SAPIENZA
UNIVERSITÀ DI ROMA

Tensor decomposition in mortality: identifying subgroups, modeling, forecasting and exploring causes of death

Scuola di dottorato in Scienze Statistiche
Dottorato di Ricerca in Scienze Attuariali – XXXV Ciclo

Candidate
Giovanni Cardillo
ID number 1534618

Thesis Advisors
Prof. Susanna Levantesi
Prof. Paolo Giordani

October 2022

Thesis defended on 14 February 2023
in front of a Board of Examiners composed by:

Prof. Nino Savelli (chairman)
Prof. Marilena Sibillo
Dr. Luca Ballestra

Tensor decomposition in mortality: identifying subgroups, modeling, forecasting and exploring causes of death
Ph.D. thesis. Sapienza – University of Rome
ISBN: 000000000-0

This thesis has been typeset by LaTeX and the Sapthesis class.

Version: February 23, 2023

Author's email: giovanni.cardillo@uniroma1.it

A Mamma e Papà

Abstract

With the increasing availability of temporal data, a researcher often analyzes information stored in matrices, in which entries are replicated on different occasions. For example, in the context of underwriting, pricing, or forecast, an actuary manages a greater amount of information and could have to deal with the death rates (or with log-death rates) by age and year (or different countries). The occasions can be time-varying or refer to different conditions, and in these situations, data can be stored in a 3-way array or tensor. Also, we can consider an additional dimension (the second occasions) and the data are stored in a 4-way array or (4-way) tensor. More in general, data can be stored in a N-way array or (N-way) tensor. These data are called multi-way data and they are analysed and handled by multi-way models. The aim of this work is to illustrate the different uses of DEDICOM, Tucker and CANDECOMP/PARAFAC models in the context of mortality, such as: identifying subgroups, modeling, forecasting and exploring causes of death. To achieve this aim, we gradually approach the problem, considering respectively three and four dimensions in different order and in various applications. In particular we focus on the Tucker method to modeling and exploring data, on Canonical Polyadic Decomposition (CANDECOMP) or Parallel Factors (PARAFAC) (CANDECOMP/PARAFAC) to forecasting data and on the nonnegative 3-way DEcomposition into DIrectional COMponents (DEDICOM) method, that is a special case of Tucker decomposition, to identify subgroups in the data. Aiming at identifying subgroups, we show how the DEDICOM is able to extract meaningful relational patterns from multi population log centered death rate mortality data. Our work, by specifically describing the mesoscale interactions between countries, could help to design appropriate actions against longevity risk that may impact on the stability conditions of life assurance and pensions. Concerning the mortality modeling, firstly we refer to the three-way Lee Carter model [91], that is based on Tucker 3 decomposition, and that can be considered an extension of the classic Lee carter model [61]. The proposed approach allows us to simplify the data structure and to obtain a rank reduced representation. Then, following this line of research and focusing on the forecasting, we propose a coherent mortality forecasting using a four-way CANDECOMP/PARAFAC decomposition, hence considering another dimension. Our proposal based on the four-way structure allows managing mortality data aggregated in multi-dimensional settings, according to common demographic features: age class, time, country, and gender. We deal with four-dimensional mortality data using two main approaches proposed in the literature, the first one which works on centered mortality rates as in [28], and the second one working on compositional data as in [13]. Here, we provide two steps further on methodological developments in the field of mortality analysis and forecasting in a high-dimensional space. Firstly, compared to the current literature, we use an additional dimension, implementing a 4-way tensor decomposition. Thus, we further extend this framework including the CoDa analysis in the spirit of [14]. In the last part, we apply the Tucker 4 method to the mortality by cause of death, hence considering again four dimensions and referring to death rates. This four-way component analysis is useful for the exploratory analysis of four-way data and in this context it reveals some peculiar aspects of the mortality phenomenon. In particular,

this analysis lets us understand how the longevity improvements, witnessed in many high-income countries during the twentieth century, were determined especially by the reduction in a few specific major causes of death groups.

Ringraziamenti

In primo luogo, vorrei esprimere i miei più sinceri ringraziamenti alla Prof. Susanna Levantesi, per i suoi preziosi consigli e aiuti durante tutto il dottorato di ricerca. La sua esperienza in campo accademico e l'ampia conoscenza attuariale sono stati per me punti di riferimento. In campo statistico, ho trovato la stessa competenza nel Prof. Paolo Giordani, che mi ha assistito tecnicamente negli articoli che abbiamo scritto con Susanna. Ringrazio il Dr. Andrea Nigri per i costanti feedback durante la stesura del primo articolo e per la sua esperienza in campo demografico. Ringrazio anche il Prof. Alessandro Spelta, il cui incontro ha rappresentato un punto di svolta per il completamento di questo elaborato. Vorrei ringraziare in particolare Fausto Belliscioni, assieme agli altri ex colleghi di Generali, per essere intervenuto in prima persona per darmi la possibilità di frequentare questo dottorato. Ringrazio il Prof. Emanuele Vannucci per la fiducia dimostrata nel chiamarmi a tenere delle lezioni presso l'Università di Pisa ed avermi coinvolto in prima linea nel mondo accademico. Vorrei ringraziare anche i colleghi della Banca D'Italia per l'elevata competenza dimostrata e per avermi indotto a proseguire fino in fondo alla ricerca di una formazione continua. Ringrazio la mia famiglia per il sostegno: i nonni, mamma, papà, le mie tre sorelle e tutti gli altri. In particolare mia madre per l'aiuto con l'inglese. Vorrei ringraziare i miei amici di sempre e i nuovi per aver contributo a rendere più leggero questo percorso. Ringrazio anche Michela ed Angelo per avermi spronato a continuare durante uno dei nostri pranzi della domenica. In ultimo, ringrazio Martina, con cui ho condiviso in prima persona le difficoltà e le gioie di questi tre anni, senza il suo amore e la sua pazienza questo impegno sarebbe stato difficile rispettarlo.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The aim of this chapter is to discuss the backgrounds behind the thesis, describing the papers that led to the development of the work under consideration underlining its aim: to illustrate in the context of mortality, the different uses of DEDICOM, Tucker and CANDECOMP/PARAFAC model, such as identifying subgroups, modeling, forecasting and exploring mortality data. Then, the thesis outline will be introduced, where we describe the structure of the work, specifying the contents and describing briefly the data of each chapter. Finally, to better understand the reference context, the focus will be shift on a general description of the multi-way data and on a brief history of the development of tensor-based methods.

1.1 Backgrounds

The area of mortality is of particular interest for researcher and actuaries. This field of study includes the examination of historical mortality trends, the inspection of mortality projections, the exploitation of data suitable for the underwriting, pricing, and analysis of life assurance and pensions products [89]. The twentieth century witnessed longevity improvements in many high-income countries. These improvements were determined especially by the reduction in a few specific major causes of death groups. Against this background, it is of utmost significance to understand mortality behaviours related to countries, which share similar socio-cultural roots, in order to evaluate the impact of longevity risk and for predicting possible future impacts in the insurance sector, especially with reference to life assurance and pensions. In this background it is also important to try to model mortality. [28] summarizes the history of the development of mortality modeling into four main stages. The first stage is characterized by the creation of deterministic and one-dimensional models; the second step concerns the deterministic and two-dimensional models; the third stage relies on the development of stochastic models. The model that marks the beginning of this new era is [61]. Such kind of framework is one of the earliest stochastic mortality models in the literature for a single-population. The last stage is characterized by researches concerning multi-populations mortality models. The milestone of this research line is the innovative paper of [65], which extends the Lee & Carter model for a group of populations.

In parallel with the growing interest in the topic of mortality, there has also been a massive development of statistical methodologies to analyse multi-dimensional mortality data. Indeed, the embedding of additional sources of information and the introduction of multivariate data, i.e. age, year and countries, has called for the exploitation of multivariate mathematical methods developed years earlier in other field of research. Among these methodologies, multi-way tensor decomposition models have been applied to the topic of mortality only recently. In fact, the tensor algebra was established in the sixties, but for many years its applications were confined to psychometrics and chemometrics (see [45, 43]). Only in more recent times, have they been used in statistics, where the first attempts of using these models concerns three dimensions [54], and the analysis of the mortality data [91, 36, 28]. [91] proposed a three-way extension of the Lee-Carter [61] model by considering death rates aggregated over time, age-groups and country. The Lee-Carter model cannot be applied to multi-population mortality data with 3 dimensions, so the three-way Lee-Carter model can be applied to multi-population mortality modelling using three-way methods. [36] applied the three-way Lee-Carter model developed in [91] to a group of countries by extending the Lee-Carter model [61] to a three-way structure. [28] generalises the model used in [91] by applying different tensor decompositions and by considering death rates aggregated for age, year, and country/gender and addresses the forecasting problem of multi-population mortality. [13] extend the paper of [91] by introducing a compositional data framework to simultaneously model and forecast the compositional structure of three-way mortality data.

As emerges from these papers, a researcher, who analyses multidimensional mortality data, usually deals with statistics stored in different matrices, where the information is replicated over different occasions. For example, in the context of underwriting, pricing, or forecast, an actuary manages a greater amount of information and could have to deal with the death rates (or with the number of deaths) by age, year, and replicated for different countries. In other words, mortality data displays both a cross-sectional and a temporal dimension and can be easily stored in a 3-way array or tensor. The CANDECOMP/PARAFAC 3 (CP3) [18, 42] and Tucker3 (T3) [105, 104, 59, 54, 34] represent the most common methods for 3-way analysis in the actuarial field. In a nutshell, CANDECOMP/PARAFAC and Tucker methods represent multi-way extensions of classical Principal Component Analysis (PCA), which allow to discover the multi-way interactions among the dimensions.

Also, we can consider an additional dimension (the second occasions) and the data are stored in a 4-way array or (4-way) tensor. These data are called in general multi-way data and multi-way models were introduced to analyse them. The aim of this work is to illustrate in the context of mortality, the different uses of these methods of tensor decomposition, such as: identifying subgroups, modeling, forecasting and exploring mortality data. To reach this aim, we gradually approach the problem, considering respectively three and four dimensions in different order and in various applications. In particular we focus on the Tucker method to modeling and exploring data, on CANDECOMP/PARAFAC to forecasting data and on the nonnegative 3-way DEcomposition into DIrectional COMponents (DEDICOM) method, that is a special case of Tucker decomposition, to identify subgroups in the data.

1.2 Identifying subgroups

We have seen that it is of utmost significance to understand mortality behaviours related to countries, which share similar socio-cultural roots, in order to evaluate the impact of longevity risk and for predicting possible future impacts in the insurance sector, especially with reference to life assurance and pensions. In other words, it is important to identify persistent groups of countries with homogeneous mortality behaviours related to the evolutionary process of longevity improvements.

Aiming at identifying subgroups, we propose an in-depth inspection of country mortality data to uncover how the evolving relationships among multipopulation log-centered death rates induce distinguishable subsets associated with socio-cultural attitudes of such countries. First, we investigate the patterns of co-occurrences of log-centered death rates through time to determine similarities and differences across countries. Secondly, we exploit the nonnegative 3-way DEcomposition into DIrectional COMponents (DEDICOM), with the aim to find persistent subsets of countries with similar mortality behaviors.

The DEDICOM model was introduced for two-way data, stored in a standard matrix, by Harshman and co-authors [see 43, 44, 45]. In short, DEDICOM is an algebraic model which provides information on latent components in the data that can be regarded as “properties” or “aspects” of the objects, by also providing patterns of relationships among these components. In DEDICOM, individual objects can have substantial weights on more than one of the latent components or patterns, meaning that DEDICOM identifies types of correspondence patterns that have distinctive properties, and these are then linked to the individuals that exhibit mixtures of these patterns in their particular history. The DEDICOM model has also been formulated for three-way data [55, 68, 9, 10, 86]. In this context, one is able to better investigate the relationships among the data components by aggregating trends over time and also to study the strength of each cluster’s participation in the third mode, i.e. time. The intrinsically temporal nature of the methodology allows us also to shed light on similarity patterns among countries during various temporal phases. Our study reveals the existence of two components characterized by different mortality traits; the first component mainly identifies countries well known for their longevity, encompassing European countries, together with North America and Australia. The second component is characterized by Hungary and Denmark, which despite the geographical proximity with Northern Europe exhibit different mortality trend due to lifestyle and specific risk factors.

1.3 Modeling and Forecasting

We have seen that the Lee-Carter model [61] is one of the earliest stochastic mortality models in the literature for a single-population: it marks the beginning of a new era of mortality modeling. [91] proposes a three-way extension of the Lee-Carter [61] model by considering death rates aggregated over time, age-groups and country. So, concerning the mortality modeling, firstly we refer to the three-way Lee Carter model [91] that is based on Tucker 3 decomposition. [91] applies the method considering ten European countries; here we apply the method to one of the resulting subgroups of countries obtained from the DEDICOM application. The proposed approach allows us to simplify the data structure and to obtain a rank reduced representation. Then, following this line of research and focusing on the forecasting aim, we propose a coherent mortality forecasting using a four-way CANDECOMP/PARAFAC decomposition, hence considering another dimension. In this context we consider the same ten European countries considered in [91] for a better comparison. Our proposal based on the four-way structure allows managing mortality data aggregated in multi-dimensional settings, according to common demographic features: age class, time, country, and gender. We deal with four-dimensional mortality data using two main approaches proposed in the literature, the first one which works on centered mortality rates as in [28], and the second one working on compositional data as in [13]. Here, we provide two steps further on methodological developments in the field of mortality analysis and forecasting in a high-dimensional space. We offer a comprehensive picture of multiway decomposition on mortality analysis/forecasting. Firstly, compared to the current literature, we use an additional dimension, implementing a 4-way tensor decomposition. Thus, we further extend this framework including the CoDa analysis in the spirit of [14].

1.4 Exploring the causes of death

We have seen that since the nineteenth century, developed countries witnessed improvements in longevity and the consequent decline in mortality rates across ages and years ([80]). The decrease in mortality rates is mainly due to steps taken in preventing diseases through advances in public health, nutrition, and medicine ([106]). However, this general decline in mortality rates may overshadow periods with stagnation or deceleration in life expectancy for some groups of countries ([64]). In particular, past studies highlighted the increase in life expectancy at birth for a group of Scandinavian countries and stagnations for other ones ([67]), whereas, for example, females in France and Japan, exhibit positive improvements in life expectancy. A similar pattern has been shown in the United States and the Netherlands, which experienced a slowdown in life expectancy at age 65 from 1984 to 2000 ([75]). Recently, longevity decelerations have been extensively documented ([40]) e.g. improvements in life expectancy have slowed in the United Kingdom ([63]). A profound understanding of changing mortality patterns is important for several reasons. From a policy point of view, it is important to establish the causes of the slowdown and, in particular, what can be done to reverse these trends in order to better implement and target health and financial policies. In this perspective, knowledge about mortality trends by cause of death (CoD) and different populations can help governments manage their health care costs and financial planning, including public pensions, and social security schemes ([27],[79]). Thus a comprehensive analysis considering different mortality features, such as age, period CoD, and countries, is crucial. However, this task entails several difficulties. Indeed, mortality data are usually referred to as overall mortality for a single country and gender, displayed as a matrix in order to represent ages and periods (Lee-Carter[61] model). More recently, models have been proposed that leverage a multi-population framework, as in the case of ([65]) model, which extends the Lee-Carter model to allow for a group of populations.

They solve the higher dimensional issue, using a common factor to describe the long-term mortality trend shared by all countries within a group and use a country-specific factor to describe the short-term country-specific mortality patterns. This aspect becomes more problematic if we look at the multi-population model in the causes of death structure. The single-population mortality modeling problem can be solved by using matrix decomposition (or matrix factorisation), which is to approximate the original data matrix using low-rank matrix representations through underlying components. For example, [61] uses the SVD method to forecast the US mortality data. However, the single-population models cannot be applied to multi-population mortality data with three or more dimensions. A natural extension of the matrix decomposition (two dimensions) is tensor decomposition (three or more dimensions), which can be used for multi-population mortality modeling. A tensor is a multidimensional or N-way array, being N the number of dimensions. It is only in more recent times that methods of tensor decomposition have been used in the analysis of the mortality data [91, 36, 28], where the first attempts of using these models concern the case with $N = 3$ and the so-called Tucker3 model, originally introduced in [105], see also [58, 54]. In [91], a three-way extension of the Lee-Carter model by considering death rates aggregated over time, age-groups and country. Paper [13] proposes a compositional data adaptation of the Tucker3 model using three-dimensional arrays indexed by time, age, and population, and providing coherent forecasts of the mortality of Canadian provinces. In [36] the three-way Lee-Carter model developed in [91] is applied to a group of countries by extending the Lee-Carter model to a three-way structure. In [28] the model used in [91] is generalized by using different tensor decomposition models and considering death rates aggregated over age, year, and country/gender and addresses the forecasting problem of multi-population mortality. Papers [91, 36, 28, 66] refer to cause of death log-mortality rates, an alternative involves the use of cause of death rates: in fact, in [57] cause-specific death distributions, rather than log-mortality rates, using compositional data analysis are forecasted. Following this line of research, considering cause of death rates, we can analyze the causes of death mortality by taking into consideration an additional dimension. In this context we apply the more general Tucker4 method, which contains Candecomp/Parafac as a special case. Such a four-way analysis is useful for the exploratory analysis of four-way mortality data and it reveals some peculiar aspects of the mortality phenomenon. More in general, by the current applications, our aim is to stimulate the use of tensor decomposition models whenever four-way data are available. Regardless of the specific domain of research, any four-way analysis is composed of different steps involving crucial choices to be made. These will be carefully described and motivated providing a guidance for the practical use of N-way models.

1.5 Thesis outline

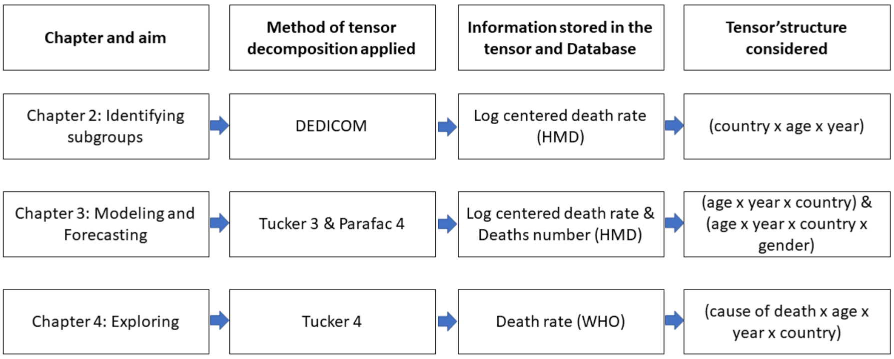
The rest of the thesis is organized as follows.

In Chapter 2, aiming at identifying subgroups, we apply the DEDICOM to the log centered death rate mortality data. The dataset considered is provided by the

Human Mortality Database (HMD) and refers to 3 dimensions: countries, age groups and years. The steps we follow are description of the database, construction of the 3-dimensional similarity tensor, the description of the decomposition method and the reporting the findings of the analysis by showing the results of the tensor decomposition.

In Chapter 3, concerning the mortality modeling, firstly in 3.1 we apply the three-way Lee Carter, that is based on Tucker3 decomposition, considering again the log centered death rate mortality data. The dataset considered is provided by the Human Mortality Database (HMD) and refers to 3 dimensions: age groups, years and countries. Here, to obtain different results compared to [91], we apply the method to one of the resulting subgroups of countries obtained from the DEDICOM application. The steps we follow are: description of the decomposition method (Tucker 3), description of the three-way extension of the Lee Carter model, identification and description of the database, construction of a 3-dimensional array (*age* × *time* × *country*), application of the decomposition method, choice of the number of components. Then, in 3.2 following this line of research and focusing on the forecasting aim, we propose a coherent mortality forecasting using a four-way CANDECOMP/PARAFAC decomposition. The dataset considered is provided by the Human Mortality Database (HMD) and refers to 4 dimensions: age groups, years, countries and gender. Here, we apply this method for simplicity and also because the gender dimension has only two different modes. We deal with four-dimensional mortality data using two main approaches proposed in the literature, the first one which works on centered mortality rates as in [28], and the second one working on compositional data as in [13]. In this context we consider the same ten European countries considered in [91] for a better comparison. The steps we follows are description of the decomposition method (CANDECOMP/PARAFAC 4), description of the four-way models, identification and description of the database, construction of a 4-dimensional array (*age* × *time* × *country* × *gender*), application of the decomposition method, choice of the number of components and forecasting. The tensor factorization can synthesize the behavior of the target variables through a few latent components and highlight the evolution of the temporal patterns. These patterns are employed to forecast future trajectories of mortality data with the Vector-Error Correction model, which accounts for the non-stationarity of the series. We carry out Monte Carlo simulations to obtain the distributions of the time component over the forecasted period 2001-2015, and we evaluate the goodness of the prediction by computing the Root Mean Square Error and the Mean Absolute Error.

In Chapter 4, we apply Tucker 4 method to the mortality by cause of death, considering again four dimensions. In this case we consider the death rates and we apply the more general Tucker4 method, which contains Candecomp/Parafac as a special case. The dataset here considered is provided by the World Health Organization (WHO) and refers to $N = 4$ dimensions: causes of death, age groups, years, and countries. Such a four-way dataset is investigated by the Tucker4 model, the four-way extension of Tucker3. We carry out the analysis by distinguishing the results by sex, first considering the males and then the females. The steps we follow are description of the methodology, identification and description of the database,



Chapter and aim	Method of tensor decomposition applied	Information stored in the tensor and Database	Tensor'structure considered
Chapter 2: Identifying subgroups	DEDICOM	Log centered death rate (HMD)	(country x age x year)
Chapter 3: Modeling and Forecasting	Tucker 3 & Parafac 4	Log centered death rate & Deaths number (HMD)	(age x year x country) & (age x year x country x gender)
Chapter 4: Exploring	Tucker 4	Death rate (WHO)	(cause of death x age x year x country)

Figure 1.1. Structure of the thesis.

construction of the four-dimensional array (causes of death, age, time, and country), application of the tensor decomposition method, choice of the number of underlying components and interpretation.

Fig. 1.1 clarifies the structure of the thesis, representing for each chapter: the aim, the method of tensor decomposition applied, the information stored in the tensor considered and the respective database, the structure of the tensor considered.

1.6 Data choices

Starting from these various applications, the versatility of the tensor decomposition methods is clear. To give coherence to the work, with reference to each dimension involved (where possible and according to the objectives of the specific chapter) we have considered the same quantities for each dimension in the different applications. In the second and third chapter the age dimension is organized in 21 classes from 0 to 99 years, with a step of 5 years, except for the first two classes, which represent respectively the individuals aged 0, and aged 1-4. In the fourth chapter age is organized like the others chapters, but to avoid sparse data of older age with reference to cause of death it is considered up to 84 years. Regarding time, we focus the analysis on the years 1961-2015 in each chapter. Regarding countries, in the second chapter we consider, from HMD, all the possible countries in relation to the chosen time window for which data is available without interruption, these 18 countries are described in the following table:

Code	Countries
AUS	Australia
AUT	Austria
CAN	Canada
USA	Usa
JAP	Japan
BEL	Belgium
DNK	Denmark
FIN	Finland
FRA	France
HUN	Hungary
IRL	Ireland
ITA	Italy
NLD	The Netherlands
NOR	Norway
ESP	Spain
SWE	Sweden
CHE	Switzerland
GBR	United Kingdom

Table 1.1. Countries considered from HMD in the second chapter.

[91] applies the Tucker 3 method considering ten European countries, because they are geographically closer and share similar social-economic status; nevertheless [28] specifies that the tensor decomposition method does not require similar mortality experience among different populations, which can be applied to other more different populations too. So, following this guideline, here to obtain different results respect to [91], here we apply the method to the resulting country's components obtained from the DEDICOM application. In particular, we consider the countries of the first component resulting from the DEDICOM application of the previous chapter. So, in the first part of the third chapter the 13 countries are the following:

Code	Countries
AUS	Australia
CAN	Canada
USA	USA
JAP	Japan
FIN	Finalnd
FRATNP	France
IRL	Ireland
ITA	Italy
NLD	Netherlands
NOR	Norway
ESP	Spain
SWE	Sweden
CHE	Switzerland

Table 1.2. Countries considered from HMD in the first part of the third chapter

In the second part of the third chapter, with reference to the four-way Lee Carter, we consider the same ten countries of [91] for a better comparison with the three-way

Lee Carter. So, we consider the following ten countries:

Code	Countries
AUT	Austria
BEL	Belgium
DNK	Denmark
FIN	Finland
FRATNP	France
ITA	Italy
NLD	Netherlands
ESP	Spain
SWE	Sweden
GBR_NP	United Kingdom

Table 1.3. Countries considered from HMD in the second part of the third chapter

Finally, in the last chapter, where we have reconstructed data from the WHO, to give coherence to the work and to deepen the demographic relationships between countries we have consider the same 18 countries of the second chapter. Data are described more in detail in each chapter, but firstly, we describe these multi-way data in general.

1.7 Multi-way Data

We have seen that with the increasing availability of temporal data, a researcher often analyzes information stored in matrices, in which entries are replicated on different occasions. The occasions can be time-varying or refer to different conditions, and in these situations, data can be stored in a 3-way array or tensor. Also, we can consider an additional dimension (the second occasions) and the data are stored in a 4-way array or (4-way) tensor. More in general, data can be stored in a N-way array or (N-way) tensor. This multi-way structure represents one of the possible formalizations of the data, in response to the need to analyze them, taking into account their exponential increase and the resulting complexity of processing. These data are called in general multi-way data. These data differ from classical multivariate analysis, where the data matrix $\mathbf{X}$ is characterized by a set of x_{ij} values in relation to J variables detected on I statistical units. In fact, multi-way data is defined on a multi-dimensional structure or modes, which are defined on various levels. As anticipated, these additional dimensions can be time-varying or refer to different condition. So, the values of a generic tensor can be represented as follows $x_{ijkl...}$ with i=1,...,I j=1,...,J, k=1,...,K, l=1,...,L where I,J,K,L are the number of levels associated with each mode. For example, we can represent 3-way data in a structure $\mathbf{X}=[x_{ijk}$: i=1,...,I j=1,...,J, k=1,...,$K]$ as in 1.2 where the three modes are represented respectively by rows, columns and tubes.

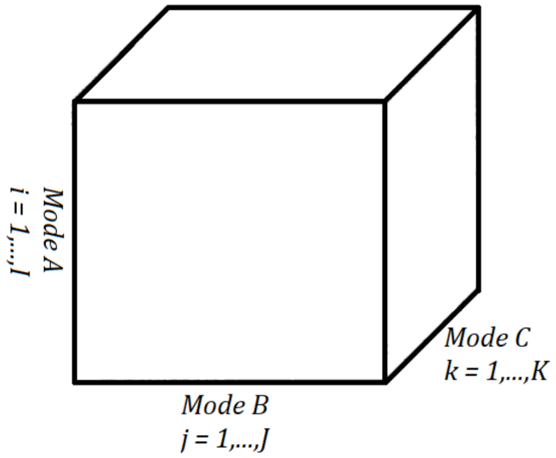


Figure 1.2. Graphical representation of a three-way data tensor

A representation of this type can be complex, for this reason the array can be represented in a simplified way through a set of matrices. These matrices can be of three types: horizontal (A), lateral (B) and frontal (C) matrices as represented in 1.3

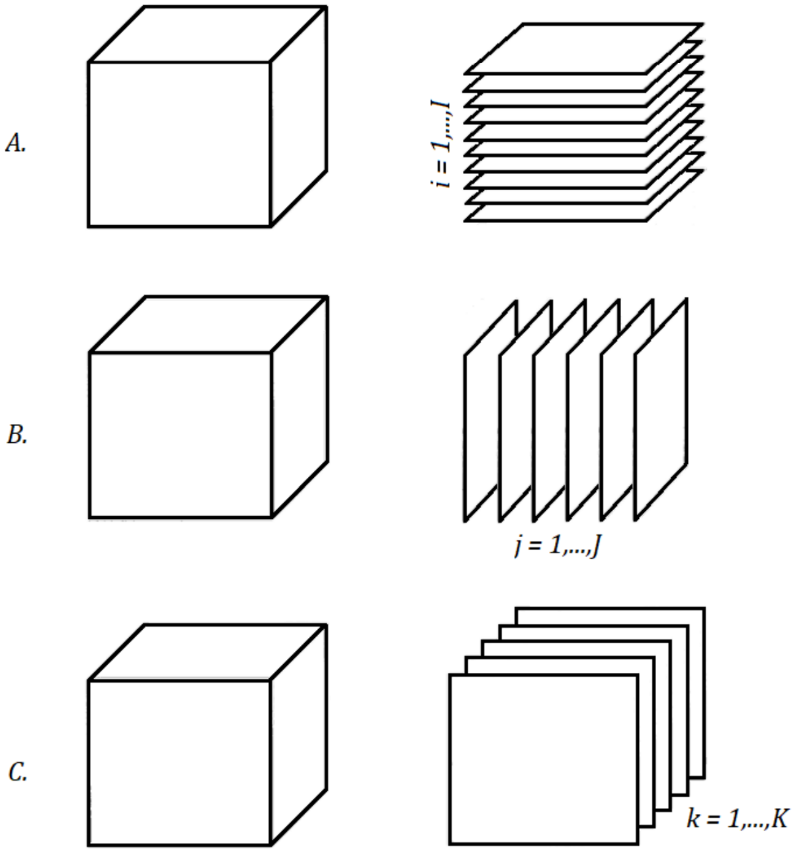


Figure 1.3. Three-way tensor divided into horizontal (A), lateral (B) and frontal (C) matrices

These aspects can be extended to arrays with more than three dimensions, even if it is not possible to provide an adequate graphical representation. An attempt of representation is given in chapter 3 with reference to four dimensions. In general, with reference to N-way tensor it is possible to define subsets of any size (n=1,...,N). To analyze arrays, a procedure called unfolding is often useful, which consists in the concatenation of all the matrices relating to a single dimension (horizontal, lateral or frontal) which are placed side by side to form a single supermatrix. With reference to a three dimensional array this procedure is represented in Fig. 1.4

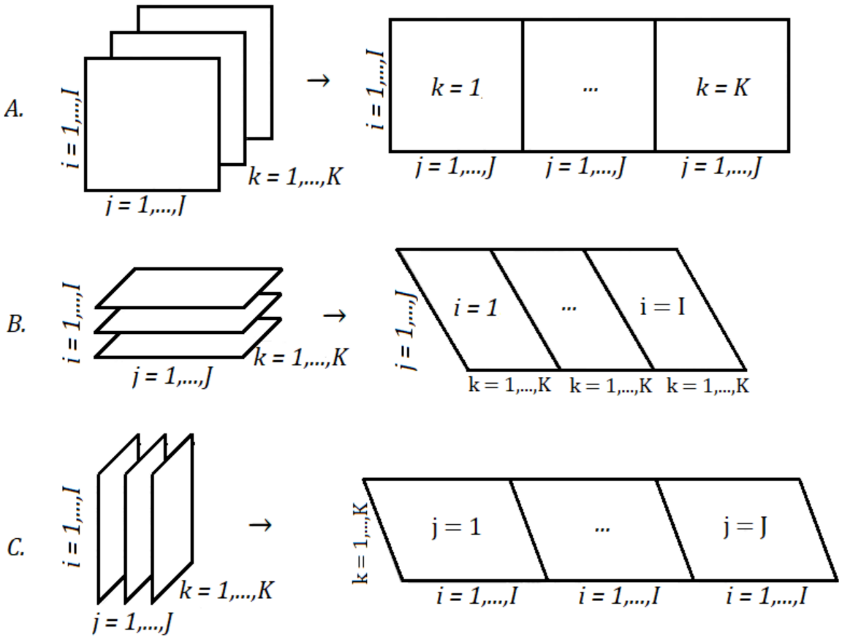


Figure 1.4. Matricization procedure of a three-way tensor starting from the subdivision into horizontal (A), lateral (B) and frontal (C) matrices

The unfolding procedure can be extended to arrays with more than three dimensions. More details about multi-way data will be defined in the next chapters; to complete the reference context, in the next subparagraph we propose a brief history of the development of tensor-based methods.

1.7.1 A brief history of the development of tensor-based methods

Since the 60s, numerous techniques have been developed to process multi-way data. Many of these began as multidimensional extensions of two-dimensional models such as principal component analysis, factor analysis, and group analysis. Most of these models have a predominantly explorative character, since they are often used with the aim of identifying particular patterns within multidimensional data, although, as we will see, over time these techniques have been used for different purposes. These models were born in response to the need to consider the multidimensional structure in the data analysis, as the analysis techniques used for the two-way data do not take these aspects into account. More in details, the study of tensor-based methods originate from the works of Hitchcock [48, 50] and Cattel [21]. The findings emerged from these works were studied during the following years, and they led to the development of tensor method such as CANDECOMP/PARAFAC (CP) [18, 42] and Tucker3 (T3) [105, 104, 59] methods. These two methods represent the most common tensor-based methods and a recent review of them is in [34]. We have also seen that the use of these methods has been confined to psychometrics and chemometrics for many years and only in more recent times, have they been used in statistics, where the first attempts of using these models concern three dimensions [54], and the analysis of the mortality data [91, 36, 28]. Moreover, other decomposition methods have also been developed over time, such as DEDICOM that can be considered a special case of Tucker model. The DEDICOM algorithm was introduced for matrix data by Harshman and co-authors (see [43, 44, 45]). The DEDICOM model has subsequently been extended to three-way data by [55, 68, 9, 10, 86]. A first approach to describe more in general these method considering N components, is in [53] , and then these aspects have been reviewed in [47]. In this work we focus on the Tucker model (T3 and T4), on the special case of it, the DEDICOM model, and on the CANDECOMP/PARAFAC model (CP4): all these models are described more in details in their respective chapters.

Chapter 2

IDENTIFYING SUBGROUPS

In this chapter, we show how the nonnegative 3-way DEcomposition into DIrectional COMponents (DEDICOM) is able to extract meaningful relational patterns from multi population log centered death rate mortality data. The dataset considered is provided by the Human Mortality Database (HMD) and refers to 3 dimensions: countries, age groups and years. The 3-dimensional decomposition technique identifies persistent cluster of countries with homogeneous mortality behaviours related to the evolutionary process of longevity improvements. Our work, by specifically describing the mesoscale interactions between countries, could help to design appropriate actions against longevity risk that may impact on the stability conditions of life assurance and pensions.

The chapter is structured as follows: Section 2.1 encompasses the identification and description of the database, the construction of the 3-dimensional similarity tensor, and the description of the decomposition method. Section 2.2 reports the findings of the analysis by showing the results of the tensor decomposition. Finally, Section 2.3 concludes.

2.1 Data Description and Methodology

2.1.1 Data Information and Pre-processing phase

We consider the dataset provided by the Human Mortality Database (HMD)[1], which is an archive of death information for several countries. The application of the methodology requires to structure the data in three dimensions: country, age and time. The age dimension is organized in 21 classes from 0 to 99 years, with a step of 5 years, except for the first two classes, which represent respectively the individuals aged 0, and aged 1-4. Regarding the time dimension, we focus the analysis on the years from 1961 to 2015 in order to consider the same time window for each country (see the 18 countries listed in Table 2.1). The analysis is developed for the total population, considering both genders. Regarding countries, we consider all the possible countries in relation to the chosen time window for which data is available without interruption.

---

[1] Human Mortality Database. University of California; Berkeley (USA); and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 08/11/21).

Code	Countries
AUS	Australia
AUT	Austria
CAN	Canada
USA	Usa
JAP	Japan
BEL	Belgium
DNK	Denmark
FIN	Finland
FRA	France
HUN	Hungary
IRL	Ireland
ITA	Italy
NLD	The Netherlands
NOR	Norway
ESP	Spain
SWE	Sweden
CHE	Switzerland
GBR	United Kingdom

Table 2.1. Countries considered from HMD.

A cleaning process is of fundamental importance before the construction and investigation of death similarities in the countries listed above. We consider an initial array with three dimensions (*country*, *age*, *time*). The dimension of the array is: 18 × 21 × 55, for a total of 20.790 observations. In particular, following [65] that generalized the Lee-Carter model, we collect (with a different order) the information related to the log central death rates in the 3-dimensional array $\mathcal{Y}$, such that each entry contains the specific information of the log central death rate for population i, at age j at time k, i.e.

$$y_{ijk} = log(m_{ijk}). \tag{2.1}$$

The central death rate for an individual in country i, of age j at time k is computed as:

$$m_{ijk} = \frac{d_{ijk}}{L_{ijk}} \tag{2.2}$$

where

$$d_{ijk} = l_{ijk} - l_{i(j+1)k} \tag{2.3}$$

is the expected number of deaths for population in country i, between age j and $j+1$ in year k and l_{ijk} is the expected number of living individuals at age j in year k in country i initially made up of l_0 individuals. The variable

$$L_{ijk} \simeq \frac{1}{2}[l_{ijk} + l_{i(j+1)k}] \tag{2.4}$$

is the risk exposure of the total population i, of age j in year k (assumed to be equal to the population at mid-year).

To investigate the similarity patterns among countries through the application of the DEDICOM model, we need to create a *country* × *country* similarity matrix $\mathbf{X}(k)$ for each year k, starting from the observations contained in the original array

$\mathcal{Y}$. To accomplish this purpose, we describe each country through the vector $\mathbf{y}(k)$ containing the logarithm of the central death rates for each class, this means that we consider the rows of each matrix *country* × *age* for each single year k. Once we have constructed the yearly vectors of countries' log central death rates $\mathbf{y}(k)$, we compute the cosine metric for measuring the distances among each pair of countries. Given two countries, whose time-k log central death rates for age classes are collected in vectors $\mathbf{y}_i(k)$ and $\mathbf{y}_l(k)$, the time-k cosine distance between country i and country l is defined as

$$\mathbf{C}_{il}(k) = 1 - \cos\left(\mathbf{y}_i(k), \mathbf{y}_l(k)\right), \tag{2.5}$$

where the cosine (cos) is the angle between the two vectors:

$$\cos\left(\mathbf{y}_i(k), \mathbf{y}_l(k)\right) = \frac{< \mathbf{y}_i(k), \mathbf{y}_l(k) >}{\|\mathbf{y}_i(k)\| \, \|\mathbf{y}_l(k)\|}, \tag{2.6}$$

with the symbol $< \circ >$ and $\| \circ \|$ indicating the inner product and the Frobenious norm, respectively. We rely on the cosine distance because our primary aim is to consider the gap among countries' mortality rate vectors in terms of orientation and not by their length. Finally, from the cosine distance we produce the similarity tensor by collecting the matrices

$$\mathbf{X}_{il}(k) = \max(\mathbf{C}(k)) - \mathbf{C}_{il}(k), \tag{2.7}$$

where $\max(\mathbf{C}(k))$ is the maximum value of the time-k cosine distance matrix.

In a nutshell, the stronger the similarity (i.e., the force that connects two countries mortality characteristics), the shorter the length of the links connecting the countries. In other words, pairs of countries that are similar receive higher weights since they are placed near by from each other, while values approaching zero are assigned to pairs with highly dissimilar characteristics.

2.1.2 DEcomposition into DIrectional COMponents (DEDICOM)

Starting from the similarity matrix $\mathbf{X}(k)$ obtained at each year k, we build the three-way tensor [see 9, 7, 97, 85], $\mathcal{X} \in \mathbb{R}^{N \times N \times K}$, where $N = 18$ represents the number of countries and $K = 55$ denotes the number of years. Thus, the tensor is composed of the 55 slices $\mathbf{X}(k) \in \mathbb{R}^{18 \times 18}$. The tensor $\mathcal{X}$ encompasses both the cross-sectional and temporal information of the evolving similarities that characterize the countries. To reveal the hidden structure of such traits and their activity patterns, lower-dimensional factors need to be identified and extracted from the data. In other words, in order to retrieve the mesoscale structure of our temporal similarity tensor, we exploit the DEcomposition into DIrectional COMponents (DEDICOM) model [see 43, 9, 10]. The formulation of DEDICOM with P components, given the three-way tensor $\mathcal{X}$, is as follows:

$$\mathbf{X}_k \approx \mathbf{A}\mathbf{D}_k\mathbf{R}\mathbf{D}_k\mathbf{A}^T, k = 1, 2, \ldots, K, \tag{2.8}$$

where we use $\mathbf{X}_k = \mathbf{X}(k)$ to simplify the notation. This linear algebra model maps the country-to-country structure of the similarities onto a community-level structure of P latent components and comprises:

- **A**: $N \times P$ matrix, where each value a_{ip} reflects the strength of country i belonging to component p;

- $\mathbf{D}_k$: $P \times P$ diagonal matrix with diagonal elements describing the role of each component at time k;

- **R**: $P \times P$ matrix, where each value r_{pq} is a measure of the interaction between component p and q.

The DEDICOM model is very general and contains several models as special cases. First of all, DEDICOM can be applied to both symmetric and asymmetric similarities. In the asymmetric case, **R** is an asymmetric matrix, whereas, in the symmetric case, **R** is a symmetric matrix and DEDICOM coincides with the so-called PARAFAC2 model [41]. When $\mathbf{R} = \mathbf{I}$, i.e., no information on the relationships among the components is provided by the model, the DEDICOM model reduces to the INDSCAL one [19]. All the previous models analyze three-way data stored in the tensor $\mathcal{X}$. If $K = 1$, data are available for only one year and $\mathcal{X}$ is nothing but a standard (two-way) matrix, say $\mathbf{X} \in \mathbb{R}^{N \times N}$. In such a case, the matrices $\mathbf{D}_k$ disappear and the model, called DEDICOM for two-way data, can be formalized as

$$\mathbf{X} = \mathbf{A}\mathbf{R}\mathbf{A}^T. \tag{2.9}$$

So, in this case the idea behind this family of decompositions introduced by Harshman, is as follows. Suppose that we have I objects and a matrix $\mathbf{X} \in \mathbb{R}^{I \times I}$ that describes the asymmetric relationships between them. For instance, the objects might be countries and $x_i j$ represents the value of exports from country i to country j. Typical factor analysis techniques either do not account for the fact that the two modes of a matrix may correspond to the same entities or that there may be directed interactions between them. DEDICOM, on the other hand, attempts to group the I objects into R latent components (or groups) and describe their pattern of interactions by computing $\mathbf{A} \in \mathbb{R}^{I \times I}$ and $\mathbf{R} \in \mathbb{R}^{R \times R}$ such that 2.9. Each column in **A** corresponds to a latent component such that air indicates the participation of object i in group r. The matrix **R** indicates the interaction between the different components, e.g., r_{ij} represents the exports from group i to group j. Three-way DEDICOM [43] is a higher-order extension of the DEDICOM model that incorporates a third mode of the data. Adding a third dimension gives this decomposition stronger uniqueness properties. The decomposition is illustrated in 2.8. Here **A** and **R** are as in 2.9, except that **A** is not necessarily orthogonal. The matrices $\mathbf{D}_k \in \mathbb{R}^{R \times R}$ are diagonal. Three-way DEDICOM is illustrated in Figure 2.1.

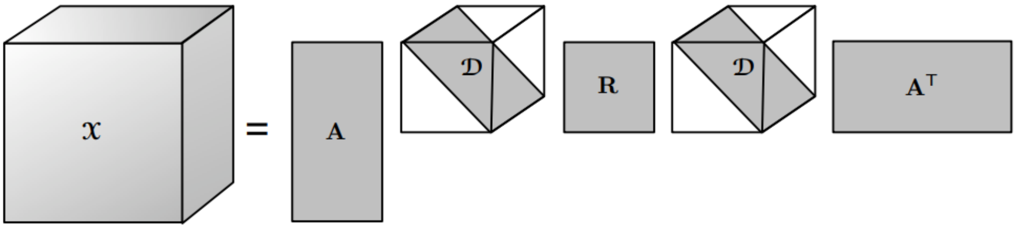


Figure 2.1. Three-way DEDICOM model

The main advantage of DEDICOM in this context is that it is able to extract meaningful relational patterns from multi population log centered death rate mortality data. In 2.2 is represented the relation of the DEDICOM with the Tucker model (described in 3). Starting from the HMD mortality data representing log central death rates for population in country i, at age j in time k, yearly country-by-country cosine similarity matrices $\mathbf{X}_k$ are created. Those matrices are embedded into a three-way tensor of similarities $\mathcal{X}$ of dimensions country × country × year. The tensor is thus decomposed by the DEDICOM model, which is able to provide information on latent components in the matrix $\mathbf{R}$ and to associate individual countries with component weights stored in matrix $\mathbf{A}$. Moreover, thanks to the relationship of this decomposition with the Tucker model, the methodology yields also information on the strength of each component's activity at time k (matrix $\mathbf{D}_k$). In a nutshell, the stronger the similarity (i.e., the force that connects two countries mortality characteristics), the shorter the length of the links connecting the countries. In other words, pairs of countries that are similar receive higher weights since they are placed near by from each other, while values approaching zero are assigned to pairs with highly dissimilar characteristics. Moreover, the DEDICOM model, by taking into consideration both the cross-sectional and temporal features of country similarities, allows us to investigate how components respond to events affecting the logarithm of the central death rates.

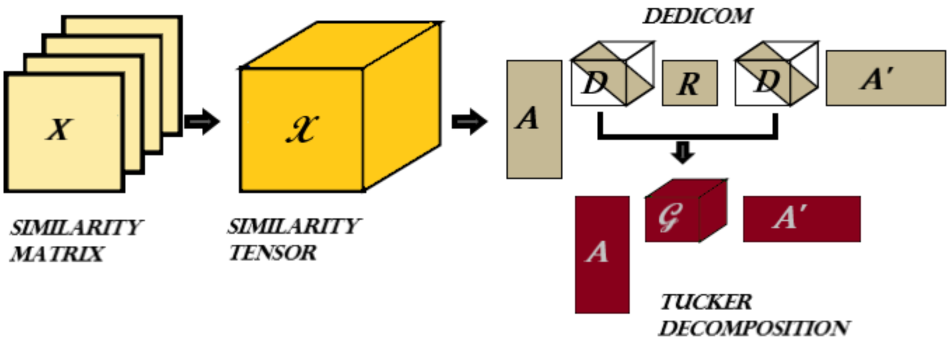


Figure 2.2. Graphical representation of the model.

Computing DEDICOM

[107] explain that there are a number of algorithms for computing the two-way DEDICOM model, e.g., [37], and for variations such as constrained DEDICOM [38, 39]. For three-way DEDICOM, see [55] and [10]. Because **A**, and **D**, appear on both the left and right, fitting three-way DEDICOM is a difficult nonlinear optimization problem with many local minima. [55] presents an alternating least squares (ALS) algorithm that is efficient on small tensors. Each column of **A** is updated with its own least-squares solution while holding the others fixed. Each subproblem to compute one column of **A** involves a full eigendecomposition of a dense $I \times I$ matrix, which makes this procedure expensive for large, sparse X. In a similar alternating fashion, the elements of **D** are updated one at a time by minimizing a fourth degree polynomial. The best **R** for a given **A** and **D** is found from a least-squares solution using the pseudo-inverse of an $I^2 \times R^2$ matrix, which can be simplified to the inverse of an $R^2 \times R^2$ matrix.

[10] have proposed an algorithm called Alternating Simultaneous Approximation, Least Squares, and Newton (ASALSAN). The approach relies on the same update for **A** as in [55] but uses different methods for updating **A** and **D**. Instead of solving for **A** column-wise, ASALSAN solves for all columns of **A** simultaneously using an approximate least-squares solution. Because there are RK elements of **D**, which is not likely to be many, Newton's method is used to find all elements of **D** simultaneously. The same paper [10] introduces a nonnegative variant.

DEDICOM applications

[107] explain that most of the applications of DEDICOM in the literature have focused on two-way (matrix) data, but there are some three-way applications. [87] analyzed asymmetric measures of yearly trade (import-export) among a set of nations over a period of 10 years. [68] presented an application of three-way DEDICOM to skew-symmetric data for paired preference ratings of treatments for chronic back pain, though they note that they needed to impose some constraints to get meaningful results. [10] recently applied their ASALSAN method for computing DEDICOM on email communication graphs over time. In this case, x_{ijk} corresponded to the number of email messages sent from person i to person j in month k.

2.2 Results

We report here a walkthrough of our results: in the first place, we display the results of the tensor factorization and then the roles of the countries inside the subgroups.

2.2.1 The DEDICOM components

We start our analysis by providing an overview of the model fitting for different number of components P. In order to choose the number of clusters we have computed the model fitting for an increasing number of components. Moreover, as many tensor decomposition, also the DEDICOM model could be affected by the presence of local minima. To account for this feature, for each P we have run 100 times the DEDICOM model using different starting points at each run. Figure 2.3 shows the percentage fits for different number of components. Each colored circle refers to a model run with different starting conditions, while the black line shows the maximum fit for each component.

In order to discover homogeneous traits at the component level that may reveal countries' different mortality patterns, we employ the DEDICOM model with $P = 2$ components, which explains about 96% of the original data variability. We have opted for two components to have a balancing between model fit and stability of the results.

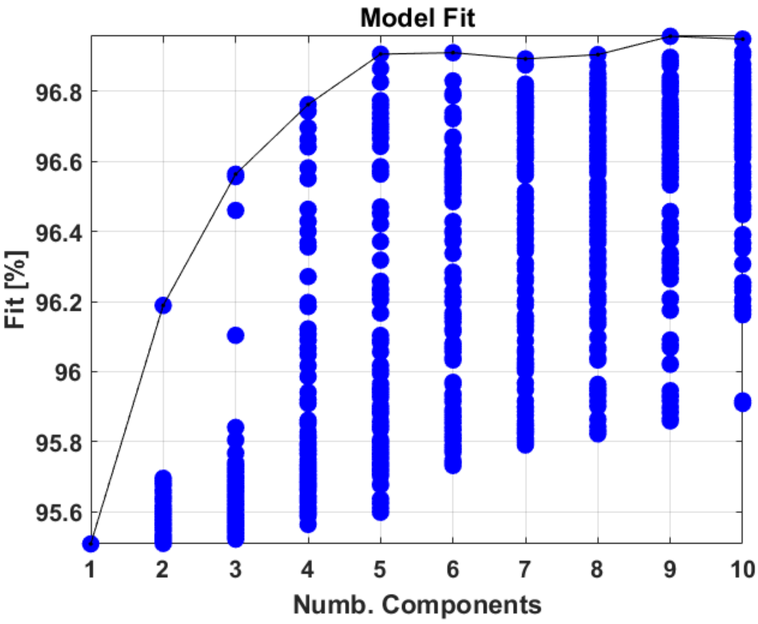


Figure 2.3. Model Fit. The figure reports the percentage fit of the DEDICOM model while varying the number of components P and also for different starting conditions as shown by the blue circles. The black line shows the maximum fit for each component. At $P = 2$, the DEDICOM results in a fit equal to 96.2%.

Next, we are interested in discovering relevant patterns, at the component level, that may reveal homogeneous mortality behaviours related to the evolutionary process of longevity improvements in different countries. Against this background, we normalize the loading matrix $\mathbf{A}$ so that the columns have unit length. In this way, the elements of such a normalized matrix, denoted by $\hat{\mathbf{A}}$, have comparable sizes and describe membership degrees relating the N countries to the P components. Moreover, we also normalize the interaction matrix $\mathbf{R}$ so that its diagonal elements are equal to one. Finally, we compensated these two normalizations in the matrices $\mathbf{D}_k$, $k = 1, \ldots, K$. In formulae, setting

$$\mathbf{T} = \text{diag}\left\{||\mathbf{a}_1||^{-1/2}, \cdots, ||\mathbf{a}_P||^{-1/2}\right\}, \tag{2.10}$$

and

$$\mathbf{V} = \text{diag}\{r_{11}^{-1/2}, \cdots, r_{PP}^{-1/2}\} \tag{2.11}$$

where diag$\{\circ\}$ is the operator which creates a diagonal matrix with the given argument values along the diagonal and $\mathbf{a}_p$ $(p = 1, \ldots, P)$ is the p-th column of $\mathbf{A}$, we have:

$$\hat{\mathbf{A}} = \mathbf{A}\mathbf{T}, \tag{2.12}$$

$$\hat{\mathbf{R}} = \mathbf{V}\mathbf{R}\mathbf{V}, \tag{2.13}$$

$$\hat{\mathbf{D}}_k = \mathbf{T}^{-1}\mathbf{D}_k\mathbf{V}^{-1} = \mathbf{V}^{-1}\mathbf{D}_k\mathbf{T}^{-1}, k = 1, \ldots, K. \tag{2.14}$$

Once the information on latent components in the data is extracted, the DEDICOM model associates each country to more than one of the latent components or patterns with different weights. In a nutshell, DEDICOM identifies the communities by grouping countries that have mortality profiles that are similar in ages and temporal dimensions.

Results are reported in Figure 2.4 and Figure 2.5 where, in the first figure, the upper left sub-plot displays the normalized component matrix $\hat{\mathbf{A}}$, i.e, for each country, the panel shows the degree of belonging to one of the two identified components. The lower left panel shows the temporal profiles embedded in the matrix $\hat{\mathbf{D}}$. This matrix describes the level of synchronization of death similarities within each component, i.e., a high value of an element of $\hat{\mathbf{D}}$ denotes the importance of a component for a year, i.e., in such a year, similarities among countries are strongly explained by the component involved. The second figure refers to the interaction matrix $\hat{\mathbf{R}}$ among the two components.

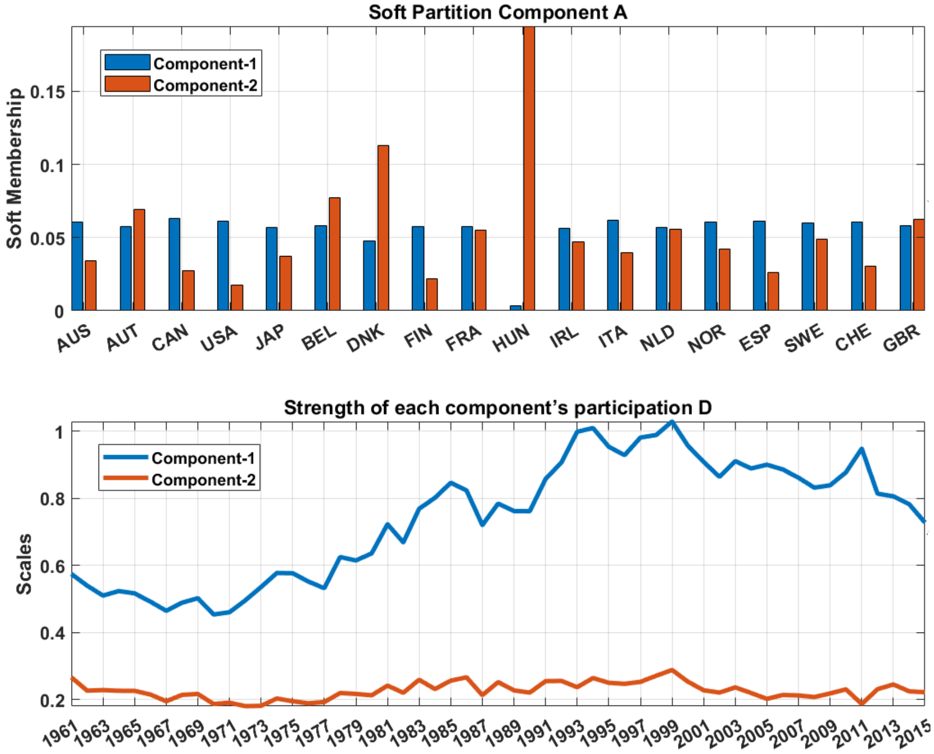


Figure 2.4. Country component membership degree $\hat{\mathbf{A}}$ and temporal activity patterns $\hat{\mathbf{D}}$. The upper left sub-plot reports, for each country, the degree of belonging to one of the two components. The lower left panel shows the level of synchronization of death similarities within each component.

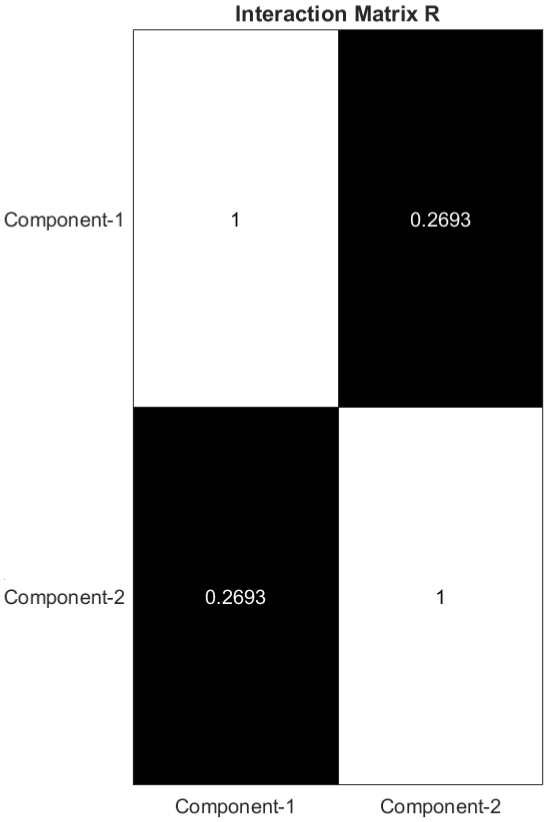


Figure 2.5. Latent components' interaction $\hat{\mathbf{R}}$. The figure refers to the interaction of patterns among the two components.

The DEDICOM model provides a soft membership distribution of countries inside components. Specifically, subgroups of countries can be retrieved independently from each other and countries can belong to more than one community. On the other hand, we also provide a hard partition scheme, where overlapping is not allowed, by assigning countries to the component in which they have the highest degree of membership. Figure 2.6 reports the WordCloud chats of such a hard partition scheme, where each colored country is assigned to the component in which it has the highest value of $\hat{\mathbf{A}}$. In particular, the right panel refers to component-1, and the left panel identifies component-2. The sizes of the words correspond to the country component membership degrees in $\hat{\mathbf{A}}$ and the colored countries are those having the highest membership degrees in the given component.

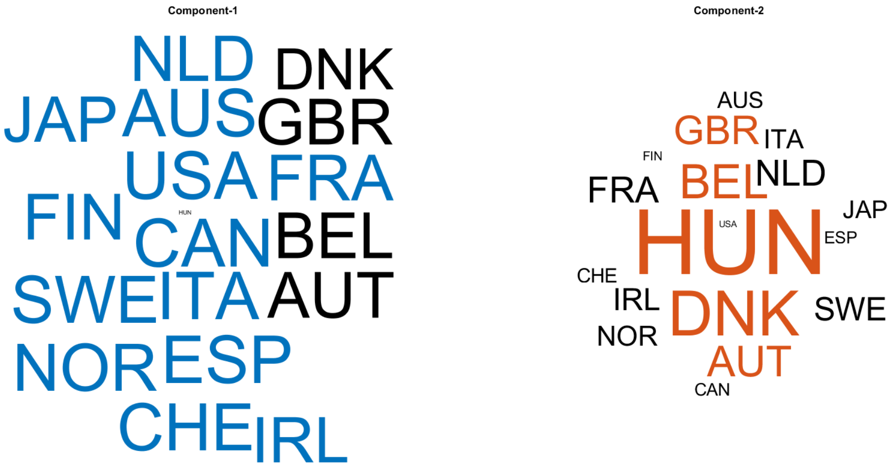


Figure 2.6. WordCloud chats of the hard partition scheme. The figure reports the hard partition scheme, where each colored country is assigned to the component in which it has the highest degree of membership. In particular, the left panel refers to component-1 and the the right panel identifies component-2. The sizes of the words correspond to the country component membership degrees in $\hat{\mathbf{A}}$ and the colored countries are those characterized by the highest membership degrees in the given component.

To better understand the results obtained, firstly, it is interesting to observe the composition of the dataset. We have considered 14 countries from Europe, and 4 countries from the rest of the world (Japan, USA, Canada, and Australia).

The first subgroup related to component-1 is characterized by countries well-known for their longevity, such as Italy, France, and Spain as underlined in [77] as emerging longevous populations, and some extra-EU countries, such as Japan. Northern countries exhibit remarkable levels of life expectancy at birth as observed by [22]. Indeed, in our study, Norway, Sweden, Finland (also with the Netherlands) fall into this component. Furthermore, it should be noted that this component is also characterized by the presence of two North American countries.

The results of component-2 can be led back to longevity patterns related to European countries located in the central/east area, mainly represented by Hungary, where the improvement in social economic conditions had reflection in the healthcare system [see 100]. This component also reveals similarity with Denmark, which *a priori* could be expected to form a cluster with the other Northern countries, well-known to be leaders in life expectancy. On the other hand, despite the geographical proximity, [22] highlight the different mortality trend between Denmark and Sweden, chosen as representative of the group of Northern countries, and explain some potential reasons such as lifestyle, and specific risk factors (smoking and alcohol consumption).

Moreover, the DEDICOM model, by taking into consideration both the cross-sectional and temporal features of country similarities, allows us to investigate how components respond to events affecting the logarithm of the central death rates. Indeed, the investigation of the scale matrix $\hat{\mathbf{D}}$, reported in Figure 2.4, reveals how the similarities summarized by the components have changed during time, i.e. the level of synchronization among components' members. While countries belonging to component-1 display an increasing level of synchronization up to the mid-nineties, component-2 has only a marginal level of synchronization. In order to shed light on the component dynamics, we identify the behavioral characteristics of each subgroup by examining its features values over time. The features of each subgroup have been identified as the average of the log central death rates of the countries that form such a subgroup, after having associated each country to each component according to the hard partition scheme. Results are reported in Figure 2.7, which shows the characteristics of the components and how they have been changing over time. Specifically, in the figure, the average log central death rate of each component is displayed as a dashed line, while standard errors are shown as colored areas and identify the variability of component features.

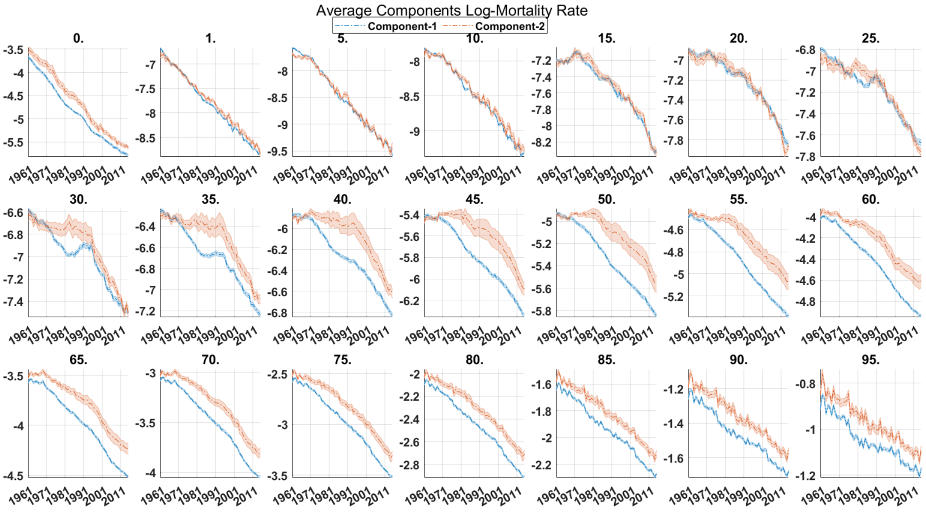


Figure 2.7. Death characteristics of the components over time. The figure shows the average log central death rate of each component, together with the standard errors and how they have been changing over time. In these panels, countries have been associated to each component according to the hard partition scheme.

Figure 2.7, clearly explains the different behavior of the two components, which shows distinct dynamics, especially for the age classes ranging from 30 to 80 years. Moreover, the figure reveals some interesting aggregate patterns at age classes. For instance, for 30 and 35 years one can notice the increasing log mortality rate associated with component-1 while from 30 to 55 component-2 shows a stable log mortality rate up to the mid-nineties, which decreases afterwards. Instead, for older classes, despite the two components display different log mortality rates, their decreasing trend is similar as suggested by the slope of the curves.

2.3 Conclusion and discussion

In this chapter we have applied the DEDICOM model to the similarity tensor computed on the multi population log centered death rate mortality data, starting from an initial array of three dimensions (*country* × *age* × *time*) obtained from the Human Mortality Database. The DEDICOM is an algebraic model, which provides information on latent components in the data that can be regarded as "properties" or "aspects" of the objects, by also supplying patterns of relationships among these components. In the DEDICOM model, individual objects can have substantial weights on more than one of the latent components or pattern, meaning that DEDICOM identifies types of correspondence patterns that have distinctive properties, and these are then linked to the individuals that exhibit mixtures of these patterns in their particular history. We have created yearly distance matrices, computing the cosine metric for measuring the similarities (distances) among each pair of countries defined by log central death rates vectors. We have shown the main advantage of DEDICOM: the model is able to extract meaningful relational patterns by reporting a walkthrough of our results: in the first place, we have displayed results of the tensor factorization and then membership of the countries designed to the components. In a nutshell, the stronger the similarity (i.e., the force that connects two countries mortality characteristics), the shorter the length of the links connecting the countries. In other words, pairs of countries that are similar receive higher weights since they are placed near by from each other, while values approaching zero are assigned to pairs with highly dissimilar characteristics. Moreover, the DEDICOM model, by taking into consideration both the cross-sectional and temporal features of country similarities, allows us to investigate how components respond to events affecting the logarithm of the central death rates. The 3-dimensional decomposition identifies persistent subgroups of countries with homogeneous mortality behaviours related to the evolutionary process of longevity improvements. In particular, we found two different clusters that reveal some peculiar aspects of the mortality phenomenon. The first community is mainly represented by longevous populations mixed with North American countries and Australia. The results of component-2 can be led back to longevity patterns related to European countries located in the central/east area, mainly represented by Hungary, where the improvement in social economic conditions had reflection in the healthcare system [see 100]. This component also reveals similarity with Denmark, which *a priori* could be expected to form a cluster with the other Northern countries, well-known to be leaders in life expectancy. On the other hand, despite the geographical proximity, [22] highlight the different mortality trend between Denmark and Sweden, chosen as representative of the group of Northern countries, and explain some potential reasons such as lifestyle, and specific risk factors (smoking and alcohol consumption).

Our work, by specifically describing the mesoscale interactions between subgroups and their evolution in time, could help to design appropriate actions against longevity risk that may impact on the stability conditions of life assurance and pensions.

Chapter 3

MODELING AND FORECASTING

In this chapter, firstly in 3.1 we refer to the three-way Lee Carter model [91], that is based on Tucker 3 decomposition, and that can be considered an extension of the classic Lee carter model [61]. The proposed approach allows us to simplify the data structure and to obtain a rank reduced representation. [91] applies the method considering ten European countries; here, to obtain different results, we apply the method to one of the resulting country's component obtained from the DEDICOM application. Then, in 3.2 following this line of research and focusing on the forecasting aim, we propose a coherent mortality forecasting using a four-way CANDECOMP/PARAFAC decomposition. In this context we consider the same ten European countries considered in [91] to make a better comparison. Our proposal based on the four-way structure allows managing mortality data aggregated in multi-dimensional settings, according to common demographic features: age class, time, country, and gender. We deal with four-dimensional mortality data using two main approaches proposed in the literature, the first one which works on centered mortality rates as in [28], and the second one working on compositional data as in [13]. Here, we provide two steps further on methodological developments in the field of mortality analysis and forecasting in a high-dimensional space. Firstly, compared to the current literature, we use an additional dimension, implementing a 4-way tensor decomposition. Thus, we further extend this framework including the CoDa analysis in the spirit of [14]. In the first part of the chapter the steps we follow are: description of the three-way Lee Carter model, description of the decomposition method (Tucker 3), identification and description of the database, construction of the 3-dimensional array ($age \times time \times country$), application of the decomposition method, choice of the number of components. In the second part the steps we follow are: description of the decomposition method (CANDECOMP/PARAFAC 4), description of the four-way models, identification and description of the database, construction of a 4-dimensional array ($age \times time \times country \times gender$), application of the decomposition method, choice of the number of components and forecasting. The tensor factorization can synthesize the behavior of the target variables through a few latent components and highlight the evolution of the temporal patterns. These patterns are employed to forecast future trajectories of mortality data with the

Vector-Error Correction model, which accounts for the non-stationarity of the series. We carry out Monte Carlo simulations to obtain the distributions of the time component over the forecasted period 2001-2015, and we evaluate the goodness of the prediction by computing the Root Mean Square Error and the Mean Absolute Error.

3.1 Mortality modeling: Three-way Lee Carter

The Lee-Carter model, developed in the early 1990s, is one of the earliest stochastic mortality models in the literature for a single-population and represents an extrapolative method based on a multiplicative two-way model which is used with time series analysis in forecasting mortality. [101] have highlighted the statistical features of this method as an association model for a two-way data matrix. The model is based, from both a statistical and an algebraic point of view, on a Singular Value Decomposition (SVD – [29]). The model used for mortality data is a principal component approach (PCA detailed in sec. 3.1.1), based on the log-death centered rates at age i at time j: $log(m_{ij})$. The death centered rates, computed from the mortality table provided by the HMD, for an individual of age i at time j is denoted by

$$m_{ij} = \frac{d_{ij}}{L_{ij}} \tag{3.1}$$

where

$$d_{ij} = l_{ij} - l_{i(j+1)} \tag{3.2}$$

is the expected number of deaths of a population between age i and $i+1$ in the year and j and l_{ij} is the expected number of living individuals at age i in the year j in a population initially made up of l_0 individuals, and

$$L_{ij} \simeq \frac{1}{2}[l_{ij} + l_{i(j+1)}] \tag{3.3}$$

is the risk exposure of the total population of age i in year k (assumed to be equal to the population at mid-year).

Following the notation of [91, 54], but unlike it, considering the age classes on the rows (i=1,...,I) and the years on the columns (j=1,...,J) as in [28], the Lee-Carter model is written as

$$ln(m_{ij}) = \alpha_i + \beta_i * k_j + e_{ij}, \quad i = 1, ..., I, \quad j = 1, ..., J \tag{3.4}$$

where α_i is the age-specific log-mortality average that is independent of time, β_i is an age-pattern of mortality change at age i that represents how rapidly or slowly, mortality at each age varies when the general level of mortality changes, and k_j reflects the trend in the level of mortality in the year j. The e_{ij} component represents the error term, which is assumed to be homoscedastic and normally distributed. Usually the model is expressed by considering the mean centred log-mortality rates:

$$\widetilde{m}_{ij} = ln(m_{ij}) - \alpha_i = \beta_i * k_j + e_{ij}, \quad i = 1, ..., I, \quad j = 1, ..., J \tag{3.5}$$

The values of α_i are calculated by averaging the log-mortality rates across the J years, for each age-class i = 1, ..., I. The values $\widetilde{m}_{ij}$ can be organized in a matrix **X**.

The SVD of the matrix **X** is:

$$\mathbf{X} = \mathbf{U}_{[I\times h]}\mathbf{S}_{[h\times h]}\mathbf{V'}_{[h\times J]}, \quad h \leq (I,J) \tag{3.6}$$

The scalar h is the rank of $\mathbf{X}$, $\mathbf{S}$ is a diagonal matrix $(h \times h)$ holding the positive singular values of $\mathbf{X}$. $\mathbf{U}$ is a matrix $(I \times h)$ and $\mathbf{V}$ is a matrix $(h \times J)$ that hold respectively, the left and right singular vectors $\mathbf{u}_i$ and $\mathbf{v}_j$ forming orthogonal bases for the I rows and the J columns of the matrix $\mathbf{X}$. The estimates of k_j and β_i in the Lee Carter model are obtained as:

$$\hat{\beta}_i = u_{i1}; \quad \hat{k}_j = \lambda_1 v'_{1j} \tag{3.7}$$

and these estimates minimize

$$\sum_{i=1}^{I}\sum_{j=1}^{J}[\widetilde{m}_{ij} - (\beta_i * k_j)]^2 \tag{3.8}$$

The last stage in the development of mortality modelling is characterized by research on multi-populations mortality models; the innovative paper is the Li and Lee model developed in 2005 [65], which extends the Lee and Carter model for a group of populations. In recent years, the introduction of further information and of three dimensions (such as, age, year and countries) has allowed scholars to use statistical methods that are based on tensor decomposition, developed years earlier. [91] proposes a three-way extension of the Lee-Carter model by considering death rates aggregated for time, age-groups and country. The Lee-Carter model cannot be applied to multi-population mortality data with 3 dimensions, so the three-way Lee-Carter model can be applied to multi-population mortality modeling using three-way methods. The three way Lee-Carter is:

$$\widetilde{m}_{ijl} = ln(m_{ijl}) - \alpha_{il} = \beta_i * k_j * \gamma_l + e_{ijl}, \tag{3.9}$$

$$i = 1, ..., I, \quad j = 1, ..., J \quad l = 1, ..., L$$

where α_{il}, are computed by averaging the log-mortality rates across the J years, for each age-group i=1, 2,...,I and for each country l=1,...,L, β_i and k_j have the same interpretation of the Lee Carter model, while γ_l (l=1,...,L) represents the term associated to the generic element of the third criterion, in [91] it represents the different countries. [91] points out that all of the parameters share information common to the three ways, that is the β_i estimates are influenced both by time and country effects. To estimate the parameters the authors proposed, as natural extension of the SVD in the three-way framework, the Tucker3 model (detailed in sec. 3.1.1).

3.1.1 Methodology

Having introduced the three way Lee Carter, to better understand it, we describe the decomposition method involved, the Tucker 3, starting from the more general context of the three-way component analysis techniques. Three-way component analysis techniques carry out a descriptive analysis of 3-way data that can emerge in different contexts, our context is mortality. Three-way data refers to data that can be arranged in a three-dimensional array, in our example (*age* × *time* × *country*). The three sets of entities associated with this four-way data set are the dimensions (or "modes") of the array. The idea of this method is to efficiently summarize all the information in the three-way data, also considering the various interactions (two-way and three-way); in particular, the method summarizes the entities of each mode through a few components (P,Q and R) and the relation between these components (Core tensor $\underline{\mathbf{G}}$). As we have seen, different techniques of three-way component analysis have been proposed, one of the best known is the Tucker3 (T3) analysis. We can consider this multi-way method a generalization of Principal Component Analysis (PCA), which is a technique to explore the relationship among two-way data. For convenience, PCA is briefly presented.

Principal Component Analysis

PCA is a statistical method that can be applied when the available information can be collected in a matrix, say $\mathbf{X}$. The generic element of this matrix, x_{ij}, expresses the score of the i-th observation unit (i=1,...,I) with respect to the j-th variable (j=1,...,J). We can consider $\mathbf{X}$ as a two-way tensor in $\Re^{I\times J}$ where the two modes are the observation units (from now on simply called "units") and the variables. For example, as seen in the Lee Carter model, the observation units are the age classes and the variables are the years. PCA with S ($\leq \min(I, J)$) components can be formulated as

$$x_{ij} = \sum_{s=1}^{S} a_{is} b_{js} + e_{ij}, \quad i = 1, ..., I, \quad j = 1, ..., J \tag{3.10}$$

where a_{is} and b_{js} are the component score of unit i for component s (i=1,...,I and s=1,...,S) and the component loading of variable j for component s (j=1,...,J and s=1,...,S), respectively, S denoting the number of components, and e_{ij} is the error term, the generic element of the matrix $\mathbf{E}$. In matrix form PCA can be rewritten as

$$\mathbf{X} = \mathbf{AB}' + \mathbf{E}, \tag{3.11}$$

where $\mathbf{A}$ ($I \times S$) with generic element a_{is} and $\mathbf{B}$ ($J \times S$) with generic element b_{js} denote the component score and component loading matrices, respectively, and $\mathbf{E}$ ($I \times J$) is the error matrix. PCA finds estimates for $\mathbf{A}$ and $\mathbf{B}$ by minimizing

$$||\mathbf{E}||^2 = ||\mathbf{X} - \mathbf{AB}'||^2, \tag{3.12}$$

with respect to $\mathbf{A}$ and $\mathbf{B}$, where the symbol $||\cdot||$ denotes the Frobenius norm of matrices. The estimates of $\mathbf{A}$ and $\mathbf{B}$ are found by computing the SVD of $\mathbf{X}$, thus highlighting a strong connection between PCA and the Lee-Carter[61] model.

Tucker3

[107] summarizes the different names for the Tucker decomposition as in 3.1.

Name	Proposed by
Three-mode factor analysis (3MFA/Tucker3)	Tucker [105]
Three-mode principal component analysis (3MPCA)	Kroonenberg and De Leeuw [58]
N-mode principal components analysis	Kapteyn et al. [53]
Higher-order SVD (HOSVD)	De Lathauwer et al. [26]

Table 3.1. Different names for the Tucker decomposition

As we have seen, PCA is a statistical method to explore the relationship among two-way data. However, the information is often replicated in different (time) occasions, in this situation there is an additional mode (labelled occasions), and the data is stored in a three-way array, a three-way tensor in $\underline{\mathbf{X}}$ in $\Re^{I\times J\times L}$ with generic element x_{ijl} expressing the score of observation unit i for variable j at occasion l (l=1,...,L). In our example these time occasions are the different countries. The Tucker3 model summarizes a three-way tensor as:

$$x_{ijl} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b_{jq}c_{lr}g_{pqr} + e_{ijl}, \quad i=1,...,I, \quad j=1,...,J, \quad l=1,...,L. \tag{3.13}$$

Differently from PCA, different sets of components are determined for every dimension. In particular, P ($\leq I$) components for the units, Q ($\leq J$) components for the variables and R ($\leq L$) components for the occasions are sought. In (3.13), a_{ip} and b_{jq} express the relations between the units and the components for the units and the variables and the components for the variables. With respect to PCA, the Tucker3 model adds new loadings, denoted by c_{lr} (l=1,...,L and r=1,...,R), linking occasions and components for the occasions. The triple interactions between the components of all the dimensions are provided by g_{pqr}, the generic element of the core tensor $\underline{\mathbf{G}}$ of order $(P \times Q \times R)$. A high value of g_{pqr} in absolute sense suggests a strong relation among these components. Finally, e_{ijl} is the generic error term belonging to the error term tensor $\underline{\mathbf{E}}$.

To further compare PCA and Tucker3, it is convenient to express the Tucker3 model in matrix notation. We get:

$$\mathbf{X}_A = \mathbf{A}\mathbf{G}_A(\mathbf{C}\otimes\mathbf{B})' + \mathbf{E}_A, \tag{3.14}$$

where $\mathbf{X}_A$ is the matrix of order $(I \times JL)$, representing the so-called unit mode matricization of $\underline{\mathbf{X}}$, i.e., the matrix obtained by juxtaposing the matrices corresponding to the K occasions that are juxtaposed next to each other. The matrices $\mathbf{A}$ $\mathbf{B}$ and $\mathbf{C}$ of order $(I \times P)$, $(J \times Q)$ and $(L \times R)$, respectively, are the component matrices for the units, variables and occasions, respectively. It can be shown that these matrices can be constrained to be columnwise orthnormal without loss of fit (see, e.g., [58]). Furthermore, $\mathbf{G}_A$ of order $(P \times QR)$ and $\mathbf{E}_A$ of order $(I \times JL)$ denote the unit mode matricizations of $\underline{\mathbf{G}}$ and $\underline{\mathbf{E}}$, respectively. Finally, the symbol $\otimes$ denotes the Kronecker product of matrices. By comparing (3.11) and (3.14) we can observe the idea underlying the Tucker3 model. The Tucker3 model can be seen as a particular PCA of $\mathbf{X}_A$ where the component loadings are constrained to be $\mathbf{G}_A(\mathbf{C}\otimes\mathbf{B})'$. Such a particular structure allows for properly exploiting the three-way structure of the data. In principle, we might apply standard PCA to $\mathbf{X}_A$, but in this way the component loadings fully ignore that the same variables are observed in different occasions.

So, the Tucker decomposition is a form of higher-order principal component analysis. It decomposes a tensor into a core tensor multiplied (or transformed) by a matrix along each mode. The Tucker decomposition is illustrated in the 3.1

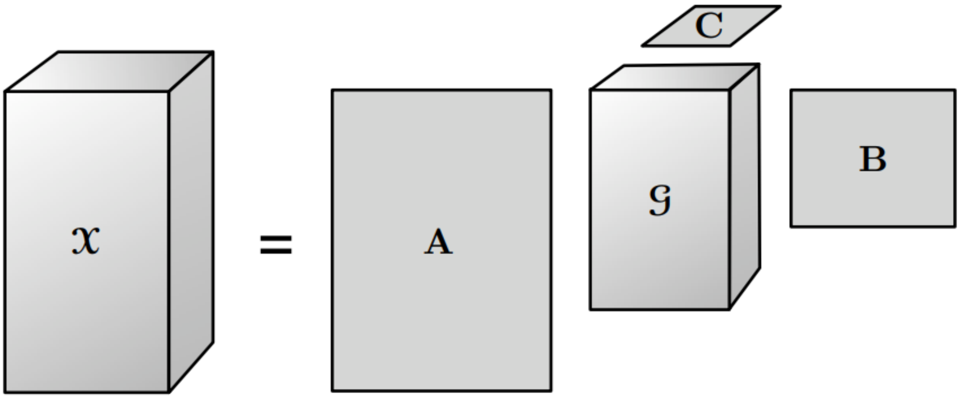


Figure 3.1. Tucker decomposition of a three way array

Most fitting algorithms (discussed in §4.2) assume that the factor matrices are columnwise orthonormal, but it is not required. In fact, CANDECOM/PARAFAC (described in 3.2.1 can be viewed as a special case of Tucker where the core tensor is superdiagonal and P = Q = R.

For completeness we illustrate also the tensor notation for the Tucker3 as in [51]:

$$\underline{\mathbf{X}} = \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_2 \mathbf{C} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} (\mathbf{a}_p \circ \mathbf{b}_q \circ \mathbf{c}_r) \tag{3.15}$$

For further details on the comparison between the notations, see [51].

As expresses in [99] one of the main advantage of the Tucker model is that we can observe the *full core array* $\underline{\mathbf{G}}$ that explain the relation between the components. This will be more clear in the 4 where it help us to identify the main relationship between the components in a more complex situation with four dimension. Another advantages of the model emerge in [99] that explains the Tucker3 model possesses some of the properties of the singular value decomposition; in particular, it allows for a complete decomposition of the three-way array. It is also possible, given orthonormal components, to make a complete component-wise partitioning of the variability. The model has subspace uniqueness. Properties this model lacks are component uniqueness (any transformation of the components and/or core array with the inverse transformation of the appropriate other matrices will not affect the fit) and nesting of solutions.

Estimation in Tucker3 is carried out by following the least squares approach. Thus, the parameter estimates are found by minimizing the sum of squared errors:

$$||\mathbf{E}_A||^2 = ||\mathbf{X}_A - \mathbf{A}\mathbf{G}_A(\mathbf{C}\otimes\mathbf{B})'||^2, \tag{3.16}$$

with respect to **A**, **B**, **C**, and $\mathbf{G}_A$.

Once the estimates of **A**, **B**, **C** and $\underline{\mathbf{G}}$ are determined, the fit percentage of the Tucker3 model can be estimated by

$$[1 - (||\mathbf{E}_A||^2/||\mathbf{X}_A||^2)]100. \tag{3.17}$$

The closer to 100, the better the fit of the model, and the fit percentage is used to select the optimal number of components.

[107] summarizes some of the different alghoritms to compute the Tucker decomposition starting from 1966, when [105] introduced three methods for computing a Tucker decomposition, but he was somewhat hampered by the computing ability of the day, stating that calculating the eigendecomposition for a 300×300 matrix "may exceed computer capacity." The basic idea is to find those components that best capture the variation in mode n, independent of the other modes. Tucker presented it only for the three-way case, but the generalization to N ways is straightforward. This is sometimes referred to as the "Tucker1" method, though it is not clear whether this is because a Tucker1 factorization is computed for each mode or it was Tucker's first method. Today, this method is better known as the higher-order SVD (HOSVD) from [26], who showed that the HOSVD is a convincing generalization of the matrix SVD and discussed ways to efficiently compute the singular vectors of $\mathbf{X}(\mathrm{n})$. When $R_n < rank_n(\underline{\mathbf{X}})$ for one or more n, the decomposition is called the *truncated* HOSVD. The truncated HOSVD is not optimal, but it is a good starting point for an iterative alternating least squares (ALS) algorithm. In 1980 [58] developed an ALS algorithm, called TUCKALS3, for computing a Tucker decomposition for three-way arrays. [53] later extended TUCKALS3 to N-way arrays for N > 3. [26] called it the *Higher-order Orthogonal Iteration* (HOOI). [24] consider methods for speeding up the HOOI algorithm such as how to do the computations, how to initialize the method, and how to compute the singular vectors. Improvement software and algorithms are still being studied today, for example [76]developed a package in software R that provides a set of tools for fitting the Tucker3 models to multidimensional arrays by use of classical, robust and also compositional estimating procedures.

Application of Tucker

[107] summarize some applications of the Tucker method. Several examples of using the Tucker decomposition in chemical analysis are provided by Henrion [47] as part of a tutorial on N-way PCA. Examples from psychometrics are provided by Kiers and Van Mechelen [54] in their overview of three-way component analysis techniques. The overview is a good introduction to three-way methods, explaining when to use three-way techniques rather than two-way (based on an ANOVA test), how to preprocess the data, guidance on choosing the rank of the decomposition and an appropriate rotation, and methods for presenting the results. De Lathauwer and Vandewalle [26] consider applications of the Tucker decomposition to signal processing. Vasilescu and Terzopoulos [71] pioneered the use of Tucker decompositions in computer vision with TensorFaces. In data mining, [8]] applied the HOSVD to the problem of identifying handwritten digits.

Tucker3 in the Three-way Lee-Carter model

So, with reference to the three-way Lee Carter model [91], the Tucker 3 model factorizes the three-way array $\mathbf{X}$ and can be expressed in this generic form:

$$\widetilde{m}_{ijl} = ln(m_{ijl}) - \alpha_{il} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b_{jq}c_{lr}g_{pqr} + e_{ijl} \tag{3.18}$$

$$i = 1, ..., I, \quad j = 1, ..., J, \quad l = 1, ..., L$$

with specific reference to the notation of the Lee Carter model we obtain:

$$\widetilde{m}_{ijl} = ln(m_{ijl}) - \alpha_{il} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} \beta_{ip}k_{jq}\gamma_{lr}g_{pqr} + e_{ijl} \tag{3.19}$$

$$i = 1, ..., I, \quad j = 1, ..., J, \quad l = 1, ..., L$$

The left-hand side of equation holds the log-mortality rates expressed as deviations from the average of each age-group i. The terms I, J and L denote the numbers of elements in each of the three modes, respectively age classes, years, and countries; P, Q, R are the numbers of components selected in the approximation. The coefficients β_{ip}, k_{jq} and γ_{lr} are the entries of the component matrices for the first, second and third mode. The g_{pqr} are the elements of the three-way core matrix. By setting p, q, r equal to 1, we obtain the estimates for the three way Lee Carter model 3.1, assuming the remaining components as residual terms.

Choice of the numbers of components

To choose the number of components we balanced fit and parsimony. Those criteria involve subjective decisions based on objective measures, and the decisions should be explicit. Regarding balancing fitting, we use this measure to compare different models, but not in an absolute sense. Another important aspect is that if we based our choice only on fitting, we would probably choose the most complex model. Therefore, we should consider a compromise between fitting and parsimony. The balance between these two objectives is based on the purpose of the analysis and cannot be determined on the basis of other criteria. For this choice, we can choose one from different solutions based on the same total number of components $(P + Q + R)$ that are compared and the best of these are selected. Then the selected solutions are ordered with respect to the total numbers of components. From these different ordered solutions we can search that corresponding to a considerable fit increase compared with the best solution with one component less but for which adding more components gives a relatively small increase. As we will see, we can base our decision also on the basis of the objective of the analysis, in particular on interpretability [54]. These aspect will be further analysed in the following chapter 4.

3.1.2 Application

[91] applies the method considering ten European countries, because they are geographically closer and share similar social-economic status; nevertheless [28]

specifies that the tensor decomposition method does not require similar mortality experience among different populations, which can be applied to other more different populations too. So, following this guideline, here to obtain different results respect to [91], here we apply the method to the resulting country's components obtained from the DEDICOM application and in the second part of the chapter, with reference to the four-way Lee Carter, we consider the same ten countries of [91] for a better comparison with the three-way Lee Carter. In particular, we consider the countries of the first component resulting from the DEDICOM application of the previous chapter. So, we consider the dataset provided by the Human Mortality Database (HMD) [1], which is an archive of death information for several countries. The application of the methodology requires to structure the data in four dimensions: age, time, country, sex. The age dimension is organized in 21 classes from 0 to 99 years, with a step of 5 years, except for the first two classes, which represent respectively the individuals aged 0, and aged 1-4. Regarding the time dimension, we focus the analysis on the years from 1961 to 2015 in order to consider the same time window for each country (see the thirteen countries listed in Table 3.2). The analysis is developed considering the total population.

Code	Countries
AUS	Australia
CAN	Canada
USA	USA
JAP	Japan
FIN	Finalnd
FRATNP	France
IRL	Ireland
ITA	Italy
NLD	Netherlands
NOR	Norway
ESP	Spain
SWE	Sweden
CHE	Switzerland

Table 3.2. Countries considered from HMD

In order to apply the multi way method, we organize the data in an array of three dimensions (age × time × country). The dimension of the array is: $21 \times 55 \times 13$ for a total of 15.015 entries and we collect in it the $\widetilde{m}_{ijl}$.

Results

The results in this paragraph are obtained through the use of the R packages called Three-way [35].

Figure 3.2 gives fit values (%) of all solutions plotted against number of free parameters. Following [91] the possible solution is P=2, Q=2, R=2 with the fit value of 94.45%; this solution represents a good compromise between goodness-of-fit

---

[1] (Human Mortality Database. University of California; Berkeley (USA); and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 08/11/21).

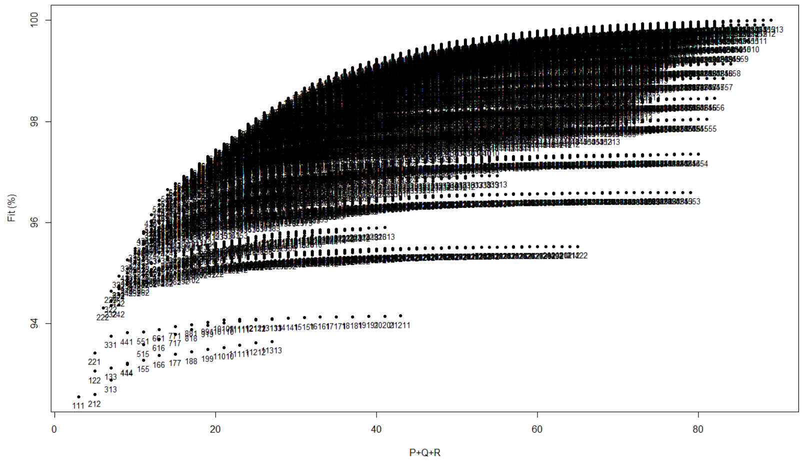


Figure 3.2. Possible solutions

and parsimony. In fact, the solutions with higher fit values have more additional components for a limited increase of fit.

Table 3.3 represents the tabular representation of tensor core, where the higher in absolute value an element of the core, the stronger the interaction among the components involved. The value 43.55 indicates the higher contribution made by the first SVD components for the three modes.

Age components	Country1		Country2	
	Year1	Year2	Year1	Year2
Age1	2.28	-1.74	**4.47**	0.01
Age2	3.22	0.02	-0.01	**43.55**

Table 3.3. Interaction

The image 3.3 represents $\beta_{i,p=1,2}$, the two components for the first dimension, the classes of age.

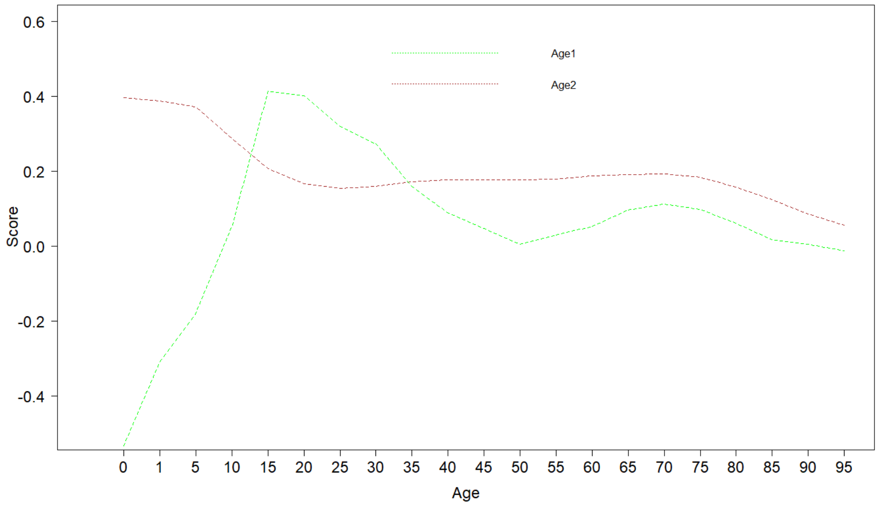


Figure 3.3. Age components

The first component has a peak at 15-20 years and it could describe the youth mortality linked to road accidents. As in [91] the second component of the class of age, highlighted in brown, has a regular pattern from class 20 to class 80. Also the component is affected by the different contributions of the younger age-groups.

The image 3.4 represents $k_{j,q=1,2}$, the two components for the second dimension, the years.

Figure 3.4. Year components

The first component increases until 1994 and then start to decrease, instead as in [91] the second component of the years, highlighted in red, has a linear trend.

The image 3.5 represents $\gamma_{l,r=1,2}$ for the third dimension, the countries.

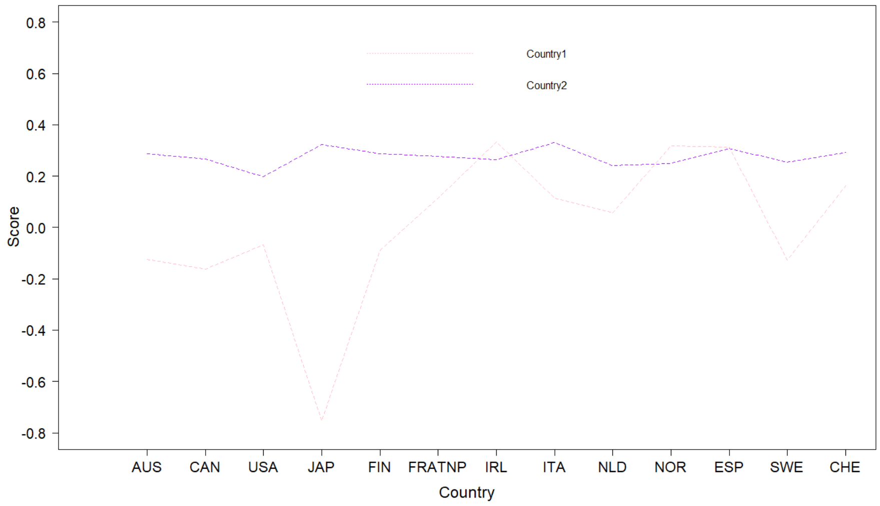


Figure 3.5. Country components

The first component may describe the mortality of Japan respect to the other countries, in fact Japan is well known for its longevity pattern. The second component has a regular pattern and we may think that it describes all the thirteen countries considered from the database (HMD).

Finally, after the application of the method, [91] suggest how the model proposed in this paper may be used to deal with forecasting; we focus on this problem in the second part of the chapter, considering another dimension.

3.1.3 Conclusion

In this first part of the chapter we have applied the three way Lee Carter Model, that extends the Lee Carter model and specifies a three-way model structure which looks at the mortality data aggregated according to three criteria: age class, time, country. The model incorporates a further component in the decomposition of the log-mortality rates, specifically, we refer to the Tucker 3 decomposition method. The proposed approach allows us to simplify the data structure and to obtain a rank reduced representation. Through the results derived by the Tucker3 decomposition we have compared, in a numerical and graphical way, the relationships among all three modes. For example we have discover that the first component of the age has a peak at 15-20 years and it could describe the youth mortality linked to road accidents and the second component has a regular pattern from class 20 to class 80 and it is also affected by the different contributions of the younger age-groups; or that the first component of the countries may describe the mortality of Japan respect to the other countries and that the second component of the countries has a regular pattern and we may think that it describes all the thirteen countries considered from the database. In this sense, the model is useful for understanding some aspects of mortality considering the three dimensions under consideration. Finally, [91] suggest how the model proposed in this paper may be used to deal with forecasting. To address this problem, in the second part of this chapter, we propose a coherent mortality forecasting using a four-way CANDECOMP/PARAFAC decomposition, hence considering another dimension. The idea comes from the fact that [28] generalises the model used in [91] using different tensor decompositions by considering mortality data aggregated for age, year, and country/gender. Relying on this background, the authors solve the multi-population mortality forecasting problem based on 3-way data (age, year, and population/gender). They formulate a mortality model by applying both the CPD method and different forms of the Tucker method to allow for different ranks in different dimensions. In this second part, we provide two steps further on methodological developments in the field of mortality analysis and forecasting in a high-dimensional space. Firstly, compared to the current literature, we use an additional dimension, implementing a 4-way tensor decomposition. Thus, we further extend this framework including the CoDa analysis in the spirit of [14].

3.2 Mortality forecasting using a four-way CANDECOMP/PARAFAC decomposition

In this second part of the chapter, we propose a coherent mortality forecasting using a four-way CANDECOMP/PARAFAC decomposition. Our proposal based on the four-way structure allows managing mortality data aggregated in multi-dimensional settings, according to common demographic features: age class, time, population, and gender. Here, we apply Parafac, instead of Tucker, for simplicity and also because the gender dimension has only two different modes. Understanding mortality dynamics in a high-dimensional framework is crucial for all the analyses involving demographic assessments. The fact that people are living longer causes, for instance: i) an increase in the time spent in retirement that poses challenges to the sustainability of pension systems; ii) an increase in the use of medical and assistance treatments that enhances health care expenditure. This has forced policymakers, on the one hand, to rethink the design of pension plans to mitigate the burden of increased longevity and, on the other hand, to make significant efforts in containing health care costs.

We deal with four-dimensional mortality data using two main approaches proposed in the literature, the first one which works on centered mortality rates as in [28], and the second one working on compositional data as in [13]. In this context we consider the same ten European countries considered in [91] for a better comparison. The steps we follow are: description of the decomposition method (CANDECOMP/PARAFAC 4), description of the four-way models, identification and description of the database, construction of a 4-dimensional array ($age \times time \times country \times gender$), application of the decomposition method, choice of the number of components and forecasting. The tensor factorization can synthesize the behavior of the target variables through a few latent components and highlight the evolution of the temporal patterns. These patterns are employed to forecast future trajectories of mortality data with the Vector-Error Correction model, which accounts for the non-stationarity of the series. We carry out Monte Carlo simulations to obtain the distributions of the time component over the forecasted period 2001-2015, and we evaluate the goodness of the prediction by computing the Root Mean Square Error and the

Mean Absolute Error.The steady reduction in death rates at all ages represents an enormous achievement that is the result of progress in almost all sectors of modern society; this fueled the improvement of statistical-mathematical methods devoted to mortality modeling and forecasting. To this aim, the prevalent framework (for a complete discussion see [16]) is the extrapolative approach, which found regularities in historical death rates during the estimation phase, and projects them forward using econometric models in the forecasting step. This approach includes the stochastic models such as the Lee-Carter [61] described in 3.1 where we have seen that the method suggests using principal component analysis to describe and forecast mortality trends. Following the approach based on the matrix decomposition, [82] introduced the Compositional Data Analysis (CoDa), the method by [2], into the Lee-Carter model. The compositional framework occurs whenever we need to partition a whole measurement into percentage contributions from its components, thus with the unit-sum constraint, which represents an important advantage in demographic context, ensuring coherence of the estimated outcome. This coherence might be related to causes, different populations, or gender. Indeed, longevity changes are often projected separately, decreasing the projection reliability between groups in the long run. In this framework, [14] provide the first attempt to combine the multi-way analysis considering the compositional nature of data in a unified framework. Indeed, the authors propose a methodology based on simultaneously modeling and forecasting extending the three-way technique to a CoDa framework for multiple populations. To test this approach, they carried out a Tucker3 analysis on deaths counts of Canadian provinces and territories treated as compositions. More recently, [28] formulate a mortality model by applying both the CPD method and different forms of the Tucker method to allow for different ranks in different dimensions. The Authors underline that tensor decomposition allows researchers to obtain combined estimates for the three modes (age, time and different populations) and see relationships among them. In this second part, we provide two steps further on methodological developments in the field of mortality analysis and forecasting in a high-dimensional space. Firstly, compared to the current literature, we use an additional dimension, implementing a 4-way tensor decomposition. Thus, we further extend this framework including the CoDa analysis in the spirit of [14]. To test this approach, the CANDECOMP/PARAFAC analysis will be carried out on both mortality rates and death counts of the European countries included in the HMD.

3.2.1 Methodology

We have seen that multi-way methods can be considered a generalization of the PCA for exploring the relationship among multi-way data. In this paper, we use the Canonical Polyadic Decomposition (CANDECOMP) or Parallel Factors (PARAFAC) model. The CANDECOMP model was proposed by [18], while the PARAFAC model by [42]. As pointed out by [34], although CANDECOMP/PARAFAC is probably the most widely used name, some authors, such as [28], refer to it as the Canonical Polyadic Decomposition (CPD) to emphasize the preceedings papers of [48] and [49]. To better understand this, [107] summarizes the different names for the CP decomposition as in 3.4.

Name	Proposed by
Polyadic Form of a Tensor	Hitchcock [48]
PARAFAC (Parallel Factors)	Harshman [42]
CANDECOMP or CAND (Canonical decomposition)	Carroll and Chang [18]
CP (CANDECOMP/PARAFAC)	Kiers [56]

Table 3.4. Different names for the CP decomposition

Decomposition of three-way data using CANDECOMP/PARAFAC (CP).

We have seen that when dealing with three-dimensional data, matrix $\mathbf{X}$ becomes a tensor $\underline{\mathbf{X}}$, whose generic element is denoted by x_{ijl}. $\underline{\mathbf{X}}$ is a three-way tensor in $\Re^{I\times J\times L}$, where the three ways are the sets of indexes i=1,...,I, j=1,...,J and l=1,...,L. CP decomposes a tensor into a sum of component rank-one tensors. Elementwise, this can be illustrated in the 3.6

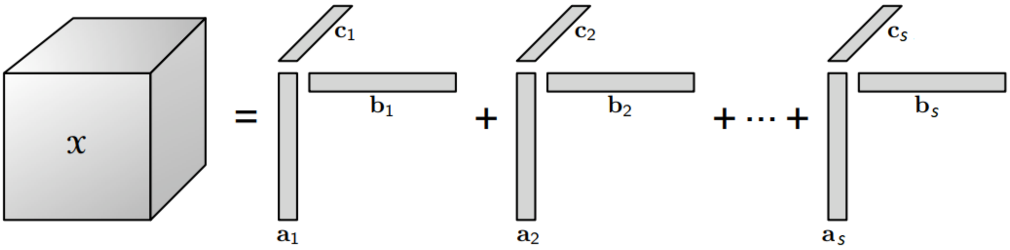


Figure 3.6. CP decomposition of a three way array

So, tensor $\underline{\mathbf{X}}$ can be decomposed using the CP model, which adds new scores (denoted by c_{ls} ($l = 1, \ldots, L$, $s = 1, \ldots, S$), linking occasions and components, to Eq. 3.10. In scalar form, the CP model can be formulated as:

$$x_{ijl} = \sum_{s=1}^{S} a_{is}b_{js}c_{ls} + e_{ijl}, \quad i = 1, ..., I, \quad j = 1, ..., J \quad l = 1, ..., L, \tag{3.20}$$

where e_{ijl} is the generic error term belonging to the error term tensor $\underline{\mathbf{E}}$. The CP model can also be expressed in tensor formulation, but it is more common to consider the matrix formulation. We get:

$$\mathbf{X}_A = \mathbf{A}(\mathbf{C} \odot \mathbf{B})' + \mathbf{E}_A \tag{3.21}$$

where $\mathbf{X}_A$, and $\mathbf{E}_A$ represent the unit mode matricizations of $\underline{\mathbf{X}}$, and $\underline{\mathbf{E}}$, respectively (see, e.g., [34]). $\mathbf{A}$ ($I \times S$), $\mathbf{B}$ ($J \times S$), and $\mathbf{C}$ ($I \times S$) are the component matrices for three modes with generic elements a_{is}, b_{js} and c_{ls}, respectively, and $\odot$ denotes the so-called Khatri-Rao product.

The three-way model is sometimes written in terms of the frontal slices of $\underline{\mathbf{X}}$ (see figures 1.3 and 1.4).

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}^{(k)}\mathbf{B}' + \mathbf{E}_k \tag{3.22}$$

where $\mathbf{D}^{(k)} = diag(c_{k:})$ for $k = 1, \ldots, K$. Analogous equations can be written for the horizontal and lateral slices. In general, though, slice-wise expressions do not easily extend beyond three dimensions.

For completeness we illustrate also the tensor notation for the CP3 as in [51]:

$$\underline{\mathbf{X}} = \underline{\mathbf{I}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_2 \mathbf{C} = \sum_{s=1}^{S} (\mathbf{a}_s \circ \mathbf{b}_s \circ \mathbf{c}_s) \tag{3.23}$$

where $\circ$ is the outer product. For further details on the comparison between the notations, see [51].

Decomposition of four-way data using CANDECOMP/PARAFAC.

When dealing with four-dimensional data, the available information can be collected in the four-way tensor $\underline{\mathbf{X}}$ in $\Re^{I \times J \times L \times N}$, whose generic element is denoted by x_{ijln}. The sets of indexes are in this case i=1,...,I, j=1,...,J, l=1,...,L and n=1,...,N. In scalar notation, the four-way CP can be formulated as follows:

$$x_{ijln} = \sum_{s=1}^{S} a_{is} b_{js} c_{ls} d_{ns} + e_{ijln}, \quad i = 1, ..., I, \quad j = 1, ..., J \quad l = 1, ..., L \quad n = 1, ..., N, \tag{3.24}$$

where e_{ijln} is the generic error term, and a_{is}, b_{js}, c_{ls}, and d_{ns} are the generic elements of the four underlying component matrices **A**, **B**, **C** and **D**, respectively, where **D** refers to the additional, fourth mode. Similarly to the three-way model, the CP model expresses the four-way array as the sum of S rank-1 four-way arrays. It can be shown that the rank of $\underline{\mathbf{X}}$ corresponds to the minimum number of rank-1 four-way arrays needed to perfectly decompose $\underline{\mathbf{X}}$. In matrix formulation, we get:

$$\mathbf{X}_A = \mathbf{A}(\mathbf{D} \odot \mathbf{C} \odot \mathbf{B})' + \mathbf{E}_A \tag{3.25}$$

Estimation is carried out in the least-square sense by minimizing:

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{l=1}^{L} \sum_{n=1}^{N} e_{ijln}^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{l=1}^{L} \sum_{n=1}^{N} (x_{ijln} - \sum_{s=1}^{S} a_{is} b_{js} c_{ls} d_{ns})^2 \tag{3.26}$$

with respect to **A**, **B**, **C**, and **D**. The previous expression can also be used to evaluate the fit percentage of the CP solution:

$$\left[1 - \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{l=1}^{L} \sum_{n=1}^{N} e_{ijln}^2}{\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{l=1}^{L} \sum_{n=1}^{N} x_{ijln}^2} \right] 100. \tag{3.27}$$

The closer to 100, the better the fit of the CP model. As for PCA, the fit percentage in Eq. 3.27 helps select the optimal number of components.

Advantage of the model emerge in [99] that also explains some difference with the Tucker model. [99] explains the the nature of the model is such that the components do not need to be constrained to orthonormality in order to find a solution, and the components are correlated in most applications. A further characteristic of the model is that, like the Tucker3 model, it is symmetric in all modes and that, unlike the Tucker3 model, the numbers of components in all modes are necessarily the same. The simple superdiagonal core array facilitates the interpretation of the model compared to the Tucker models, but on the other hand the correlated components in the modes may complicate the interpretation. Moreover the model possesses some of the properties of the singular value decomposition, but not all. In particular, each component of a mode is uniquely related to only one component in each of the other modes. A very special characteristic of the Parafac model is its component uniqueness: any transformation of the components will affect the fit of the model to the data (called the "intrinsic axis property" by Harshman). More technically, the model is identified, generally even overidentified due to the restrictions, so that it is possible to test whether the model fits or not. This feature, together with the relative simplicity of the model, makes it extremely attractive, certainly from a theoretical point of view.

[107] illustrated tha the first issue that arises in computing a CP decomposition is how to chose the number of rank-one components. Most procedures fit multiple CP decompositions with different numbers of components until one is "good". If the data is noise-free data, then the procedure can compute the CP model for S = 1, 2, 3, . . . and stop at the first value that gives a fit of 100 percentage. Assuming the number of components is fixed, there are many algorithms to compute a CP decomposition. We consider the alternating least squares (ALS) method proposed in the original papers by Carroll and Chang [18] and Harshman [42]. For convenience, we only illustrate the idea behind the method in the third-order case. The alternating least squares approach fixes **B** and **C** to solve for **A**, then fixes **A** and **C** to solve for **B**, then fixes **A** and **B** to solve for **C**, and continues to repeat the entire procedure until some convergence criterion is satisfied. Having fixed all but one matrix, the problem reduces to a linear least squares problem. [107] illustrate the full procedure for a N-way tensor.

Application of CP

[107] summarize some applications of the CP method starting from the first application. We have seen that the first application where in the filed of psychometrics in 1970. [18] introduced CANDECOMP in the context of analyzing multiple similarity or dissimilarity matrices from a variety of subjects. The idea was that simply averaging the data for all the subjects annihilated different points of view on the data. They applied the method to one data set on auditory tones from Bell Labs and to another data set of comparisons of countries. [42] introduced PARAFAC because it eliminated the ambiguity associated with two-dimensional PCA and thus has better uniqueness properties. Other authors have used CP decompositions in neuroscience. For example [30] apply CP to a time-varying EEG spectrum arranged as a three-dimensional array with modes corresponding to time, frequency, and channel.

The first application of tensors in data mining was by [1] who applied different tensor decompositions, including CP, to the problem of discussion detanglement in online chatrooms. In text analysis, [11] used CP for automatic conversation detection in email over time using a term-byauthor-by-time array. The method is still applied today, both to different fields and as a starting point for studying new algorithms and new estimation procedures, for example [70] developed a package in software R that provides a set of tools for fitting the CP3 models to multidimensional arrays by use of classical, robust and also compositional estimating procedures; and [93] implement a 4th-order extension of the effcient trilinear procedure INT-2 to tackle estimating setbacks and test it in a simulation study.

3.2.2 Mortality four-way decomposition

Following the methodology previously described, we now illustrate the four-way decomposition of mortality data arranged in a four-way array. Our analysis concerns both the decomposition of log-centered mortality rates (log-CMR) and compositional data (CoDa) obtained as the *clr* transformation of the deaths counts as described in the following paragraphs.

Decomposition of four-way data using CANDECOMP/PARAFAC.

Let us denote as $m_{x,t,p,g}$ the centered mortality rate for age x, time t, population p, and gender g. First, our objective is to decompose the following tensor, representing the four-way log centered mortality rates:

$$m_{x,t,p,g} = \log\left(m_{x,t,p,g}\right) - \alpha_{x,p,g} \tag{3.28}$$

where $\alpha_{x,p,g}$ is the average value of $\log m_{x,t,p,g}$ over time t.
Applying the model expressed in Eq. 3.24, we obtain the four-way CP of the matrix of log centered mortality rates, where the generic element $m_{x,t,p,g}$ can be represented as follows:

$$m_{x,t,p,g} = \sum_{s=1}^{S} a_{xs} b_{ts} c_{ps} d_{gs} + e_{xtpg} \tag{3.29}$$

Note the sets of indexes are in this case x, t, p and g instead of i, j, l and n in Eq. 3.24.

Logic behind Compositional data (CoDa)

We have already seen that the compositional framework occurs whenever we need to partition a whole measurement into percentage contributions from its components, thus with the unit-sum constraint, which represents an important advantage in demographic context, ensuring coherence of the estimated outcome. In the introduction of [94] is well explained the logic behind the CoDa. Here, we summarize the main aspects. In the literature CoDa are defined as non-negative vectors with a biased covariance structure [3, 2]. The compositional parts of such vectors describe the proportions of a whole; so, they are bounded by the total magnitude of the phenomenon they express. This bias from a statistical point, can only be overcome if the magnitude is set aside by focusing on ratios among parts rather than absolute scores. Therefore these vectors are said to carry relative information. Thus a classical analysis of compositions is relevant only in terms of magnitude differences among vectors; however, it is distorted with respect to the variability patterns among parts. In geometrical terms, CoDa are bounded in a subspace of the real space, called simplex, with its own rules and operations [4, 5]. These methods follow either a staying-in-the-simplex approach, which deals with compositions directly into their sample space, or a log-ratio coordinates approach, preferred due to simplicity. In this case, CoDa are projected onto real space by transforming them into log-ratios among compositional parts through a simple change of basis. Different transformations in log-ratio coordinates have been proposed in the literature to achieve this goal [3, 2, 31]. In recent years, the preferred way to proceed has been to express CoDa with respect to an orthonormal basis. This method ensures defined properties and can simplify interpretation if a meaningful basis is identified. One way to build an orthonormal basis is to find balances with a sequential binary partition (SBP) [32]. From a J -part composition, this procedure yields $J\ -1$ binary sets of non-intersecting mutually exclusive sub-compositions. A balance is then associated with each partition of two contrasting sub-compositions, together with the corresponding balancing element for the new orthonormal basis. This arrangement of parts also determines a complete partition of the total variance. The new coordinate system obtained in this fashion is easily interpretable only if the contrasts emerging from the SBP can be assigned a relevant meaning with respect to the phenomenon at hand. Manual selection of contrasts with expert knowledge is the best option, but seldom available outside of natural sciences. An exploratory approach is thus necessary. Principal balance analysis (PBA) is a sequential procedure identifying sets of J -1 principal balances (PB) that successively maximizes the explained variance in the data. Most studies on tridimensional data still use different types of log-ratio coordinates than orthonormal ones to deal with CoDA. Specifically, the centered log-ratio coordinate system is the preferred method due to its simplicity (for example in [95]). Also here, we consider this transformation.

Four-way decomposition of compositional data (CoDa)

When considering multi-population mortality data, one should bear in mind the dependence among mortality levels of the single populations for simultaneously modeling and forecasting them. Compositional data (CoDa), which are vectors of relative information constrained to sum a constant, enable to achieve a coherent mortality forecasting of multiple populations. Aiming to reach coherent forecasting, we introduce the CoDa transformation of mortality data. As we have explained in the previous paragraph, CoDa literature offers different transformations, such as *alr*, *ilr* or *plr* [2]. Here, we consider the most common approach used in mortality analysis (see, e.g., [82] and [13]), the centered log-ratio coordinates *clr*, which is more easily interpretable than the other existing transformations.

Defining $d_{x,t,p,g}$ as the deaths for age x, time t, population p, and gender g, the *clr* transformation is the logarithm of each observation $d_{x,t,p,g}$ divided by its geometric mean $g_{x,t,p,g}$:

$$clr(d_{x,t,p,g}) = \log\left(\frac{d_{x,t,p,g}}{g_{x,t,p,g}}\right) \tag{3.30}$$

We aim at decomposing the centered tensor of $d_{x,t,p,g}$ expressed in *clr* coordinates, i.e.:

$$clr(d_{x,t,p,g} \ominus \tilde{\alpha}_{x,p,g}) \tag{3.31}$$

where $\tilde{\alpha}_{x,p,g}$ is the age-specific geometric mean of $d_{x,t,p,g}$ over time t, and the operator $\ominus$ measures the distance between $d_{x,t,p,g}$ and $\tilde{\alpha}_{x,p,g}$ in compositional data similarly to subtraction in some data on the real axis [57].
Applying the model in Eq. 3.24, we obtain the four-way CP of the matrix of CoDa, where the generic element $clr(d_{x,t,p,g} \ominus \tilde{\alpha}_{x,p,g})$ can be expressed as follows:

$$clr(d_{x,t,p,g} \ominus \tilde{\alpha}_{x,p,g}) = \sum_{s=1}^{S} a_{xs} b_{ts} c_{ps} d_{gs} + e_{xtpg} \tag{3.32}$$

Note that the sets of indexes are as for the decomposition of log centered rates x, t, p and g instead of i, j, l and n in Eq. 3.24.

3.2.3 Application

Dataset description and model fit

The study focuses on both male and female mortality of 10 Western European countries, listed in Table 3.5. The data is provided by the Human Mortality Database (HMD)[2], which is an archive of death information for several countries. The application of the methodology requires to structure the data in four dimensions: age, time, population, sex. The age dimension is organized in 21 classes from 0 to 99 years, with a step of 5 years, except for the first two classes, which represent respectively the individuals aged 0, and aged 1-4. Regarding the time, we focus the analysis on the years 1961-2015 to consider the same time window for each population.

Code	Countries
AUT	Austria
BEL	Belgium
DNK	Denmark
FIN	Finland
FRATNP	France
ITA	Italy
NLD	Netherlands
ESP	Spain
SWE	Sweden
GBR_NP	United Kingdom

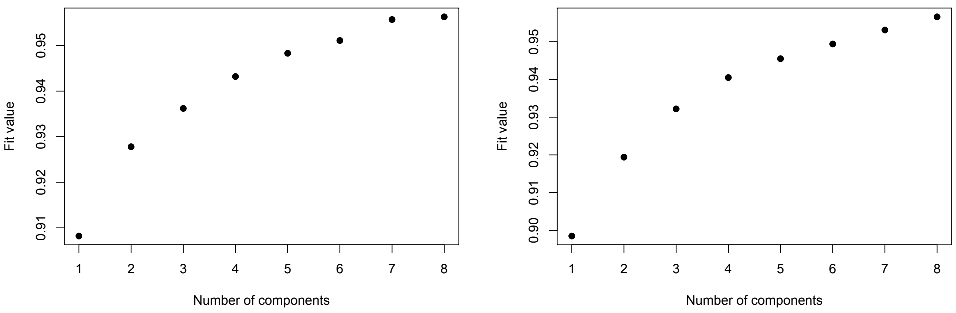
Table 3.5. List of countries (and their codes) considered in the study.

In order to apply the multi way method, we organize the data in an array of four dimensions (age × time × population × sex). The dimension of the array is: 18 × 55 × 10 × 2. We apply the four-way CP model and to choose the best solutions, we use up to eight components ($S = 1, 2, ..., 8$) and compute the associated fit values.

In Fig. 3.7, we report the fit values by the number of components S. We choose the solution S=2 with the fit value (see Eq. 3.27) of 92.78% for the log-centered mortality rates and 91.94% for CoDa.

---

[2] Human Mortality Database. University of California; Berkeley (USA); and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 08/11/21).



(a) Log-centered mortality rates

(b) CoDa

Figure 3.7. Possibile PARAFAC solutions and corresponding fit values.

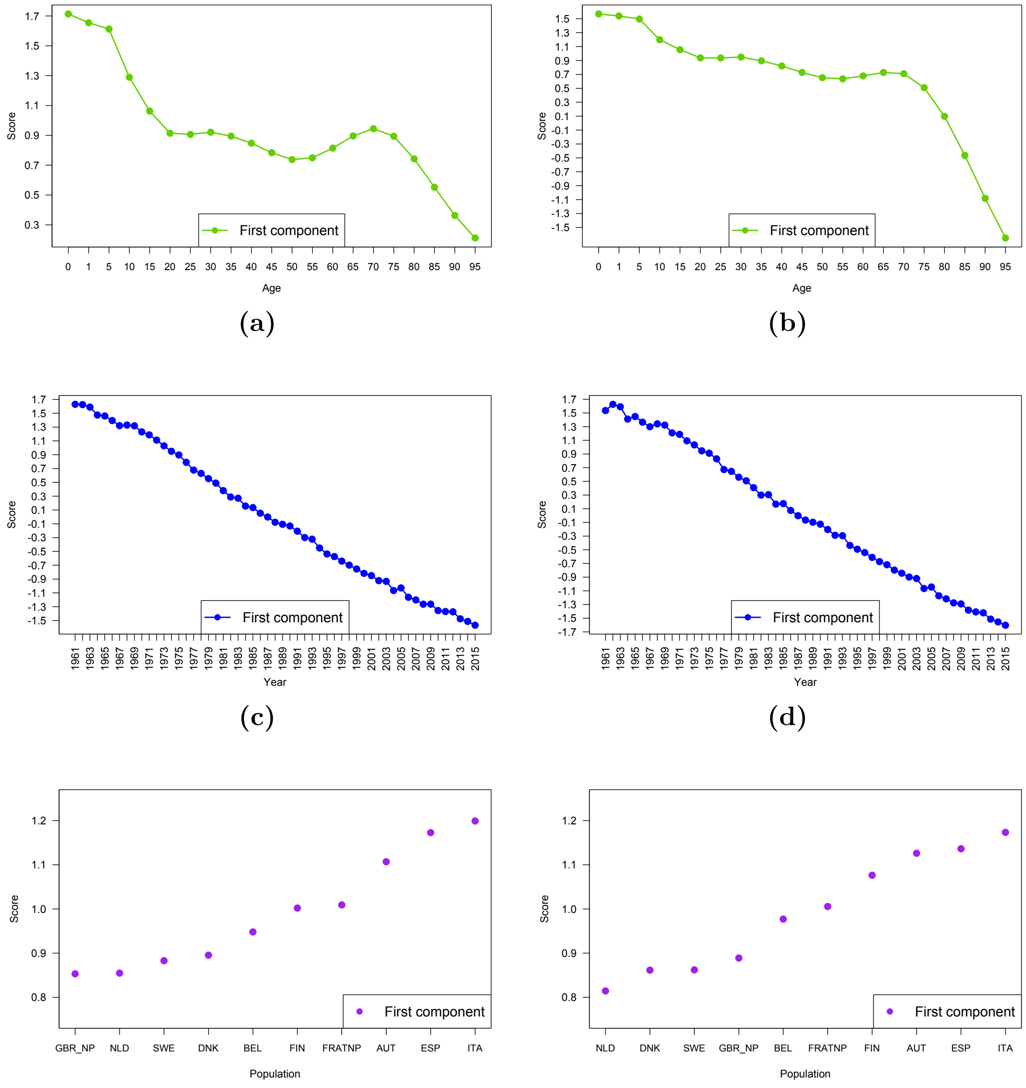
Demographic significance

We show the scores of the two components for each of the four data dimensions. Fig. 3.8 depicts the values of the first component for the first three dimensions: age class, year and population.

While Lee-Carter parameters are easy to interpret, thanks to their direct reflection on death rates on a log scale, its CoDa counterpart may be difficult to use. Therefore, as a service to readers, it is worth creating a bridge between the two frameworks, proposing the description of the demographic significance of both models' parameters. The difference is in the process to be modeled. Indeed, the idea behind CoDa relies on the life-saving process, in which the life table deaths shift from one age group to another over time. On the contrary, in processes based on log-CMR, the outputs are decreasing in a log-linear way over time. According to [57]; [13] and [82], the main consequence of the different processes is found in the interpretation of the age-specific parameter. In the CoDa framework, the age pattern shows how deaths are transferred from one age group to another in relative terms, in practice from younger ages to older ages. These deaths are not redistributed randomly but towards the ages and other dimensions where they are most likely to occur according to the parameter estimates in the model. Finally, the other parameters, the α and the ones related to time, country, and gender, have a common interpretation for mortality rates and CoDa methods, with the only difference that α-CoDa is the age-specific geometric mean of the deaths. Figure 3.8 provides the log-CMR (left panels) and CoDa (right panels) model parameters. The time pattern can be interpreted as an index of the general level of mortality over time, which shows a linear behavior for both models (Fig. 3.8, panels (c) and (d)), thus facilitating the forecasting phase. The population-specific indexes can be interpreted, as the speed of mortality changes that each country is experiencing. The greater the magnitude, the faster the decline in mortality. For instance, in both models, Italy has experienced a more rapid mortality deceleration than the other European countries in the study period. This can be the result of improvements in old-age survival along with the fall in infant mortality, after World War II.

Looking at figure 3.8, panels (e), (f), we can distinguish different countries' groups with respect to longevity levels, according to the adopted model. Classifying countries by looking at their magnitude levels, we can easily interpret three groups, namely: 1) slow-rate countries: countries that experienced improvements in longevity with a smaller pace; 2) medium-rate: populations that show a faster decline in mortality compared to the slow-rate group; and 3) fast-rate: those countries that show remarkable improvements.

Based on the estimation provided by log-CMR, we can outline the following groups: 1) slow-rate: GBR, NLD, SWE, DNK. 2) medium-rate: BEL, FIN, FRA. 3) fast-rate: AUT, ESP, ITA. Using CoDa, we identify the following groups: 1) slow-rate: NLD, DNK, SWE, GBR. 2) medium-rate: BEL, FRA. 3) fast-rate: FIN, AUT, ESP, ITA.



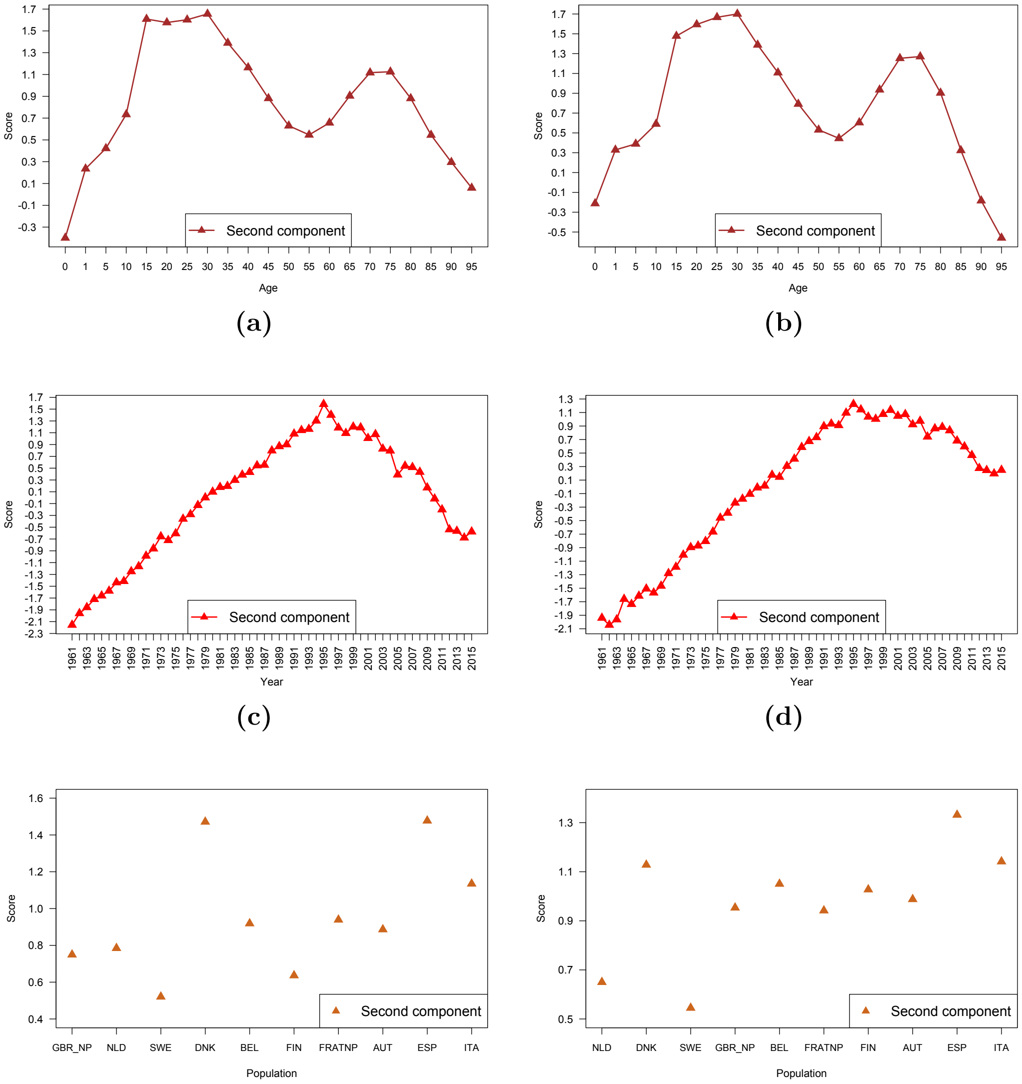
(a) (b)

(c) (d)

(e) (f)

Figure 3.8. First component of age: panels (a) and (b); first component of year: panels (c) and (d); first component of population: panels (e) and (f). Log-centered mortality rates (left panels); CoDa (right panels).

Although the first component accounted for most of the phenomenon, the second component can be useful to capture residual effects, shedding light on specific longevity dynamics. In Fig. 3.9, it is worth highlighting a great mortality stagnation between the ages 15-30 and 70-75 on log-CMR for Denmark and Spain. Similarly, on CoDa, we observe this effect mainly in Spain, and, with a smaller magnitude, in Denmark and Italy. For both models (log-CMR and CoDa), the temporal dynamics described by the second component show a reverting trend in the year 1995. Indeed, in 1961-1995, the second component shows an opposite trend compared with the first one, while in the subsequent period 1996-2015, it shows a decreasing trend, coherently with the first component. In fact, despite the geographical proximity, [22] highlight the different mortality trend between Denmark and Sweden, chosen as representative of the group of Northern countries, and explain some potential reasons such as lifestyle, and specific risk factors (smoking and alcohol consumption).



(a) (b)

(c) (d)

(e) (f)

Figure 3.9. Second component of age: panels (a) and (b); second component of year: panels (c) and (d); second component of population: panels (e) and (f). Log-centered mortality rates (left panels); CoDa (right panels).

Finally, concerning sex, the first component scores are reported in Table 3.6.

We can observe that for the first component the scores are very similar to each other, especially for the log-centered mortality rates (respectively 0.40 and 0.38 with a relative difference of 0.02 and for CoDa of 0.06).

Sex	Log-centered mortality rates	CoDa
Male	0.40	0.45
Female	0.38	0.39

Table 3.6. First component of sex.

The second component scores are reported in Table 3.7 we can observe that for the first component the scores are higher for the Male case.

Sex	Log-centered mortality rates	CoDa
Male	0.11	0.14
Female	0.05	0.04

Table 3.7. Second component of sex.

Forecasting

In this section, we focus on the temporal component, represented by profile matrix $\mathbf{B}$, whose generic element is b_{ts} with s=1,2. In particular we denote $b_t = \begin{bmatrix} b_{t1} \\ b_{t2} \end{bmatrix}$. We consider mortality data from 1961 to 2000 and develop out-of-sample forecasts for the coming 15 years, i.e. from 2001 to 2015. We propose to exploit Vector-Error Correction (VEC) models to produce consistent forecast of $\mathbf{B}$ since the methodology is able to address non-stationarity in multivariate time series resulting from co-movements of multiple response series. Indeed both temporal patterns related to components s=1,2, i.e.m $b_{\cdot s}$, display a unit-root according to the Augmented Dickey-Fuller test.

The estimation of the VEC model proceeds in two steps. The first one is often referred to as "Johansen step" and it consists of estimating the cointegrating relations r, i.e. the number of independent linear combinations for which $\mathbf{B}$ is stationary. These combinations can be interpreted as the number of long-run equilibrium relations in $\mathbf{B}$. In other words, although the component series of the profile matrix $\mathbf{B}$ are individually non-stationary, various linear combinations of them are stationary. The second step is often referred to as "VARX step", and concerns the estimation of the VAR model in differences using the cointegrating relations of the first step as predictors. The detailed procedure can be summarized as follows:

1. 1.1 Determine the lag length P of the model;
   1.2 determine the cointegrating rank r;
   1.3 determine a model of cointegration that best captures the deterministic terms of the data;

2. 2.1 Estimate the unrestricted VARX model to serve as a baseline model;
   2.2 determine and impose suitable constraints on the parameters of the VARX

model, including the constant, regression coefficients, and autoregressive coefficients.

Because the data clearly exhibit time trends, we examine the possibility that the data could be described by either of two Johansen parametric forms that incorporate linear time trends.
The more general of the two forms is the Johansen H^* model, in which a component of the overall constant appears both inside (χ_0) and outside (χ_1) the cointegrating relations, while the time trend (ι_0) is restricted to the cointegrating relations:

$$\Delta b_t = \Theta(\Upsilon' b_{t-1}+\chi_0+\iota_0 t)+\chi_1+\sum_{i=1}^{p-1}\Gamma_i\Delta b_{t-i}+\varepsilon_t = (\Theta\chi_0+\chi_1)+\Theta(\Upsilon' b_{t-1}+\iota_0 t)+\sum_{i=1}^{p-1}\Gamma_i\Delta b_{t-i}+\varepsilon_t. \tag{3.33}$$

Because a component of the constant appears inside and outside the cointegrating relations, the overall constant ($\chi = \Theta\chi_0 + \chi_1$) is unrestricted and the H^* model becomes:

$$\Delta b_t = \chi + \Theta(\Upsilon' b_{t-1} + \iota_0 t) + \sum_{i=1}^{p-1}\Gamma_i\Delta b_{t-i} + \varepsilon_t. \tag{3.34}$$

It is important to observe that the VEC model is expressed in differences Δb_t, and so the unrestricted constant (χ) represents linear trends in the corresponding levels b_{ts}. The second model is the Johansen *H1* model, in which the model constant is also unrestricted, but the cointegrating relations contain no time trend:

$$\Delta b_t = \Theta(\Upsilon' b_{t-1}+\chi_0)+\chi_1+\sum_{i=1}^{p-1}\Gamma_i\Delta b_{t-i}+\varepsilon_t = (\Theta\chi_0+\chi_1)+\Theta\Upsilon' b_{t-1}+\sum_{i=1}^{p-1}\Gamma_i\Delta b_{t-i}+\varepsilon_t. \tag{3.35}$$

The *H1* model emphasizes the unrestricted nature of the constant[3]:

$$\Delta b_t = \chi + \Theta\Upsilon' b_{t-1} + \sum_{i=1}^{p-1}\Gamma_i\Delta b_{t-i} + \varepsilon_t. \tag{3.36}$$

To determine the cointegrating rank, we use the Johansen's trace test. It appears that, if $r = 1$, cointegrating relations are reasonable. Using $r = 1$ as the cointegrating rank, we determine which of the two models better describes the data by using a likelihood-ratio test, where the H^* model is the unrestricted model and the *H1* model is the restricted model with r restrictions. In summary, the hypothesis test results suggest that the cointegrating rank is 1 and the model has the *H1* Johansen parametric form. The estimation of the cointegrating relations is complete.

Now, we express the VEC(1) model as the VARX(1) model in differences in which the cointegrating relations are predictors, $X_t = \Upsilon' b_{t-1}$, and the adjustment matrix (Θ) is the corresponding regression coefficient:

$$\Delta b_t = \chi + \Theta(\Upsilon' b_{t-1}) + \Gamma_1\Delta b_{t-1} + \varepsilon_t = \chi + \Gamma_1\Delta b_{t-1} + \Theta X_t + \varepsilon_t. \tag{3.37}$$

---

[3] Observe that the *H1* model is a restricted parameterization of the H^* model in that the *H1* model imposes an additional r restrictions on the parameters of an otherwise unrestricted H^* model. Specifically, the *H1* model excludes time trends in the cointegrating relations by restricting $\iota_0 = 0$. This situation that can arise when the cointegrating relations drift along a common linear trend and the trend slope is the same for series with a linear trend.

To determine constraints on the parameters of the VARX model, it is important to understand that the Johansen cointegration parameters (Υ) estimated in the first step converge at a rate proportional to the sample size T. Therefore, using the cointegrating relations obtained in the first step as predictors in the second step does not affect the distribution of the second step VARX parameters. In contrast, the VARX parameters of the second step are asymptotically normally distributed and converge at the usual rate of $\sqrt{T}$, and so their t-statistics can be interpreted in the usual way.

After fitting the restricted *H1* model, we develop out-of-sample forecasts. To test the forecasts' accuracy, we iteratively compute out-of-sample forecasts. Firstly, we estimate the VEC model over the initial sample period. Then, we forecast the model one year into the future and re-estimate the model by adding the data of the next year to the sample, thereby increasing the sample size available for estimation. We repeat the latter two steps, accumulating a time series of forecasts until the end of the sample.

Results

Fig. 3.10 represents the result of the VEC model with a prediction interval of 10%-90% calculated on the simulated values for the first component score of the years. The left panel refers to the log-centered mortality rates, while the right panel to CoDa.

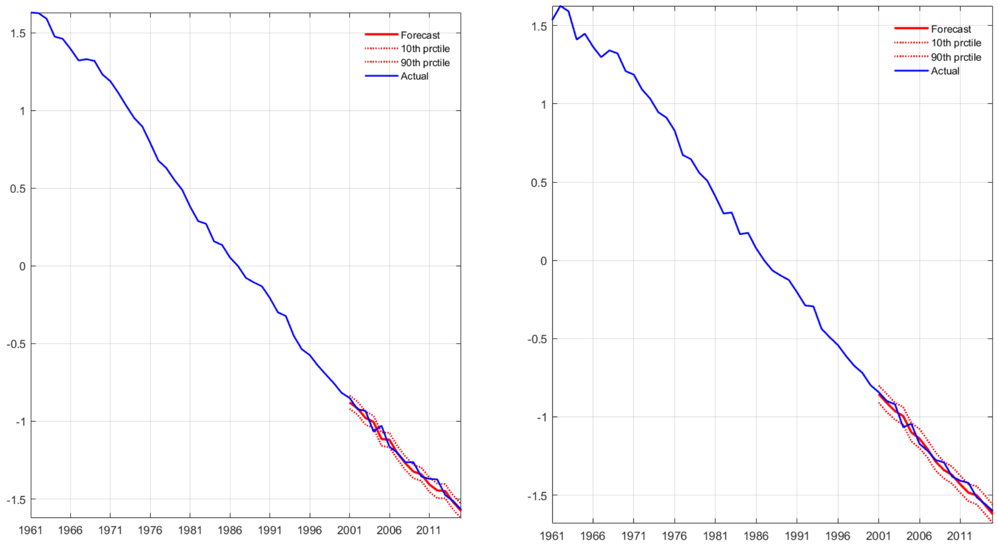


Figure 3.10. Forecast of the year first component scores. Left panel: log-centered mortality rates; right panel: CoDa.

Fig. 3.11 represents the the corresponding plot for the second component of the years.

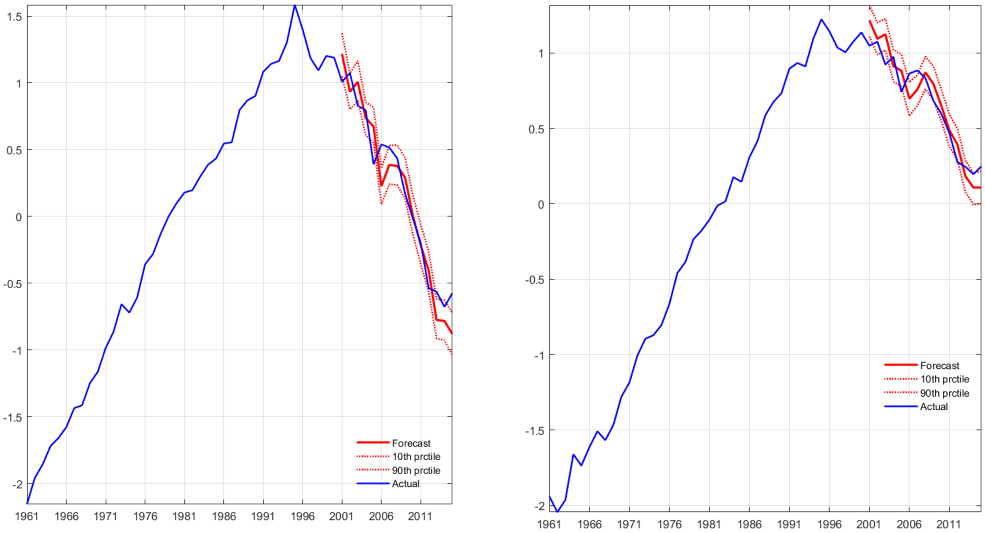


Figure 3.11. Forecast of the second component scores of the years. Left panel: log-centered mortality rates; right panel: CoDa.

We also carried out a Monte Carlo simulation to obtain the distributions of the year component over the forecasted period 2001-2015. In particular, for each one-step-ahead forecast, we simulated 1000 runs and then performed a kernel smoothing with 50 points. Fig. 3.12 illustrates the distributions of the forecast value of the first component of the years over the period 2001-2015 for log-centered mortality rates (left panel) and CoDa (right panel). Fig. 3.13) is the corresponding plot for the second component of the years. Figures 3.10-3.13 highlight the ability of the model to correctly forecast the decreasing patterns of the first component of both the log-centered mortality rates and the CoDa. Moreover, from Fig. 3.13, which displays kernel distributions of the Monte Carlo simulations, clearly emerges the homogeneity of the forecast through different time periods suggesting the stability of the data-generating-process.

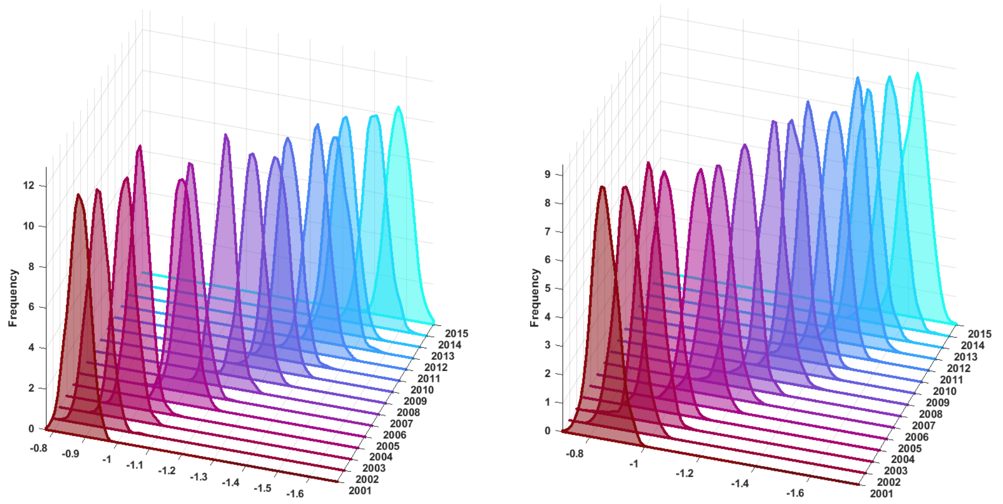


Figure 3.12. Montecarlo distributions of the first component scores for the years. Left panel: log-centered mortality rates; right panel: CoDa.

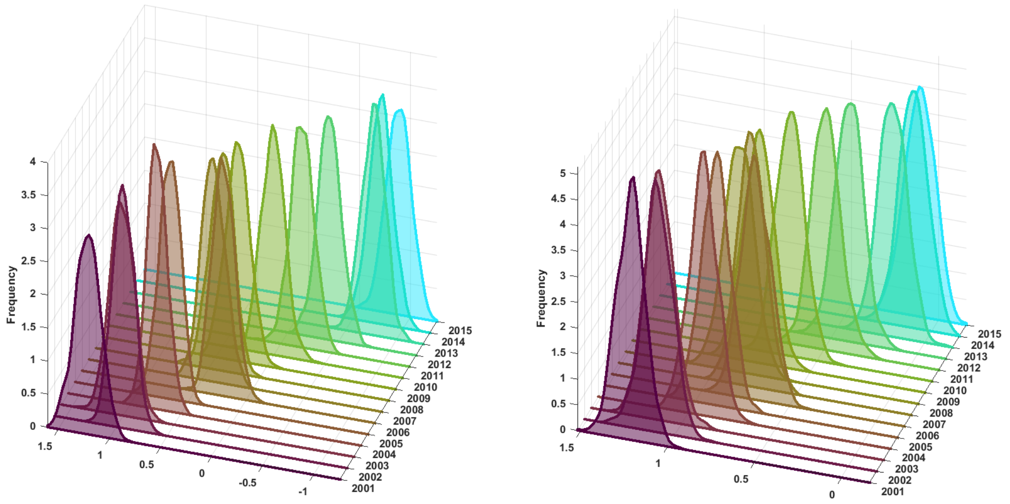


Figure 3.13. Montecarlo distributions of the second component scores pf the years. Left panel: log-centered mortality rates; right panel: CoDa.

Error measures

To evaluate the goodness of the prediction, we compute the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE), that are respectively defined as:

$$RMSE = \sqrt{\frac{1}{IJLN}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{n=1}^{N}(x_{ijln} - \hat{x}_{ijln})^2}, \tag{3.38}$$

$$MAE = \frac{1}{IJLN} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{l=1}^{L} \sum_{n=1}^{N} |x_{ijln} - \hat{x}_{ijln}|. \tag{3.39}$$

To calculate these error measures, we reconstruct the array starting from the component scores obtained from the tensor decomposition and taking into account the predicted values on the temporal components. The resulting values are reported in Table 3.8.

Model	RMSE	MAE
Log-centered mortality rates	0.0087	0.0033
CoDa	0.0116	0.0044

Table 3.8. Overall error measures.

We also calculate the values of RMSE and MAE by gender and countries (see Tables 3.9 and 3.10 for the log-centered mortality rates and the CoDa, respectively). In the case of log-CMR, Spain, Denmark, and Italy show the highest errors for both genders, while for CoDa, we find Spain, Italy, and Finland.

Countries	Male		Female	
	RMSE	MAE	RMSE	MAE
Austria	0.0077	0.0031	0.0498	0.0384
Belgium	0.0090	0.0036	0.0512	0.0393
Denmark	0.0129	0.0051	0.0835	0.0629
Finland	0.0081	0.0032	0.0369	0.0283
France	0.0093	0.0037	0.0524	0.0403
Italy	0.0112	0.0045	0.0633	0.0486
The Netherlands	0.0078	0.0031	0.0438	0.0337
Spain	0.0133	0.0053	0.0827	0.0629
Sweden	0.0070	0.0027	0.0307	0.0234
United Kingdom	0.0077	0.0031	0.0419	0.0322
Total	**0.0098**	**0.0038**	**0.0563**	**0.0410**

Table 3.9. Error measures for log-centered mortality rates.

Countries	Male		Female	
	RMSE	MAE	RMSE	MAE
Austria	0.0150	0.0061	0.0778	0.0606
Belgium	0.0157	0.0062	0.0819	0.0635
Denmark	0.0150	0.0058	0.0788	0.0608
Finland	0.0165	0.0066	0.0858	0.0666
France	0.0139	0.0056	0.0720	0.0561
Italy	0.0167	0.0067	0.0866	0.0674
The Netherlands	0.0096	0.0040	0.0498	0.0388
Spain	0.0183	0.0073	0.0960	0.0743
Sweden	0.0091	0.0038	0.0475	0.0365
United Kingdom	0.0147	0.0058	0.0771	0.0597
Total	**0.0147**	**0.0058**	**0.0767**	**0.0584**

Table 3.10. Error measures for CoDa.

3.2.4 Conclusions

To design appropriate pension or insurance plans, an unrealistic prediction/estimation of changes in death rates could result in inadequate funding of pension obligations. In this framework, understanding mortality dynamics in a high-dimensional framework is crucial for all the analyses involving demographic assessments. To this aim, we have proposed a coherent mortality forecasting methodology, which leverages the four-way CANDECOMP/PARAFAC and Vector-Error Correction models. We have examined how age groups, years, countries, and gender impact on target variables, namely log-centered mortality rates and compositional transformation of mortality data from the Human Mortality Database.

The need for a four-way analysis is justified by the fact that the obtained results properly take into account all the double, triple and quadruple interactions among the four dimensions. Certainly, four-way data sets might be analysed by three-way, or even two-way, methods. This can be done by aggregating over one or two dimensions or performing separate three- or two-way analyses. Nevertheless, results would be incomplete at best, as such alternative strategies do not exploit the relevant information in the data because some of the interactions are arbitrarily skipped. For instance, if one is tempted by considering two separate three-way analyses distinguished by gender, all the interactions involving gender do not play an active role in detecting the underlying structure of the data. In our studies, we decided to consider $S = 2$ components because they offered an interesting description of the phenomenon of longevity. In fact, Component 1 allowed us to discover its common features. In addition, a deeper insight was offered by Component 2 highlighting some specific peculiarities in longevity changes.

The (four-way) CP model is usually considered an exploratory tool without probabilistic assumptions. Nevertheless, it is interesting to get an insight into uncertainties associated with the obtained components. In the paper, this was done by assessing stability through a split-half analysis. We discovered that the underlying structures for both log-CMR and CoDa were very stable and, therefore, the estimated component matrices appeared to be reasonably good and affected by a low level of uncertainty. The inspection of such matrices, in particular the scores of Component 2, allowed us to reveal some differences among countries. For instance, we found that Spain and, to a lesser extent, Denmark present peculiar mortality rates with respect to the other countries, in particular, those from Northern Europe, especially for young people in the years around 1995.

The two-component CP model enables us to synthesize the behavior of the target variables and highlight the evolution of the temporal patterns, which have been employed to forecast future trajectories of mortality data for the years 2001-2015 by using the Vector-Error Correction model. The associated uncertainty has been evaluated by Monte Carlo simulations.

The results we found are satisfactory, helping to rethink the design of pension plans to mitigate the burden of increased longevity and make significant efforts in containing health care costs.

Chapter 4

EXPLORING THE CAUSES OF DEATH

In this chapter, we apply a tensor-based method to mortality by cause of death, hence considering four dimensions. Again we analyze a 4-way array, but we use the Tucker 4 method. In this context we apply the more general Tucker4 method, which contains Candecomp/Parafac as a special case. The dataset here considered is provided by the World Health Organization and refers to $N = 4$ dimensions: causes of death, age groups, years, and countries. Such a four-way dataset is investigated by the Tucker4 model, the four-way extension of Tucker3. We carry out the analysis by distinguishing the results by sex, first considering the males and then the females. The steps we follow are: description of the methodology, identification and description of the database, construction of the four-dimensional array (causes of death, age, time, and country), application of the tensor decomposition method, choice of the number of underlying components and interpretation. This four-way component analysis is useful for the exploratory analysis of four-way mortality data and it reveals some peculiar aspects of the mortality phenomenon. More in general, by the current applications, our aim is to stimulate the use of tensor decomposition methods whenever four-way data are available. Regardless of the specific domain of research, any four-way analysis is composed of different steps involving choices to be made. These will be carefully described and motivated providing a guidance for the practical application of N-way methods.

4.1 Methodology

In this chapter we apply the same methodology (Tucker) of the previous paragraph 3.2.1 but we consider four dimensions. Here, we refer again to a a four-dimensional array $\underline{\mathbf{X}}$, in this case “Causes of death × Age × Time × Countries”.

Hence, the Tucker4 model can be formulated as

$$x_{ijln} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R}\sum_{s=1}^{S} a_{ip}b_{jq}c_{lr}d_{ns}g_{pqrs} + e_{ijln}, \tag{4.1}$$

$$i = 1,...,I, \quad j = 1,...,J, \quad l = 1,...,L, \quad n = 1,...,N$$

where there are the new loadings d_{ns} linking the conditions and components for the conditions and e_{ijln} denotes the generic error term pertaining to the four-way array $\underline{\mathbf{E}}$. Therefore, the Tucker4 model summarizes every dimension by a limited number of dimension-specific components. In particular, P, Q, R and S $(\leq N)$ denote the number of such components for the units, the variables, the occasions and the conditions, respectively. As a natural extension of Tucker3, the interactions between these components are given by the core $\underline{\mathbf{G}}$, which is now a four-way array of order $(P \times Q \times R \times S)$, with generic element g_{pqrs}. Obviously, its elements express the quadruple interaction among component p of the unit mode, component q of the variable mode, component r of the first occasion mode and component s of the second occasion mode and high values of g_{pqrs} in absolute sense suggest strong relations among the quadruple of components involved. In matrix notation, the Tucker4 model can be formalized as:

$$\mathbf{X}_A = \mathbf{A}\mathbf{G}_A(\mathbf{D}\otimes\mathbf{C}\otimes\mathbf{B})' + \mathbf{E}_A \tag{4.2}$$

where $\mathbf{X}_A$ is now a matrix of order $(I \times JLN)$ and represents the unit mode matricization of the four-way array $\underline{\mathbf{X}}$. In this case, $\mathbf{X}_A$ is obtained by juxtaposing next to each other the previously-defined matricizations of the three-way arrays pertaining to all the conditions. $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are the component matrices for the four modes and $\mathbf{G}_A$ of order $(P \times QRS)$ is the unit mode matricization of the core tensor $\underline{\mathbf{G}}$. Finally, $\mathbf{E}_A$ is the unit mode matricization of $\underline{\mathbf{E}}$. With similar reasoning as for Tucker3, the Tucker4 captures the four-way structure of the data by constraining the loadings to take the form $\mathbf{G}_A(\mathbf{D}\otimes\mathbf{C}\otimes\mathbf{B})'$.

Estimation is done following the least squares approach. In fact, we look for $\mathbf{A}$ $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{G}_A$ such that the sum of squared errors

$$||\mathbf{E}_A||^2 = ||\mathbf{X}_A - \mathbf{A}\mathbf{G}_A(\mathbf{D}\otimes\mathbf{C}\otimes\mathbf{B})'||^2 \tag{4.3}$$

is minimized. To this purpose, an alternating least squares algorithm can be implemented, which in turn minimizes the loss function with respect to one of the parameter matrices upon convergence. More than one (random) start of the algorithm is usually recommended in order to limit the risk of local optima.

The fit of the model can be evaluated according to (3.17), provided that the unit mode matricizations of the four-way arrays $\underline{\mathbf{E}}$ and $\underline{\mathbf{X}}$ are used.

The Tucker4 solution is not unique. Equally well fitting solutions can be found by postmultiplying the component matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ by non-singular square rotation matrices and compensating such rotations in the core. For instance, $\mathbf{A}$ can be rotated by using the rotation matrix $\mathbf{T}$ of order $(P \times P)$ so as to obtain the rotated matrix $\mathbf{A}_R = \mathbf{AT}$. The rotation must be compensated into the core leading to $\mathbf{G}_{AR} = \mathbf{T}^{-1}\mathbf{G}_A$. The model fit is not affected because $\mathbf{A}_R\mathbf{G}_{AR}(\mathbf{D}\otimes\mathbf{C}\otimes\mathbf{B})' = \mathbf{A}\mathbf{T}\mathbf{T}^{-1}\mathbf{G}_A(\mathbf{D}\otimes\mathbf{C}\otimes\mathbf{B})' = \mathbf{A}\mathbf{G}_A(\mathbf{D}\otimes\mathbf{C}\otimes\mathbf{B})'$. To exploit this rotational freedom, the solution can be rotated to a simple form. In this paper, we achieve simplicity by varimax ([52]) rotating $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$.

In this chapter, we observe also that multiway models, such as Tucker4, can require some preprocessing of the data. In particular, prior to fitting the model to the data, these can be centered and/or normalized. By centering, the aim is to remove offset terms.

By normalizing, unwanted differences in scale are removed. Differently from the standard two-way case, where the preprocessing step is performed by centering and/or normalizing every variable, in the multiway context it is not obvious how to proceed. In fact, centering and/or scaling can be done with respect to a dimension or even a combination of dimensions. For instance, one might center the data *across* the occasion mode (single dimension) as

$$\tilde{x}_{ijkl} = x_{ijkl} - \bar{x}_{ij \cdot l}, \tag{4.4}$$

where $\bar{x}_{ij \cdot l} = \frac{\sum_{k=1}^{K} x_{ijkl}}{K}$, or *across* the occasion and condition modes (combination of dimensions) as

$$\tilde{x}_{ijkl} = x_{ijkl} - \bar{x}_{ij \cdot \cdot}. \tag{4.5}$$

Similarly, scaling can be done with respect to a single dimension or a combination of dimensions. Suppose to scale the data *within* the unit mode. In this case, we have

$$\tilde{x}_{ijkl} = \frac{x_{ijkl}}{\nu_i}, \tag{4.6}$$

where ν_i is the normalizing factor computed within the units. A common choice is based on the square root of the sum of squares:

$$\nu_i = \sqrt{\sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} x_{ijkl}^2}. \tag{4.7}$$

In the three-way case, refer to [56] for more details.

Furthermore, we have seen that to choose the number of components we balance fit and parsimony and that we can base our decision also on the basis of the objective of the analysis, in particular on interpretability [54]. With the aim to explore our data, here, we choose a solution also to be able to interpret and report the results. First, we interpret the components pertaining to each dimension, i.e., the component matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$; next, the core array $\underline{\mathbf{G}}$ summarizing the main interactions in the data. For each component matrix, every component is interpreted by inspecting the values of every column separately. In this sense, interpretation differs from classic PCA where component loadings can be related to correlations between variables and components. To intepret the core, one should bear in mind that its values describe the full four-way data, reduced to the summarizing descriptions given by the components for the four modes. Thus, the core array summarizes the information in the original four-way array and contains main effects and two-, three-, and four-way interactions present in the original array [54]. The core elements can be compared with each other. Roughly speaking, high core elements in absolute sense denote strong associations between the components for the four modes.

When the chosen solution is not well interpretable, a new solution with a different number of components $P + Q + R + S$ is investigated.

Finally, in this chapter is clear one of the main advantage of the Tucker model as expresses in [99], we can observe the *full core array* $\underline{\mathbf{G}}$ that explain the relation between the components. It helps us to identify the main relationship between the components in this more complex situation with four dimensions.

4.2 Application

We analyse a dataset provided by the World Health Organization (WHO) mortality database, which is an archive of causes of death information for several countries. The longest time series starts in 1950, however, for many countries, the information starts from 1959. The application of the methodology requires the data for each dimension: causes of death, age, time, country. We consider the following $I = 8$ causes of death: Infectious diseases, Smoking-related cancer, Non-Smoking-related cancer, Diabetes, Cardiovascular diseases (hereinafter CVD), Respiratory diseases, External causes of death, Other causes of death. It is worth noting that the classification of causes of death has greatly changed since 1959, passing from ICD 7th to ICD 10th revision. Age is organized in classes as follows: age groups 0, 1-4, 5-9, 10-14, 15-19, ..., 80-84 ($J = 18$). Only the countries for which data is available in this time frame are included in the analysis (see the $L = 18$ countries listed in Table 4.4) and because of this, regarding time, we focus the analysis on the years 1961-2015 (hence, $K = 55$) to consider the same time window for each country. The analysis is developed distinguishing the results by sex, first considering the males and then the females. Therefore, with respect to the previous papers [91, 36, 28, 66], we introduce a new dimension, passing from the Tucker3 model to the Tucker4 one that, as far as we saw, was never used for the analysis of mortality data. Regarding countries, in this last chapter, where we have reconstructed data from the WHO, to give coherence to the work and to deepen the demographic relationships between countries we have consider the same 18 countries of the second chapter.

In order to carry out the four-way analysis, we organize the data in an array of four dimensions (causes of death × age classes × times × countries). The dimension of the array is: 8 × 18 × 55 × 18, for a total of 142,560 entries. In particular, the generic element of the array is the death rate for cause of death i, at age class j in year k, in country l: d_{ijkl}.

The death rate[1] for an individual for cause of death i, at age class j in year k, in country l is $d_{ijkl} = \frac{D_{ijkl}}{E_{jkl}}$, where D_{ijkl} is the number of deaths for cause of death i, at age class j during year k, in country l and E_{jkl} is the risk exposure of the population of age j in year k, in country l.

It is noteworthy that it is not possible to obtain causes-specific Exposures, thus the causes-specific mortality rates are computed by using the Number of Deaths specific for each cause, and the Exposures, that are aggregated on the overall mortality.

4.2.1 Male results

To get more reliable results, before fitting the model to the data, we decide to scale them in such a way that the data related to each country and each age class plays

---

[1] In Actuarial Science, Demography, Epidemiology, and Biostatistics as well, it is common to deal with the so-called Hazard function (or Force of Mortality), which is, in its continuous formulation, the instantaneous rate of death at age j, conditioned upon surviving to that age, (in a given year k). In practice, researchers work on its discrete counterpart, expressed by the mortality rate as the ratio of the Number of Death/Population (or Exposure to risks). Often, this rate, and thus both Deaths and Exposures, are specific for age and years. Nevertheless, they might also be specified for age classes (j), years (k), and country (l).

the same role in the analysis. In particular, with reference to the age classes, such a scaling avoids that the underlying components strongly depend on the older age classes, obviously characterised by the highest death rates. In this way, as we shall see, we are able to extract relevant components involving patterns related to children and even infants that, otherwise, would have been overlooked by the model. To this purpose, the observed death rates are divided by

$$\nu_{jl} = \sqrt{\sum_{i=1}^{I} \sum_{k=1}^{K} d_{ijkl}^2}. \tag{4.8}$$

Moreover, we decided not to center the data because a meaningful zero point already exists. In fact, $d_{ijkl} = 0$ means absence of death due to cause i at age j, in year k, in country l.

To choose among the multitude of possible Tucker4 solutions, we run the Tucker4 method by varying P, Q, R and S from 2 to 8 ($= \min(I, J, K, L)$) and compute the associated fit value. In this way, we consider all the Tucker4 solutions with a total number of components P+Q+R+S ranging from 8 to 32. In Table 4.1 we listed the best solutions for each total number of components and the corresponding fit expressed as a percentage.

P	Q	R	S	$P+Q+R+S$	Fit value (%)
2	2	2	2	8	83.66%
2	3	2	2	9	85.79%
3	3	2	2	10	88.44%
3	4	2	2	11	90.47%
4	4	2	2	12	91.36%
5	4	2	2	13	92.07%
5	5	2	2	14	92.77%
5	4	3	3	15	93.43%
5	5	3	3	16	94.16%
5	5	3	4	17	94.54%
5	5	3	5	18	94.87%
5	5	4	5	19	95.23%
5	6	4	5	20	95.60%
5	6	4	6	21	95.94%
5	6	5	6	22	96.29%
6	6	5	6	23	96.60%
6	6	5	7	24	96.88%
6	7	5	7	25	97.14%
6	7	5	8	26	97.35%
6	8	5	8	27	97.54%
6	8	6	8	28	97.67%
7	8	6	8	29	97.76%
7	8	7	8	30	97.83%
7	8	8	8	31	97.87%
8	8	8	8	32	97.98%

Table 4.1. Best retained solutions in terms of fit values varying P+Q+R+S from 8 to 32 . % with number of components and fit values (male analysis)

By inspecting Table 4.1, we conclude that the solution corresponding to the best compromise between fit and parsimony is the one with $P = 3$, $Q = 4$, $R = 2$, $S = 2$ with the fit value equal to 90.5%. This choice is also corroborated by the next interesting interpretation of the underlying components. As already noted, all the component matrices are varimax-rotated, compensating such rotations into the core.

In Table 4.2, we report the component scores for the causes of death mortality (component matrix **A**). Component 1 for the causes of death, labeled "External", mainly depends on external causes (0.98), frequently related to young mortality, road accidents, and violence. In other words, it identifies a behavior of mortality characterised by external diseases leading to a higher incidence of mortality with respect to the other causes of death. The meaning will be clearer when we will study the core elements expressing the interactions between the components of the four modes. Component 2, labeled "CVD+Cancer", is strongly associated with CVD and Cancer (even if, to a lesser extent, with Non-smoking-related cancer). Component, "Other", is strongly related to other diseases (0.97).

Cause of death	External	CVD+Cancer	Other
Infectious	0.08	0.04	0.13
Smoking-related cancer	-0.03	**0.41**	0.08
Non-smoking-related cancer	0.13	0.15	0.12
Diabetes	0.00	0.03	0.01
CVD	0.05	**0.89**	-0.11
Respiratory	0.00	0.10	0.02
External	**0.98**	-0.06	0.05
Other	-0.07	0.04	**0.97**

Table 4.2. Component matrix for the causes of death (bold values indicate scores greater than 0.25). Male analysis.

Table 4.3 shows the component scores for the age classes (component matrix **B**). Component 1 for the age classes, labeled "Adults+Old", reflects the adult and old mortality, showing the typical regularity well described by the linear Gompertz law of mortality. We can relate Component 2, labeled "Young+Early Adults", to the behavior of mortality of the young and the early adults. In particular, this component captures the excess of mortality for age classes from 15-20 to 40-44. Component 3, "Infants", depicts the infant mortality. Finally, Component 4, labeled "Children+Young" reflects the behaviors of mortality of children and young. It is positively associated with age classes 1-4, 5-9, 10-14 and 15-19.

Age class	Adults+Old	Young+Early Adults	Infants	Children+Young
0	-0.05	0.01	**0.95**	0.04
1-4	0.01	-0.06	0.14	**0.54**
5-9	0.03	-0.07	-0.06	**0.62**
10-14	0.01	-0.06	-0.03	**0.48**
15-19	-0.05	**0.29**	-0.15	**0.25**
20-24	-0.07	**0.39**	-0.13	0.11
25-29	-0.07	**0.46**	-0.05	0.00
30-34	-0.03	**0.48**	0.03	-0.05
35-39	0.05	**0.42**	0.08	-0.06
40-44	0.16	**0.30**	0.11	-0.04
45-49	**0.26**	0.18	0.10	-0.02
50-54	**0.31**	0.07	0.06	0.00
55-59	**0.34**	0.01	0.02	0.02
60-64	**0.36**	-0.03	-0.02	0.02
65-69	**0.37**	-0.04	-0.04	0.02
70-74	**0.38**	-0.04	-0.05	0.01
75-79	**0.37**	-0.03	-0.02	-0.01
80-84	**0.35**	-0.01	0.03	-0.04

Table 4.3. Component matrix for the age classes (bold values indicate scores greater than 0.25). Male analysis.

The components for the years (component matrix **C**), displayed in Figure 4.1, offer an interesting interpretation in the light of longevity literature. We can see two different components: Component 1 (in blue), labeled "Late", represents the mortality dynamics after the "converging period" in the 1980s. Component 2 (in red), labeled "Early", shows higher values before 1980. Indeed, the recent historical worldwide longevity dynamics highlight the first decrease of mortality after the 50s, albeit, with high heterogeneity levels, that flattened in the decades around 75s-85s, which might be described as the global convergence ([80], [81], [78]). Subsequently, another improvement emerges quite globally around the world.

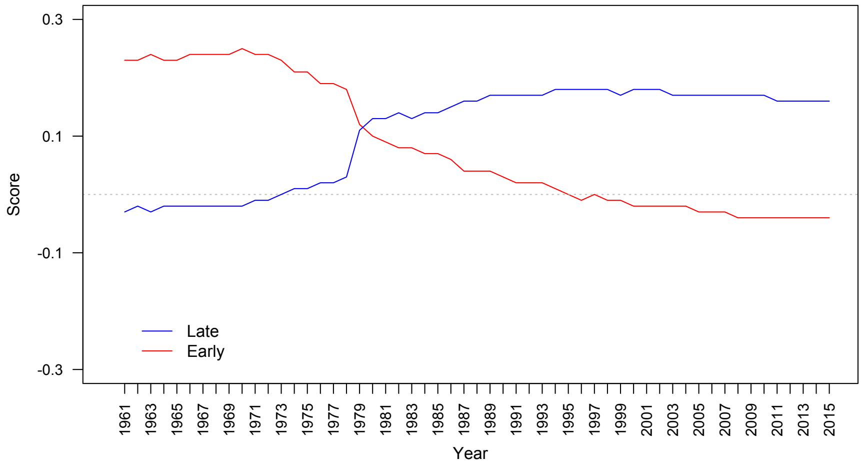


Figure 4.1. Component matrix for the years (male analysis)

The components for the countries (component matrix **D**) are given in Table 4.4. Component 1 for the countries, labeled "High Welfare", is mainly associated with Japan, but also with other European countries characterized by geographical proximity (France, Austria, Switzerland, and Italy) and with two North European countries (Sweden and Finland) that have recently shown relevant improvements in life expectancy leading the global records as the case of Sweden [80]. In general, the component mainly identifies countries well-known for their longevity and characterized by a high/medium welfare status. This component underlines how, as countries wealth grows, the raising rate of return on human capital causes a demographic transition [103]. This highlights the relation between economic and health status as suggested by Smith [96]. Component 2, labeled "Anglophone", is mainly composed by countries usually considered similar in culture and values: Ireland, UK, USA and Australia, which share the anglophone culture.

Country	High Welfare	Anglophone
AUS	0.06	**0.32**
AUT	**0.31**	-0.01
CAN	0.22	-0.11
USA	0.02	**0.36**
JAP	**0.46**	-0.22
BEL	0.14	0.22
DNK	0.22	0.11
FIN	**0.27**	0.03
FRATNP	**0.32**	-0.03
HUN	0.24	0.07
IRL	-0.18	**0.62**
ITA	**0.25**	0.05
NLD	0.23	0.09
NOR	0.12	0.24
ESP	0.16	0.16
SWE	**0.26**	0.05
CHE	**0.29**	0.01
GBR_NP	0.00	**0.40**

Table 4.4. Component matrix for the countries (bold values indicate scores greater than 0.25). Male analysis.

The previous component matrices provide a good description of the cause-of-death mortality scenario. Nevertheless, a piece of more detailed information could be achieved by interpreting the core tensor **G**, which provides the interaction term. Table 4.5 reports the tabular representation of the core where, once again, we highlight that the higher in absolute value an element of the core, the stronger the interaction among the quadruple of components involved.

Period: Late								
CoD components	Countries: High Welfare				Countries: Anglophone			
	Ad+Old	Y+E.Ad	Infants	Child+Y	Ad+Old	Y+E.Ad	Infants	Child+Y
External	1.44	**5.08**	-0.11	1.84	0.76	**4.46**	-0.22	1.85
CVD+Cancer	**6.40**	1.24	0.46	0.11	**4.88**	0.99	0.37	0.08
Other	2.13	1.96	**2.68**	1.32	1.16	1.58	2.18	1.29
Period: Early								
CoD components	Countries: High Welfare				Countries: Anglophone			
	Ad+Old	Y+E.Ad	Infants	Child+Y	Ad+Old	Y+E.Ad	Infants	Child+Y
External	1.15	**5.02**	0.13	**5.25**	0.70	**2.78**	0.00	**3.66**
CVD+Cancer	**5.77**	1.35	0.44	0.26	**4.98**	1.15	0.39	0.18
Other	0.26	0.83	0.79	1.73	0.00	0.41	0.56	1.26

Table 4.5. Core tensor (bold values indicate scores greater than 2.5). Male analysis.

The highest core element ($g_{2111} = 6.40$) refers to the interaction term between CVD and cancer (Component 2 for the causes of death) that occurred at adult and old ages (Component 1 for the age classes) in the late period (Component 1 for the years) for the high welfare countries considered by HMD (Component 1 for the countries). It means that during this period, for adults and elderly people, CVD and cancer lead to particularly high mortality in the countries considered. This is explained by the recent trends regarding longevity we are experiencing, especially in the high welfare countries. The interaction between Component 2 for the causes of death and Component 1 for the age classes is however uniformly high for all country and year components ($g_{2121} = 5.77$, $g_{2122} = 4.98$, $g_{2112} = 4.88$), thus highlighting its relevance in the analysis of mortality patterns. In the late period, for the high welfare countries, another high core element refers to the interaction linking to external causes of death and young and early adults mortality ($g_{1211} = 5.08$); it means that during the phase of convergence of longevity, in all the countries considered, young and adult individuals experienced high mortality for external causes. Once again, high interactions are also observed for different countries and/or years, as witnessed by core elements $g_{1221} = 5.02$, $g_{1212} = 4.46$, and, to a lesser extent, $g_{1222} = 2.78$. Moreover, in the early period there is a prevalence of deaths caused by external causes for children and young in the high welfare countries ($g_{1421} = 5.25$). Such a finding is also observed for the anglophone countries, even if the corresponding core element is lower ($g_{1422} = 3.66$), hence the interaction is weaker. Finally, in the high welfare countries, we can note the pretty high interaction between infants and other causes of death in the late period ($g_{3311} = 2.68$).

4.2.2 Female results

To get more reliable results, before fitting the model to the data, these are preprocessed. Similarly to the male analysis, we scale the data in such a way that the information pertaining to the countries and the age classes plays the same role.

In Table 4.6 we listed the best solution for the total number of components $P+Q+R+S$ with the corresponding goodness of fit index, expressed as a percentage.

P	Q	R	S	$P+Q+R+S$	Fit value (%)
2	2	2	2	8	80.07%
3	2	2	2	9	81.89%
3	3	2	2	10	84.68%
4	3	2	2	11	86.15%
4	4	2	2	12	88.85%
4	5	2	2	13	90.20%
5	5	2	2	14	91.48%
5	5	3	2	15	92.29%
5	5	3	3	16	92.93%
5	5	4	3	17	93.43%
5	5	4	4	18	93.93%
5	5	4	5	19	94.34%
5	6	4	5	20	94.74%
6	6	4	5	21	95.01%
6	6	4	6	22	95.40%
6	7	4	6	23	95.69%
6	6	5	7	24	95.98%
6	7	5	7	25	96.29%
6	7	5	8	26	96.57%
6	8	5	8	27	96.74%
6	8	6	8	28	96.87%
6	8	7	8	29	96.95%
6	8	7	8	30	97.02%
7	8	8	8	31	97.06%
8	8	8	8	32	97.42%

Table 4.6. Best retained solutions in terms of fit values varying P+Q+R+S from 8 to 32 . % with number of components and fit values (female analysis)

We select the solution with P=4, Q=4, R=2, S=2 that provides a fit value equal to 88.85%. Therefore, in comparison with the male analysis, we consider one more component for the causes of death. To motivate this choice, in Table 4.7, we give the component scores for the causes of death (matrix $\mathbf{A}$). We can see that, to a certain extent, the interpretation of the extracted components resembles the one of the male analysis. Component 1, labeled "CVD vs Other", mainly depends on CVD. Similarly to the male analysis, Component 2 is mainly associated with external diseases, while Component 3 is positively related to other causes of death. Finally, Component 4 is associated to non-smoking- and smoking-related cancer (0.84 and 0.52, respectively), portraying the higher incidence of cancer diseases among the populations. Such a component allows us to distinguish cancer diseases from the other causes, therefore legitimating the choice of having one more component for such a dimension compared to the male analysis.

Cause of death	CVD vs Other	External	Other	Cancer
Infectious	0.09	0.16	0.18	-0.13
Smoking-related cancer	0.12	-0.12	0.00	**0.52**
Non-smoking-related cancer	0.06	0.10	0.01	**0.84**
Diabetes	0.05	0.00	0.00	0.02
CVD	**0.98**	-0.01	0.02	-0.10
Respiratory	0.05	0.00	0.04	0.03
External	0.00	**0.97**	-0.01	0.00
Other	**-0.35**	-0.02	**0.98**	0.02

Table 4.7. Component matrix for the causes of death (bold values indicate scores greater than 0.25). Female analysis.

Table 4.8 reports the component score matrix for the age classes (matrix **B**). Component 1 reflects the adult mortality (age classes from 35-49 to 60-64). Component 2 is positively associated with all the classes from 1-4 to 30-35, capturing high mortality of children and young individuals. Finally, Component 3 reflects the mortality behavior of old people (age classes from 65-69 to 80-84) and Component 4 depicts the infant mortality (age classes 0 and 1-4).

Age	Adults	Child+Young	Old	Infants
0	0.02	-0.12	0.00	**0.91**
1-4	-0.06	**0.29**	0.01	**0.31**
5-9	0.02	**0.34**	-0.05	0.16
10-14	0.04	**0.34**	-0.04	0.15
15-19	-0.08	**0.47**	0.01	-0.09
20-24	-0.06	**0.45**	0.03	-0.08
25-29	0.03	**0.37**	0.03	-0.04
30-34	0.17	**0.27**	0.00	-0.01
35-39	**0.32**	0.16	-0.04	0.00
40-44	**0.41**	0.06	-0.05	0.00
45-49	**0.45**	0.00	-0.04	0.00
50-54	**0.44**	-0.04	0.00	0.00
55-59	**0.39**	-0.06	0.08	-0.01
60-64	**0.29**	-0.05	0.21	-0.02
65-69	0.16	-0.03	**0.34**	-0.03
70-74	0.04	0.00	**0.45**	-0.03
75-79	-0.06	0.02	**0.53**	0.01
80-84	-0.14	0.03	**0.58**	0.07

Table 4.8. Component matrix for the age classes (bold values indicate scores greater than 0.25). Female analysis.

Figure 4.2, displaying the component matrix for the years (**C**), confirms the interesting interpretation previously found in the male analysis with Component 1 (in red), labeled "Early", and Component 2 (in blue), labeled "Late". Note that, in the two analyses, the interpretation of the components remains the same, but the labels of the components are switched.

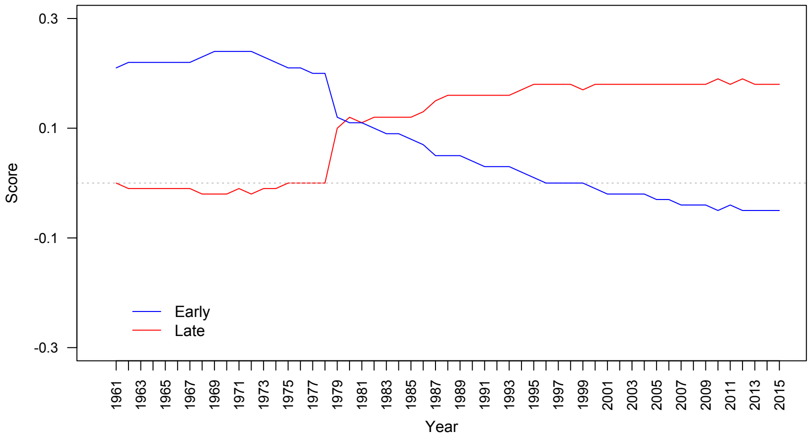


Figure 4.2. Component matrix for the years. Female analysis.

Table 4.9 shows the component matrix for the countries (**D**). As for the male analysis, Component 1 is associated to high welfare countries (France, Switzerland, and Sweden emerges as in the corresponding component in the male analysis), but some differences are visible. Component 2, labeled "Other", includes the two largest countries of Southern Europe, Italy and Spain, but also United Kingdom, Ireland and Hungary. Italy and Spain are usually considered similar in culture, values, and economic patterns; for example, both are Catholic and Latin countries, they are latecomers to European industrialization, and are medium- to large-sized countries with significant regional differences. The paper in [33] explains the common demographic and economic factors between Italy and Spain. What is more, Ireland is often compared to Spain and Italy for its pattern of mortality, for example in terms of life expectancy [90].

Countries	High Welfare	Other
AUS	0.20	0.14
AUT	0.20	0.13
CAN	**0.33**	-0.04
USA	0.18	0.15
JAP	0.08	0.24
BEL	**0.27**	0.04
DNK	**0.40**	-0.13
FIN	0.22	0.10
FRATNP	**0.34**	-0.05
HUN	-0.08	**0.48**
IRL	0.01	**0.37**
ITA	0.02	**0.35**
NLD	0.23	0.10
NOR	0.24	0.07
ESP	-0.06	**0.45**
SWE	**0.34**	-0.05
CHE	**0.38**	-0.10
GBR_NP	0.06	**0.32**

Table 4.9. Component matrix for the countries (bold values indicate scores greater than 0.25). Female analysis.

Further information can be revealed by the core tensor $\bar{\mathbf{G}}$ reported in Table 4.10.

Period: Early								
CoD components	Countries: High Welfare				Countries: Other			
	Ad	Child+Y	Old	Inf	Ad	Child+Y	Old	Inf
CVD vs Other	**3.43**	0.97	**4.86**	0.15	**3.46**	1.49	**3.75**	0.22
External	1.45	**5.79**	0.17	1.44	0.76	**3.69**	0.09	0.94
Other	0.83	1.52	0.37	1.11	0.79	1.79	0.28	1.37
Cancer	**3.29**	1.66	0.66	0.23	2.17	1.25	0.24	0.01
Period: Late								
CoD components	Countries: High Welfare				Countries: Other			
	Ad	Child+Y	Old	Inf	Ad	Child+Y	Old	Inf
CVD vs Other	2.31	0.74	**3.90**	0.05	2.22	0.81	**3.39**	0.03
External	1.94	**4.30**	0.23	0.20	1.07	**3.14**	0.18	0.09
Other	2.18	1.96	1.64	**3.03**	1.66	1.95	1.11	2.00
Cancer	**4.38**	1.20	1.80	0.13	**3.14**	1.15	1.09	0.09

Table 4.10. Core tensor (bold values indicate scores greater than 2.5). Female analysis.

The highest core element ($g_{2211} = 5.79$) refers to the interaction term between external causes of death that occur for children and at young ages in the early period for the high welfare countries. It means that during this period, for these age classes, the external causes of death lead to high mortality in such countries. In the early period, for the same group of countries the element linked to CVD vs Other ($g_{1311} = 4.86$) reveals that during the early phase, for old people, there is a higher incidence of deaths by CVD compared to Other causes of death. Finally, we can observe, for adults, high levels of mortality caused by CVD compared to Other ($g_{1111} = 3.43$), and Cancer ($g_{4111} = 3.29$).

In the late period, still focusing our attention on the high welfare countries, high mortality for adults due to cancer emerges ($g_{4121} = 4.38$). As for the early period and for the elderly, we can see high mortality due to CVD compared to Other ($g_{1321} = 3.90$). The core element $g_{3421} = 3.03$ allows us to discover the strong level of mortality due to Other causes of death. Finally, $g_{2221} = 4.30$ refers to the interaction between external causes of deaths that occur to children and young individuals in the late period for the high welfare countries.

With respect to the core elements involving Other countries (Component 2 for the countries), in the early period, high mortality due to external causes for children and young people ($g_{2212} = 3.69$) is registered. We also observe high mortality for CVD vs Other in adults and the elderly ($g_{1112} = 3.46$, $g_{1312} = 3.75$). Note that such results are consistent with the male analysis. Furthermore, in the late period, CVD vs Other is relevant for old people $g_{1322} = 3.39$). Finally, we can discover high mortality due to Cancer and External causes for, respectively, adults and children and young individuals (respectively, $g_{4122} = 3.14$, $g_{2222} = 3.14$).

4.3 Conclusions

Modeling mortality represents a crucial issue in various fields: public health, pension schemes and financial planning are prime examples. In recent years, researchers were prioritizing the investigation of the overall longevity. Nevertheless, more specific mortality data might provide a piece of crucial information for designing national decision planning. Our approach represents an advance in longevity modeling, showing a more informative view of mortality levels specific for causes of death, in a multi-population framework. Leveraging the Tucker4 method, our contribution to the literature consists of a multi-population and causes of death analysis in a unified framework. Indeed, different populations with similar socioeconomic statuses share similar mortality trends to some extent. Such common information could be extracted and exploited in a multi-population framework to help model mortality of individual populations. The dataset we have considered has been provided by the World Health Organization, arranged into a four-way array structure composed by causes of death, age, years, and countries. For both genders separately, we have shown how the Tucker4 model is able to extract meaningful demographic insights. In fact, the aim of this model is to efficiently summarize all the information in the four-way data, also considering the various interactions (two-, three-, and four-way). In particular, it helps us understand how the rises in survival, witnessed in many high-income and developed countries, during the twentieth century, were determined especially by the reduction in a few specific major causes of death groups, which led to the longevity improvements.

In this chapter is clear one of the main advantage of the Tucker model as expresses in [99], we can observe the *full core array* $\underline{\mathbf{G}}$ that explain the relation between the components. This helped us to identify the main relationship between the components in this more complex situation with four dimensions.

A final consideration is related to possible limitations in the data. For the HMD, with reference to the fact that in the selected time window a limited number of countries were selected, probably through a broadening of the data base the models could provide further significant demographic relationships. With reference to the causes of death, the main limitation is linked to the need to manually reconstruct the data, also considering the different classifications of the causes of death that have occurred over time. This is linked to one of the actuarial themes that is recurrent today: the quality of the data. Only through a joint economic-political effort can we converge towards a global interest in collecting demographic data for actuarial, demographic and pension analysis purposes.

Possible further steps in the future researches, may concern the use of different convergence algorithms for the identification of solutions, the introduction and development of further analyzes to support the choice of solutions, the use of different models to compare the results, the introduction of further significant dimensions of analysis and finally the possibility of changing the field of application. For example, in non-life insurance to calculate the claims reserve, identifying a data structure that can be constructed in a multidimensional array, exploiting tensor decomposition methods to obtain further information useful to insurance companies.

Bibliography

[1] Acar, Evrim Camtepe, Seyit A. and Krishnamoorthy, Mukkai S. and Yener, Bulent, Modeling and Multiway Analysis of Chatroom Tensors, Intelligence and Security Informatics, 2005,256-258

[2] Aitchison, J. 1986. The Statistical Analysis of Compositional Data. London: Chapman & Hall."

[3] Aitchison, J. (1982). The statistical analysis of compositional data. Journal of the Royal Statistical Society: Series B (Methodological), 44(2), 139–160.

[4] Aitchison, J. (1994). Principles of compositional data analysis. Multivariate analysis and its applications. In T. W. Anderson and I. Olkin (Eds.), IMS Lectures notes-monograph series (Vol. 24, pp. 73–81). Institute of Mathematical Statistics.

[5] Aitchison, J. (2005). A concise guide to compositional data analysis. In 2nd Compositional data analysis workshop. Girona, Spain. Retrieved from http://ima.udg.edu/Activitats/CoDaWork05/.

[6] Andersson, C.A. & Bro, R. (2000) The N-way Toolbox for MATLAB. Chemometrics and Intelligent Laboratory Systems, 52, 1-4. Available at: http://www.models.life.ku.dk/nwaytoolbox"

[7] Avdjiev, Stefan and Giudici, Paolo and Spelta, Alessandro (2019). Measuring contagion risk in international banking. Journal of Financial Stability 42, 36–51"

[8] B. Savas, Analyses and tests of handwritten digit recognition algorithms, master's thesis, Link¨oping University, Sweden, 2003.

[9] Bader, Brett W. and Kolda, Tamara G. and Harshman, Richard A. (2006). Temporal analysis of social networks using three-way DEDICOM. Sandia National Lab.(SNL-NM), Albuquerque, NM (United States)"

[10] Bader, Brett W. and Harshman, Richard A. and Kolda, Tamara G. (2007). Temporal analysis of semantic graphs using ASALSAN. Seventh IEEE international conference on data mining (ICDM 2007), 33-42"

[11] Bader, Brett W. and Berry, Mike and Browne, Murray, Discussion Tracking in Enron Email using PARAFAC, 2007, 4

[12] Balarajan, R. and McDowall, ME (1988). Regional socioeconomic differences in mortality among men in Great Britain today. Public Health 102(1), 33–43"

[13] Bergeron-Boucher, Marie-Pier and Canudas-Romo, Vladimir and Oeppen, Jim and Vaupel, James W (2017). Coherent forecasts of mortality with compositional data analysis. Demographic Research, 37: 527-566, JSTOR."

[14] Bergeron-Boucher, Marie-Pier., Simonacci, Violetta., Oeppen, Jim., Gallo, Michele. 2018. Coherent modeling and forecasting of mortality patterns for subpopulations using multiway analysis of compositions: An application to Canadian provinces and territories. Journal North American Actuarial Journal, 92-118, 22."

[15] Bobak, Martin and Marmot, Michael (1996). East-West mortality divide and its potential explanations: proposed research agenda. Bmj 312(7028), 421–425"

[16] Booth, H., and Tickle L. (2008). Mortality Modelling and Forecasting: A Review of Methods. Annals of Actuarial Science 3: 3–43."

[17] Bro, Rasmus and Kiers, Henk AL. (2003). A new efficient method for determining the number of components in PARAFAC models. Journal of Chemometrics: A Journal of the Chemometrics Society 17(5), 274-286"

[18] Carroll, J. Douglas and Chang, Jih-Jie (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika 35(3), 283-319"

[19] Carroll, John Douglas, and Chang, Jih Jie (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition.Psychometrika, 35, 283-319."

[20] Caselli, Graziella and Meslé, France and Vallin, Jacques (2002). Epidemiologic transition theory exceptions. Genus, 9–51."

[21] Cattell, Raymond B. (1944). "Parallel proportional profiles" and other principles for determining the choice of factors by rotation. Psychometrika 9(4), 267-283"

[22] Christensen, Kaare and Davidsen, Michael and Juel, Knud and Mortensen, Laust Hvas and Rau, Roland and Vaupel, James W (2010). The divergent life-expectancy trends in Denmark and Sweden-and some potential explanations. International differences in mortality at older ages: Dimensions and sources, 385–407"

[23] Cichocki, Andrzej and Zdunek, Rafal and Phan, Anh Huy and Amari, Shun-ichi (2009). Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation. John Wiley & Sons"

[24] Claus A. Andersson, Rasmus Bro, Improving the speed of multi-way algorithms:: Part I. Tucker3, Chemometrics and Intelligent Laboratory Systems, Volume 42, Issues 1–2,1998,Pages 93-103,

[25] Crimmins E., Kim JK., Vasunilashorn S (2010) Biodemography: new approaches to understanding trends and differences in population health and mortality. Demography 47(Suppl 1):S41"

[26] De Lathauwer et al., A multilinear singular value decomposition, SIAM J. Matrix Anal. A., 21 (2000), pp. 1253–1278.

[27] Devolder, P., Levantesi, S. & Menzietti, M. (2021). Automatic Balance Mechanisms for Notional Defined Contribution pension systems guaranteeing social adequacy and financial sustainability: an application to the Italian pension system. Annals of Operations Research, 299, 765-795. DOI: 10.1007/s10479-020-03819-x"

[28] Dong, Y., Huang, F., Yu, H., Haberman, S. (2020). Multi-population mortality forecasting using tensor decomposition. Scandinavian Actuarial Journal, 8: 754-775, DOI: 10.1080/03461238.2020.1740314"

[29] Eckart, C., Young, G., (1936). The approximation of one matrix by another of lower rank. Psychometrika 1: 211-218"

[30] Eduardo Martınez-Montes and Pedro A. Valdés-Sosa and Fumikazu Miwakeichi and Robin I. Goldman and Mark S. Cohen, Concurrent EEG/fMRI analysis by multiway Partial Least Squares, NeuroImage, 22,3, 1023-1034, 2004

[31] Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barcelo-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. Mathematical Geology, 35(3), 279–300.

[32] Egozcue, J. J., and Pawlowsky-Glahn, V. (2005). Groups of parts and their balances in compositional data analysis. Mathematical Geology, 37(7), 795–828.

[33] Emanuele Felice, Josep Pujol Andreu, Carlo D'Ippoliti (2016). GDP and life expectancy in Italy and Spain over the long run: A time-series approach. DEMOGRAPHIC RESEARCH VOLUME 35, ARTICLE 28, PAGES 813−86. PUBLISHED 22 SEPTEMBER 2016. DOI: 10.4054/DemRes.2016.35.28"

[34] Giordani P, Kiers HAL. (2018). A review of tensor based methods and their application to hospital care data. Statistics in Medicine. 37:137–156. https://doi.org/10.1002/sim.7514"

[35] Giordani, Paolo and Kiers, Henk AL. and Del Ferraro, Maria Antonietta (2014). Three-way component analysis using the R package ThreeWay. Journal of Statistical Software 57(1), 1–23"

[36] Giordano, G., Haberman, S., Russolillo, M. (2019). Coherent modeling of mortality patterns for age-specific subgroups. Decisions in Economics and Finance, 42: 189-204. DOI: https://doi.org/10.1007/s10203-019-00245-y"

[37] H. A. L. Kiers, J. M. F. ten Berge, Y. Takane, and J. de Leeuw, A generalization of Takane's algorithm for DEDICOM, Psychometrika, 55 (1990), pp. 151–158.

[38] H. A. L. Kiers and Y. Takane, Constrained DEDICOM, Psychometrika, 58 (1993), pp. 339–355.

[39] H. A. L. Kiers, J. M. F. Ten Berge, and R. Bro, PARAFAC2 Part I. A direct fitting algorithm for the PARAFAC2 model, J. Chemometr., 13 (1999), pp. 275–294.

[40] Ho, J. Y. and Hendi, A. S. (2018). Recent trends in life expectancy across high income countries: retrospective observational study. BMJ 362, k2562"

[41] Harshman, Richard Allan (1972). PARAFAC2: Mathematical and technical notes. UCLA Working Papers in Phonetics, 22, 30-47."

[42] Harshman, Richard A. and others (1970). Foundations of the PARAFAC procedure: Models and conditions for an"" explanatory"" multimodal factor analysis. University of California at Los Angeles Los Angeles, CA"

[43] Harshman, Richard A. (1978). Models for analysis of asymmetrical relationships among N objects or stimuli. First Joint Meeting of the Psychometric Society and the Society of Mathematical Psychology, Hamilton, Ontario, 1978"

[44] Harshman, Richard A. and Green, Paul E and Wind, Yoram and Lundy, Margaret E (1982). A model for the analysis of asymmetric data in marketing research. Marketing Science 1(2), 205-242"

[45] Harshman, RA. and Kiers, HAL (1987). Algorithms for DEDICOM analysis of asymmetric data. European Meeting of the Psychometric Society, Enschede 1987"

[46] Helwig, N.E. (2019). multiway: Component Models for Multi-Way Data. R package version 1.0-6. https://CRAN.R-project.org/package=multiway"

[47] Henrion, René (1994). N-way principal component analysis theory, algorithms and applications. Chemometrics and intelligent laboratory systems 25(1), 1-23"

[48] Hitchcock (1927). The expression of a tensor or a polyadic as a sum of products. Journal of Mathematics and Physics 6(1-4), 164-189"

[49] Hitchcock, F.L. (1927). Multiple invariants and generalized rank of a p-way matrix or tensor. Journal of Mathematics and Physics 7, 39-79."

[50] Hitchcock (1928). Multiple invariants and generalized rank of a p-way matrix or tensor. Journal of Mathematics and Physics 7(1-4), 39-79"

[51] John Wiley and Sons, Ltd, isbn: 9780470747278, Introduction – Problem Statements and Models, Nonnegative Matrix and Tensor Factorizations,1, pages:1-80, doi: https://doi.org/10.1002/9780470747278.ch1, url https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470747278.ch1, 2009

[52] Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis. Psychometrika, 23, 187-200."

[53] Kapteyn, Arie and Neudecker, Heinz and Wansbeek, Tom (1986). An approach ton-mode components analysis. Psychometrika 51(2), 269-275"

[54] Kiers, and Mechelen. (2001). Three Way Component Analysis: Principles and Illustrative Application. Psychological Methods. Vol6, No.1 84-110. doi: 10.1037//1082-989X.6.1.84"

[55] Kiers, Henk AL (1982). An alternating least squares algorithm for PARAFAC2 and three-way DEDICOM. Computational Statistics & Data Analysis 16(1), 103-118"

[56] Kiers, H.A.L. (2000). Towards a standardized notation and terminology in multiway analysis. Journal of Chemometrics, 14, 105-122."

[57] Kjærgaard, S., Ergemen, Y.E., Kallestrup-Lamb, M., Oeppen, J., Lindahl-Jacobsen, R. (2019). Forecasting causes of death by using compositional data analysis: the case of cancer deaths. J. R. Stat. Soc. C, 68: 1351-1370. DOI: https://doi.org/10.1111/rssc.12357"

[58] Kroonenberg, P.M. & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. Psychometrika, 45(1), 69-97. "

[59] Kroonenberg P.M. (1983). Three-mode principal component analysis: Theory and applications. DSWO Press: Leiden, 1983.

[60] Kroonenberg, P.M. (2016). My multiway analysis: from Jan de Leeuw to TWPack and back. Journal of Statistical Software, 73, 1-22.

[61] Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. Journal of the American Statistical Association 87(419), 659–671"

[62] Lee, Daniel D and Seung, H Sebastian (1999). Learning the parts of objects by nonnegative matrix factorization. Nature 401(6755), 788–791"

[63] Leon, D. A.,Jdanov, D. A., and Shkolnikov, V. M.(2019). Trends in life expectancy and age specific mortality in England and Wales, 1970-2016, in comparison with a set of 22 high income countries: An analysis of vital statistics data. The Lancet Public Health, 4, E575-E582"

[64] Levantesi, S., Nigri, A., Piscopo, G. (2021). Clustering-based simultaneous forecasting of life expectancy time series through Long-Short Term Memory Neural Networks. International Journal of Approximate Reasoning. https://doi.org/10.1016/j.ijar.2021.10.008"

[65] Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee–Carter method. Demography 42(3), 575–594."

[66] Li, H., Li, H., Lu, Y., Panagiotelis, A. (2019). A forecast reconciliation approach to cause-of-death mortality modeling. Insurance: Mathematics and Economics, 86: 122-133."

[67] Lindahl Jacobsen, R., Rau, R., Jeune, B., Canudas Romo, V., Lenart, A., Christensen, K., & Vaupel, J. W. (2016). Rise, stagnation, and rise of Danish women's life expectancy. Proceedings of the National Academy of Sciences, 113, 4015-4020"

[68] Lundy, Margaret E and Harshman, Richard A and Paatero, Pentti and Swartzman, Leora C (2003). Application of the 3-way Dedicom model to skew-symmetric data for paired preference ratings of treatments for chronic back pain. TRICAP 2003 Meeting, Lexington, Kentucky. Citeseer"

[69] Pintao Lyu, Anja De Waegenaere and Bertrand Melenberg (2020): A Multipopulation Approach to Forecasting All-Cause Mortality Using Cause-of-Death Mortality Data, North American Actuarial Journal, DOI: 10.1080/10920277.2019.1662316"

[70] M. A. Di Palma and P. Filzmoser and M. Gallo and K. Hron (2018) A robust Parafac model for compositional data, vol. 45, n. 8,1347-1369, Taylor and Francis, doi: 10.1080/02664763.2017.1381669

[71] M. A. O. Vasilescu and D. Terzopoulos, Multilinear analysis of image ensembles: TensorFaces, in ECCV 2002: 7th European Conference on Computer Vision, vol. 2350 of Lecture Notes in Computer Science, Springer, 2002, pp. 447–460.

[72] Mehta, N.K., Abrams, L.R., Myrskyla, M. (2020). US life expectancy stalls due to cardiovascular disease, not drug deaths. Proc National Acad Sci, 117(13): 6998-7000"

[73] Meslé, France (2004). Mortality in Central and Eastern Europe: long-term trends and recent upturns. Demographic Research 2, 45–70."

[74] Meslé, France and Vallin, Jacques (2017). The end of East-West divergence in European life expectancies? An introduction to the special issue. European Journal of Population 33(5), 615–627"

[75] Mesle, F., and Vallin, J. (2006). Diverging trends in female old age mortality: The United States and The Netherlands versus France and Japan. Population and Development Review, 32, 123-145."

[76] Michele Gallo (2015) Tucker3 Model for Compositional Data, Communications in Statistics - Theory and Methods, 44:21, 4441-4453, DOI: 10.1080/03610926.2013.798664

[77] Nigri, Andrea and Barbi, Elisabetta and Levantesi, Susanna(2022). The relay for human longevity: country-specific contributions to the increase of the best-practice life expectancy. Springer, Quality and Quantity, https://doi.org/10.1007/s11135-021-01298-1"

[78] Nigri, A., Barbi, E., Levantesi, S. (2021). The relationship between longevity and lifespan variation. Statistical Methods and Applications. https://doi.org/10.1007/s10260-021-00584-4"

[79] Nigri, A., Levantesi, S. & Piscopo, G. (2022). Causes-of-Death Specific Estimates from Synthetic Health Measure: A Methodological Framework. Social Indicators Research. DOI: https://doi.org/10.1007/s11205-021-02870-w"

[80] Oeppen, J., Vaupel, J.W. (2002). Broken Limits to Life Expectancy. Science, 296(5570), 1029-1031. https://doi.org/10.1126/science.1069675"

[81] Oeppen, J., Vaupel, J.W. (2006). The Linear Rise in the Number of Our Days. Social Insurance. Studies, 3. The Linear Rise in Life Expectancy: History and Prospects"

[82] Oeppen, J. W. (2008). Coherent forecasting of multiple-decrement life tables: a test using Japanese cause of death data. Paper presented at the European Population Conference 2008, Barcelona, Spain, July 9–12, 2008."

[83] Olshansky, S Jay and Ault, A Brian (1986). The fourth stage of the epidemiologic transition: the age of delayed degenerative diseases. The Milbank Quarterly, 355–391"

[84] Pecora, Nicolò and Rovira Kaltwasser, Pablo and Spelta, Alessandro (2016). Discovering sifis in interbank communities. PloS one 11(12)"

[85] Pecora, Nicolò and Spelta, Alessandro (2017). A multi-way analysis of international bilateral claims. Social Networks 49, 81-92"

[86] Phan, Anh Huy and Cichocki, Andrzej and Vu-Dinh, Thanh (2010). Nonnegative DEDICOM based on tensor decompositions for social networks exploration. Australian Journal of Intelligent Information Processing Systems (ICONIP'10) 12(1), 10-15"

[87] R. A. Harshman and M. E. Lundy, Three-way DEDICOM: Analyzing multiple matrices of asymmetric relationships. Paper presented at the Annual Meeting of the North American Psychometric Society, 1992.

[88] Remund, A., Camarda, C.G., Riffe, T. (2018). A Cause-of-Death Decomposition of Young Adult Excess Mortality. Demography. 957-978. doi: 10.1007/s13524-018-0680-9."

[89] Ridsdale, and Gallop. (2010). Mortality by cause of death and by socio-economic and demographic stratification, ICA2010."

[90] Roser, Max and Ortiz-Ospina, Esteban and Ritchie, Hannah (2013). Life expectancy. Our World in Data"

[91] Russolillo, M., Giordano, G., Haberman, S. (2011). Extending the Lee–Carter model: a three-way decomposition. Scandinavian Actuarial Journal, 2011(2): 96–117."

[92] Shkolnikov V, McKee M, Leon DA. (2001). Changes in life expectancy in Russia in the mid-1990s. Lancet, 357: 917-921"

[93] Simonacci, V., Gallo, M. On four-way CP model estimation efficiency. Comput Stat (2022). https://doi.org/10.1007/s00180-022-01271-y

[94] Simonacci, V., Gallo, M. Three-way principal balance analysis: algorithm and interpretation. Ann Oper Res (2022). https://doi.org/10.1007/s10479-022-04782-5

[95] Simonacci, V., and Gallo, M. (2017). Statistical tools for student evaluation of academic educational quality. Quality and Quantity, 51(2), 565–579.

[96] Smith, James P (1999). Healthy bodies and thick wallets: the dual relation between health and economic status. Journal of Economic perspectives 13(2), 145–166"

[97] Spelta, Alessandro and Flori, Andrea and Pammolli, Fabio (2018). Investment communities: behavioral attitudes and economic dynamics. Social Networks 55, 170-188"

[98] Spelta, A and Pecora, N and Kaltwasser, P Rovira (2019). Identifying systemically important banks: A temporal approach for macroprudential policies. Journal of policy modeling 42(1), 197–218"

[99] Pieter M. Kroonenberg, Component Models for Fully-Crossed Designs, Chapter 4, 16 July 2007 https://doi.org/10.1002/9780470238004.ch4

[100] Tóth, Csaba G. (2021). Multi-population models to handle mortality crises in forecasting mortality: A case study from Hungary. Society and Economy 43(2), 128–146."

[101] Tabeau, E.; van den Berg Jeths, A.; Heathcote, C. eds. (2001). Forecasting mortality in developed countries. Dordrecht: Kluwer Academic."

[102] Takashi, Inoguchi and Jain, Purnendra (2000). Japanese foreign policy today: A reader. Springer"

[103] Tamura, Robert (1996). From decay to growth: a demographic transition to economic growth. Journal of Economic Dynamics and Control 20(6-7), 1237–1261"

[104] Tucker, Ledyard R (1963). Implications of factor analysis of three-way matrices for measurement of change. Problems in measuring change 15(122-137), 3"

[105] Tucker, Ledyard R (1966). Some mathematical notes on three-mode factor analysis. Psychometrika 31(3), 279-311"

[106] Wilmoth, J. R. (2000). Demography of longevity: past, present, and future trends. Experimental Gerontology 35(9–10), 1111–1129. https://doi.org/10.1016/S0531-5565(00)00194-7"

[107] Tamara G. Kolda and Brett W. Bader. Tensor Decompositions and Applications. SIAM Rev, 2009,51,455-500.