10^{-2}	7.700×10^{-1}	4.4×10^{-1}	
$\kappa^{(2)}_{1973}$	3.354×10^{-2}	4.268×10^{-2}	7.860×10^{-1}	4.3×10^{-1}	
$\kappa^{(2)}_{1974}$	3.543×10^{-2}	4.163×10^{-2}	8.510×10^{-1}	3.9×10^{-1}	
$\kappa^{(2)}_{1975}$	3.540×10^{-2}	4.059×10^{-2}	8.720×10^{-1}	3.8×10^{-1}	
$\kappa^{(2)}_{1976}$	3.400×10^{-2}	3.955×10^{-2}	8.600×10^{-1}	3.9×10^{-1}	
$\kappa^{(2)}_{1977}$	3.566×10^{-2}	3.851×10^{-2}	9.260×10^{-1}	3.5×10^{-1}	
$\kappa^{(2)}_{1978}$	3.623×10^{-2}	3.747×10^{-2}	9.670×10^{-1}	3.3×10^{-1}	
$\kappa^{(2)}_{1979}$	3.601×10^{-2}	3.643×10^{-2}	9.890×10^{-1}	3.2×10^{-1}	
$\kappa^{(2)}_{1980}$	3.650×10^{-2}	3.539×10^{-2}	1.031	3.0×10^{-1}	
$\kappa^{(2)}_{1981}$	3.607×10^{-2}	3.434×10^{-2}	1.050	2.9×10^{-1}	
$\kappa^{(2)}_{1982}$	3.752×10^{-2}	3.330×10^{-2}	1.127	2.6×10^{-1}	
$\kappa^{(2)}_{1983}$	3.540×10^{-2}	3.226×10^{-2}	1.097	2.7×10^{-1}	
$\kappa^{(2)}_{1984}$	3.664×10^{-2}	3.122×10^{-2}	1.174	2.4×10^{-1}	
$\kappa^{(2)}_{1985}$	3.484×10^{-2}	3.018×10^{-2}	1.154	2.5×10^{-1}	
$\kappa^{(2)}_{1986}$	3.504×10^{-2}	2.914×10^{-2}	1.202	2.3×10^{-1}	
$\kappa^{(2)}_{1987}$	3.565×10^{-2}	2.810×10^{-2}	1.269	2.0×10^{-1}	
$\kappa^{(2)}_{1988}$	3.333×10^{-2}	2.705×10^{-2}	1.232	2.2×10^{-1}	
$\kappa^{(2)}_{1989}$	3.240×10^{-2}	2.601×10^{-2}	1.246	2.1×10^{-1}	
$\kappa^{(2)}_{1990}$	3.208×10^{-2}	2.497×10^{-2}	1.285	2.0×10^{-1}	
$\kappa^{(2)}_{1991}$	3.239×10^{-2}	2.393×10^{-2}	1.353	1.8×10^{-1}	
$\kappa^{(2)}_{1992}$	3.231×10^{-2}	2.289×10^{-2}	1.412	1.6×10^{-1}	
$\kappa^{(2)}_{1993}$	2.906×10^{-2}	2.185×10^{-2}	1.330	1.8×10^{-1}	
$\kappa^{(2)}_{1994}$	2.970×10^{-2}	2.081×10^{-2}	1.427	1.5×10^{-1}	
$\kappa^{(2)}_{1995}$	2.939×10^{-2}	1.977×10^{-2}	1.487	1.4×10^{-1}	
$\kappa^{(2)}_{1996}$	2.711×10^{-2}	1.873×10^{-2}	1.448	1.5×10^{-1}	
$\kappa^{(2)}_{1997}$	2.570×10^{-2}	1.768×10^{-2}	1.453	1.5×10^{-1}	
$\kappa^{(2)}_{1998}$	2.561×10^{-2}	1.664×10^{-2}	1.538	1.2×10^{-1}	
$\kappa^{(2)}_{1999}$	2.312×10^{-2}	1.560×10^{-2}	1.482	1.4×10^{-1}	
$\kappa^{(2)}_{2000}$	2.256×10^{-2}	1.456×10^{-2}	1.549	1.2×10^{-1}	
$\kappa^{(2)}_{2001}$	2.060×10^{-2}	1.352×10^{-2}	1.523	1.3×10^{-1}	
$\kappa^{(2)}_{2002}$	1.809×10^{-2}	1.248×10^{-2}	1.449	1.5×10^{-1}	
$\kappa^{(2)}_{2003}$	1.757×10^{-2}	1.144×10^{-2}	1.535	1.2×10^{-1}	
$\kappa^{(2)}_{2004}$	1.719×10^{-2}	1.040×10^{-2}	1.653	9.8×10^{-2}	.
$\kappa^{(2)}_{2005}$	1.583×10^{-2}	9.364×10^{-3}	1.690	9.1×10^{-2}	.
$\kappa^{(2)}_{2006}$	1.295×10^{-2}	8.326×10^{-3}	1.555	1.2×10^{-1}	
$\kappa^{(2)}_{2007}$	1.092×10^{-2}	7.289×10^{-3}	1.499	1.3×10^{-1}	
$\kappa^{(2)}_{2008}$	8.670×10^{-3}	6.253×10^{-3}	1.387	1.7×10^{-1}	
$\kappa^{(2)}_{2009}$	8.382×10^{-3}	5.220×10^{-3}	1.606	1.1×10^{-1}	
$\kappa^{(2)}_{2010}$	7.052×10^{-3}	4.191×10^{-3}	1.683	9.2×10^{-2}	.
$\kappa^{(2)}_{2011}$	3.920×10^{-3}	3.171×10^{-3}	1.236	2.2×10^{-1}	
$\kappa^{(2)}_{2012}$	1.955×10^{-3}	2.169×10^{-3}	9.020×10^{-1}	3.7×10^{-1}	
$\kappa^{(2)}_{2013}$	1.382×10^{-3}	1.205×10^{-3}	1.147	2.5×10^{-1}	

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z \rvert)$	Signif. code
γ_{1800}	1.024	1.109	9.230×10^{-1}	3.6×10^{-1}	
γ_{1801}	1.178	1.042	1.131	2.6×10^{-1}	
γ_{1802}	1.525	1.021	1.493	1.4×10^{-1}	
γ_{1803}	1.158	1.014	1.142	2.5×10^{-1}	
γ_{1804}	1.110	1.022	1.087	2.8×10^{-1}	
γ_{1805}	1.168	1.039	1.124	2.6×10^{-1}	
γ_{1806}	1.126	1.061	1.061	2.9×10^{-1}	
γ_{1807}	1.069	1.088	9.830×10^{-1}	3.3×10^{-1}	
γ_{1808}	1.152	1.118	1.030	3.0×10^{-1}	
γ_{1809}	1.218	1.153	1.056	2.9×10^{-1}	
γ_{1810}	1.212	1.190	1.018	3.1×10^{-1}	
γ_{1811}	1.258	1.229	1.023	3.1×10^{-1}	
γ_{1812}	1.286	1.271	1.012	3.1×10^{-1}	
γ_{1813}	1.338	1.313	1.018	3.1×10^{-1}	
γ_{1814}	1.341	1.357	9.880×10^{-1}	3.2×10^{-1}	
γ_{1815}	1.387	1.402	9.890×10^{-1}	3.2×10^{-1}	
γ_{1816}	1.434	1.448	9.900×10^{-1}	3.2×10^{-1}	
γ_{1817}	1.452	1.494	9.720×10^{-1}	3.3×10^{-1}	
γ_{1818}	1.443	1.541	9.370×10^{-1}	3.5×10^{-1}	
γ_{1819}	1.501	1.587	9.460×10^{-1}	3.4×10^{-1}	
γ_{1820}	1.486	1.634	9.090×10^{-1}	3.6×10^{-1}	
γ_{1821}	1.530	1.681	9.110×10^{-1}	3.6×10^{-1}	
γ_{1822}	1.553	1.727	8.990×10^{-1}	3.7×10^{-1}	
γ_{1823}	1.547	1.774	8.720×10^{-1}	3.8×10^{-1}	
γ_{1824}	1.589	1.820	8.730×10^{-1}	3.8×10^{-1}	
γ_{1825}	1.624	1.865	8.710×10^{-1}	3.8×10^{-1}	
γ_{1826}	1.644	1.910	8.610×10^{-1}	3.9×10^{-1}	
γ_{1827}	1.660	1.955	8.490×10^{-1}	4.0×10^{-1}	
γ_{1828}	1.645	1.999	8.230×10^{-1}	4.1×10^{-1}	
γ_{1829}	1.728	2.042	8.460×10^{-1}	4.0×10^{-1}	
γ_{1830}	1.713	2.085	8.220×10^{-1}	4.1×10^{-1}	
γ_{1831}	1.727	2.127	8.120×10^{-1}	4.2×10^{-1}	
γ_{1832}	1.715	2.169	7.910×10^{-1}	4.3×10^{-1}	
γ_{1833}	1.746	2.209	7.900×10^{-1}	4.3×10^{-1}	
γ_{1834}	1.801	2.249	8.010×10^{-1}	4.2×10^{-1}	
γ_{1835}	1.801	2.288	7.870×10^{-1}	4.3×10^{-1}	
γ_{1836}	1.794	2.327	7.710×10^{-1}	4.4×10^{-1}	
γ_{1837}	1.838	2.364	7.770×10^{-1}	4.4×10^{-1}	
γ_{1838}	1.849	2.401	7.700×10^{-1}	4.4×10^{-1}	
γ_{1839}	1.815	2.437	7.450×10^{-1}	4.6×10^{-1}	
γ_{1840}	1.888	2.472	7.640×10^{-1}	4.4×10^{-1}	
γ_{1841}	1.894	2.506	7.560×10^{-1}	4.5×10^{-1}	
γ_{1842}	1.915	2.539	7.540×10^{-1}	4.5×10^{-1}	
γ_{1843}	1.903	2.571	7.400×10^{-1}	4.6×10^{-1}	
γ_{1844}	1.936	2.603	7.440×10^{-1}	4.6×10^{-1}	

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
γ_{1845}	1.951	2.633	7.410×10^{-1}	4.6×10^{-1}	
γ_{1846}	1.955	2.663	7.340×10^{-1}	4.6×10^{-1}	
γ_{1847}	1.965	2.692	7.300×10^{-1}	4.7×10^{-1}	
γ_{1848}	1.969	2.719	7.240×10^{-1}	4.7×10^{-1}	
γ_{1849}	1.981	2.746	7.210×10^{-1}	4.7×10^{-1}	
γ_{1850}	2.016	2.772	7.270×10^{-1}	4.7×10^{-1}	
γ_{1851}	1.981	2.797	7.080×10^{-1}	4.8×10^{-1}	
γ_{1852}	2.023	2.821	7.170×10^{-1}	4.7×10^{-1}	
γ_{1853}	1.980	2.844	6.960×10^{-1}	4.9×10^{-1}	
γ_{1854}	2.024	2.866	7.060×10^{-1}	4.8×10^{-1}	
γ_{1855}	2.034	2.887	7.040×10^{-1}	4.8×10^{-1}	
γ_{1856}	2.046	2.907	7.040×10^{-1}	4.8×10^{-1}	
γ_{1857}	2.028	2.926	6.930×10^{-1}	4.9×10^{-1}	
γ_{1858}	2.056	2.944	6.980×10^{-1}	4.8×10^{-1}	
γ_{1859}	2.054	2.962	6.940×10^{-1}	4.9×10^{-1}	
γ_{1860}	2.078	2.978	6.980×10^{-1}	4.9×10^{-1}	
γ_{1861}	2.061	2.993	6.890×10^{-1}	4.9×10^{-1}	
γ_{1862}	2.048	3.007	6.810×10^{-1}	5.0×10^{-1}	
γ_{1863}	2.070	3.020	6.850×10^{-1}	4.9×10^{-1}	
γ_{1864}	2.065	3.033	6.810×10^{-1}	5.0×10^{-1}	
γ_{1865}	2.039	3.044	6.700×10^{-1}	5.0×10^{-1}	
γ_{1866}	2.064	3.054	6.760×10^{-1}	5.0×10^{-1}	
γ_{1867}	2.059	3.063	6.720×10^{-1}	5.0×10^{-1}	
γ_{1868}	2.058	3.072	6.700×10^{-1}	5.0×10^{-1}	
γ_{1869}	2.032	3.079	6.600×10^{-1}	5.1×10^{-1}	
γ_{1870}	2.046	3.085	6.630×10^{-1}	5.1×10^{-1}	
γ_{1871}	2.033	3.090	6.580×10^{-1}	5.1×10^{-1}	
γ_{1872}	2.029	3.094	6.560×10^{-1}	5.1×10^{-1}	
γ_{1873}	2.016	3.097	6.510×10^{-1}	5.2×10^{-1}	
γ_{1874}	2.010	3.100	6.490×10^{-1}	5.2×10^{-1}	
γ_{1875}	1.983	3.101	6.400×10^{-1}	5.2×10^{-1}	
γ_{1876}	1.985	3.101	6.400×10^{-1}	5.2×10^{-1}	
γ_{1877}	1.969	3.100	6.350×10^{-1}	5.3×10^{-1}	
γ_{1878}	1.955	3.098	6.310×10^{-1}	5.3×10^{-1}	
γ_{1879}	1.933	3.095	6.250×10^{-1}	5.3×10^{-1}	
γ_{1880}	1.922	3.091	6.220×10^{-1}	5.3×10^{-1}	
γ_{1881}	1.895	3.086	6.140×10^{-1}	5.4×10^{-1}	
γ_{1882}	1.894	3.080	6.150×10^{-1}	5.4×10^{-1}	
γ_{1883}	1.850	3.073	6.020×10^{-1}	5.5×10^{-1}	
γ_{1884}	1.831	3.065	5.970×10^{-1}	5.5×10^{-1}	
γ_{1885}	1.796	3.056	5.880×10^{-1}	5.6×10^{-1}	
γ_{1886}	1.789	3.046	5.870×10^{-1}	5.6×10^{-1}	
γ_{1887}	1.747	3.035	5.750×10^{-1}	5.6×10^{-1}	
γ_{1888}	1.747	3.023	5.780×10^{-1}	5.6×10^{-1}	
γ_{1889}	1.693	3.010	5.620×10^{-1}	5.7×10^{-1}	

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\|z\|)$	Signif. code
γ_{1890}	1.664	2.996	5.550×10^{-1}	5.8×10^{-1}	
γ_{1891}	1.635	2.981	5.480×10^{-1}	5.8×10^{-1}	
γ_{1892}	1.591	2.965	5.360×10^{-1}	5.9×10^{-1}	
γ_{1893}	1.547	2.947	5.250×10^{-1}	6.0×10^{-1}	
γ_{1894}	1.523	2.929	5.200×10^{-1}	6.0×10^{-1}	
γ_{1895}	1.499	2.910	5.150×10^{-1}	6.1×10^{-1}	
γ_{1896}	1.449	2.890	5.010×10^{-1}	6.2×10^{-1}	
γ_{1897}	1.405	2.868	4.900×10^{-1}	6.2×10^{-1}	
γ_{1898}	1.383	2.846	4.860×10^{-1}	6.3×10^{-1}	
γ_{1899}	1.332	2.823	4.720×10^{-1}	6.4×10^{-1}	
γ_{1900}	1.285	2.798	4.590×10^{-1}	6.5×10^{-1}	
γ_{1901}	1.238	2.773	4.470×10^{-1}	6.6×10^{-1}	
γ_{1902}	1.193	2.746	4.340×10^{-1}	6.6×10^{-1}	
γ_{1903}	1.142	2.719	4.200×10^{-1}	6.7×10^{-1}	
γ_{1904}	1.103	2.690	4.100×10^{-1}	6.8×10^{-1}	
γ_{1905}	1.058	2.661	3.980×10^{-1}	6.9×10^{-1}	
γ_{1906}	1.015	2.630	3.860×10^{-1}	7.0×10^{-1}	
γ_{1907}	9.703×10^{-1}	2.599	3.730×10^{-1}	7.1×10^{-1}	
γ_{1908}	9.361×10^{-1}	2.566	3.650×10^{-1}	7.2×10^{-1}	
γ_{1909}	8.800×10^{-1}	2.532	3.480×10^{-1}	7.3×10^{-1}	
γ_{1910}	8.486×10^{-1}	2.497	3.400×10^{-1}	7.3×10^{-1}	
γ_{1911}	8.010×10^{-1}	2.462	3.250×10^{-1}	7.4×10^{-1}	
γ_{1912}	7.632×10^{-1}	2.425	3.150×10^{-1}	7.5×10^{-1}	
γ_{1913}	7.104×10^{-1}	2.387	2.980×10^{-1}	7.7×10^{-1}	
γ_{1914}	6.810×10^{-1}	2.348	2.900×10^{-1}	7.7×10^{-1}	
γ_{1915}	6.531×10^{-1}	2.308	2.830×10^{-1}	7.8×10^{-1}	
γ_{1916}	6.003×10^{-1}	2.267	2.650×10^{-1}	7.9×10^{-1}	
γ_{1917}	5.617×10^{-1}	2.226	2.520×10^{-1}	8.0×10^{-1}	
γ_{1918}	5.279×10^{-1}	2.183	2.420×10^{-1}	8.1×10^{-1}	
γ_{1919}	4.635×10^{-1}	2.139	2.170×10^{-1}	8.3×10^{-1}	
γ_{1920}	4.871×10^{-1}	2.093	2.330×10^{-1}	8.2×10^{-1}	
γ_{1921}	4.029×10^{-1}	2.047	1.970×10^{-1}	8.4×10^{-1}	
γ_{1922}	4.019×10^{-1}	2.000	2.010×10^{-1}	8.4×10^{-1}	
γ_{1923}	3.468×10^{-1}	1.952	1.780×10^{-1}	8.6×10^{-1}	
γ_{1924}	3.193×10^{-1}	1.903	1.680×10^{-1}	8.7×10^{-1}	
γ_{1925}	2.845×10^{-1}	1.853	1.540×10^{-1}	8.8×10^{-1}	
γ_{1926}	2.349×10^{-1}	1.801	1.300×10^{-1}	9.0×10^{-1}	
γ_{1927}	2.328×10^{-1}	1.749	1.330×10^{-1}	8.9×10^{-1}	
γ_{1928}	1.769×10^{-1}	1.696	1.040×10^{-1}	9.2×10^{-1}	
γ_{1929}	1.731×10^{-1}	1.642	1.050×10^{-1}	9.2×10^{-1}	
γ_{1930}	1.420×10^{-1}	1.586	9.000×10^{-2}	9.3×10^{-1}	
γ_{1931}	1.390×10^{-1}	1.530	9.100×10^{-2}	9.3×10^{-1}	
γ_{1932}	1.140×10^{-1}	1.472	7.700×10^{-2}	9.4×10^{-1}	
γ_{1933}	9.003×10^{-2}	1.414	6.400×10^{-2}	9.5×10^{-1}	
γ_{1934}	6.759×10^{-2}	1.354	5.000×10^{-2}	9.6×10^{-1}	

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z \rvert)$	Signif. code
γ_{1935}	6.640×10^{-2}	1.294	5.100×10^{-2}	9.6×10^{-1}	
γ_{1936}	6.942×10^{-2}	1.232	5.600×10^{-2}	9.6×10^{-1}	
γ_{1937}	3.580×10^{-2}	1.169	3.100×10^{-2}	9.8×10^{-1}	
γ_{1938}	2.861×10^{-2}	1.106	2.600×10^{-2}	9.8×10^{-1}	
γ_{1939}	2.456×10^{-2}	1.041	2.400×10^{-2}	9.8×10^{-1}	
γ_{1940}	5.342×10^{-2}	9.754×10^{-1}	5.500×10^{-2}	9.6×10^{-1}	
γ_{1941}	1.988×10^{-2}	9.086×10^{-1}	2.200×10^{-2}	9.8×10^{-1}	
γ_{1942}	6.984×10^{-2}	8.409×10^{-1}	8.300×10^{-2}	9.3×10^{-1}	
γ_{1943}	3.099×10^{-2}	7.721×10^{-1}	4.000×10^{-2}	9.7×10^{-1}	
γ_{1944}	6.416×10^{-2}	7.023×10^{-1}	9.100×10^{-2}	9.3×10^{-1}	
γ_{1945}	5.577×10^{-2}	6.314×10^{-1}	8.800×10^{-2}	9.3×10^{-1}	
γ_{1946}	5.471×10^{-2}	5.597×10^{-1}	9.800×10^{-2}	9.2×10^{-1}	
γ_{1947}	7.494×10^{-2}	4.869×10^{-1}	1.540×10^{-1}	8.8×10^{-1}	
γ_{1948}	4.887×10^{-2}	4.132×10^{-1}	1.180×10^{-1}	9.1×10^{-1}	
γ_{1949}	5.038×10^{-2}	3.387×10^{-1}	1.490×10^{-1}	8.8×10^{-1}	
γ_{1950}	5.379×10^{-2}	2.637×10^{-1}	2.040×10^{-1}	8.4×10^{-1}	
γ_{1951}	6.443×10^{-2}	1.887×10^{-1}	3.410×10^{-1}	7.3×10^{-1}	
γ_{1952}	-1.025×10^{-2}	1.166×10^{-1}	-8.800×10^{-2}	9.3×10^{-1}	

B.1.3 | M7 Model

Table B.3: Regression table of the M7 model for Swedish females. 438 of 498 parameters (≈ 87%) are significant on the 5% level. p-value significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' . ' 0.1 ' ' 1.

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z \rvert)$	Signif. code
$\kappa^{(1)}_{1900}$	-1.313×10^{1}	1.440	−9.115	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1901}$	-1.375×10^{1}	1.487	−9.249	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1902}$	-1.422×10^{1}	1.534	−9.272	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1903}$	-1.474×10^{1}	1.581	−9.324	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1904}$	-1.508×10^{1}	1.628	−9.268	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1905}$	-1.559×10^{1}	1.674	−9.311	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1906}$	-1.604×10^{1}	1.721	−9.319	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1907}$	-1.637×10^{1}	1.767	−9.265	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1908}$	-1.675×10^{1}	1.813	−9.239	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1909}$	-1.719×10^{1}	1.859	−9.251	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1910}$	-1.750×10^{1}	1.904	−9.193	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1911}$	-1.785×10^{1}	1.948	−9.164	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1912}$	-1.812×10^{1}	1.992	−9.097	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1913}$	-1.848×10^{1}	2.035	−9.080	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1914}$	-1.873×10^{1}	2.077	−9.018	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1915}$	-1.894×10^{1}	2.119	−8.936	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1916}$	-1.932×10^{1}	2.160	−8.945	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1917}$	-1.954×10^{1}	2.201	−8.878	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1918}$	-1.976×10^{1}	2.240	−8.821	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1919}$	-1.996×10^{1}	2.279	−8.759	$<2.0 \times 10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(1)}_{1920}$	-2.022×10^{1}	2.317	−8.729	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1921}$	-2.038×10^{1}	2.354	−8.661	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1922}$	-2.048×10^{1}	2.390	−8.571	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1923}$	-2.076×10^{1}	2.425	−8.563	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1924}$	-2.088×10^{1}	2.459	−8.489	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1925}$	-2.101×10^{1}	2.493	−8.426	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1926}$	-2.116×10^{1}	2.525	−8.377	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1927}$	-2.115×10^{1}	2.557	−8.270	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1928}$	-2.136×10^{1}	2.588	−8.253	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1929}$	-2.143×10^{1}	2.617	−8.186	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1930}$	-2.150×10^{1}	2.646	−8.123	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1931}$	-2.144×10^{1}	2.674	−8.019	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1932}$	-2.161×10^{1}	2.701	−8.001	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1933}$	-2.165×10^{1}	2.727	−7.940	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1934}$	-2.168×10^{1}	2.752	−7.877	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1935}$	-2.165×10^{1}	2.776	−7.799	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1936}$	-2.164×10^{1}	2.799	−7.731	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1937}$	-2.164×10^{1}	2.821	−7.671	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1938}$	-2.169×10^{1}	2.842	−7.631	2.3×10^{-14}	***
$\kappa^{(1)}_{1939}$	-2.158×10^{1}	2.862	−7.540	4.7×10^{-14}	***
$\kappa^{(1)}_{1940}$	-2.155×10^{1}	2.881	−7.479	7.5×10^{-14}	***
$\kappa^{(1)}_{1941}$	-2.152×10^{1}	2.899	−7.423	1.1×10^{-13}	***
$\kappa^{(1)}_{1942}$	-2.165×10^{1}	2.916	−7.423	1.1×10^{-13}	***
$\kappa^{(1)}_{1943}$	-2.156×10^{1}	2.932	−7.352	2.0×10^{-13}	***
$\kappa^{(1)}_{1944}$	-2.143×10^{1}	2.947	−7.272	3.6×10^{-13}	***
$\kappa^{(1)}_{1945}$	-2.136×10^{1}	2.962	−7.214	5.4×10^{-13}	***
$\kappa^{(1)}_{1946}$	-2.123×10^{1}	2.975	−7.137	9.6×10^{-13}	***
$\kappa^{(1)}_{1947}$	-2.108×10^{1}	2.987	−7.057	1.7×10^{-12}	***
$\kappa^{(1)}_{1948}$	-2.109×10^{1}	2.998	−7.036	2.0×10^{-12}	***
$\kappa^{(1)}_{1949}$	-2.093×10^{1}	3.008	−6.957	3.5×10^{-12}	***
$\kappa^{(1)}_{1950}$	-2.079×10^{1}	3.017	−6.892	5.5×10^{-12}	***
$\kappa^{(1)}_{1951}$	-2.065×10^{1}	3.025	−6.828	8.6×10^{-12}	***
$\kappa^{(1)}_{1952}$	-2.057×10^{1}	3.032	−6.786	1.2×10^{-11}	***
$\kappa^{(1)}_{1953}$	-2.042×10^{1}	3.038	−6.724	1.8×10^{-11}	***
$\kappa^{(1)}_{1954}$	-2.026×10^{1}	3.043	−6.658	2.8×10^{-11}	***
$\kappa^{(1)}_{1955}$	-2.017×10^{1}	3.046	−6.619	3.6×10^{-11}	***
$\kappa^{(1)}_{1956}$	-1.999×10^{1}	3.049	−6.557	5.5×10^{-11}	***
$\kappa^{(1)}_{1957}$	-1.975×10^{1}	3.051	−6.473	9.6×10^{-11}	***
$\kappa^{(1)}_{1958}$	-1.963×10^{1}	3.052	−6.430	1.3×10^{-10}	***
$\kappa^{(1)}_{1959}$	-1.945×10^{1}	3.052	−6.373	1.9×10^{-10}	***
$\kappa^{(1)}_{1960}$	-1.924×10^{1}	3.051	−6.307	2.9×10^{-10}	***
$\kappa^{(1)}_{1961}$	-1.909×10^{1}	3.049	−6.261	3.8×10^{-10}	***
$\kappa^{(1)}_{1962}$	-1.885×10^{1}	3.045	−6.188	6.1×10^{-10}	***
$\kappa^{(1)}_{1963}$	-1.865×10^{1}	3.041	−6.132	8.7×10^{-10}	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(1)}_{1964}$	-1.846×10^{1}	3.036	-6.083	1.2×10^{-9}	***
$\kappa^{(1)}_{1965}$	-1.823×10^{1}	3.029	-6.019	1.8×10^{-9}	***
$\kappa^{(1)}_{1966}$	-1.802×10^{1}	3.022	-5.963	2.5×10^{-9}	***
$\kappa^{(1)}_{1967}$	-1.782×10^{1}	3.014	-5.912	3.4×10^{-9}	***
$\kappa^{(1)}_{1968}$	-1.754×10^{1}	3.004	-5.838	5.3×10^{-9}	***
$\kappa^{(1)}_{1969}$	-1.731×10^{1}	2.994	-5.783	7.4×10^{-9}	***
$\kappa^{(1)}_{1970}$	-1.713×10^{1}	2.982	-5.744	9.3×10^{-9}	***
$\kappa^{(1)}_{1971}$	-1.690×10^{1}	2.970	-5.689	1.3×10^{-8}	***
$\kappa^{(1)}_{1972}$	-1.662×10^{1}	2.956	-5.622	1.9×10^{-8}	***
$\kappa^{(1)}_{1973}$	-1.637×10^{1}	2.942	-5.564	2.6×10^{-8}	***
$\kappa^{(1)}_{1974}$	-1.613×10^{1}	2.926	-5.512	3.6×10^{-8}	***
$\kappa^{(1)}_{1975}$	-1.585×10^{1}	2.909	-5.449	5.1×10^{-8}	***
$\kappa^{(1)}_{1976}$	-1.555×10^{1}	2.892	-5.379	7.5×10^{-8}	***
$\kappa^{(1)}_{1977}$	-1.534×10^{1}	2.873	-5.340	9.3×10^{-8}	***
$\kappa^{(1)}_{1978}$	-1.508×10^{1}	2.853	-5.285	1.3×10^{-7}	***
$\kappa^{(1)}_{1979}$	-1.479×10^{1}	2.832	-5.221	1.8×10^{-7}	***
$\kappa^{(1)}_{1980}$	-1.449×10^{1}	2.811	-5.156	2.5×10^{-7}	***
$\kappa^{(1)}_{1981}$	-1.422×10^{1}	2.788	-5.100	3.4×10^{-7}	***
$\kappa^{(1)}_{1982}$	-1.398×10^{1}	2.764	-5.057	4.3×10^{-7}	***
$\kappa^{(1)}_{1983}$	-1.369×10^{1}	2.739	-4.998	5.8×10^{-7}	***
$\kappa^{(1)}_{1984}$	-1.342×10^{1}	2.713	-4.948	7.5×10^{-7}	***
$\kappa^{(1)}_{1985}$	-1.310×10^{1}	2.686	-4.876	1.1×10^{-6}	***
$\kappa^{(1)}_{1986}$	-1.283×10^{1}	2.658	-4.828	1.4×10^{-6}	***
$\kappa^{(1)}_{1987}$	-1.255×10^{1}	2.629	-4.775	1.8×10^{-6}	***
$\kappa^{(1)}_{1988}$	-1.221×10^{1}	2.599	-4.697	2.6×10^{-6}	***
$\kappa^{(1)}_{1989}$	-1.199×10^{1}	2.568	-4.668	3.0×10^{-6}	***
$\kappa^{(1)}_{1990}$	-1.168×10^{1}	2.535	-4.605	4.1×10^{-6}	***
$\kappa^{(1)}_{1991}$	-1.139×10^{1}	2.502	-4.552	5.3×10^{-6}	***
$\kappa^{(1)}_{1992}$	-1.111×10^{1}	2.468	-4.500	6.8×10^{-6}	***
$\kappa^{(1)}_{1993}$	-1.079×10^{1}	2.433	-4.437	9.1×10^{-6}	***
$\kappa^{(1)}_{1994}$	-1.059×10^{1}	2.396	-4.420	9.9×10^{-6}	***
$\kappa^{(1)}_{1995}$	-1.030×10^{1}	2.359	-4.366	1.3×10^{-5}	***
$\kappa^{(1)}_{1996}$	-1.001×10^{1}	2.321	-4.314	1.6×10^{-5}	***
$\kappa^{(1)}_{1997}$	-9.751	2.281	-4.275	1.9×10^{-5}	***
$\kappa^{(1)}_{1998}$	-9.494	2.241	-4.237	2.3×10^{-5}	***
$\kappa^{(1)}_{1999}$	-9.197	2.199	-4.182	2.9×10^{-5}	***
$\kappa^{(1)}_{2000}$	-8.945	2.157	-4.148	3.4×10^{-5}	***
$\kappa^{(1)}_{2001}$	-8.686	2.113	-4.111	3.9×10^{-5}	***
$\kappa^{(1)}_{2002}$	-8.409	2.068	-4.066	4.8×10^{-5}	***
$\kappa^{(1)}_{2003}$	-8.182	2.023	-4.046	5.2×10^{-5}	***
$\kappa^{(1)}_{2004}$	-7.967	1.976	-4.032	5.5×10^{-5}	***
$\kappa^{(1)}_{2005}$	-7.721	1.928	-4.004	6.2×10^{-5}	***
$\kappa^{(1)}_{2006}$	-7.486	1.879	-3.983	6.8×10^{-5}	***
$\kappa^{(1)}_{2007}$	-7.249	1.829	-3.963	7.4×10^{-5}	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(1)}_{2008}$	−7.029	1.779	−3.952	7.8×10^{-5}	***
$\kappa^{(1)}_{2009}$	−6.833	1.727	−3.958	7.6×10^{-5}	***
$\kappa^{(1)}_{2010}$	−6.618	1.674	−3.954	7.7×10^{-5}	***
$\kappa^{(1)}_{2011}$	−6.430	1.620	−3.970	7.2×10^{-5}	***
$\kappa^{(1)}_{2012}$	−6.197	1.565	−3.961	7.5×10^{-5}	***
$\kappa^{(1)}_{2013}$	−6.025	1.508	−3.996	6.4×10^{-5}	***
$\kappa^{(1)}_{2014}$	−5.857	1.452	−4.034	5.5×10^{-5}	***
$\kappa^{(2)}_{1900}$	5.912×10^{-1}	6.376×10^{-2}	9.273	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1901}$	5.710×10^{-1}	6.270×10^{-2}	9.107	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1902}$	5.524×10^{-1}	6.162×10^{-2}	8.965	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1903}$	5.359×10^{-1}	6.055×10^{-2}	8.852	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1904}$	5.217×10^{-1}	5.947×10^{-2}	8.772	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1905}$	4.996×10^{-1}	5.840×10^{-2}	8.555	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1906}$	4.870×10^{-1}	5.732×10^{-2}	8.496	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1907}$	4.716×10^{-1}	5.625×10^{-2}	8.383	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1908}$	4.539×10^{-1}	5.518×10^{-2}	8.226	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1909}$	4.359×10^{-1}	5.411×10^{-2}	8.055	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1910}$	4.234×10^{-1}	5.304×10^{-2}	7.983	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1911}$	4.082×10^{-1}	5.198×10^{-2}	7.853	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1912}$	3.912×10^{-1}	5.091×10^{-2}	7.683	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1913}$	3.752×10^{-1}	4.985×10^{-2}	7.526	5.2×10^{-14}	***
$\kappa^{(2)}_{1914}$	3.621×10^{-1}	4.879×10^{-2}	7.421	1.2×10^{-13}	***
$\kappa^{(2)}_{1915}$	3.461×10^{-1}	4.773×10^{-2}	7.251	4.1×10^{-13}	***
$\kappa^{(2)}_{1916}$	3.293×10^{-1}	4.667×10^{-2}	7.056	1.7×10^{-12}	***
$\kappa^{(2)}_{1917}$	3.176×10^{-1}	4.561×10^{-2}	6.963	3.3×10^{-12}	***
$\kappa^{(2)}_{1918}$	2.989×10^{-1}	4.456×10^{-2}	6.708	2.0×10^{-11}	***
$\kappa^{(2)}_{1919}$	2.868×10^{-1}	4.350×10^{-2}	6.592	4.4×10^{-11}	***
$\kappa^{(2)}_{1920}$	2.740×10^{-1}	4.245×10^{-2}	6.455	1.1×10^{-10}	***
$\kappa^{(2)}_{1921}$	2.644×10^{-1}	4.140×10^{-2}	6.385	1.7×10^{-10}	***
$\kappa^{(2)}_{1922}$	2.480×10^{-1}	4.035×10^{-2}	6.147	7.9×10^{-10}	***
$\kappa^{(2)}_{1923}$	2.365×10^{-1}	3.931×10^{-2}	6.016	1.8×10^{-9}	***
$\kappa^{(2)}_{1924}$	2.226×10^{-1}	3.826×10^{-2}	5.818	6.0×10^{-9}	***
$\kappa^{(2)}_{1925}$	2.122×10^{-1}	3.722×10^{-2}	5.702	1.2×10^{-8}	***
$\kappa^{(2)}_{1926}$	1.950×10^{-1}	3.618×10^{-2}	5.390	7.0×10^{-8}	***
$\kappa^{(2)}_{1927}$	1.869×10^{-1}	3.514×10^{-2}	5.320	1.0×10^{-7}	***
$\kappa^{(2)}_{1928}$	1.700×10^{-1}	3.410×10^{-2}	4.986	6.2×10^{-7}	***
$\kappa^{(2)}_{1929}$	1.570×10^{-1}	3.307×10^{-2}	4.748	2.1×10^{-6}	***
$\kappa^{(2)}_{1930}$	1.486×10^{-1}	3.204×10^{-2}	4.638	3.5×10^{-6}	***
$\kappa^{(2)}_{1931}$	1.384×10^{-1}	3.101×10^{-2}	4.464	8.1×10^{-6}	***
$\kappa^{(2)}_{1932}$	1.242×10^{-1}	2.999×10^{-2}	4.143	3.4×10^{-5}	***
$\kappa^{(2)}_{1933}$	1.153×10^{-1}	2.897×10^{-2}	3.980	6.9×10^{-5}	***
$\kappa^{(2)}_{1934}$	1.045×10^{-1}	2.795×10^{-2}	3.741	1.8×10^{-4}	***
$\kappa^{(2)}_{1935}$	9.274×10^{-2}	2.693×10^{-2}	3.444	5.7×10^{-4}	***
$\kappa^{(2)}_{1936}$	8.220×10^{-2}	2.592×10^{-2}	3.171	1.5×10^{-3}	**

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(2)}_{1937}$	7.033×10^{-2}	2.492×10^{-2}	2.823	4.8×10^{-3}	**
$\kappa^{(2)}_{1938}$	5.932×10^{-2}	2.392×10^{-2}	2.480	1.3×10^{-2}	*
$\kappa^{(2)}_{1939}$	5.442×10^{-2}	2.292×10^{-2}	2.374	1.8×10^{-2}	*
$\kappa^{(2)}_{1940}$	4.385×10^{-2}	2.194×10^{-2}	1.999	4.6×10^{-2}	*
$\kappa^{(2)}_{1941}$	3.438×10^{-2}	2.096×10^{-2}	1.640	1.0×10^{-1}	
$\kappa^{(2)}_{1942}$	2.263×10^{-2}	2.000×10^{-2}	1.132	2.6×10^{-1}	
$\kappa^{(2)}_{1943}$	1.411×10^{-2}	1.903×10^{-2}	7.420×10^{-1}	4.6×10^{-1}	
$\kappa^{(2)}_{1944}$	3.655×10^{-3}	1.808×10^{-2}	2.020×10^{-1}	8.4×10^{-1}	
$\kappa^{(2)}_{1945}$	-7.439×10^{-3}	1.714×10^{-2}	-4.340×10^{-1}	6.6×10^{-1}	
$\kappa^{(2)}_{1946}$	-1.177×10^{-2}	1.620×10^{-2}	-7.270×10^{-1}	4.7×10^{-1}	
$\kappa^{(2)}_{1947}$	-1.911×10^{-2}	1.530×10^{-2}	-1.249	2.1×10^{-1}	
$\kappa^{(2)}_{1948}$	-3.198×10^{-2}	1.442×10^{-2}	-2.218	2.7×10^{-2}	*
$\kappa^{(2)}_{1949}$	-3.703×10^{-2}	1.355×10^{-2}	-2.733	6.3×10^{-3}	**
$\kappa^{(2)}_{1950}$	-4.490×10^{-2}	1.271×10^{-2}	-3.532	4.1×10^{-4}	***
$\kappa^{(2)}_{1951}$	-4.990×10^{-2}	1.191×10^{-2}	-4.191	2.8×10^{-5}	***
$\kappa^{(2)}_{1952}$	-6.092×10^{-2}	1.115×10^{-2}	-5.462	4.7×10^{-8}	***
$\kappa^{(2)}_{1953}$	-6.938×10^{-2}	1.045×10^{-2}	-6.642	3.1×10^{-11}	***
$\kappa^{(2)}_{1954}$	-7.330×10^{-2}	9.795×10^{-3}	-7.483	7.3×10^{-14}	***
$\kappa^{(2)}_{1955}$	-8.318×10^{-2}	9.229×10^{-3}	-9.013	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1956}$	-9.012×10^{-2}	8.742×10^{-3}	-1.031×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1957}$	-9.325×10^{-2}	8.354×10^{-3}	-1.116×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1958}$	-1.010×10^{-1}	8.102×10^{-3}	-1.246×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1959}$	-1.061×10^{-1}	7.973×10^{-3}	-1.330×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1960}$	-1.134×10^{-1}	7.982×10^{-3}	-1.421×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1961}$	-1.202×10^{-1}	8.130×10^{-3}	-1.478×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1962}$	-1.239×10^{-1}	8.400×10^{-3}	-1.476×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1963}$	-1.288×10^{-1}	8.788×10^{-3}	-1.465×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1964}$	-1.363×10^{-1}	9.283×10^{-3}	-1.469×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1965}$	-1.397×10^{-1}	9.858×10^{-3}	-1.417×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1966}$	-1.447×10^{-1}	1.051×10^{-2}	-1.377×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1967}$	-1.516×10^{-1}	1.122×10^{-2}	-1.350×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1968}$	-1.547×10^{-1}	1.198×10^{-2}	-1.291×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1969}$	-1.606×10^{-1}	1.278×10^{-2}	-1.256×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1970}$	-1.647×10^{-1}	1.362×10^{-2}	-1.210×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1971}$	-1.694×10^{-1}	1.448×10^{-2}	-1.170×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1972}$	-1.718×10^{-1}	1.536×10^{-2}	-1.118×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1973}$	-1.751×10^{-1}	1.626×10^{-2}	-1.077×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1974}$	-1.799×10^{-1}	1.718×10^{-2}	-1.047×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1975}$	-1.828×10^{-1}	1.812×10^{-2}	-1.009×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1976}$	-1.834×10^{-1}	1.906×10^{-2}	-9.620	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1977}$	-1.870×10^{-1}	2.002×10^{-2}	-9.345	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1978}$	-1.909×10^{-1}	2.098×10^{-2}	-9.098	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1979}$	-1.914×10^{-1}	2.195×10^{-2}	-8.720	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1980}$	-1.930×10^{-1}	2.293×10^{-2}	-8.416	$<2.0\times10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(2)}_{1981}$	-1.946×10^{-1}	2.391×10^{-2}	-8.137	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1982}$	-1.974×10^{-1}	2.490×10^{-2}	-7.925	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1983}$	-1.959×10^{-1}	2.590×10^{-2}	-7.563	3.9×10^{-14}	***
$\kappa^{(2)}_{1984}$	-1.983×10^{-1}	2.690×10^{-2}	-7.373	1.7×10^{-13}	***
$\kappa^{(2)}_{1985}$	-1.971×10^{-1}	2.790×10^{-2}	-7.065	1.6×10^{-12}	***
$\kappa^{(2)}_{1986}$	-1.980×10^{-1}	2.890×10^{-2}	-6.852	7.3×10^{-12}	***
$\kappa^{(2)}_{1987}$	-1.985×10^{-1}	2.991×10^{-2}	-6.638	3.2×10^{-11}	***
$\kappa^{(2)}_{1988}$	-1.957×10^{-1}	3.092×10^{-2}	-6.328	2.5×10^{-10}	***
$\kappa^{(2)}_{1989}$	-1.951×10^{-1}	3.194×10^{-2}	-6.110	9.9×10^{-10}	***
$\kappa^{(2)}_{1990}$	-1.947×10^{-1}	3.295×10^{-2}	-5.909	3.5×10^{-9}	***
$\kappa^{(2)}_{1991}$	-1.937×10^{-1}	3.397×10^{-2}	-5.701	1.2×10^{-8}	***
$\kappa^{(2)}_{1992}$	-1.926×10^{-1}	3.499×10^{-2}	-5.506	3.7×10^{-8}	***
$\kappa^{(2)}_{1993}$	-1.886×10^{-1}	3.601×10^{-2}	-5.238	1.6×10^{-7}	***
$\kappa^{(2)}_{1994}$	-1.883×10^{-1}	3.704×10^{-2}	-5.083	3.7×10^{-7}	***
$\kappa^{(2)}_{1995}$	-1.866×10^{-1}	3.806×10^{-2}	-4.902	9.5×10^{-7}	***
$\kappa^{(2)}_{1996}$	-1.822×10^{-1}	3.909×10^{-2}	-4.661	3.1×10^{-6}	***
$\kappa^{(2)}_{1997}$	-1.788×10^{-1}	4.011×10^{-2}	-4.457	8.3×10^{-6}	***
$\kappa^{(2)}_{1998}$	-1.766×10^{-1}	4.114×10^{-2}	-4.292	1.8×10^{-5}	***
$\kappa^{(2)}_{1999}$	-1.714×10^{-1}	4.217×10^{-2}	-4.065	4.8×10^{-5}	***
$\kappa^{(2)}_{2000}$	-1.679×10^{-1}	4.320×10^{-2}	-3.887	1.0×10^{-4}	***
$\kappa^{(2)}_{2001}$	-1.631×10^{-1}	4.424×10^{-2}	-3.688	2.3×10^{-4}	***
$\kappa^{(2)}_{2002}$	-1.571×10^{-1}	4.527×10^{-2}	-3.472	5.2×10^{-4}	***
$\kappa^{(2)}_{2003}$	-1.526×10^{-1}	4.630×10^{-2}	-3.295	9.8×10^{-4}	***
$\kappa^{(2)}_{2004}$	-1.484×10^{-1}	4.734×10^{-2}	-3.134	1.7×10^{-3}	**
$\kappa^{(2)}_{2005}$	-1.426×10^{-1}	4.837×10^{-2}	-2.948	3.2×10^{-3}	**
$\kappa^{(2)}_{2006}$	-1.351×10^{-1}	4.941×10^{-2}	-2.735	6.2×10^{-3}	**
$\kappa^{(2)}_{2007}$	-1.287×10^{-1}	5.044×10^{-2}	-2.552	1.1×10^{-2}	*
$\kappa^{(2)}_{2008}$	-1.211×10^{-1}	5.148×10^{-2}	-2.353	1.9×10^{-2}	*
$\kappa^{(2)}_{2009}$	-1.154×10^{-1}	5.252×10^{-2}	-2.198	2.8×10^{-2}	*
$\kappa^{(2)}_{2010}$	-1.086×10^{-1}	5.356×10^{-2}	-2.027	4.3×10^{-2}	*
$\kappa^{(2)}_{2011}$	-9.973×10^{-2}	5.460×10^{-2}	-1.827	6.8×10^{-2}	.
$\kappa^{(2)}_{2012}$	-9.134×10^{-2}	5.564×10^{-2}	-1.642	1.0×10^{-1}	
$\kappa^{(2)}_{2013}$	-8.430×10^{-2}	5.670×10^{-2}	-1.487	1.4×10^{-1}	
$\kappa^{(2)}_{2014}$	-7.583×10^{-2}	5.769×10^{-2}	-1.314	1.9×10^{-1}	
$\kappa^{(3)}_{1900}$	8.146×10^{-3}	5.629×10^{-4}	1.447×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1901}$	7.965×10^{-3}	5.630×10^{-4}	1.415×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1902}$	7.910×10^{-3}	5.616×10^{-4}	1.409×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1903}$	7.740×10^{-3}	5.611×10^{-4}	1.379×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1904}$	7.796×10^{-3}	5.593×10^{-4}	1.394×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1905}$	7.538×10^{-3}	5.589×10^{-4}	1.349×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1906}$	7.692×10^{-3}	5.584×10^{-4}	1.377×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1907}$	7.362×10^{-3}	5.574×10^{-4}	1.321×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1908}$	7.170×10^{-3}	5.567×10^{-4}	1.288×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1909}$	7.164×10^{-3}	5.563×10^{-4}	1.288×10^{1}	$<2.0 \times 10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(3)}_{1910}$	7.168×10^{-3}	5.552×10^{-4}	1.291×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1911}$	7.103×10^{-3}	5.546×10^{-4}	1.281×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1912}$	6.934×10^{-3}	5.536×10^{-4}	1.253×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1913}$	6.731×10^{-3}	5.534×10^{-4}	1.216×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1914}$	6.728×10^{-3}	5.525×10^{-4}	1.218×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1915}$	6.451×10^{-3}	5.515×10^{-4}	1.170×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1916}$	6.454×10^{-3}	5.516×10^{-4}	1.170×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1917}$	6.427×10^{-3}	5.509×10^{-4}	1.167×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1918}$	6.444×10^{-3}	5.500×10^{-4}	1.172×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1919}$	6.090×10^{-3}	5.494×10^{-4}	1.109×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1920}$	6.001×10^{-3}	5.492×10^{-4}	1.093×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1921}$	5.990×10^{-3}	5.487×10^{-4}	1.092×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1922}$	5.762×10^{-3}	5.477×10^{-4}	1.052×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1923}$	5.920×10^{-3}	5.479×10^{-4}	1.081×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1924}$	5.706×10^{-3}	5.472×10^{-4}	1.043×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1925}$	5.772×10^{-3}	5.468×10^{-4}	1.056×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1926}$	5.400×10^{-3}	5.463×10^{-4}	9.884	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1927}$	5.347×10^{-3}	5.453×10^{-4}	9.804	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1928}$	5.160×10^{-3}	5.455×10^{-4}	9.459	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1929}$	4.995×10^{-3}	5.450×10^{-4}	9.166	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1930}$	5.198×10^{-3}	5.446×10^{-4}	9.545	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1931}$	4.874×10^{-3}	5.437×10^{-4}	8.963	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1932}$	4.783×10^{-3}	5.439×10^{-4}	8.793	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1933}$	4.912×10^{-3}	5.435×10^{-4}	9.039	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1934}$	4.784×10^{-3}	5.430×10^{-4}	8.810	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1935}$	4.470×10^{-3}	5.423×10^{-4}	8.241	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1936}$	4.423×10^{-3}	5.419×10^{-4}	8.162	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1937}$	4.307×10^{-3}	5.415×10^{-4}	7.955	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1938}$	4.176×10^{-3}	5.414×10^{-4}	7.713	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1939}$	4.165×10^{-3}	5.407×10^{-4}	7.702	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1940}$	4.073×10^{-3}	5.405×10^{-4}	7.536	4.9×10^{-14}	***
$\kappa^{(3)}_{1941}$	3.892×10^{-3}	5.403×10^{-4}	7.203	5.9×10^{-13}	***
$\kappa^{(3)}_{1942}$	3.951×10^{-3}	5.409×10^{-4}	7.306	2.8×10^{-13}	***
$\kappa^{(3)}_{1943}$	3.941×10^{-3}	5.402×10^{-4}	7.296	3.0×10^{-13}	***
$\kappa^{(3)}_{1944}$	3.566×10^{-3}	5.396×10^{-4}	6.608	3.9×10^{-11}	***
$\kappa^{(3)}_{1945}$	3.407×10^{-3}	5.392×10^{-4}	6.318	2.6×10^{-10}	***
$\kappa^{(3)}_{1946}$	3.480×10^{-3}	5.386×10^{-4}	6.461	1.0×10^{-10}	***
$\kappa^{(3)}_{1947}$	3.331×10^{-3}	5.382×10^{-4}	6.190	6.0×10^{-10}	***
$\kappa^{(3)}_{1948}$	3.222×10^{-3}	5.383×10^{-4}	5.985	2.2×10^{-9}	***
$\kappa^{(3)}_{1949}$	3.197×10^{-3}	5.378×10^{-4}	5.945	2.8×10^{-9}	***
$\kappa^{(3)}_{1950}$	2.976×10^{-3}	5.374×10^{-4}	5.537	3.1×10^{-8}	***
$\kappa^{(3)}_{1951}$	2.957×10^{-3}	5.372×10^{-4}	5.504	3.7×10^{-8}	***
$\kappa^{(3)}_{1952}$	2.853×10^{-3}	5.371×10^{-4}	5.311	1.1×10^{-7}	***
$\kappa^{(3)}_{1953}$	2.674×10^{-3}	5.368×10^{-4}	4.982	6.3×10^{-7}	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z \rvert)$	Signif. code
$\kappa^{(3)}_{1954}$	2.720×10^{-3}	5.365×10^{-4}	5.070	4.0×10^{-7}	***
$\kappa^{(3)}_{1955}$	2.507×10^{-3}	5.365×10^{-4}	4.674	3.0×10^{-6}	***
$\kappa^{(3)}_{1956}$	2.366×10^{-3}	5.361×10^{-4}	4.413	1.0×10^{-5}	***
$\kappa^{(3)}_{1957}$	2.461×10^{-3}	5.355×10^{-4}	4.595	4.3×10^{-6}	***
$\kappa^{(3)}_{1958}$	2.254×10^{-3}	5.356×10^{-4}	4.208	2.6×10^{-5}	***
$\kappa^{(3)}_{1959}$	2.242×10^{-3}	5.354×10^{-4}	4.188	2.8×10^{-5}	***
$\kappa^{(3)}_{1960}$	1.964×10^{-3}	5.352×10^{-4}	3.671	2.4×10^{-4}	***
$\kappa^{(3)}_{1961}$	1.793×10^{-3}	5.352×10^{-4}	3.350	8.1×10^{-4}	***
$\kappa^{(3)}_{1962}$	1.752×10^{-3}	5.348×10^{-4}	3.275	1.1×10^{-3}	**
$\kappa^{(3)}_{1963}$	1.813×10^{-3}	5.346×10^{-4}	3.391	7.0×10^{-4}	***
$\kappa^{(3)}_{1964}$	1.750×10^{-3}	5.345×10^{-4}	3.273	1.1×10^{-3}	**
$\kappa^{(3)}_{1965}$	1.638×10^{-3}	5.343×10^{-4}	3.067	2.2×10^{-3}	**
$\kappa^{(3)}_{1966}$	1.524×10^{-3}	5.341×10^{-4}	2.853	4.3×10^{-3}	**
$\kappa^{(3)}_{1967}$	1.272×10^{-3}	5.340×10^{-4}	2.383	1.7×10^{-2}	*
$\kappa^{(3)}_{1968}$	1.251×10^{-3}	5.337×10^{-4}	2.344	1.9×10^{-2}	*
$\kappa^{(3)}_{1969}$	1.251×10^{-3}	5.335×10^{-4}	2.346	1.9×10^{-2}	*
$\kappa^{(3)}_{1970}$	1.251×10^{-3}	5.335×10^{-4}	2.345	1.9×10^{-2}	*
$\kappa^{(3)}_{1971}$	1.074×10^{-3}	5.334×10^{-4}	2.013	4.4×10^{-2}	*
$\kappa^{(3)}_{1972}$	1.074×10^{-3}	5.331×10^{-4}	2.014	4.4×10^{-2}	*
$\kappa^{(3)}_{1973}$	9.247×10^{-4}	5.330×10^{-4}	1.735	8.3×10^{-2}	.
$\kappa^{(3)}_{1974}$	8.128×10^{-4}	5.329×10^{-4}	1.525	1.3×10^{-1}	
$\kappa^{(3)}_{1975}$	6.854×10^{-4}	5.328×10^{-4}	1.287	2.0×10^{-1}	
$\kappa^{(3)}_{1976}$	6.283×10^{-4}	5.326×10^{-4}	1.180	2.4×10^{-1}	
$\kappa^{(3)}_{1977}$	5.489×10^{-4}	5.326×10^{-4}	1.030	3.0×10^{-1}	
$\kappa^{(3)}_{1978}$	3.152×10^{-4}	5.325×10^{-4}	5.920×10^{-1}	5.5×10^{-1}	
$\kappa^{(3)}_{1979}$	3.351×10^{-4}	5.324×10^{-4}	6.290×10^{-1}	5.3×10^{-1}	
$\kappa^{(3)}_{1980}$	3.037×10^{-4}	5.322×10^{-4}	5.710×10^{-1}	5.7×10^{-1}	
$\kappa^{(3)}_{1981}$	1.344×10^{-4}	5.320×10^{-4}	2.530×10^{-1}	8.0×10^{-1}	
$\kappa^{(3)}_{1982}$	2.614×10^{-5}	5.320×10^{-4}	4.900×10^{-2}	9.6×10^{-1}	
$\kappa^{(3)}_{1983}$	-2.063×10^{-5}	5.319×10^{-4}	-3.900×10^{-2}	9.7×10^{-1}	
$\kappa^{(3)}_{1984}$	-1.684×10^{-4}	5.320×10^{-4}	-3.170×10^{-1}	7.5×10^{-1}	
$\kappa^{(3)}_{1985}$	-2.607×10^{-4}	5.317×10^{-4}	-4.900×10^{-1}	6.2×10^{-1}	
$\kappa^{(3)}_{1986}$	-4.055×10^{-4}	5.318×10^{-4}	-7.630×10^{-1}	4.5×10^{-1}	
$\kappa^{(3)}_{1987}$	-4.648×10^{-4}	5.317×10^{-4}	-8.740×10^{-1}	3.8×10^{-1}	
$\kappa^{(3)}_{1988}$	-4.826×10^{-4}	5.316×10^{-4}	-9.080×10^{-1}	3.6×10^{-1}	
$\kappa^{(3)}_{1989}$	-6.827×10^{-4}	5.317×10^{-4}	-1.284	2.0×10^{-1}	
$\kappa^{(3)}_{1990}$	-8.431×10^{-4}	5.317×10^{-4}	-1.586	1.1×10^{-1}	
$\kappa^{(3)}_{1991}$	-8.395×10^{-4}	5.316×10^{-4}	-1.579	1.1×10^{-1}	
$\kappa^{(3)}_{1992}$	-9.441×10^{-4}	5.316×10^{-4}	-1.776	7.6×10^{-2}	.
$\kappa^{(3)}_{1993}$	-1.116×10^{-3}	5.315×10^{-4}	-2.099	3.6×10^{-2}	*
$\kappa^{(3)}_{1994}$	-1.298×10^{-3}	5.317×10^{-4}	-2.441	1.5×10^{-2}	*
$\kappa^{(3)}_{1995}$	-1.465×10^{-3}	5.317×10^{-4}	-2.755	5.9×10^{-3}	**
$\kappa^{(3)}_{1996}$	-1.544×10^{-3}	5.316×10^{-4}	-2.905	3.7×10^{-3}	**
$\kappa^{(3)}_{1997}$	-1.697×10^{-3}	5.316×10^{-4}	-3.192	1.4×10^{-3}	**

continued ...

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(3)}_{1998}$	-1.886×10^{-3}	5.316×10^{-4}	-3.548	3.9×10^{-4}	***
$\kappa^{(3)}_{1999}$	-2.019×10^{-3}	5.316×10^{-4}	-3.798	1.5×10^{-4}	***
$\kappa^{(3)}_{2000}$	-2.149×10^{-3}	5.315×10^{-4}	-4.042	5.3×10^{-5}	***
$\kappa^{(3)}_{2001}$	-2.374×10^{-3}	5.315×10^{-4}	-4.466	8.0×10^{-6}	***
$\kappa^{(3)}_{2002}$	-2.512×10^{-3}	5.315×10^{-4}	-4.727	2.3×10^{-6}	***
$\kappa^{(3)}_{2003}$	-2.564×10^{-3}	5.315×10^{-4}	-4.824	1.4×10^{-6}	***
$\kappa^{(3)}_{2004}$	-2.753×10^{-3}	5.315×10^{-4}	-5.179	2.2×10^{-7}	***
$\kappa^{(3)}_{2005}$	-2.866×10^{-3}	5.315×10^{-4}	-5.393	6.9×10^{-8}	***
$\kappa^{(3)}_{2006}$	-3.003×10^{-3}	5.315×10^{-4}	-5.650	1.6×10^{-8}	***
$\kappa^{(3)}_{2007}$	-3.253×10^{-3}	5.315×10^{-4}	-6.120	9.3×10^{-10}	***
$\kappa^{(3)}_{2008}$	-3.332×10^{-3}	5.315×10^{-4}	-6.269	3.6×10^{-10}	***
$\kappa^{(3)}_{2009}$	-3.468×10^{-3}	5.316×10^{-4}	-6.523	6.9×10^{-11}	***
$\kappa^{(3)}_{2010}$	-3.629×10^{-3}	5.317×10^{-4}	-6.825	8.8×10^{-12}	***
$\kappa^{(3)}_{2011}$	-3.834×10^{-3}	5.318×10^{-4}	-7.209	5.6×10^{-13}	***
$\kappa^{(3)}_{2012}$	-3.898×10^{-3}	5.318×10^{-4}	-7.331	2.3×10^{-13}	***
$\kappa^{(3)}_{2013}$	-4.023×10^{-3}	5.367×10^{-4}	-7.495	6.6×10^{-14}	***
$\kappa^{(3)}_{2014}$	-4.009×10^{-3}	5.262×10^{-4}	-7.619	2.6×10^{-14}	***
γ_{1800}	1.839	1.110	1.657	9.8×10^{-2}	.
γ_{1801}	2.760	1.043	2.647	8.1×10^{-3}	**
γ_{1802}	3.873	1.023	3.786	1.5×10^{-4}	***
γ_{1803}	4.252	1.017	4.181	2.9×10^{-5}	***
γ_{1804}	4.936	1.026	4.809	1.5×10^{-6}	***
γ_{1805}	5.704	1.045	5.456	4.9×10^{-8}	***
γ_{1806}	6.343	1.069	5.933	3.0×10^{-9}	***
γ_{1807}	6.951	1.098	6.333	2.4×10^{-10}	***
γ_{1808}	7.677	1.130	6.793	1.1×10^{-11}	***
γ_{1809}	8.365	1.166	7.171	7.4×10^{-13}	***
γ_{1810}	8.961	1.205	7.435	1.1×10^{-13}	***
γ_{1811}	9.592	1.247	7.695	$<2.0\times10^{-16}$	***
γ_{1812}	1.019×10^{1}	1.290	7.899	$<2.0\times10^{-16}$	***
γ_{1813}	1.078×10^{1}	1.334	8.083	$<2.0\times10^{-16}$	***
γ_{1814}	1.132×10^{1}	1.380	8.201	$<2.0\times10^{-16}$	***
γ_{1815}	1.187×10^{1}	1.426	8.324	$<2.0\times10^{-16}$	***
γ_{1816}	1.241×10^{1}	1.474	8.422	$<2.0\times10^{-16}$	***
γ_{1817}	1.290×10^{1}	1.521	8.482	$<2.0\times10^{-16}$	***
γ_{1818}	1.335×10^{1}	1.569	8.509	$<2.0\times10^{-16}$	***
γ_{1819}	1.385×10^{1}	1.617	8.564	$<2.0\times10^{-16}$	***
γ_{1820}	1.426×10^{1}	1.666	8.561	$<2.0\times10^{-16}$	***
γ_{1821}	1.471×10^{1}	1.713	8.585	$<2.0\times10^{-16}$	***
γ_{1822}	1.512×10^{1}	1.761	8.587	$<2.0\times10^{-16}$	***
γ_{1823}	1.549×10^{1}	1.809	8.566	$<2.0\times10^{-16}$	***
γ_{1824}	1.589×10^{1}	1.856	8.564	$<2.0\times10^{-16}$	***
γ_{1825}	1.627×10^{1}	1.902	8.553	$<2.0\times10^{-16}$	***
γ_{1826}	1.662×10^{1}	1.948	8.528	$<2.0\times10^{-16}$	***
γ_{1827}	1.694×10^{1}	1.994	8.497	$<2.0\times10^{-16}$	***
continued …					

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
γ_{1828}	1.723×10^{1}	2.039	8.448	$<2.0 \times 10^{-16}$	***
γ_{1829}	1.759×10^{1}	2.083	8.443	$<2.0 \times 10^{-16}$	***
γ_{1830}	1.784×10^{1}	2.127	8.388	$<2.0 \times 10^{-16}$	***
γ_{1831}	1.810×10^{1}	2.170	8.344	$<2.0 \times 10^{-16}$	***
γ_{1832}	1.833×10^{1}	2.212	8.287	$<2.0 \times 10^{-16}$	***
γ_{1833}	1.858×10^{1}	2.253	8.248	$<2.0 \times 10^{-16}$	***
γ_{1834}	1.885×10^{1}	2.294	8.217	$<2.0 \times 10^{-16}$	***
γ_{1835}	1.904×10^{1}	2.334	8.161	$<2.0 \times 10^{-16}$	***
γ_{1836}	1.922×10^{1}	2.372	8.101	$<2.0 \times 10^{-16}$	***
γ_{1837}	1.943×10^{1}	2.410	8.060	$<2.0 \times 10^{-16}$	***
γ_{1838}	1.960×10^{1}	2.448	8.006	$<2.0 \times 10^{-16}$	***
γ_{1839}	1.970×10^{1}	2.484	7.932	$<2.0 \times 10^{-16}$	***
γ_{1840}	1.991×10^{1}	2.519	7.901	$<2.0 \times 10^{-16}$	***
γ_{1841}	2.003×10^{1}	2.554	7.842	$<2.0 \times 10^{-16}$	***
γ_{1842}	2.015×10^{1}	2.588	7.788	$<2.0 \times 10^{-16}$	***
γ_{1843}	2.023×10^{1}	2.620	7.722	$<2.0 \times 10^{-16}$	***
γ_{1844}	2.034×10^{1}	2.652	7.672	$<2.0 \times 10^{-16}$	***
γ_{1845}	2.043×10^{1}	2.683	7.615	2.7×10^{-14}	***
γ_{1846}	2.049×10^{1}	2.713	7.553	4.3×10^{-14}	***
γ_{1847}	2.054×10^{1}	2.742	7.494	6.7×10^{-14}	***
γ_{1848}	2.058×10^{1}	2.769	7.431	1.1×10^{-13}	***
γ_{1849}	2.062×10^{1}	2.796	7.372	1.7×10^{-13}	***
γ_{1850}	2.066×10^{1}	2.822	7.321	2.5×10^{-13}	***
γ_{1851}	2.063×10^{1}	2.847	7.245	4.3×10^{-13}	***
γ_{1852}	2.066×10^{1}	2.872	7.196	6.2×10^{-13}	***
γ_{1853}	2.060×10^{1}	2.895	7.117	1.1×10^{-12}	***
γ_{1854}	2.062×10^{1}	2.917	7.068	1.6×10^{-12}	***
γ_{1855}	2.059×10^{1}	2.938	7.008	2.4×10^{-12}	***
γ_{1856}	2.055×10^{1}	2.958	6.948	3.7×10^{-12}	***
γ_{1857}	2.047×10^{1}	2.977	6.878	6.1×10^{-12}	***
γ_{1858}	2.044×10^{1}	2.995	6.823	8.9×10^{-12}	***
γ_{1859}	2.036×10^{1}	3.012	6.759	1.4×10^{-11}	***
γ_{1860}	2.030×10^{1}	3.028	6.703	2.0×10^{-11}	***
γ_{1861}	2.019×10^{1}	3.043	6.634	3.3×10^{-11}	***
γ_{1862}	2.007×10^{1}	3.057	6.566	5.2×10^{-11}	***
γ_{1863}	1.999×10^{1}	3.070	6.509	7.6×10^{-11}	***
γ_{1864}	1.986×10^{1}	3.083	6.444	1.2×10^{-10}	***
γ_{1865}	1.971×10^{1}	3.094	6.372	1.9×10^{-10}	***
γ_{1866}	1.961×10^{1}	3.104	6.317	2.7×10^{-10}	***
γ_{1867}	1.946×10^{1}	3.113	6.252	4.0×10^{-10}	***
γ_{1868}	1.931×10^{1}	3.121	6.189	6.1×10^{-10}	***
γ_{1869}	1.913×10^{1}	3.128	6.118	9.5×10^{-10}	***
γ_{1870}	1.899×10^{1}	3.134	6.059	1.4×10^{-9}	***
γ_{1871}	1.881×10^{1}	3.138	5.992	2.1×10^{-9}	***
γ_{1872}	1.863×10^{1}	3.142	5.928	3.1×10^{-9}	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
γ_{1873}	1.844×10^{1}	3.145	5.862	4.6×10^{-9}	***
γ_{1874}	1.825×10^{1}	3.147	5.797	6.7×10^{-9}	***
γ_{1875}	1.803×10^{1}	3.148	5.727	1.0×10^{-8}	***
γ_{1876}	1.783×10^{1}	3.148	5.665	1.5×10^{-8}	***
γ_{1877}	1.761×10^{1}	3.147	5.598	2.2×10^{-8}	***
γ_{1878}	1.739×10^{1}	3.144	5.531	3.2×10^{-8}	***
γ_{1879}	1.716×10^{1}	3.141	5.462	4.7×10^{-8}	***
γ_{1880}	1.693×10^{1}	3.137	5.397	6.8×10^{-8}	***
γ_{1881}	1.668×10^{1}	3.132	5.327	10.0×10^{-8}	***
γ_{1882}	1.646×10^{1}	3.125	5.265	1.4×10^{-7}	***
γ_{1883}	1.618×10^{1}	3.118	5.190	2.1×10^{-7}	***
γ_{1884}	1.593×10^{1}	3.109	5.123	3.0×10^{-7}	***
γ_{1885}	1.566×10^{1}	3.100	5.051	4.4×10^{-7}	***
γ_{1886}	1.541×10^{1}	3.090	4.987	6.1×10^{-7}	***
γ_{1887}	1.512×10^{1}	3.078	4.913	9.0×10^{-7}	***
γ_{1888}	1.488×10^{1}	3.066	4.853	1.2×10^{-6}	***
γ_{1889}	1.457×10^{1}	3.052	4.774	1.8×10^{-6}	***
γ_{1890}	1.429×10^{1}	3.037	4.705	2.5×10^{-6}	***
γ_{1891}	1.400×10^{1}	3.022	4.634	3.6×10^{-6}	***
γ_{1892}	1.370×10^{1}	3.005	4.559	5.1×10^{-6}	***
γ_{1893}	1.340×10^{1}	2.988	4.484	7.3×10^{-6}	***
γ_{1894}	1.311×10^{1}	2.969	4.415	1.0×10^{-5}	***
γ_{1895}	1.282×10^{1}	2.949	4.347	1.4×10^{-5}	***
γ_{1896}	1.250×10^{1}	2.928	4.270	2.0×10^{-5}	***
γ_{1897}	1.219×10^{1}	2.907	4.194	2.7×10^{-5}	***
γ_{1898}	1.190×10^{1}	2.884	4.126	3.7×10^{-5}	***
γ_{1899}	1.157×10^{1}	2.860	4.047	5.2×10^{-5}	***
γ_{1900}	1.126×10^{1}	2.835	3.970	7.2×10^{-5}	***
γ_{1901}	1.093×10^{1}	2.809	3.893	9.9×10^{-5}	***
γ_{1902}	1.061×10^{1}	2.782	3.815	1.4×10^{-4}	***
γ_{1903}	1.029×10^{1}	2.754	3.735	1.9×10^{-4}	***
γ_{1904}	9.972	2.725	3.660	2.5×10^{-4}	***
γ_{1905}	9.651	2.695	3.581	3.4×10^{-4}	***
γ_{1906}	9.332	2.664	3.503	4.6×10^{-4}	***
γ_{1907}	9.011	2.631	3.424	6.2×10^{-4}	***
γ_{1908}	8.700	2.598	3.348	8.1×10^{-4}	***
γ_{1909}	8.368	2.564	3.264	1.1×10^{-3}	**
γ_{1910}	8.062	2.529	3.188	1.4×10^{-3}	**
γ_{1911}	7.740	2.492	3.106	1.9×10^{-3}	**
γ_{1912}	7.430	2.455	3.026	2.5×10^{-3}	**
γ_{1913}	7.106	2.417	2.940	3.3×10^{-3}	**
γ_{1914}	6.807	2.377	2.863	4.2×10^{-3}	**
γ_{1915}	6.511	2.337	2.786	5.3×10^{-3}	**
γ_{1916}	6.193	2.295	2.698	7.0×10^{-3}	**
γ_{1917}	5.892	2.253	2.616	8.9×10^{-3}	**

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
γ_{1918}	5.598	2.209	2.534	1.1×10^{-2}	*
γ_{1919}	5.277	2.164	2.438	1.5×10^{-2}	*
γ_{1920}	5.047	2.119	2.382	1.7×10^{-2}	*
γ_{1921}	4.714	2.072	2.275	2.3×10^{-2}	*
γ_{1922}	4.468	2.024	2.207	2.7×10^{-2}	*
γ_{1923}	4.172	1.975	2.112	3.5×10^{-2}	*
γ_{1924}	3.908	1.926	2.030	4.2×10^{-2}	*
γ_{1925}	3.643	1.875	1.943	5.2×10^{-2}	.
γ_{1926}	3.368	1.823	1.848	6.5×10^{-2}	.
γ_{1927}	3.146	1.770	1.777	7.5×10^{-2}	.
γ_{1928}	2.876	1.716	1.676	9.4×10^{-2}	.
γ_{1929}	2.665	1.661	1.605	1.1×10^{-1}	
γ_{1930}	2.433	1.605	1.516	1.3×10^{-1}	
γ_{1931}	2.236	1.547	1.445	1.5×10^{-1}	
γ_{1932}	2.024	1.489	1.359	1.7×10^{-1}	
γ_{1933}	1.821	1.430	1.273	2.0×10^{-1}	
γ_{1934}	1.627	1.370	1.188	2.3×10^{-1}	
γ_{1935}	1.461	1.308	1.117	2.6×10^{-1}	
γ_{1936}	1.308	1.246	1.049	2.9×10^{-1}	
γ_{1937}	1.126	1.183	9.520×10^{-1}	3.4×10^{-1}	
γ_{1938}	9.793×10^{-1}	1.118	8.760×10^{-1}	3.8×10^{-1}	
γ_{1939}	8.437×10^{-1}	1.053	8.010×10^{-1}	4.2×10^{-1}	
γ_{1940}	7.510×10^{-1}	9.863×10^{-1}	7.610×10^{-1}	4.5×10^{-1}	
γ_{1941}	6.043×10^{-1}	9.187×10^{-1}	6.580×10^{-1}	5.1×10^{-1}	
γ_{1942}	5.498×10^{-1}	8.501×10^{-1}	6.470×10^{-1}	5.2×10^{-1}	
γ_{1943}	4.161×10^{-1}	7.805×10^{-1}	5.330×10^{-1}	5.9×10^{-1}	
γ_{1944}	3.659×10^{-1}	7.099×10^{-1}	5.150×10^{-1}	6.1×10^{-1}	
γ_{1945}	2.833×10^{-1}	6.382×10^{-1}	4.440×10^{-1}	6.6×10^{-1}	
γ_{1946}	2.203×10^{-1}	5.656×10^{-1}	3.890×10^{-1}	7.0×10^{-1}	
γ_{1947}	1.894×10^{-1}	4.920×10^{-1}	3.850×10^{-1}	7.0×10^{-1}	
γ_{1948}	1.181×10^{-1}	4.174×10^{-1}	2.830×10^{-1}	7.8×10^{-1}	
γ_{1949}	8.820×10^{-2}	3.421×10^{-1}	2.580×10^{-1}	8.0×10^{-1}	
γ_{1950}	6.991×10^{-2}	2.662×10^{-1}	2.630×10^{-1}	7.9×10^{-1}	
γ_{1951}	6.857×10^{-2}	1.903×10^{-1}	3.600×10^{-1}	7.2×10^{-1}	
γ_{1952}	-1.583×10^{-2}	1.174×10^{-1}	-1.350×10^{-1}	8.9×10^{-1}	

B.1.4 | KAN Model

Table B.4: Regression table of the KAN model for Swedish females. All 230 parameters are significant even on the 0.1% level. p-value significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' . ' 0.1 ' ' 1

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(1)}_{1900}$	−4.191	1.750×10^{-2}	-2.395×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1901}$	−4.264	1.807×10^{-2}	-2.359×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1902}$	−4.210	1.765×10^{-2}	-2.385×10^{2}	$<2.0 \times 10^{-16}$	***
continued …					

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\|z\|)$	Signif. code
$\kappa^{(1)}_{1903}$	−4.295	1.807×10^{-2}	-2.377×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1904}$	−4.231	1.744×10^{-2}	-2.426×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1905}$	−4.201	1.730×10^{-2}	-2.428×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1906}$	−4.277	1.781×10^{-2}	-2.402×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1907}$	−4.293	1.759×10^{-2}	-2.441×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1908}$	−4.273	1.739×10^{-2}	-2.458×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1909}$	−4.273	1.749×10^{-2}	-2.443×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1910}$	−4.291	1.740×10^{-2}	-2.467×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1911}$	−4.310	1.746×10^{-2}	-2.469×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1912}$	−4.235	1.689×10^{-2}	-2.508×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1913}$	−4.292	1.721×10^{-2}	-2.494×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1914}$	−4.288	1.704×10^{-2}	-2.516×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1915}$	−4.227	1.646×10^{-2}	-2.568×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1916}$	−4.270	1.692×10^{-2}	-2.524×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1917}$	−4.291	1.688×10^{-2}	-2.542×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1918}$	−4.145	1.607×10^{-2}	-2.579×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1919}$	−4.227	1.619×10^{-2}	-2.611×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1920}$	−4.305	1.656×10^{-2}	-2.600×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1921}$	−4.364	1.672×10^{-2}	-2.611×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1922}$	−4.247	1.587×10^{-2}	-2.676×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1923}$	−4.345	1.658×10^{-2}	-2.621×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1924}$	−4.313	1.617×10^{-2}	-2.667×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1925}$	−4.336	1.621×10^{-2}	-2.674×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1926}$	−4.302	1.593×10^{-2}	-2.702×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1927}$	−4.292	1.561×10^{-2}	-2.749×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1928}$	−4.296	1.588×10^{-2}	-2.706×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1929}$	−4.274	1.575×10^{-2}	-2.714×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1930}$	−4.294	1.595×10^{-2}	-2.692×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1931}$	−4.273	1.557×10^{-2}	-2.745×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1932}$	−4.310	1.597×10^{-2}	-2.700×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1933}$	−4.326	1.609×10^{-2}	-2.689×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1934}$	−4.337	1.606×10^{-2}	-2.701×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1935}$	−4.320	1.575×10^{-2}	-2.744×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1936}$	−4.296	1.557×10^{-2}	-2.759×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1937}$	−4.271	1.539×10^{-2}	-2.776×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1938}$	−4.323	1.563×10^{-2}	-2.767×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1939}$	−4.362	1.555×10^{-2}	-2.806×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1940}$	−4.349	1.539×10^{-2}	-2.826×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1941}$	−4.396	1.551×10^{-2}	-2.834×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1942}$	−4.479	1.619×10^{-2}	-2.767×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1943}$	−4.463	1.588×10^{-2}	-2.810×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1944}$	−4.444	1.540×10^{-2}	-2.885×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1945}$	−4.425	1.514×10^{-2}	-2.922×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1946}$	−4.466	1.513×10^{-2}	-2.952×10^{2}	$<2.0 \times 10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(1)}_{1947}$	−4.466	1.486×10^{-2}	-3.005×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1948}$	−4.475	1.499×10^{-2}	-2.985×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1949}$	−4.503	1.489×10^{-2}	-3.024×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1950}$	−4.531	1.483×10^{-2}	-3.056×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1951}$	−4.590	1.498×10^{-2}	-3.064×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1952}$	−4.571	1.489×10^{-2}	-3.070×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1953}$	−4.571	1.475×10^{-2}	-3.099×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1954}$	−4.629	1.490×10^{-2}	-3.107×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1955}$	−4.654	1.498×10^{-2}	-3.108×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1956}$	−4.670	1.487×10^{-2}	-3.141×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1957}$	−4.673	1.468×10^{-2}	-3.184×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1958}$	−4.727	1.486×10^{-2}	-3.181×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1959}$	−4.765	1.493×10^{-2}	-3.191×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1960}$	−4.769	1.471×10^{-2}	-3.242×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1961}$	−4.820	1.484×10^{-2}	-3.248×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1962}$	−4.841	1.471×10^{-2}	-3.292×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1963}$	−4.858	1.471×10^{-2}	-3.302×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1964}$	−4.846	1.459×10^{-2}	-3.322×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1965}$	−4.901	1.463×10^{-2}	-3.351×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1966}$	−4.932	1.462×10^{-2}	-3.373×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1967}$	−4.950	1.454×10^{-2}	-3.405×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1968}$	−4.949	1.431×10^{-2}	-3.458×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1969}$	−4.925	1.412×10^{-2}	-3.487×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1970}$	−4.983	1.437×10^{-2}	-3.467×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1971}$	−5.009	1.433×10^{-2}	-3.496×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1972}$	−5.020	1.421×10^{-2}	-3.532×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1973}$	−5.056	1.422×10^{-2}	-3.556×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1974}$	−5.055	1.414×10^{-2}	-3.575×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1975}$	−5.070	1.406×10^{-2}	-3.606×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1976}$	−5.109	1.407×10^{-2}	-3.631×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1977}$	−5.150	1.430×10^{-2}	-3.601×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1978}$	−5.151	1.420×10^{-2}	-3.627×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1979}$	−5.179	1.425×10^{-2}	-3.634×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1980}$	−5.175	1.413×10^{-2}	-3.663×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1981}$	−5.204	1.412×10^{-2}	-3.685×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1982}$	−5.216	1.419×10^{-2}	-3.676×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1983}$	−5.287	1.439×10^{-2}	-3.673×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1984}$	−5.283	1.437×10^{-2}	-3.677×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1985}$	−5.304	1.431×10^{-2}	-3.708×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1986}$	−5.329	1.442×10^{-2}	-3.696×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1987}$	−5.330	1.444×10^{-2}	-3.691×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1988}$	−5.352	1.440×10^{-2}	-3.718×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1989}$	−5.434	1.485×10^{-2}	-3.661×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1990}$	−5.420	1.471×10^{-2}	-3.685×10^{2}	$<2.0 \times 10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(1)}_{1991}$	−5.421	1.476×10^{-2}	-3.673×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1992}$	−5.423	1.479×10^{-2}	-3.668×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1993}$	−5.477	1.489×10^{-2}	-3.679×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1994}$	−5.523	1.530×10^{-2}	-3.611×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1995}$	−5.511	1.519×10^{-2}	-3.628×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1996}$	−5.552	1.532×10^{-2}	-3.624×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1997}$	−5.592	1.551×10^{-2}	-3.607×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1998}$	−5.591	1.552×10^{-2}	-3.603×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1999}$	−5.618	1.550×10^{-2}	-3.623×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2000}$	−5.625	1.556×10^{-2}	-3.615×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2001}$	−5.656	1.562×10^{-2}	-3.622×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2002}$	−5.682	1.560×10^{-2}	-3.642×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2003}$	−5.692	1.568×10^{-2}	-3.631×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2004}$	−5.694	1.564×10^{-2}	-3.640×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2005}$	−5.696	1.550×10^{-2}	-3.674×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2006}$	−5.743	1.552×10^{-2}	-3.700×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2007}$	−5.753	1.536×10^{-2}	-3.746×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2008}$	−5.791	1.538×10^{-2}	-3.765×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2009}$	−5.784	1.528×10^{-2}	-3.785×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2010}$	−5.787	1.513×10^{-2}	-3.825×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2011}$	−5.857	1.529×10^{-2}	-3.831×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2012}$	−5.857	1.510×10^{-2}	-3.880×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2013}$	−5.862	1.507×10^{-2}	-3.889×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2014}$	−5.896	1.518×10^{-2}	-3.886×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1900}$	1.118×10^{-1}	1.075×10^{-3}	1.040×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1901}$	1.090×10^{-1}	1.099×10^{-3}	9.920×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1902}$	1.071×10^{-1}	1.071×10^{-3}	9.990×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1903}$	1.089×10^{-1}	1.081×10^{-3}	1.007×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1904}$	1.099×10^{-1}	1.043×10^{-3}	1.054×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1905}$	1.063×10^{-1}	1.035×10^{-3}	1.027×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1906}$	1.056×10^{-1}	1.052×10^{-3}	1.004×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1907}$	1.117×10^{-1}	1.029×10^{-3}	1.086×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1908}$	1.119×10^{-1}	1.016×10^{-3}	1.101×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1909}$	1.070×10^{-1}	1.021×10^{-3}	1.048×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1910}$	1.091×10^{-1}	1.009×10^{-3}	1.082×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1911}$	1.088×10^{-1}	1.008×10^{-3}	1.080×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1912}$	1.087×10^{-1}	9.790×10^{-4}	1.110×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1913}$	1.093×10^{-1}	9.910×10^{-4}	1.103×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1914}$	1.099×10^{-1}	9.800×10^{-4}	1.121×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1915}$	1.125×10^{-1}	9.540×10^{-4}	1.179×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1916}$	1.068×10^{-1}	9.780×10^{-4}	1.092×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1917}$	1.087×10^{-1}	9.720×10^{-4}	1.118×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1918}$	1.016×10^{-1}	9.400×10^{-4}	1.081×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1919}$	1.083×10^{-1}	9.370×10^{-4}	1.156×10^{2}	$<2.0 \times 10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(2)}_{1920}$	1.086×10^{-1}	9.520×10^{-4}	1.140×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1921}$	1.116×10^{-1}	9.550×10^{-4}	1.169×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1922}$	1.117×10^{-1}	9.190×10^{-4}	1.215×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1923}$	1.075×10^{-1}	9.540×10^{-4}	1.127×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1924}$	1.089×10^{-1}	9.330×10^{-4}	1.167×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1925}$	1.089×10^{-1}	9.340×10^{-4}	1.166×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1926}$	1.084×10^{-1}	9.210×10^{-4}	1.177×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1927}$	1.139×10^{-1}	9.050×10^{-4}	1.259×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1928}$	1.090×10^{-1}	9.220×10^{-4}	1.182×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1929}$	1.094×10^{-1}	9.160×10^{-4}	1.195×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1930}$	1.079×10^{-1}	9.250×10^{-4}	1.166×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1931}$	1.149×10^{-1}	9.050×10^{-4}	1.270×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1932}$	1.104×10^{-1}	9.270×10^{-4}	1.192×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1933}$	1.089×10^{-1}	9.310×10^{-4}	1.170×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1934}$	1.100×10^{-1}	9.260×10^{-4}	1.187×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1935}$	1.134×10^{-1}	9.090×10^{-4}	1.248×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1936}$	1.128×10^{-1}	9.020×10^{-4}	1.251×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1937}$	1.117×10^{-1}	8.930×10^{-4}	1.251×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1938}$	1.110×10^{-1}	9.030×10^{-4}	1.229×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1939}$	1.163×10^{-1}	8.940×10^{-4}	1.301×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1940}$	1.156×10^{-1}	8.880×10^{-4}	1.301×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1941}$	1.174×10^{-1}	8.930×10^{-4}	1.315×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1942}$	1.104×10^{-1}	9.260×10^{-4}	1.192×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1943}$	1.104×10^{-1}	9.050×10^{-4}	1.220×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1944}$	1.146×10^{-1}	8.760×10^{-4}	1.309×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1945}$	1.137×10^{-1}	8.630×10^{-4}	1.318×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1946}$	1.163×10^{-1}	8.590×10^{-4}	1.354×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1947}$	1.195×10^{-1}	8.470×10^{-4}	1.411×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1948}$	1.136×10^{-1}	8.570×10^{-4}	1.326×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1949}$	1.167×10^{-1}	8.480×10^{-4}	1.375×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1950}$	1.194×10^{-1}	8.440×10^{-4}	1.415×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1951}$	1.216×10^{-1}	8.510×10^{-4}	1.429×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1952}$	1.177×10^{-1}	8.490×10^{-4}	1.387×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1953}$	1.184×10^{-1}	8.420×10^{-4}	1.406×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1954}$	1.202×10^{-1}	8.460×10^{-4}	1.421×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1955}$	1.185×10^{-1}	8.480×10^{-4}	1.398×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1956}$	1.196×10^{-1}	8.390×10^{-4}	1.426×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1957}$	1.214×10^{-1}	8.270×10^{-4}	1.469×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1958}$	1.217×10^{-1}	8.320×10^{-4}	1.462×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1959}$	1.218×10^{-1}	8.320×10^{-4}	1.465×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1960}$	1.241×10^{-1}	8.170×10^{-4}	1.520×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1961}$	1.242×10^{-1}	8.180×10^{-4}	1.519×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1962}$	1.263×10^{-1}	8.070×10^{-4}	1.565×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1963}$	1.248×10^{-1}	8.050×10^{-4}	1.551×10^{2}	$<2.0 \times 10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(2)}_{1964}$	1.219×10^{-1}	7.970×10^{-4}	1.530×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1965}$	1.246×10^{-1}	7.910×10^{-4}	1.575×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1966}$	1.250×10^{-1}	7.860×10^{-4}	1.591×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1967}$	1.252×10^{-1}	7.770×10^{-4}	1.611×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1968}$	1.263×10^{-1}	7.620×10^{-4}	1.657×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1969}$	1.233×10^{-1}	7.520×10^{-4}	1.640×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1970}$	1.215×10^{-1}	7.590×10^{-4}	1.602×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1971}$	1.220×10^{-1}	7.500×10^{-4}	1.626×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1972}$	1.224×10^{-1}	7.390×10^{-4}	1.657×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1973}$	1.236×10^{-1}	7.330×10^{-4}	1.686×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1974}$	1.221×10^{-1}	7.250×10^{-4}	1.683×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1975}$	1.229×10^{-1}	7.160×10^{-4}	1.717×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1976}$	1.255×10^{-1}	7.100×10^{-4}	1.767×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1977}$	1.236×10^{-1}	7.160×10^{-4}	1.727×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1978}$	1.236×10^{-1}	7.060×10^{-4}	1.750×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1979}$	1.242×10^{-1}	7.030×10^{-4}	1.768×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1980}$	1.239×10^{-1}	6.930×10^{-4}	1.787×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1981}$	1.248×10^{-1}	6.880×10^{-4}	1.814×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1982}$	1.228×10^{-1}	6.870×10^{-4}	1.788×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1983}$	1.252×10^{-1}	6.880×10^{-4}	1.819×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1984}$	1.238×10^{-1}	6.830×10^{-4}	1.812×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1985}$	1.261×10^{-1}	6.740×10^{-4}	1.870×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1986}$	1.258×10^{-1}	6.740×10^{-4}	1.865×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1987}$	1.247×10^{-1}	6.710×10^{-4}	1.858×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1988}$	1.274×10^{-1}	6.640×10^{-4}	1.920×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1989}$	1.278×10^{-1}	6.760×10^{-4}	1.890×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1990}$	1.282×10^{-1}	6.670×10^{-4}	1.923×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1991}$	1.270×10^{-1}	6.650×10^{-4}	1.910×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1992}$	1.265×10^{-1}	6.620×10^{-4}	1.911×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1993}$	1.301×10^{-1}	6.600×10^{-4}	1.972×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1994}$	1.281×10^{-1}	6.720×10^{-4}	1.907×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1995}$	1.280×10^{-1}	6.630×10^{-4}	1.929×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1996}$	1.297×10^{-1}	6.630×10^{-4}	1.956×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1997}$	1.304×10^{-1}	6.660×10^{-4}	1.959×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1998}$	1.296×10^{-1}	6.620×10^{-4}	1.956×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1999}$	1.317×10^{-1}	6.570×10^{-4}	2.004×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2000}$	1.310×10^{-1}	6.570×10^{-4}	1.995×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2001}$	1.324×10^{-1}	6.550×10^{-4}	2.021×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2002}$	1.341×10^{-1}	6.510×10^{-4}	2.060×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2003}$	1.325×10^{-1}	6.520×10^{-4}	2.031×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2004}$	1.310×10^{-1}	6.490×10^{-4}	2.019×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2005}$	1.308×10^{-1}	6.410×10^{-4}	2.041×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2006}$	1.325×10^{-1}	6.380×10^{-4}	2.075×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2007}$	1.334×10^{-1}	6.300×10^{-4}	2.117×10^{2}	$<2.0 \times 10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(2)}_{2008}$	1.342×10^{-1}	6.290×10^{-4}	2.134×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2009}$	1.325×10^{-1}	6.240×10^{-4}	2.121×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2010}$	1.325×10^{-1}	6.170×10^{-4}	2.147×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2011}$	1.347×10^{-1}	6.200×10^{-4}	2.172×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2012}$	1.357×10^{-1}	6.120×10^{-4}	2.216×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2013}$	1.345×10^{-1}	6.110×10^{-4}	2.201×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2014}$	1.343×10^{-1}	6.140×10^{-4}	2.186×10^{2}	$<2.0 \times 10^{-16}$	***

B.1.5 | KAN:2 Model

Table B.5: Regression table of the KAN:2 model for Swedish females. 317 of 345 parameters (≈ 92%) are significant on the 5% level. p-value significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' . ' 0.1 ' ' 1

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(1)}_{1900}$	−4.036	2.478×10^{-2}	-1.628×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1901}$	−4.148	2.578×10^{-2}	-1.609×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1902}$	−4.076	2.500×10^{-2}	-1.631×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1903}$	−4.173	2.575×10^{-2}	-1.620×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1904}$	−4.067	2.461×10^{-2}	-1.652×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1905}$	−4.077	2.454×10^{-2}	-1.661×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1906}$	−4.114	2.513×10^{-2}	-1.637×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1907}$	−4.149	2.529×10^{-2}	-1.641×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1908}$	−4.147	2.513×10^{-2}	-1.650×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1909}$	−4.143	2.505×10^{-2}	-1.654×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1910}$	−4.130	2.474×10^{-2}	-1.670×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1911}$	−4.145	2.480×10^{-2}	-1.671×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1912}$	−4.083	2.408×10^{-2}	-1.696×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1913}$	−4.163	2.474×10^{-2}	-1.683×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1914}$	−4.140	2.433×10^{-2}	-1.702×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1915}$	−4.100	2.357×10^{-2}	-1.739×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1916}$	−4.145	2.409×10^{-2}	-1.721×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1917}$	−4.156	2.404×10^{-2}	-1.729×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1918}$	−4.000	2.252×10^{-2}	-1.777×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1919}$	−4.117	2.300×10^{-2}	-1.790×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1920}$	−4.194	2.351×10^{-2}	-1.784×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1921}$	−4.225	2.374×10^{-2}	-1.780×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1922}$	−4.124	2.261×10^{-2}	-1.824×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1923}$	−4.196	2.341×10^{-2}	-1.793×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1924}$	−4.187	2.301×10^{-2}	-1.820×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1925}$	−4.176	2.293×10^{-2}	-1.821×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1926}$	−4.196	2.281×10^{-2}	-1.840×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1927}$	−4.147	2.238×10^{-2}	-1.853×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1928}$	−4.195	2.303×10^{-2}	-1.821×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1929}$	−4.182	2.305×10^{-2}	-1.815×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1930}$	−4.157	2.314×10^{-2}	-1.797×10^{2}	$<2.0 \times 10^{-16}$	***

continued ...

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(1)}_{1931}$	−4.147	2.279×10^{-2}	-1.820×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1932}$	−4.207	2.332×10^{-2}	-1.804×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1933}$	−4.185	2.318×10^{-2}	-1.806×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1934}$	−4.198	2.308×10^{-2}	-1.819×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1935}$	−4.215	2.279×10^{-2}	-1.850×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1936}$	−4.187	2.238×10^{-2}	-1.871×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1937}$	−4.167	2.198×10^{-2}	-1.896×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1938}$	−4.237	2.233×10^{-2}	-1.897×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1939}$	−4.240	2.214×10^{-2}	-1.915×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1940}$	−4.233	2.190×10^{-2}	-1.933×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1941}$	−4.292	2.223×10^{-2}	-1.931×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1942}$	−4.384	2.313×10^{-2}	-1.895×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1943}$	−4.350	2.266×10^{-2}	-1.919×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1944}$	−4.375	2.236×10^{-2}	-1.957×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1945}$	−4.366	2.201×10^{-2}	-1.984×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1946}$	−4.367	2.188×10^{-2}	-1.996×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1947}$	−4.368	2.160×10^{-2}	-2.023×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1948}$	−4.408	2.178×10^{-2}	-2.024×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1949}$	−4.412	2.164×10^{-2}	-2.039×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1950}$	−4.460	2.176×10^{-2}	-2.049×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1951}$	−4.496	2.196×10^{-2}	-2.048×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1952}$	−4.496	2.180×10^{-2}	-2.062×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1953}$	−4.522	2.175×10^{-2}	-2.079×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1954}$	−4.543	2.185×10^{-2}	-2.079×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1955}$	−4.607	2.211×10^{-2}	-2.084×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1956}$	−4.630	2.200×10^{-2}	-2.104×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1957}$	−4.589	2.155×10^{-2}	-2.130×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1958}$	−4.669	2.198×10^{-2}	-2.124×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1959}$	−4.695	2.207×10^{-2}	-2.127×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1960}$	−4.727	2.197×10^{-2}	-2.152×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1961}$	−4.801	2.231×10^{-2}	-2.152×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1962}$	−4.802	2.208×10^{-2}	-2.174×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1963}$	−4.794	2.194×10^{-2}	-2.185×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1964}$	−4.786	2.174×10^{-2}	-2.202×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1965}$	−4.837	2.191×10^{-2}	-2.207×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1966}$	−4.880	2.202×10^{-2}	-2.216×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1967}$	−4.929	2.210×10^{-2}	-2.231×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1968}$	−4.909	2.172×10^{-2}	-2.260×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1969}$	−4.865	2.126×10^{-2}	-2.289×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1970}$	−4.914	2.160×10^{-2}	-2.275×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1971}$	−4.957	2.174×10^{-2}	-2.280×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1972}$	−4.946	2.154×10^{-2}	-2.296×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1973}$	−4.983	2.170×10^{-2}	-2.296×10^{2}	$<2.0\times10^{-16}$	***
$\kappa^{(1)}_{1974}$	−4.988	2.165×10^{-2}	-2.304×10^{2}	$<2.0\times10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(1)}_{1975}$	−5.003	2.167×10^{-2}	-2.309×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1976}$	−5.008	2.174×10^{-2}	-2.304×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1977}$	−5.047	2.215×10^{-2}	-2.279×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1978}$	−5.086	2.229×10^{-2}	-2.282×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1979}$	−5.071	2.229×10^{-2}	-2.276×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1980}$	−5.030	2.188×10^{-2}	-2.299×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1981}$	−5.061	2.184×10^{-2}	-2.317×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1982}$	−5.078	2.190×10^{-2}	-2.318×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1983}$	−5.119	2.226×10^{-2}	-2.300×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1984}$	−5.119	2.228×10^{-2}	-2.298×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1985}$	−5.118	2.225×10^{-2}	-2.300×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1986}$	−5.140	2.253×10^{-2}	-2.281×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1987}$	−5.121	2.256×10^{-2}	-2.270×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1988}$	−5.092	2.247×10^{-2}	-2.266×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1989}$	−5.193	2.345×10^{-2}	-2.214×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1990}$	−5.177	2.336×10^{-2}	-2.217×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1991}$	−5.128	2.322×10^{-2}	-2.208×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1992}$	−5.111	2.325×10^{-2}	-2.199×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1993}$	−5.148	2.364×10^{-2}	-2.178×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1994}$	−5.224	2.448×10^{-2}	-2.134×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1995}$	−5.208	2.437×10^{-2}	-2.137×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1996}$	−5.216	2.452×10^{-2}	-2.127×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1997}$	−5.249	2.482×10^{-2}	-2.115×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1998}$	−5.259	2.481×10^{-2}	-2.120×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1999}$	−5.257	2.464×10^{-2}	-2.133×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2000}$	−5.251	2.450×10^{-2}	-2.143×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2001}$	−5.293	2.462×10^{-2}	-2.149×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2002}$	−5.285	2.430×10^{-2}	-2.175×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2003}$	−5.274	2.391×10^{-2}	-2.206×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2004}$	−5.289	2.354×10^{-2}	-2.246×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2005}$	−5.271	2.294×10^{-2}	-2.298×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2006}$	−5.303	2.280×10^{-2}	-2.326×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2007}$	−5.336	2.262×10^{-2}	-2.359×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2008}$	−5.340	2.249×10^{-2}	-2.375×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2009}$	−5.350	2.239×10^{-2}	-2.390×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2010}$	−5.353	2.231×10^{-2}	-2.400×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2011}$	−5.434	2.294×10^{-2}	-2.369×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2012}$	−5.408	2.278×10^{-2}	-2.374×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2013}$	−5.427	2.299×10^{-2}	-2.360×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2014}$	−5.419	2.316×10^{-2}	-2.340×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1900}$	9.262×10^{-2}	3.684×10^{-3}	2.514×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1901}$	9.731×10^{-2}	3.783×10^{-3}	2.572×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1902}$	9.185×10^{-2}	3.681×10^{-3}	2.495×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1903}$	9.612×10^{-2}	3.740×10^{-3}	2.570×10^{1}	$<2.0 \times 10^{-16}$	***

continued ...

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(2)}_{1904}$	8.934×10^{-2}	3.601×10^{-3}	2.481×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1905}$	9.294×10^{-2}	3.585×10^{-3}	2.593×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1906}$	8.490×10^{-2}	3.644×10^{-3}	2.330×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1907}$	9.552×10^{-2}	3.626×10^{-3}	2.635×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1908}$	9.894×10^{-2}	3.592×10^{-3}	2.754×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1909}$	9.303×10^{-2}	3.581×10^{-3}	2.598×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1910}$	8.959×10^{-2}	3.528×10^{-3}	2.540×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1911}$	8.859×10^{-2}	3.520×10^{-3}	2.517×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1912}$	9.070×10^{-2}	3.427×10^{-3}	2.647×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1913}$	9.578×10^{-2}	3.483×10^{-3}	2.750×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1914}$	9.312×10^{-2}	3.427×10^{-3}	2.717×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1915}$	9.960×10^{-2}	3.340×10^{-3}	2.982×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1916}$	9.372×10^{-2}	3.404×10^{-3}	2.753×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1917}$	9.410×10^{-2}	3.387×10^{-3}	2.778×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1918}$	8.405×10^{-2}	3.235×10^{-3}	2.598×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1919}$	9.807×10^{-2}	3.258×10^{-3}	3.010×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1920}$	9.819×10^{-2}	3.313×10^{-3}	2.964×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1921}$	9.657×10^{-2}	3.337×10^{-3}	2.894×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1922}$	9.939×10^{-2}	3.213×10^{-3}	3.093×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1923}$	8.999×10^{-2}	3.314×10^{-3}	2.716×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1924}$	9.590×10^{-2}	3.256×10^{-3}	2.946×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1925}$	8.964×10^{-2}	3.250×10^{-3}	2.758×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1926}$	9.898×10^{-2}	3.215×10^{-3}	3.079×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1927}$	9.794×10^{-2}	3.168×10^{-3}	3.092×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1928}$	1.006×10^{-1}	3.234×10^{-3}	3.110×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1929}$	1.028×10^{-1}	3.219×10^{-3}	3.195×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1930}$	9.313×10^{-2}	3.234×10^{-3}	2.879×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1931}$	1.028×10^{-1}	3.185×10^{-3}	3.229×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1932}$	1.019×10^{-1}	3.240×10^{-3}	3.145×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1933}$	9.363×10^{-2}	3.224×10^{-3}	2.905×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1934}$	9.498×10^{-2}	3.201×10^{-3}	2.967×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1935}$	1.048×10^{-1}	3.149×10^{-3}	3.329×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1936}$	1.034×10^{-1}	3.115×10^{-3}	3.319×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1937}$	1.030×10^{-1}	3.075×10^{-3}	3.350×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1938}$	1.054×10^{-1}	3.113×10^{-3}	3.386×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1939}$	1.047×10^{-1}	3.099×10^{-3}	3.379×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1940}$	1.048×10^{-1}	3.088×10^{-3}	3.395×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1941}$	1.093×10^{-1}	3.128×10^{-3}	3.494×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1942}$	1.031×10^{-1}	3.240×10^{-3}	3.183×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1943}$	9.978×10^{-2}	3.177×10^{-3}	3.141×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1944}$	1.125×10^{-1}	3.117×10^{-3}	3.611×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1945}$	1.133×10^{-1}	3.069×10^{-3}	3.692×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1946}$	1.089×10^{-1}	3.051×10^{-3}	3.571×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1947}$	1.127×10^{-1}	3.020×10^{-3}	3.731×10^{1}	$<2.0\times10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\|z\|)$	Signif. code
$\kappa^{(2)}_{1948}$	1.117×10^{-1}	3.037×10^{-3}	3.678×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1949}$	1.109×10^{-1}	3.009×10^{-3}	3.684×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1950}$	1.174×10^{-1}	3.004×10^{-3}	3.908×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1951}$	1.157×10^{-1}	3.020×10^{-3}	3.831×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1952}$	1.148×10^{-1}	2.995×10^{-3}	3.835×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1953}$	1.201×10^{-1}	2.972×10^{-3}	4.040×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1954}$	1.155×10^{-1}	2.977×10^{-3}	3.881×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1955}$	1.205×10^{-1}	2.988×10^{-3}	4.033×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1956}$	1.231×10^{-1}	2.957×10^{-3}	4.162×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1957}$	1.176×10^{-1}	2.905×10^{-3}	4.046×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1958}$	1.222×10^{-1}	2.942×10^{-3}	4.152×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1959}$	1.202×10^{-1}	2.945×10^{-3}	4.083×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1960}$	1.275×10^{-1}	2.917×10^{-3}	4.372×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1961}$	1.315×10^{-1}	2.936×10^{-3}	4.477×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1962}$	1.304×10^{-1}	2.902×10^{-3}	4.492×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1963}$	1.244×10^{-1}	2.885×10^{-3}	4.313×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1964}$	1.221×10^{-1}	2.854×10^{-3}	4.278×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1965}$	1.244×10^{-1}	2.851×10^{-3}	4.362×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1966}$	1.267×10^{-1}	2.845×10^{-3}	4.454×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1967}$	1.319×10^{-1}	2.830×10^{-3}	4.663×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1968}$	1.300×10^{-1}	2.779×10^{-3}	4.678×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1969}$	1.236×10^{-1}	2.726×10^{-3}	4.535×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1970}$	1.203×10^{-1}	2.750×10^{-3}	4.374×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1971}$	1.235×10^{-1}	2.740×10^{-3}	4.509×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1972}$	1.203×10^{-1}	2.702×10^{-3}	4.452×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1973}$	1.217×10^{-1}	2.696×10^{-3}	4.515×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1974}$	1.212×10^{-1}	2.675×10^{-3}	4.532×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1975}$	1.221×10^{-1}	2.653×10^{-3}	4.603×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1976}$	1.197×10^{-1}	2.640×10^{-3}	4.533×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1977}$	1.176×10^{-1}	2.665×10^{-3}	4.412×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1978}$	1.231×10^{-1}	2.650×10^{-3}	4.645×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1979}$	1.176×10^{-1}	2.632×10^{-3}	4.467×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1980}$	1.118×10^{-1}	2.578×10^{-3}	4.338×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1981}$	1.132×10^{-1}	2.553×10^{-3}	4.433×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1982}$	1.118×10^{-1}	2.543×10^{-3}	4.394×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1983}$	1.104×10^{-1}	2.556×10^{-3}	4.320×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1984}$	1.095×10^{-1}	2.546×10^{-3}	4.303×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1985}$	1.091×10^{-1}	2.523×10^{-3}	4.322×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1986}$	1.085×10^{-1}	2.535×10^{-3}	4.282×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1987}$	1.048×10^{-1}	2.526×10^{-3}	4.151×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1988}$	1.011×10^{-1}	2.507×10^{-3}	4.032×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1989}$	1.044×10^{-1}	2.580×10^{-3}	4.047×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1990}$	1.047×10^{-1}	2.556×10^{-3}	4.096×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1991}$	9.712×10^{-2}	2.539×10^{-3}	3.825×10^{1}	$<2.0\times10^{-16}$	***

continued ...

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(2)}_{1992}$	9.421×10^{-2}	2.531×10^{-3}	3.722×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1993}$	9.645×10^{-2}	2.548×10^{-3}	3.785×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1994}$	9.839×10^{-2}	2.609×10^{-3}	3.772×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1995}$	9.794×10^{-2}	2.580×10^{-3}	3.795×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1996}$	9.601×10^{-2}	2.580×10^{-3}	3.721×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1997}$	9.613×10^{-2}	2.590×10^{-3}	3.712×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1998}$	9.681×10^{-2}	2.573×10^{-3}	3.762×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1999}$	9.577×10^{-2}	2.545×10^{-3}	3.762×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2000}$	9.336×10^{-2}	2.524×10^{-3}	3.698×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2001}$	9.634×10^{-2}	2.519×10^{-3}	3.824×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2002}$	9.414×10^{-2}	2.484×10^{-3}	3.790×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2003}$	8.948×10^{-2}	2.452×10^{-3}	3.649×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2004}$	8.905×10^{-2}	2.418×10^{-3}	3.683×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2005}$	8.600×10^{-2}	2.368×10^{-3}	3.632×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2006}$	8.570×10^{-2}	2.355×10^{-3}	3.638×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2007}$	8.937×10^{-2}	2.340×10^{-3}	3.819×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2008}$	8.578×10^{-2}	2.338×10^{-3}	3.668×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2009}$	8.576×10^{-2}	2.337×10^{-3}	3.670×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2010}$	8.568×10^{-2}	2.333×10^{-3}	3.673×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2011}$	9.017×10^{-2}	2.381×10^{-3}	3.786×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2012}$	8.797×10^{-2}	2.373×10^{-3}	3.707×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2013}$	8.879×10^{-2}	2.393×10^{-3}	3.710×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2014}$	8.349×10^{-2}	2.409×10^{-3}	3.465×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1900}$	6.675×10^{-4}	1.235×10^{-4}	5.405	6.5×10^{-8}	***
$\kappa^{(3)}_{1901}$	4.031×10^{-4}	1.254×10^{-4}	3.215	1.3×10^{-3}	**
$\kappa^{(3)}_{1902}$	5.261×10^{-4}	1.221×10^{-4}	4.308	1.7×10^{-5}	***
$\kappa^{(3)}_{1903}$	4.356×10^{-4}	1.227×10^{-4}	3.551	3.8×10^{-4}	***
$\kappa^{(3)}_{1904}$	7.042×10^{-4}	1.187×10^{-4}	5.933	3.0×10^{-9}	***
$\kappa^{(3)}_{1905}$	4.570×10^{-4}	1.178×10^{-4}	3.880	1.0×10^{-4}	***
$\kappa^{(3)}_{1906}$	6.993×10^{-4}	1.186×10^{-4}	5.899	3.7×10^{-9}	***
$\kappa^{(3)}_{1907}$	5.419×10^{-4}	1.172×10^{-4}	4.623	3.8×10^{-6}	***
$\kappa^{(3)}_{1908}$	4.321×10^{-4}	1.158×10^{-4}	3.731	1.9×10^{-4}	***
$\kappa^{(3)}_{1909}$	4.667×10^{-4}	1.151×10^{-4}	4.055	5.0×10^{-5}	***
$\kappa^{(3)}_{1910}$	6.497×10^{-4}	1.131×10^{-4}	5.745	9.2×10^{-9}	***
$\kappa^{(3)}_{1911}$	6.710×10^{-4}	1.123×10^{-4}	5.974	2.3×10^{-9}	***
$\kappa^{(3)}_{1912}$	5.967×10^{-4}	1.096×10^{-4}	5.446	5.2×10^{-8}	***
$\kappa^{(3)}_{1913}$	4.455×10^{-4}	1.104×10^{-4}	4.034	5.5×10^{-5}	***
$\kappa^{(3)}_{1914}$	5.536×10^{-4}	1.087×10^{-4}	5.092	3.5×10^{-7}	***
$\kappa^{(3)}_{1915}$	4.286×10^{-4}	1.067×10^{-4}	4.018	5.9×10^{-5}	***
$\kappa^{(3)}_{1916}$	4.309×10^{-4}	1.080×10^{-4}	3.989	6.6×10^{-5}	***
$\kappa^{(3)}_{1917}$	4.804×10^{-4}	1.073×10^{-4}	4.478	7.5×10^{-6}	***
$\kappa^{(3)}_{1918}$	5.845×10^{-4}	1.036×10^{-4}	5.642	1.7×10^{-8}	***
$\kappa^{(3)}_{1919}$	3.399×10^{-4}	1.036×10^{-4}	3.281	1.0×10^{-3}	**
$\kappa^{(3)}_{1920}$	3.408×10^{-4}	1.047×10^{-4}	3.255	1.1×10^{-3}	**

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(3)}_{1921}$	4.932×10^{-4}	1.052×10^{-4}	4.688	2.8×10^{-6}	***
$\kappa^{(3)}_{1922}$	4.071×10^{-4}	1.022×10^{-4}	3.982	6.8×10^{-5}	***
$\kappa^{(3)}_{1923}$	5.749×10^{-4}	1.045×10^{-4}	5.499	3.8×10^{-8}	***
$\kappa^{(3)}_{1924}$	4.256×10^{-4}	1.028×10^{-4}	4.140	3.5×10^{-5}	***
$\kappa^{(3)}_{1925}$	6.327×10^{-4}	1.026×10^{-4}	6.166	7.0×10^{-10}	***
$\kappa^{(3)}_{1926}$	3.105×10^{-4}	1.012×10^{-4}	3.068	2.2×10^{-3}	**
$\kappa^{(3)}_{1927}$	5.256×10^{-4}	1.003×10^{-4}	5.241	1.6×10^{-7}	***
$\kappa^{(3)}_{1928}$	2.753×10^{-4}	1.016×10^{-4}	2.711	6.7×10^{-3}	**
$\kappa^{(3)}_{1929}$	2.159×10^{-4}	1.008×10^{-4}	2.141	3.2×10^{-2}	*
$\kappa^{(3)}_{1930}$	4.792×10^{-4}	1.012×10^{-4}	4.734	2.2×10^{-6}	***
$\kappa^{(3)}_{1931}$	3.968×10^{-4}	1.004×10^{-4}	3.952	7.7×10^{-5}	***
$\kappa^{(3)}_{1932}$	2.777×10^{-4}	1.016×10^{-4}	2.734	6.3×10^{-3}	**
$\kappa^{(3)}_{1933}$	4.994×10^{-4}	1.012×10^{-4}	4.935	8.0×10^{-7}	***
$\kappa^{(3)}_{1934}$	4.891×10^{-4}	1.005×10^{-4}	4.865	1.1×10^{-6}	***
$\kappa^{(3)}_{1935}$	2.795×10^{-4}	9.911×10^{-5}	2.820	4.8×10^{-3}	**
$\kappa^{(3)}_{1936}$	3.111×10^{-4}	9.870×10^{-5}	3.152	1.6×10^{-3}	**
$\kappa^{(3)}_{1937}$	2.893×10^{-4}	9.789×10^{-5}	2.955	3.1×10^{-3}	**
$\kappa^{(3)}_{1938}$	1.846×10^{-4}	9.868×10^{-5}	1.870	6.1×10^{-2}	.
$\kappa^{(3)}_{1939}$	3.843×10^{-4}	9.870×10^{-5}	3.894	9.9×10^{-5}	***
$\kappa^{(3)}_{1940}$	3.570×10^{-4}	9.878×10^{-5}	3.615	3.0×10^{-4}	***
$\kappa^{(3)}_{1941}$	2.702×10^{-4}	9.984×10^{-5}	2.706	6.8×10^{-3}	**
$\kappa^{(3)}_{1942}$	2.386×10^{-4}	1.023×10^{-4}	2.332	2.0×10^{-2}	*
$\kappa^{(3)}_{1943}$	3.474×10^{-4}	1.001×10^{-4}	3.471	5.2×10^{-4}	***
$\kappa^{(3)}_{1944}$	6.796×10^{-5}	9.795×10^{-5}	6.940×10^{-1}	4.9×10^{-1}	
$\kappa^{(3)}_{1945}$	1.306×10^{-5}	9.625×10^{-5}	1.360×10^{-1}	8.9×10^{-1}	
$\kappa^{(3)}_{1946}$	2.394×10^{-4}	9.556×10^{-5}	2.505	1.2×10^{-2}	*
$\kappa^{(3)}_{1947}$	2.237×10^{-4}	9.487×10^{-5}	2.358	1.8×10^{-2}	*
$\kappa^{(3)}_{1948}$	6.046×10^{-5}	9.491×10^{-5}	6.370×10^{-1}	5.2×10^{-1}	
$\kappa^{(3)}_{1949}$	1.880×10^{-4}	9.385×10^{-5}	2.003	4.5×10^{-2}	*
$\kappa^{(3)}_{1950}$	6.500×10^{-5}	9.331×10^{-5}	6.970×10^{-1}	4.9×10^{-1}	
$\kappa^{(3)}_{1951}$	1.890×10^{-4}	9.366×10^{-5}	2.018	4.4×10^{-2}	*
$\kappa^{(3)}_{1952}$	9.287×10^{-5}	9.271×10^{-5}	1.002	3.2×10^{-1}	
$\kappa^{(3)}_{1953}$	-5.627×10^{-5}	9.184×10^{-5}	-6.130×10^{-1}	5.4×10^{-1}	
$\kappa^{(3)}_{1954}$	1.488×10^{-4}	9.183×10^{-5}	1.621	1.1×10^{-1}	
$\kappa^{(3)}_{1955}$	-6.398×10^{-5}	9.170×10^{-5}	-6.980×10^{-1}	4.9×10^{-1}	
$\kappa^{(3)}_{1956}$	-1.096×10^{-4}	9.046×10^{-5}	-1.211	2.3×10^{-1}	
$\kappa^{(3)}_{1957}$	1.233×10^{-4}	8.921×10^{-5}	1.382	1.7×10^{-1}	
$\kappa^{(3)}_{1958}$	-1.482×10^{-5}	8.989×10^{-5}	-1.650×10^{-1}	8.7×10^{-1}	
$\kappa^{(3)}_{1959}$	5.019×10^{-5}	8.974×10^{-5}	5.590×10^{-1}	5.8×10^{-1}	
$\kappa^{(3)}_{1960}$	-1.088×10^{-4}	8.874×10^{-5}	-1.227	2.2×10^{-1}	
$\kappa^{(3)}_{1961}$	-2.281×10^{-4}	8.871×10^{-5}	-2.572	1.0×10^{-2}	*
$\kappa^{(3)}_{1962}$	-1.284×10^{-4}	8.761×10^{-5}	-1.465	1.4×10^{-1}	
$\kappa^{(3)}_{1963}$	1.134×10^{-5}	8.698×10^{-5}	1.300×10^{-1}	9.0×10^{-1}	
$\kappa^{(3)}_{1964}$	-5.113×10^{-6}	8.582×10^{-5}	-6.000×10^{-2}	9.5×10^{-1}	

continued ...

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\|z\|)$	Signif. code
$\kappa^{(3)}_{1965}$	7.130×10^{-6}	8.516×10^{-5}	8.400×10^{-2}	9.3×10^{-1}	
$\kappa^{(3)}_{1966}$	-5.356×10^{-5}	8.447×10^{-5}	-6.340×10^{-1}	5.3×10^{-1}	
$\kappa^{(3)}_{1967}$	-2.085×10^{-4}	8.345×10^{-5}	-2.499	1.2×10^{-2}	*
$\kappa^{(3)}_{1968}$	-1.136×10^{-4}	8.185×10^{-5}	-1.388	1.6×10^{-1}	
$\kappa^{(3)}_{1969}$	-8.323×10^{-6}	8.028×10^{-5}	-1.040×10^{-1}	9.2×10^{-1}	
$\kappa^{(3)}_{1970}$	3.801×10^{-5}	8.032×10^{-5}	4.730×10^{-1}	6.4×10^{-1}	
$\kappa^{(3)}_{1971}$	-4.695×10^{-5}	7.935×10^{-5}	-5.920×10^{-1}	5.5×10^{-1}	
$\kappa^{(3)}_{1972}$	6.250×10^{-5}	7.792×10^{-5}	8.020×10^{-1}	4.2×10^{-1}	
$\kappa^{(3)}_{1973}$	5.428×10^{-5}	7.714×10^{-5}	7.040×10^{-1}	4.8×10^{-1}	
$\kappa^{(3)}_{1974}$	2.556×10^{-5}	7.609×10^{-5}	3.360×10^{-1}	7.4×10^{-1}	
$\kappa^{(3)}_{1975}$	2.443×10^{-5}	7.491×10^{-5}	3.260×10^{-1}	7.4×10^{-1}	
$\kappa^{(3)}_{1976}$	1.690×10^{-4}	7.406×10^{-5}	2.282	2.2×10^{-2}	*
$\kappa^{(3)}_{1977}$	1.733×10^{-4}	7.412×10^{-5}	2.338	1.9×10^{-2}	*
$\kappa^{(3)}_{1978}$	1.627×10^{-5}	7.302×10^{-5}	2.230×10^{-1}	8.2×10^{-1}	
$\kappa^{(3)}_{1979}$	1.893×10^{-4}	7.211×10^{-5}	2.625	8.7×10^{-3}	**
$\kappa^{(3)}_{1980}$	3.413×10^{-4}	7.052×10^{-5}	4.839	1.3×10^{-6}	***
$\kappa^{(3)}_{1981}$	3.268×10^{-4}	6.952×10^{-5}	4.701	2.6×10^{-6}	***
$\kappa^{(3)}_{1982}$	3.104×10^{-4}	6.884×10^{-5}	4.509	6.5×10^{-6}	***
$\kappa^{(3)}_{1983}$	4.091×10^{-4}	6.860×10^{-5}	5.964	2.5×10^{-9}	***
$\kappa^{(3)}_{1984}$	3.941×10^{-4}	6.802×10^{-5}	5.794	6.9×10^{-9}	***
$\kappa^{(3)}_{1985}$	4.676×10^{-4}	6.705×10^{-5}	6.973	3.1×10^{-12}	***
$\kappa^{(3)}_{1986}$	4.709×10^{-4}	6.692×10^{-5}	7.037	2.0×10^{-12}	***
$\kappa^{(3)}_{1987}$	5.384×10^{-4}	6.634×10^{-5}	8.116	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1988}$	7.095×10^{-4}	6.567×10^{-5}	1.080×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1989}$	6.246×10^{-4}	6.680×10^{-5}	9.351	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1990}$	6.233×10^{-4}	6.594×10^{-5}	9.453	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1991}$	7.901×10^{-4}	6.536×10^{-5}	1.209×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1992}$	8.498×10^{-4}	6.489×10^{-5}	1.310×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1993}$	8.785×10^{-4}	6.486×10^{-5}	1.354×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1994}$	7.688×10^{-4}	6.568×10^{-5}	1.171×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1995}$	7.730×10^{-4}	6.463×10^{-5}	1.196×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1996}$	8.599×10^{-4}	6.428×10^{-5}	1.338×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1997}$	8.679×10^{-4}	6.412×10^{-5}	1.354×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1998}$	8.269×10^{-4}	6.339×10^{-5}	1.304×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1999}$	9.044×10^{-4}	6.255×10^{-5}	1.446×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2000}$	9.440×10^{-4}	6.189×10^{-5}	1.525×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2001}$	8.997×10^{-4}	6.146×10^{-5}	1.464×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2002}$	9.977×10^{-4}	6.064×10^{-5}	1.645×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2003}$	1.075×10^{-3}	6.000×10^{-5}	1.792×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2004}$	1.051×10^{-3}	5.922×10^{-5}	1.775×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2005}$	1.126×10^{-3}	5.820×10^{-5}	1.935×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2006}$	1.174×10^{-3}	5.790×10^{-5}	2.028×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2007}$	1.107×10^{-3}	5.752×10^{-5}	1.924×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2008}$	1.218×10^{-3}	5.757×10^{-5}	2.115×10^{1}	$<2.0 \times 10^{-16}$	***

continued ...

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(3)}_{2009}$	1.173×10^{-3}	5.748×10^{-5}	2.040×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2010}$	1.174×10^{-3}	5.727×10^{-5}	2.050×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2011}$	1.107×10^{-3}	5.799×10^{-5}	1.909×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2012}$	1.182×10^{-3}	5.777×10^{-5}	2.046×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2013}$	1.129×10^{-3}	5.799×10^{-5}	1.948×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2014}$	1.247×10^{-3}	5.816×10^{-5}	2.144×10^{1}	$<2.0 \times 10^{-16}$	***

B.1.6 | KAN:3 Model

Table B.6: Regression table of the KAN:3 model for Swedish females. 385 of 460 parameters ($\approx$ 84%) are significant on the 5% level. p-value significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' . ' 0.1 ' ' 1

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(1)}_{1900}$	-4.026	3.499×10^{-2}	-1.151×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1901}$	-4.180	3.679×10^{-2}	-1.136×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1902}$	-4.063	3.515×10^{-2}	-1.156×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1903}$	-4.132	3.597×10^{-2}	-1.149×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1904}$	-4.057	3.445×10^{-2}	-1.178×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1905}$	-4.085	3.452×10^{-2}	-1.183×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1906}$	-4.131	3.550×10^{-2}	-1.164×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1907}$	-4.155	3.588×10^{-2}	-1.158×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1908}$	-4.168	3.575×10^{-2}	-1.166×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1909}$	-4.169	3.547×10^{-2}	-1.176×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1910}$	-4.133	3.470×10^{-2}	-1.191×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1911}$	-4.126	3.464×10^{-2}	-1.191×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1912}$	-4.085	3.394×10^{-2}	-1.203×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1913}$	-4.185	3.522×10^{-2}	-1.188×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1914}$	-4.147	3.434×10^{-2}	-1.208×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1915}$	-4.127	3.344×10^{-2}	-1.234×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1916}$	-4.137	3.387×10^{-2}	-1.222×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1917}$	-4.215	3.433×10^{-2}	-1.228×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1918}$	-4.022	3.166×10^{-2}	-1.270×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1919}$	-4.156	3.244×10^{-2}	-1.281×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1920}$	-4.218	3.302×10^{-2}	-1.277×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1921}$	-4.224	3.330×10^{-2}	-1.268×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1922}$	-4.116	3.180×10^{-2}	-1.295×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1923}$	-4.176	3.275×10^{-2}	-1.275×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1924}$	-4.221	3.268×10^{-2}	-1.292×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1925}$	-4.198	3.247×10^{-2}	-1.293×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1926}$	-4.211	3.230×10^{-2}	-1.304×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1927}$	-4.133	3.159×10^{-2}	-1.309×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1928}$	-4.205	3.291×10^{-2}	-1.278×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1929}$	-4.159	3.277×10^{-2}	-1.269×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1930}$	-4.215	3.340×10^{-2}	-1.262×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1931}$	-4.163	3.243×10^{-2}	-1.284×10^{2}	$<2.0 \times 10^{-16}$	***

continued ...

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\|z\|)$	Signif. code
$\kappa^{(1)}_{1932}$	−4.212	3.293×10^{-2}	-1.279×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1933}$	−4.229	3.285×10^{-2}	-1.287×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1934}$	−4.219	3.247×10^{-2}	-1.300×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1935}$	−4.248	3.214×10^{-2}	-1.321×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1936}$	−4.241	3.173×10^{-2}	-1.336×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1937}$	−4.201	3.095×10^{-2}	-1.357×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1938}$	−4.252	3.134×10^{-2}	-1.357×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1939}$	−4.273	3.131×10^{-2}	-1.364×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1940}$	−4.261	3.102×10^{-2}	-1.374×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1941}$	−4.321	3.162×10^{-2}	-1.366×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1942}$	−4.410	3.288×10^{-2}	-1.341×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1943}$	−4.359	3.205×10^{-2}	-1.360×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1944}$	−4.423	3.201×10^{-2}	-1.382×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1945}$	−4.359	3.098×10^{-2}	-1.407×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1946}$	−4.366	3.074×10^{-2}	-1.420×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1947}$	−4.392	3.055×10^{-2}	-1.437×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1948}$	−4.434	3.077×10^{-2}	-1.441×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1949}$	−4.428	3.057×10^{-2}	-1.449×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1950}$	−4.484	3.092×10^{-2}	-1.450×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1951}$	−4.514	3.118×10^{-2}	-1.448×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1952}$	−4.503	3.088×10^{-2}	-1.458×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1953}$	−4.558	3.111×10^{-2}	-1.465×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1954}$	−4.549	3.103×10^{-2}	-1.466×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1955}$	−4.636	3.157×10^{-2}	-1.469×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1956}$	−4.663	3.143×10^{-2}	-1.484×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1957}$	−4.647	3.094×10^{-2}	-1.502×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1958}$	−4.684	3.126×10^{-2}	-1.499×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1959}$	−4.730	3.152×10^{-2}	-1.501×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1960}$	−4.750	3.134×10^{-2}	-1.516×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1961}$	−4.796	3.162×10^{-2}	-1.517×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1962}$	−4.812	3.144×10^{-2}	-1.531×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1963}$	−4.824	3.134×10^{-2}	-1.539×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1964}$	−4.786	3.081×10^{-2}	-1.554×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1965}$	−4.818	3.098×10^{-2}	-1.555×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1966}$	−4.893	3.145×10^{-2}	-1.556×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1967}$	−4.911	3.138×10^{-2}	-1.565×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1968}$	−4.908	3.102×10^{-2}	-1.582×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1969}$	−4.826	3.000×10^{-2}	-1.609×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1970}$	−4.894	3.062×10^{-2}	-1.598×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1971}$	−4.949	3.104×10^{-2}	-1.594×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1972}$	−4.973	3.102×10^{-2}	-1.603×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1973}$	−4.982	3.112×10^{-2}	-1.601×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1974}$	−4.998	3.119×10^{-2}	-1.602×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1975}$	−4.992	3.115×10^{-2}	-1.602×10^{2}	$<2.0 \times 10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(1)}_{1976}$	−4.948	3.099×10^{-2}	-1.597×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1977}$	−4.973	3.154×10^{-2}	-1.577×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1978}$	−5.042	3.204×10^{-2}	-1.574×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1979}$	−5.032	3.206×10^{-2}	-1.569×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1980}$	−4.967	3.103×10^{-2}	-1.600×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1981}$	−4.988	3.071×10^{-2}	-1.624×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1982}$	−5.023	3.093×10^{-2}	-1.624×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1983}$	−5.080	3.167×10^{-2}	-1.604×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1984}$	−5.037	3.154×10^{-2}	-1.597×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1985}$	−5.069	3.188×10^{-2}	-1.590×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1986}$	−5.051	3.213×10^{-2}	-1.572×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1987}$	−5.038	3.233×10^{-2}	-1.558×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1988}$	−5.075	3.277×10^{-2}	-1.549×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1989}$	−5.131	3.396×10^{-2}	-1.511×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1990}$	−5.123	3.391×10^{-2}	-1.511×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1991}$	−5.064	3.351×10^{-2}	-1.511×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1992}$	−5.057	3.358×10^{-2}	-1.506×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1993}$	−5.102	3.430×10^{-2}	-1.488×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1994}$	−5.109	3.498×10^{-2}	-1.461×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1995}$	−5.095	3.478×10^{-2}	-1.465×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1996}$	−5.178	3.553×10^{-2}	-1.457×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1997}$	−5.216	3.586×10^{-2}	-1.454×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1998}$	−5.166	3.516×10^{-2}	-1.469×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{1999}$	−5.169	3.485×10^{-2}	-1.484×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2000}$	−5.135	3.427×10^{-2}	-1.498×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2001}$	−5.202	3.468×10^{-2}	-1.500×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2002}$	−5.197	3.408×10^{-2}	-1.525×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2003}$	−5.210	3.343×10^{-2}	-1.558×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2004}$	−5.189	3.242×10^{-2}	-1.600×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2005}$	−5.180	3.154×10^{-2}	-1.642×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2006}$	−5.225	3.150×10^{-2}	-1.659×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2007}$	−5.270	3.150×10^{-2}	-1.673×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2008}$	−5.237	3.120×10^{-2}	-1.678×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2009}$	−5.283	3.154×10^{-2}	-1.675×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2010}$	−5.254	3.145×10^{-2}	-1.670×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2011}$	−5.345	3.277×10^{-2}	-1.631×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2012}$	−5.371	3.314×10^{-2}	-1.621×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2013}$	−5.410	3.373×10^{-2}	-1.604×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(1)}_{2014}$	−5.372	3.371×10^{-2}	-1.593×10^{2}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1900}$	7.297×10^{-2}	8.820×10^{-3}	8.273	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1901}$	9.136×10^{-2}	9.171×10^{-3}	9.962	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1902}$	7.101×10^{-2}	8.816×10^{-3}	8.054	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1903}$	6.556×10^{-2}	8.927×10^{-3}	7.344	2.1×10^{-13}	***
$\kappa^{(2)}_{1904}$	6.983×10^{-2}	8.580×10^{-3}	8.139	$<2.0 \times 10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(2)}_{1905}$	7.910×10^{-2}	8.584×10^{-3}	9.216	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1906}$	7.569×10^{-2}	8.757×10^{-3}	8.643	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1907}$	8.091×10^{-2}	8.751×10^{-3}	9.246	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1908}$	8.931×10^{-2}	8.683×10^{-3}	1.029×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1909}$	8.570×10^{-2}	8.650×10^{-3}	9.908	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1910}$	7.497×10^{-2}	8.472×10^{-3}	8.850	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1911}$	6.679×10^{-2}	8.434×10^{-3}	7.919	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1912}$	7.585×10^{-2}	8.277×10^{-3}	9.164	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1913}$	8.708×10^{-2}	8.507×10^{-3}	1.024×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1914}$	7.962×10^{-2}	8.318×10^{-3}	9.571	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1915}$	9.161×10^{-2}	8.136×10^{-3}	1.126×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1916}$	7.509×10^{-2}	8.222×10^{-3}	9.133	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1917}$	9.732×10^{-2}	8.275×10^{-3}	1.176×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1918}$	7.655×10^{-2}	7.788×10^{-3}	9.829	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1919}$	9.425×10^{-2}	7.854×10^{-3}	1.200×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1920}$	8.955×10^{-2}	7.969×10^{-3}	1.124×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1921}$	7.972×10^{-2}	8.040×10^{-3}	9.915	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1922}$	7.999×10^{-2}	7.734×10^{-3}	1.034×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1923}$	6.800×10^{-2}	7.947×10^{-3}	8.557	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1924}$	9.083×10^{-2}	7.905×10^{-3}	1.149×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1925}$	8.170×10^{-2}	7.882×10^{-3}	1.037×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1926}$	8.722×10^{-2}	7.758×10^{-3}	1.124×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1927}$	7.710×10^{-2}	7.603×10^{-3}	1.014×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1928}$	8.742×10^{-2}	7.808×10^{-3}	1.120×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1929}$	7.897×10^{-2}	7.701×10^{-3}	1.025×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1930}$	9.632×10^{-2}	7.848×10^{-3}	1.227×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1931}$	9.116×10^{-2}	7.627×10^{-3}	1.195×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1932}$	8.712×10^{-2}	7.732×10^{-3}	1.127×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1933}$	9.207×10^{-2}	7.748×10^{-3}	1.188×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1934}$	8.618×10^{-2}	7.680×10^{-3}	1.122×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1935}$	9.824×10^{-2}	7.580×10^{-3}	1.296×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1936}$	1.039×10^{-1}	7.569×10^{-3}	1.373×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1937}$	9.685×10^{-2}	7.433×10^{-3}	1.303×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1938}$	9.283×10^{-2}	7.503×10^{-3}	1.237×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1939}$	9.794×10^{-2}	7.546×10^{-3}	1.298×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1940}$	9.635×10^{-2}	7.526×10^{-3}	1.280×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1941}$	1.004×10^{-1}	7.629×10^{-3}	1.316×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1942}$	9.465×10^{-2}	7.871×10^{-3}	1.203×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1943}$	8.626×10^{-2}	7.665×10^{-3}	1.125×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1944}$	1.099×10^{-1}	7.584×10^{-3}	1.449×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1945}$	9.267×10^{-2}	7.335×10^{-3}	1.263×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1946}$	9.079×10^{-2}	7.270×10^{-3}	1.249×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1947}$	1.018×10^{-1}	7.254×10^{-3}	1.404×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1948}$	1.022×10^{-1}	7.281×10^{-3}	1.404×10^{1}	$<2.0 \times 10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(2)}_{1949}$	9.818×10^{-2}	7.230×10^{-3}	1.358×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1950}$	1.065×10^{-1}	7.256×10^{-3}	1.468×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1951}$	1.029×10^{-1}	7.311×10^{-3}	1.408×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1952}$	9.904×10^{-2}	7.230×10^{-3}	1.370×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1953}$	1.128×10^{-1}	7.242×10^{-3}	1.557×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1954}$	9.917×10^{-2}	7.229×10^{-3}	1.372×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1955}$	1.111×10^{-1}	7.271×10^{-3}	1.528×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1956}$	1.144×10^{-1}	7.184×10^{-3}	1.592×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1957}$	1.169×10^{-1}	7.096×10^{-3}	1.648×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1958}$	1.081×10^{-1}	7.113×10^{-3}	1.520×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1959}$	1.125×10^{-1}	7.147×10^{-3}	1.575×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1960}$	1.153×10^{-1}	7.065×10^{-3}	1.632×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1961}$	1.106×10^{-1}	7.081×10^{-3}	1.563×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1962}$	1.140×10^{-1}	7.043×10^{-3}	1.618×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1963}$	1.150×10^{-1}	7.018×10^{-3}	1.638×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1964}$	1.036×10^{-1}	6.894×10^{-3}	1.503×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1965}$	1.001×10^{-1}	6.886×10^{-3}	1.453×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1966}$	1.120×10^{-1}	6.940×10^{-3}	1.615×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1967}$	1.078×10^{-1}	6.865×10^{-3}	1.570×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1968}$	1.110×10^{-1}	6.794×10^{-3}	1.634×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1969}$	9.388×10^{-2}	6.603×10^{-3}	1.422×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1970}$	9.670×10^{-2}	6.689×10^{-3}	1.446×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1971}$	1.032×10^{-1}	6.709×10^{-3}	1.539×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1972}$	1.108×10^{-1}	6.656×10^{-3}	1.665×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1973}$	1.042×10^{-1}	6.628×10^{-3}	1.572×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1974}$	1.069×10^{-1}	6.605×10^{-3}	1.619×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1975}$	1.021×10^{-1}	6.536×10^{-3}	1.562×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1976}$	8.653×10^{-2}	6.470×10^{-3}	1.338×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1977}$	8.107×10^{-2}	6.539×10^{-3}	1.240×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1978}$	9.497×10^{-2}	6.533×10^{-3}	1.454×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1979}$	9.136×10^{-2}	6.505×10^{-3}	1.404×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1980}$	7.939×10^{-2}	6.310×10^{-3}	1.258×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1981}$	7.784×10^{-2}	6.223×10^{-3}	1.251×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1982}$	8.175×10^{-2}	6.230×10^{-3}	1.312×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1983}$	8.506×10^{-2}	6.314×10^{-3}	1.347×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1984}$	7.317×10^{-2}	6.289×10^{-3}	1.164×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1985}$	8.183×10^{-2}	6.289×10^{-3}	1.301×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1986}$	7.124×10^{-2}	6.303×10^{-3}	1.130×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1987}$	6.976×10^{-2}	6.312×10^{-3}	1.105×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1988}$	8.327×10^{-2}	6.356×10^{-3}	1.310×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1989}$	7.542×10^{-2}	6.501×10^{-3}	1.160×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1990}$	7.781×10^{-2}	6.451×10^{-3}	1.206×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1991}$	6.822×10^{-2}	6.384×10^{-3}	1.069×10^{1}	$<2.0\times10^{-16}$	***
$\kappa^{(2)}_{1992}$	6.832×10^{-2}	6.372×10^{-3}	1.072×10^{1}	$<2.0\times10^{-16}$	***

continued ...

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(2)}_{1993}$	7.244×10^{-2}	6.441×10^{-3}	1.125×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1994}$	5.790×10^{-2}	6.512×10^{-3}	8.892	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1995}$	5.811×10^{-2}	6.444×10^{-3}	9.018	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1996}$	7.449×10^{-2}	6.538×10^{-3}	1.139×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1997}$	7.589×10^{-2}	6.564×10^{-3}	1.156×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1998}$	6.186×10^{-2}	6.442×10^{-3}	9.603	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{1999}$	6.204×10^{-2}	6.400×10^{-3}	9.695	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2000}$	5.253×10^{-2}	6.325×10^{-3}	8.305	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2001}$	6.167×10^{-2}	6.369×10^{-3}	9.683	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2002}$	5.972×10^{-2}	6.294×10^{-3}	9.488	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2003}$	6.109×10^{-2}	6.237×10^{-3}	9.794	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2004}$	5.096×10^{-2}	6.107×10^{-3}	8.345	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2005}$	4.990×10^{-2}	6.004×10^{-3}	8.312	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2006}$	5.315×10^{-2}	6.014×10^{-3}	8.837	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2007}$	5.967×10^{-2}	6.006×10^{-3}	9.935	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2008}$	4.666×10^{-2}	5.996×10^{-3}	7.782	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2009}$	5.643×10^{-2}	6.033×10^{-3}	9.354	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2010}$	4.857×10^{-2}	6.002×10^{-3}	8.092	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2011}$	5.578×10^{-2}	6.155×10^{-3}	9.062	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2012}$	6.730×10^{-2}	6.185×10^{-3}	1.088×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2013}$	7.289×10^{-2}	6.236×10^{-3}	1.169×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa^{(2)}_{2014}$	6.095×10^{-2}	6.217×10^{-3}	9.803	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{1900}$	2.141×10^{-3}	6.306×10^{-4}	3.396	6.9×10^{-4}	***
$\kappa^{(3)}_{1901}$	8.286×10^{-4}	6.503×10^{-4}	1.274	2.0×10^{-1}	
$\kappa^{(3)}_{1902}$	2.092×10^{-3}	6.261×10^{-4}	3.341	8.3×10^{-4}	***
$\kappa^{(3)}_{1903}$	2.737×10^{-3}	6.284×10^{-4}	4.355	1.3×10^{-5}	***
$\kappa^{(3)}_{1904}$	2.147×10^{-3}	6.056×10^{-4}	3.545	3.9×10^{-4}	***
$\kappa^{(3)}_{1905}$	1.482×10^{-3}	6.050×10^{-4}	2.450	1.4×10^{-2}	*
$\kappa^{(3)}_{1906}$	1.345×10^{-3}	6.115×10^{-4}	2.199	2.8×10^{-2}	*
$\kappa^{(3)}_{1907}$	1.601×10^{-3}	6.082×10^{-4}	2.633	8.5×10^{-3}	**
$\kappa^{(3)}_{1908}$	1.123×10^{-3}	6.026×10^{-4}	1.864	6.2×10^{-2}	.
$\kappa^{(3)}_{1909}$	9.846×10^{-4}	6.005×10^{-4}	1.640	1.0×10^{-1}	
$\kappa^{(3)}_{1910}$	1.703×10^{-3}	5.876×10^{-4}	2.899	3.7×10^{-3}	**
$\kappa^{(3)}_{1911}$	2.256×10^{-3}	5.822×10^{-4}	3.875	1.1×10^{-4}	***
$\kappa^{(3)}_{1912}$	1.663×10^{-3}	5.724×10^{-4}	2.906	3.7×10^{-3}	**
$\kappa^{(3)}_{1913}$	1.061×10^{-3}	5.841×10^{-4}	1.816	6.9×10^{-2}	.
$\kappa^{(3)}_{1914}$	1.515×10^{-3}	5.715×10^{-4}	2.651	8.0×10^{-3}	**
$\kappa^{(3)}_{1915}$	9.945×10^{-4}	5.626×10^{-4}	1.768	7.7×10^{-2}	.
$\kappa^{(3)}_{1916}$	1.787×10^{-3}	5.641×10^{-4}	3.167	1.5×10^{-3}	**
$\kappa^{(3)}_{1917}$	2.066×10^{-4}	5.658×10^{-4}	3.650×10^{-1}	7.1×10^{-1}	
$\kappa^{(3)}_{1918}$	1.103×10^{-3}	5.381×10^{-4}	2.050	4.0×10^{-2}	*
$\kappa^{(3)}_{1919}$	5.985×10^{-4}	5.385×10^{-4}	1.111	2.7×10^{-1}	
$\kappa^{(3)}_{1920}$	9.560×10^{-4}	5.437×10^{-4}	1.758	7.9×10^{-2}	.
$\kappa^{(3)}_{1921}$	1.701×10^{-3}	5.479×10^{-4}	3.104	1.9×10^{-3}	**

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(3)}_{1922}$	1.821×10^{-3}	5.306×10^{-4}	3.432	6.0×10^{-4}	***
$\kappa^{(3)}_{1923}$	2.157×10^{-3}	5.415×10^{-4}	3.984	6.8×10^{-5}	***
$\kappa^{(3)}_{1924}$	7.676×10^{-4}	5.396×10^{-4}	1.423	1.5×10^{-1}	
$\kappa^{(3)}_{1925}$	1.173×10^{-3}	5.386×10^{-4}	2.178	2.9×10^{-2}	*
$\kappa^{(3)}_{1926}$	1.154×10^{-3}	5.269×10^{-4}	2.190	2.9×10^{-2}	*
$\kappa^{(3)}_{1927}$	2.029×10^{-3}	5.183×10^{-4}	3.915	9.1×10^{-5}	***
$\kappa^{(3)}_{1928}$	1.218×10^{-3}	5.265×10^{-4}	2.314	2.1×10^{-2}	*
$\kappa^{(3)}_{1929}$	1.940×10^{-3}	5.163×10^{-4}	3.757	1.7×10^{-4}	***
$\kappa^{(3)}_{1930}$	2.115×10^{-4}	5.276×10^{-4}	4.010×10^{-1}	6.9×10^{-1}	
$\kappa^{(3)}_{1931}$	1.224×10^{-3}	5.165×10^{-4}	2.369	1.8×10^{-2}	*
$\kappa^{(3)}_{1932}$	1.340×10^{-3}	5.226×10^{-4}	2.565	1.0×10^{-2}	*
$\kappa^{(3)}_{1933}$	5.773×10^{-4}	5.252×10^{-4}	1.099	2.7×10^{-1}	
$\kappa^{(3)}_{1934}$	1.102×10^{-3}	5.215×10^{-4}	2.114	3.5×10^{-2}	*
$\kappa^{(3)}_{1935}$	7.465×10^{-4}	5.147×10^{-4}	1.450	1.5×10^{-1}	
$\kappa^{(3)}_{1936}$	2.501×10^{-4}	5.179×10^{-4}	4.830×10^{-1}	6.3×10^{-1}	
$\kappa^{(3)}_{1937}$	7.269×10^{-4}	5.096×10^{-4}	1.426	1.5×10^{-1}	
$\kappa^{(3)}_{1938}$	1.098×10^{-3}	5.113×10^{-4}	2.148	3.2×10^{-2}	*
$\kappa^{(3)}_{1939}$	8.569×10^{-4}	5.169×10^{-4}	1.658	9.7×10^{-2}	.
$\kappa^{(3)}_{1940}$	9.609×10^{-4}	5.172×10^{-4}	1.858	6.3×10^{-2}	.
$\kappa^{(3)}_{1941}$	9.105×10^{-4}	5.220×10^{-4}	1.744	8.1×10^{-2}	.
$\kappa^{(3)}_{1942}$	8.418×10^{-4}	5.324×10^{-4}	1.581	1.1×10^{-1}	
$\kappa^{(3)}_{1943}$	1.311×10^{-3}	5.174×10^{-4}	2.533	1.1×10^{-2}	*
$\kappa^{(3)}_{1944}$	2.560×10^{-4}	5.113×10^{-4}	5.010×10^{-1}	6.2×10^{-1}	
$\kappa^{(3)}_{1945}$	1.523×10^{-3}	4.937×10^{-4}	3.086	2.0×10^{-3}	**
$\kappa^{(3)}_{1946}$	1.549×10^{-3}	4.892×10^{-4}	3.167	1.5×10^{-3}	**
$\kappa^{(3)}_{1947}$	1.006×10^{-3}	4.917×10^{-4}	2.046	4.1×10^{-2}	*
$\kappa^{(3)}_{1948}$	7.525×10^{-4}	4.913×10^{-4}	1.532	1.3×10^{-1}	
$\kappa^{(3)}_{1949}$	1.102×10^{-3}	4.880×10^{-4}	2.259	2.4×10^{-2}	*
$\kappa^{(3)}_{1950}$	8.552×10^{-4}	4.879×10^{-4}	1.753	8.0×10^{-2}	.
$\kappa^{(3)}_{1951}$	1.108×10^{-3}	4.914×10^{-4}	2.254	2.4×10^{-2}	*
$\kappa^{(3)}_{1952}$	1.236×10^{-3}	4.845×10^{-4}	2.552	1.1×10^{-2}	*
$\kappa^{(3)}_{1953}$	4.767×10^{-4}	4.839×10^{-4}	9.850×10^{-1}	3.2×10^{-1}	
$\kappa^{(3)}_{1954}$	1.320×10^{-3}	4.827×10^{-4}	2.735	6.2×10^{-3}	**
$\kappa^{(3)}_{1955}$	6.170×10^{-4}	4.815×10^{-4}	1.281	2.0×10^{-1}	
$\kappa^{(3)}_{1956}$	5.173×10^{-4}	4.731×10^{-4}	1.093	2.7×10^{-1}	
$\kappa^{(3)}_{1957}$	1.584×10^{-4}	4.688×10^{-4}	3.380×10^{-1}	7.4×10^{-1}	
$\kappa^{(3)}_{1958}$	9.848×10^{-4}	4.669×10^{-4}	2.109	3.5×10^{-2}	*
$\kappa^{(3)}_{1959}$	5.908×10^{-4}	4.679×10^{-4}	1.263	2.1×10^{-1}	
$\kappa^{(3)}_{1960}$	7.662×10^{-4}	4.616×10^{-4}	1.660	9.7×10^{-2}	.
$\kappa^{(3)}_{1961}$	1.250×10^{-3}	4.600×10^{-4}	2.717	6.6×10^{-3}	**
$\kappa^{(3)}_{1962}$	1.035×10^{-3}	4.580×10^{-4}	2.260	2.4×10^{-2}	*
$\kappa^{(3)}_{1963}$	6.771×10^{-4}	4.556×10^{-4}	1.486	1.4×10^{-1}	
$\kappa^{(3)}_{1964}$	1.291×10^{-3}	4.462×10^{-4}	2.893	3.8×10^{-3}	**
$\kappa^{(3)}_{1965}$	1.699×10^{-3}	4.433×10^{-4}	3.833	1.3×10^{-4}	***

continued ...

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa_{1966}^{(3)}$	9.652×10^{-4}	4.447×10^{-4}	2.171	3.0×10^{-2}	*
$\kappa_{1967}^{(3)}$	1.474×10^{-3}	4.368×10^{-4}	3.375	7.4×10^{-4}	***
$\kappa_{1968}^{(3)}$	1.199×10^{-3}	4.327×10^{-4}	2.771	5.6×10^{-3}	**
$\kappa_{1969}^{(3)}$	2.040×10^{-3}	4.209×10^{-4}	4.847	1.3×10^{-6}	***
$\kappa_{1970}^{(3)}$	1.643×10^{-3}	4.232×10^{-4}	3.883	1.0×10^{-4}	***
$\kappa_{1971}^{(3)}$	1.330×10^{-3}	4.212×10^{-4}	3.159	1.6×10^{-3}	**
$\kappa_{1972}^{(3)}$	6.959×10^{-4}	4.157×10^{-4}	1.674	9.4×10^{-2}	.
$\kappa_{1973}^{(3)}$	1.224×10^{-3}	4.115×10^{-4}	2.975	2.9×10^{-3}	**
$\kappa_{1974}^{(3)}$	9.759×10^{-4}	4.080×10^{-4}	2.392	1.7×10^{-2}	*
$\kappa_{1975}^{(3)}$	1.343×10^{-3}	4.010×10^{-4}	3.350	8.1×10^{-4}	***
$\kappa_{1976}^{(3)}$	2.339×10^{-3}	3.952×10^{-4}	5.919	3.3×10^{-9}	***
$\kappa_{1977}^{(3)}$	2.551×10^{-3}	3.968×10^{-4}	6.429	1.3×10^{-10}	***
$\kappa_{1978}^{(3)}$	1.828×10^{-3}	3.921×10^{-4}	4.662	3.1×10^{-6}	***
$\kappa_{1979}^{(3)}$	1.862×10^{-3}	3.887×10^{-4}	4.790	1.7×10^{-6}	***
$\kappa_{1980}^{(3)}$	2.412×10^{-3}	3.776×10^{-4}	6.387	1.7×10^{-10}	***
$\kappa_{1981}^{(3)}$	2.587×10^{-3}	3.716×10^{-4}	6.961	3.4×10^{-12}	***
$\kappa_{1982}^{(3)}$	2.213×10^{-3}	3.697×10^{-4}	5.984	2.2×10^{-9}	***
$\kappa_{1983}^{(3)}$	1.991×10^{-3}	3.714×10^{-4}	5.360	8.3×10^{-8}	***
$\kappa_{1984}^{(3)}$	2.667×10^{-3}	3.691×10^{-4}	7.224	5.0×10^{-13}	***
$\kappa_{1985}^{(3)}$	2.142×10^{-3}	3.662×10^{-4}	5.850	4.9×10^{-9}	***
$\kappa_{1986}^{(3)}$	2.765×10^{-3}	3.648×10^{-4}	7.579	3.5×10^{-14}	***
$\kappa_{1987}^{(3)}$	2.677×10^{-3}	3.634×10^{-4}	7.365	1.8×10^{-13}	***
$\kappa_{1988}^{(3)}$	1.762×10^{-3}	3.647×10^{-4}	4.832	1.4×10^{-6}	***
$\kappa_{1989}^{(3)}$	2.353×10^{-3}	3.692×10^{-4}	6.372	1.9×10^{-10}	***
$\kappa_{1990}^{(3)}$	2.219×10^{-3}	3.649×10^{-4}	6.080	1.2×10^{-9}	***
$\kappa_{1991}^{(3)}$	2.496×10^{-3}	3.610×10^{-4}	6.914	4.7×10^{-12}	***
$\kappa_{1992}^{(3)}$	2.369×10^{-3}	3.593×10^{-4}	6.593	4.3×10^{-11}	***
$\kappa_{1993}^{(3)}$	2.272×10^{-3}	3.610×10^{-4}	6.293	3.1×10^{-10}	***
$\kappa_{1994}^{(3)}$	3.137×10^{-3}	3.621×10^{-4}	8.665	$<2.0 \times 10^{-16}$	***
$\kappa_{1995}^{(3)}$	3.098×10^{-3}	3.571×10^{-4}	8.676	$<2.0 \times 10^{-16}$	***
$\kappa_{1996}^{(3)}$	2.083×10^{-3}	3.608×10^{-4}	5.774	7.8×10^{-9}	***
$\kappa_{1997}^{(3)}$	2.013×10^{-3}	3.609×10^{-4}	5.579	2.4×10^{-8}	***
$\kappa_{1998}^{(3)}$	2.839×10^{-3}	3.539×10^{-4}	8.021	$<2.0 \times 10^{-16}$	***
$\kappa_{1999}^{(3)}$	2.840×10^{-3}	3.521×10^{-4}	8.066	$<2.0 \times 10^{-16}$	***
$\kappa_{2000}^{(3)}$	3.298×10^{-3}	3.483×10^{-4}	9.467	$<2.0 \times 10^{-16}$	***
$\kappa_{2001}^{(3)}$	2.886×10^{-3}	3.492×10^{-4}	8.264	$<2.0 \times 10^{-16}$	***
$\kappa_{2002}^{(3)}$	2.969×10^{-3}	3.458×10^{-4}	8.585	$<2.0 \times 10^{-16}$	***
$\kappa_{2003}^{(3)}$	2.689×10^{-3}	3.439×10^{-4}	7.820	$<2.0 \times 10^{-16}$	***
$\kappa_{2004}^{(3)}$	3.247×10^{-3}	3.375×10^{-4}	9.622	$<2.0 \times 10^{-16}$	***
$\kappa_{2005}^{(3)}$	3.205×10^{-3}	3.327×10^{-4}	9.633	$<2.0 \times 10^{-16}$	***
$\kappa_{2006}^{(3)}$	3.036×10^{-3}	3.328×10^{-4}	9.122	$<2.0 \times 10^{-16}$	***
$\kappa_{2007}^{(3)}$	2.798×10^{-3}	3.312×10^{-4}	8.449	$<2.0 \times 10^{-16}$	***
$\kappa_{2008}^{(3)}$	3.457×10^{-3}	3.308×10^{-4}	1.045×10^{1}	$<2.0 \times 10^{-16}$	***
$\kappa_{2009}^{(3)}$	2.826×10^{-3}	3.311×10^{-4}	8.537	$<2.0 \times 10^{-16}$	***

continued …

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(3)}_{2010}$	3.276×10^{-3}	3.280×10^{-4}	9.988	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2011}$	3.035×10^{-3}	3.330×10^{-4}	9.115	$<2.0 \times 10^{-16}$	***
$\kappa^{(3)}_{2012}$	2.307×10^{-3}	3.337×10^{-4}	6.913	4.7×10^{-12}	***
$\kappa^{(3)}_{2013}$	1.978×10^{-3}	3.347×10^{-4}	5.910	3.4×10^{-9}	***
$\kappa^{(3)}_{2014}$	2.469×10^{-3}	3.332×10^{-4}	7.411	1.3×10^{-13}	***
$\kappa^{(4)}_{1900}$	-3.125×10^{-5}	1.309×10^{-5}	-2.387	1.7×10^{-2}	*
$\kappa^{(4)}_{1901}$	-8.920×10^{-6}	1.341×10^{-5}	-6.650×10^{-1}	5.1×10^{-1}	
$\kappa^{(4)}_{1902}$	-3.294×10^{-5}	1.291×10^{-5}	-2.552	1.1×10^{-2}	*
$\kappa^{(4)}_{1903}$	-4.802×10^{-5}	1.285×10^{-5}	-3.736	1.9×10^{-4}	***
$\kappa^{(4)}_{1904}$	-3.020×10^{-5}	1.242×10^{-5}	-2.432	1.5×10^{-2}	*
$\kappa^{(4)}_{1905}$	-2.146×10^{-5}	1.239×10^{-5}	-1.732	8.3×10^{-2}	.
$\kappa^{(4)}_{1906}$	-1.337×10^{-5}	1.241×10^{-5}	-1.077	2.8×10^{-1}	
$\kappa^{(4)}_{1907}$	-2.189×10^{-5}	1.233×10^{-5}	-1.775	7.6×10^{-2}	.
$\kappa^{(4)}_{1908}$	-1.425×10^{-5}	1.221×10^{-5}	-1.167	2.4×10^{-1}	
$\kappa^{(4)}_{1909}$	-1.068×10^{-5}	1.213×10^{-5}	-8.810×10^{-1}	3.8×10^{-1}	
$\kappa^{(4)}_{1910}$	-2.169×10^{-5}	1.185×10^{-5}	-1.831	6.7×10^{-2}	.
$\kappa^{(4)}_{1911}$	-3.246×10^{-5}	1.168×10^{-5}	-2.781	5.4×10^{-3}	**
$\kappa^{(4)}_{1912}$	-2.180×10^{-5}	1.150×10^{-5}	-1.896	5.8×10^{-2}	.
$\kappa^{(4)}_{1913}$	-1.254×10^{-5}	1.166×10^{-5}	-1.076	2.8×10^{-1}	
$\kappa^{(4)}_{1914}$	-1.951×10^{-5}	1.141×10^{-5}	-1.710	8.7×10^{-2}	.
$\kappa^{(4)}_{1915}$	-1.158×10^{-5}	1.131×10^{-5}	-1.024	3.1×10^{-1}	
$\kappa^{(4)}_{1916}$	-2.754×10^{-5}	1.122×10^{-5}	-2.454	1.4×10^{-2}	*
$\kappa^{(4)}_{1917}$	5.538×10^{-6}	1.124×10^{-5}	4.930×10^{-1}	6.2×10^{-1}	
$\kappa^{(4)}_{1918}$	-1.057×10^{-5}	1.075×10^{-5}	-9.830×10^{-1}	3.3×10^{-1}	
$\kappa^{(4)}_{1919}$	-5.239×10^{-6}	1.071×10^{-5}	-4.890×10^{-1}	6.2×10^{-1}	
$\kappa^{(4)}_{1920}$	-1.241×10^{-5}	1.076×10^{-5}	-1.153	2.5×10^{-1}	
$\kappa^{(4)}_{1921}$	-2.430×10^{-5}	1.083×10^{-5}	-2.244	2.5×10^{-2}	*
$\kappa^{(4)}_{1922}$	-2.871×10^{-5}	1.056×10^{-5}	-2.720	6.5×10^{-3}	**
$\kappa^{(4)}_{1923}$	-3.181×10^{-5}	1.068×10^{-5}	-2.979	2.9×10^{-3}	**
$\kappa^{(4)}_{1924}$	-6.869×10^{-6}	1.068×10^{-5}	-6.430×10^{-1}	5.2×10^{-1}	
$\kappa^{(4)}_{1925}$	-1.089×10^{-5}	1.066×10^{-5}	-1.022	3.1×10^{-1}	
$\kappa^{(4)}_{1926}$	-1.695×10^{-5}	1.038×10^{-5}	-1.633	1.0×10^{-1}	
$\kappa^{(4)}_{1927}$	-3.040×10^{-5}	1.026×10^{-5}	-2.962	3.1×10^{-3}	**
$\kappa^{(4)}_{1928}$	-1.891×10^{-5}	1.034×10^{-5}	-1.829	6.7×10^{-2}	.
$\kappa^{(4)}_{1929}$	-3.447×10^{-5}	1.009×10^{-5}	-3.415	6.4×10^{-4}	***
$\kappa^{(4)}_{1930}$	5.347×10^{-6}	1.035×10^{-5}	5.170×10^{-1}	6.1×10^{-1}	
$\kappa^{(4)}_{1931}$	-1.667×10^{-5}	1.023×10^{-5}	-1.630	1.0×10^{-1}	
$\kappa^{(4)}_{1932}$	-2.135×10^{-5}	1.031×10^{-5}	-2.071	3.8×10^{-2}	*
$\kappa^{(4)}_{1933}$	-1.571×10^{-6}	1.037×10^{-5}	-1.510×10^{-1}	8.8×10^{-1}	
$\kappa^{(4)}_{1934}$	-1.236×10^{-5}	1.031×10^{-5}	-1.200	2.3×10^{-1}	
$\kappa^{(4)}_{1935}$	-9.431×10^{-6}	1.019×10^{-5}	-9.260×10^{-1}	3.5×10^{-1}	
$\kappa^{(4)}_{1936}$	1.236×10^{-6}	1.031×10^{-5}	1.200×10^{-1}	9.0×10^{-1}	
$\kappa^{(4)}_{1937}$	-8.903×10^{-6}	1.015×10^{-5}	-8.770×10^{-1}	3.8×10^{-1}	
$\kappa^{(4)}_{1938}$	-1.843×10^{-5}	1.012×10^{-5}	-1.821	6.9×10^{-2}	.
continued …					

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(4)}_{1939}$	-9.596×10^{-6}	1.029×10^{-5}	-9.320×10^{-1}	3.5×10^{-1}	
$\kappa^{(4)}_{1940}$	-1.227×10^{-5}	1.033×10^{-5}	-1.188	2.3×10^{-1}	
$\kappa^{(4)}_{1941}$	-1.301×10^{-5}	1.040×10^{-5}	-1.251	2.1×10^{-1}	
$\kappa^{(4)}_{1942}$	-1.209×10^{-5}	1.047×10^{-5}	-1.154	2.5×10^{-1}	
$\kappa^{(4)}_{1943}$	-1.929×10^{-5}	1.016×10^{-5}	-1.898	5.8×10^{-2}	.
$\kappa^{(4)}_{1944}$	-3.770×10^{-6}	1.006×10^{-5}	-3.750×10^{-1}	7.1×10^{-1}	
$\kappa^{(4)}_{1945}$	-3.035×10^{-5}	9.692×10^{-6}	-3.131	1.7×10^{-3}	**
$\kappa^{(4)}_{1946}$	-2.626×10^{-5}	9.604×10^{-6}	-2.734	6.3×10^{-3}	**
$\kappa^{(4)}_{1947}$	-1.579×10^{-5}	9.731×10^{-6}	-1.622	1.0×10^{-1}	
$\kappa^{(4)}_{1948}$	-1.386×10^{-5}	9.658×10^{-6}	-1.435	1.5×10^{-1}	
$\kappa^{(4)}_{1949}$	-1.831×10^{-5}	9.595×10^{-6}	-1.909	5.6×10^{-2}	.
$\kappa^{(4)}_{1950}$	-1.578×10^{-5}	9.573×10^{-6}	-1.649	9.9×10^{-2}	.
$\kappa^{(4)}_{1951}$	-1.837×10^{-5}	9.635×10^{-6}	-1.907	5.7×10^{-2}	.
$\kappa^{(4)}_{1952}$	-2.279×10^{-5}	9.456×10^{-6}	-2.411	1.6×10^{-2}	*
$\kappa^{(4)}_{1953}$	-1.056×10^{-5}	9.432×10^{-6}	-1.120	2.6×10^{-1}	
$\kappa^{(4)}_{1954}$	-2.322×10^{-5}	9.391×10^{-6}	-2.472	1.3×10^{-2}	*
$\kappa^{(4)}_{1955}$	-1.343×10^{-5}	9.307×10^{-6}	-1.443	1.5×10^{-1}	
$\kappa^{(4)}_{1956}$	-1.232×10^{-5}	9.107×10^{-6}	-1.353	1.8×10^{-1}	
$\kappa^{(4)}_{1957}$	-6.895×10^{-7}	9.056×10^{-6}	-7.600×10^{-2}	9.4×10^{-1}	
$\kappa^{(4)}_{1958}$	-1.956×10^{-5}	8.972×10^{-6}	-2.180	2.9×10^{-2}	*
$\kappa^{(4)}_{1959}$	-1.055×10^{-5}	8.976×10^{-6}	-1.176	2.4×10^{-1}	
$\kappa^{(4)}_{1960}$	-1.713×10^{-5}	8.855×10^{-6}	-1.935	5.3×10^{-2}	.
$\kappa^{(4)}_{1961}$	-2.871×10^{-5}	8.778×10^{-6}	-3.271	1.1×10^{-3}	**
$\kappa^{(4)}_{1962}$	-2.267×10^{-5}	8.752×10^{-6}	-2.591	9.6×10^{-3}	**
$\kappa^{(4)}_{1963}$	-1.297×10^{-5}	8.689×10^{-6}	-1.492	1.4×10^{-1}	
$\kappa^{(4)}_{1964}$	-2.512×10^{-5}	8.475×10^{-6}	-2.964	3.0×10^{-3}	**
$\kappa^{(4)}_{1965}$	-3.260×10^{-5}	8.387×10^{-6}	-3.888	1.0×10^{-4}	***
$\kappa^{(4)}_{1966}$	-1.953×10^{-5}	8.382×10^{-6}	-2.330	2.0×10^{-2}	*
$\kappa^{(4)}_{1967}$	-3.221×10^{-5}	8.185×10^{-6}	-3.935	8.3×10^{-5}	***
$\kappa^{(4)}_{1968}$	-2.503×10^{-5}	8.116×10^{-6}	-3.083	2.0×10^{-3}	**
$\kappa^{(4)}_{1969}$	-3.909×10^{-5}	7.885×10^{-6}	-4.958	7.1×10^{-7}	***
$\kappa^{(4)}_{1970}$	-3.035×10^{-5}	7.868×10^{-6}	-3.858	1.1×10^{-4}	***
$\kappa^{(4)}_{1971}$	-2.592×10^{-5}	7.785×10^{-6}	-3.329	8.7×10^{-4}	***
$\kappa^{(4)}_{1972}$	-1.185×10^{-5}	7.652×10^{-6}	-1.549	1.2×10^{-1}	
$\kappa^{(4)}_{1973}$	-2.182×10^{-5}	7.539×10^{-6}	-2.895	3.8×10^{-3}	**
$\kappa^{(4)}_{1974}$	-1.764×10^{-5}	7.441×10^{-6}	-2.371	1.8×10^{-2}	*
$\kappa^{(4)}_{1975}$	-2.430×10^{-5}	7.270×10^{-6}	-3.342	8.3×10^{-4}	***
$\kappa^{(4)}_{1976}$	-3.990×10^{-5}	7.137×10^{-6}	-5.591	2.3×10^{-8}	***
$\kappa^{(4)}_{1977}$	-4.346×10^{-5}	7.121×10^{-6}	-6.104	1.0×10^{-9}	***
$\kappa^{(4)}_{1978}$	-3.277×10^{-5}	6.979×10^{-6}	-4.696	2.7×10^{-6}	***
$\kappa^{(4)}_{1979}$	-3.017×10^{-5}	6.892×10^{-6}	-4.377	1.2×10^{-5}	***
$\kappa^{(4)}_{1980}$	-3.735×10^{-5}	6.692×10^{-6}	-5.581	2.4×10^{-8}	***
$\kappa^{(4)}_{1981}$	-4.075×10^{-5}	6.570×10^{-6}	-6.203	5.5×10^{-10}	***
$\kappa^{(4)}_{1982}$	-3.406×10^{-5}	6.496×10^{-6}	-5.243	1.6×10^{-7}	***

continued ...

Covariate	Estimate	Std. Error	z value	$\mathbb{P}(>\lvert z\rvert)$	Signif. code
$\kappa^{(4)}_{1983}$	-2.808×10^{-5}	6.478×10^{-6}	−4.334	1.5×10^{-5}	***
$\kappa^{(4)}_{1984}$	-4.023×10^{-5}	6.419×10^{-6}	−6.267	3.7×10^{-10}	***
$\kappa^{(4)}_{1985}$	-2.946×10^{-5}	6.334×10^{-6}	−4.650	3.3×10^{-6}	***
$\kappa^{(4)}_{1986}$	-4.019×10^{-5}	6.274×10^{-6}	−6.407	1.5×10^{-10}	***
$\kappa^{(4)}_{1987}$	-3.727×10^{-5}	6.222×10^{-6}	−5.990	2.1×10^{-9}	***
$\kappa^{(4)}_{1988}$	-1.830×10^{-5}	6.238×10^{-6}	−2.934	3.3×10^{-3}	**
$\kappa^{(4)}_{1989}$	-2.980×10^{-5}	6.260×10^{-6}	−4.761	1.9×10^{-6}	***
$\kappa^{(4)}_{1990}$	-2.746×10^{-5}	6.172×10^{-6}	−4.450	8.6×10^{-6}	***
$\kappa^{(4)}_{1991}$	-2.932×10^{-5}	6.100×10^{-6}	−4.806	1.5×10^{-6}	***
$\kappa^{(4)}_{1992}$	-2.608×10^{-5}	6.055×10^{-6}	−4.307	1.7×10^{-5}	***
$\kappa^{(4)}_{1993}$	-2.383×10^{-5}	6.064×10^{-6}	−3.929	8.5×10^{-5}	***
$\kappa^{(4)}_{1994}$	-4.012×10^{-5}	6.032×10^{-6}	−6.651	2.9×10^{-11}	***
$\kappa^{(4)}_{1995}$	-3.931×10^{-5}	5.930×10^{-6}	−6.630	3.4×10^{-11}	***
$\kappa^{(4)}_{1996}$	-2.060×10^{-5}	5.977×10^{-6}	−3.447	5.7×10^{-4}	***
$\kappa^{(4)}_{1997}$	-1.924×10^{-5}	5.955×10^{-6}	−3.230	1.2×10^{-3}	**
$\kappa^{(4)}_{1998}$	-3.368×10^{-5}	5.827×10^{-6}	−5.780	7.5×10^{-9}	***
$\kappa^{(4)}_{1999}$	-3.238×10^{-5}	5.798×10^{-6}	−5.584	2.3×10^{-8}	***
$\kappa^{(4)}_{2000}$	-3.928×10^{-5}	5.725×10^{-6}	−6.861	6.8×10^{-12}	***
$\kappa^{(4)}_{2001}$	-3.305×10^{-5}	5.714×10^{-6}	−5.784	7.3×10^{-9}	***
$\kappa^{(4)}_{2002}$	-3.282×10^{-5}	5.659×10^{-6}	−5.799	6.7×10^{-9}	***
$\kappa^{(4)}_{2003}$	-2.683×10^{-5}	5.628×10^{-6}	−4.768	1.9×10^{-6}	***
$\kappa^{(4)}_{2004}$	-3.648×10^{-5}	5.514×10^{-6}	−6.616	3.7×10^{-11}	***
$\kappa^{(4)}_{2005}$	-3.455×10^{-5}	5.437×10^{-6}	−6.355	2.1×10^{-10}	***
$\kappa^{(4)}_{2006}$	-3.086×10^{-5}	5.425×10^{-6}	−5.688	1.3×10^{-8}	***
$\kappa^{(4)}_{2007}$	-2.799×10^{-5}	5.385×10^{-6}	−5.198	2.0×10^{-7}	***
$\kappa^{(4)}_{2008}$	-3.700×10^{-5}	5.374×10^{-6}	−6.886	5.7×10^{-12}	***
$\kappa^{(4)}_{2009}$	-2.720×10^{-5}	5.356×10^{-6}	−5.078	3.8×10^{-7}	***
$\kappa^{(4)}_{2010}$	-3.449×10^{-5}	5.289×10^{-6}	−6.521	7.0×10^{-11}	***
$\kappa^{(4)}_{2011}$	-3.144×10^{-5}	5.334×10^{-6}	−5.895	3.8×10^{-9}	***
$\kappa^{(4)}_{2012}$	-1.833×10^{-5}	5.346×10^{-6}	−3.428	6.1×10^{-4}	***
$\kappa^{(4)}_{2013}$	-1.380×10^{-5}	5.348×10^{-6}	−2.580	9.9×10^{-3}	**
$\kappa^{(4)}_{2014}$	-1.983×10^{-5}	5.313×10^{-6}	−3.732	1.9×10^{-4}	***

Publications

Scientific publications

1. Anderson, William, Anna-Maria Hamm, Thomas Knispel, Maren Liese, **Thomas Salfeld** and Stefan Weber (2013): 'Liquidity-adjusted risk measures'. *Mathematics and Financial Economics*, vol. 7(1): pp. 69–91.
2. Hamm, Anna-Maria, **Thomas Salfeld** and Stefan Weber (2013): 'Stochastic root finding for optimized certainty equivalents'. *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*. IEEE Press: pp. 922–932.

Acknowledgments

I would like to thank my supervisor Prof. Dr. Stefan Weber for giving me the opportunity to write my PhD thesis at the Institut for Probability and Statistics, for his support, and professional advise. Further, I wish to thank Prof. Dr. Matthias Scherer for being a referee of my thesis, as well as Prof. Dr. Marc Steinbach for taking the chair of the examination board.

Additionally, I want to thank all my colleagues for their companionship, especially Dr. Daniel Gaigall, Anna-Maria Hamm, and Kerstin Weske for reviewing the thesis and their thoughtful comments.

Furthermore, I would like to thank my family for their support. Especially, I am deeply grateful to my wife Ina for all her love and endless support.