Year : 2021

CAUSE-SPECIFIC MORTALITY INTERACTIONS

Glushko Viktoriya

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE SCIENCES ACTUARIELLES

CAUSE-SPECIFIC MORTALITY INTERACTIONS

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteure ès Sciences Actuarielles

par

Viktoriya GLUSHKO

Directrice de thèse
Prof. Séverine Arnold

Jury

Prof. Paul André, Président
Prof. François Dufresne, expert interne
Prof. Ana María Debón Aucejo, experte externe
Prof. Stéphane Loisel, expert externe

LAUSANNE
2021

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE SCIENCES ACTUARIELLES

CAUSE-SPECIFIC MORTALITY INTERACTIONS

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteure ès Sciences Actuarielles

par

Viktoriya GLUSHKO

Directrice de thèse
Prof. Séverine Arnold

Jury

Prof. Paul André, Président
Prof. François Dufresne, expert interne
Prof. Ana María Debón Aucejo, experte externe
Prof. Stéphane Loisel, expert externe

LAUSANNE
2021

UNIL | Université de Lausanne
HEC Lausanne
Le Décanat
Bâtiment Internef
CH-1015 Lausanne

IMPRIMATUR

Sans se prononcer sur les opinions de l'autrice, la Faculté des Hautes Etudes Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Madame Viktoriya GLUSHKO, titulaire d'un bachelor et d'un master en mathématiques appliquées de l'Université Nationale Technique d'Ukraine, et d'un master ès sciences actuarielles de l'Université de Lausanne, en vue de l'obtention du grade de docteure ès sciences actuarielles.

La thèse est intitulée :

CAUSE-SPECIFIC MORTALITY INTERACTIONS

Lausanne, le 28.08.2021

La doyenne

p.o.

Marianne Schmid Mast

Members of the Thesis Committee

Professor Séverine Arnold: Thesis supervisor, Full Professor at the Department of Actuarial Science, Faculty of Business and Economics (HEC Lausanne), University of Lausanne, Switzerland.

Professor François Dufresne: Internal expert, Full Professor at the Department of Actuarial Science, Faculty of Business and Economics (HEC Lausanne), University of Lausanne, Switzerland.

Professor Ana Maria Debón Aucejo: External expert, Associated Professor at the Department of Applied Statistics and Operational Research and Quality, Faculty of Business Administration and Management, Polytechnic University of Valencia, Spain.

Professor Stéphane Loisel: External expert, Directeur du Laboratoire de Science Actuarielle et Financière, Institut de Science Financière et d'Assurance, University of Lyon 1, France.

Professor Paul André: Jury president, Full Professor at the Department of Accounting and Control, Faculty of Business and Economics (HEC Lausanne), University of Lausanne, Switzerland.

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Viktoriya GLUSHKO

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: _____ Date: 24.08.2021

Prof. Séverine ARNOLD
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Viktoriya GLUSHKO

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: 20 July 2021

Prof. François DUFRESNE
Internal member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Viktoriya GLUSHKO

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: ___26.08.2021_____

Prof. Ana María DEBÓN AUCEJO
External member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Viktoriya GLUSHKO

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ____  _____ Date: __24/8/2021_____

Prof. Stéphane LOISEL
External member of the doctoral committee

Declaration of Authorship

I, Viktoriya Glushko, declare that this thesis titled, Cause-specific mortality interactions and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this university.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this university or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Parts of this thesis appeared in the publications that are indicated in the relevant chapters.
- Where the thesis is based on work done by myself jointly with my supervisor, I have made clear exactly what was done by others and what I have contributed myself.

_____

Signature

_____

Date

Acknowledgements

This work would not have been possible without the precious support of my thesis supervisor, Prof. Séverine Arnold. I would like to thank her for the guidance, availability and patience that she granted me with for more than five years. It was an honor and a priveledge to be her PhD student. I would also like to thank the anonymous referees that helped to improve those parts of this work that have been already published. To my colleagues at the department, thank you for the best company during my PhD one could wish for.

My special thanks and gratitude go to my husband, Sasha, and my daughter, Katya, who supported me throughout my thesis adventure and shared with me its ups and downs. All of my accomplishments I owe to their uncoditional faith in me.

I also gratefully acknowledge the financial support received from the Swiss National Science Foundation for the project "Cause-Specific Mortality Interactions" (project number 100018_16289).

Contents

List of Figures

List of Tables

Synthesis Report

The rising life expectancy is one of the biggest challenges the insurance industry has ever faced. This work aims to contribute to a better understanding of the past development of the mortality rates by first, disaggregating the mortality rates by cause of death, and second, by studying the relations between the cause-specific mortality rates through cointegration techniques. This approach helps to complement the current knowledge on cause-of-death mortality dependence that is essential for setting and testing mortality assumptions and scenarios. The present thesis provides answers to the raised question through tree essays, each corresponding to a chapter.

Chapter 1: Short- and Long-Term Dynamics of Cause-specific Mortality Rates Using Cointegration Analysis

This chapter is based on the following article:

Séverine Arnold & Viktoriya Glushko (2021) Short- and Long-Term Dynamics of Cause-Specific Mortality Rates Using Cointegration Analysis, North American Actuarial Journal, https://doi.org/10.1080/10920277.2021.1874421

This paper applies cointegration analysis and vector error correction models to model the short- and long-run relationships between cause-specific mortality rates. We work with the data from five developed countries (USA, Japan, France, England and Wales, and Australia) and split the mortality rates into five main causes of death (Infectious&Parasitic, Cancer, Circulatory diseases, Respiratory diseases, and External causes). We successively adopt the short- and long-term perspective, and analyze how each cause-specific mortality rate impacts and reacts to the shocks received from the rest of the causes. We observe that the cause-specific mortality rates are closely linked to each other, apart from the External causes that show an entirely independent behavior, and hence, could be considered as truly exogenous. We summarize our findings with the aim to help practitioners set more informed assumptions concerning the future development of mortality.

Chapter 2: Cause-Specific Mortality Rates: Common Trends and Differences

This chapter is based on the following article:

Séverine Arnold & Viktoriya Glushko (2021) Cause-specific mortality rates: Common trends and differences, Insurance: Mathematics and Economics, Volume 99, 2021, Pages 294-308, ISSN 0167-6687, https://doi.org/10.1016/j.insmatheco.2021.03.027.

In this paper, we continue to study the past development of cause-specific mortality. We work with the data from five developed countries (USA, Japan, France, England and Wales, and Australia), two sexes, and split the mortality rates into five main groups of causes of death (Infectious&Parasitic, Cancer, Circulatory diseases, Respiratory diseases, and External causes). As it was shown in Arnold and Sherris (2016), these time series of cause-specific mortality rates are cointegrated and so, there exist long-run equilibrium relationships between them. While the previous research focused on the stationary part of the system of cause-specific mortality rates, in the present paper we study its non-stationary part. For this, we explicitly extract common stochastic trends from the original variables and compare them across the different datasets. By testing cointegration assumptions about these trends, we are able to get a better representation and understanding of how cause-specific death rates are evolving. We believe that common patterns emerging from such analysis could indicate a link to more fundamental biological processes such as aging.

Chapter 3: Forecasting Cause-Specific Mortality Rates Using the Insights from the Cointegration Analysis

This chapter is based on the following working paper:

Séverine Arnold & Viktoriya Glushko (2021) Forecasting Cause-Specific Mortality Rates Using the Insights from the Cointegration Analysis. *working paper*.

Much like the all-cause mortality, cause-specific mortality rates in countries with similar socio-economic characteristics are likely to follow comparable development patterns. They are also not expected to substantially diverge in the future. We propose to assess the coherence of the past country-specific experiences by the means of the cointegration analysis applied to the mortality time trends extracted by country and cause of death. Indeed, should the time trends of two countries be cointegrated, this would indicate there

existed a long-run stationary relation between them, and so, the mortality patterns of these countries were linked to each other in their long-term development. We analyze the data from five developed Western European countries (France, Italy, Netherlands, Spain, and England and Wales), two sexes, and split the mortality rates into five main groups of causes of death (Infectious&Parasitic, Cancer, Circulatory diseases, Respiratory diseases, and External causes). We observe that while in many cases the cause-specific time trends are indeed cointegrated, this is not always the case in spite of the closeness of the studied countries. Further, once we include the countries having the cointegrated time trends in a multipopulational context, such as the Li-Lee mortality model, the forecast results are improved in comparison with the basic Lee-Carter approach.

Arnold, S. and M. Sherris, 2016, International Cause-Specific Mortality Rates: New Insights from a Cointegration Analysis., *ASTIN Bulletin*, 46(1):9-38.

Chapter 1

Short- and Long-Term Dynamics of Cause-specific Mortality Rates Using Cointegration Analysis

1.1 Introduction

It is commonly known that the mortality rates have been decreasing for many decades now. Although a joyful development *per se*, these changes pose serious problems for insurance companies, pension funds, and social security schemes, as they need to know if the observed decline will continue, slow down or, on the contrary, speed up. In this work, we will not venture to forecast the prospective evolution of mortality rates, but provide new insights on the past developments. We believe that once we understand better the past, we will be able to make better prognoses about the future.

Numerous parametric models have been developed in order to take into account such characteristics of mortality rates development as age, year of birth, and rate of improvement. For a review thereof we direct the interested reader to Booth and Tickle (2008), Cairns (2013) and Debón et al. (2006) including their references. For our part, we want to gain additional insight into the past development of mortality rates by concentrating on a more detailed breakdown of mortality data, namely by causes of death. Indeed, just from an eye inspection of the cause-specific mortality rates, it becomes clear that these rates showed strikingly divergent trends over the last 50 to 60 years. These phenomena have already been extensively studied and described (e.g., Himes 1994; Horiuchi and Wilmoth 1997; Costa 2005; Cutler et al. 2006).

However, it is much more difficult to integrate cause-specific mortality rates into a model, as they are dependent, and this dependence is, strictly speaking, not observable. Indeed, given a death event at a young age from an accident, for example, it is impossible to say what the chances of this person would be to die later from cancer or any other cause,

had he or she remained alive. Among theories and methods trying to take into account the dependency structure between the cause-specific mortality rates one can cite models incorporating individual risk factors (e.g., Manton and Poss 1979; Manton et al. 1991), models employing multiple cause-of-death data (e.g., Mackenbach et al. 1999; Manton and Myers 1987); and more recently, copulas (e.g., Lo and Wilke 2010; Dimitrova et al. 2013).

Although possible theoretically, models that take into account the dependency between the causes of death are problematic to use in practice, as they require a significant amount of additional data that are not readily available. For this reason, the most widely used approach is still based on the assumption of independence between the causes of death that was developed more than 50 years ago (Chiang, 1968). In this study, we want to look at the connections between the causes from a different angle. Without trying to describe exactly the dependency structure between the rates of death, we propose an approach based on cointegration analysis that complements the methods and practices mentioned above. In a nutshell, two non-stationary time series are said to be cointegrated if there exists such a linear combination of them that is stationary. Consequently, these time series are linked to each other in the long run and are subject to common stochastic trends. Cointegration analysis thus provides new insights on how cause-specific mortality rates depend from each other and interact in the long run.

Cointegration analysis was first introduced in the seminal paper of (Engle and Granger, 1987) and received a lot of attention from researchers in the years that followed. Numerous tests allowing one to check for the existence of cointegrated relations between the time variables were developed, those conceived by Søren Johansen (1988) being among the most widely used. Cointegration analysis and the Vector Error-Correction Models (VECMs) based on it quickly became popular in the field of econometrics as they permitted establishing the long-run relationships between such variables as interest rates, consumption, income etc. (e.g., Baillie and Selover 1987; Clarida 1992; Johansen and Juselius 1992).

To the best of our knowledge, cointegration analysis was first applied to the cause-specific mortality rates in Arnold and Sherris (2013, 2015, 2016). We want to go further and extend the analysis by applying a wider range of cointegration and VECM tools to the cause-of-death mortality rates. We aim to identify new relationships and development patterns which were not covered by the pre-cited authors.

Namely, we want to understand the way the cause-specific mortality rates interact between each other. Using the additional tools offered by the VECMs, we study the short- and long-term impacts that a change in a particular death rate produces in other cause-specific mortality rates. As we do not have prior knowledge about the precise way the cause-specific mortality rates interact, our study is exploratory in nature and gains new insight by observing the historical data from the perspective of cointegration

analysis. Indeed, there are multiple ways in which the cause-specific mortality rates can impact each other. On the one hand, being subject to the same trends (e.g. improving health care systems, changes in nutrition and lifestyle), the cause-specific mortality rates can show similar responses and so, be positively correlated. On the other hand, due to the presence of competing risks the reduction in one cause-specific mortality rate will necessarily and at least partially be compensated by increase in other causes and so, the cause-specific mortality rates will show negative correlation. In absence of a theoretical model for the relations between the causes, we think that the data analysis can reveal the end sum of such interactions.

At the same time, once a certain pattern is revealed in one country, it is impossible to say if this pattern is a reflection of that country's particularities or corresponds to some more fundamental processes and hence, can be generalized to other countries and datasets. For this reason, we start with the gender-specific statistics of deaths-by-cause from five highly populated countries with similar socioeconomic characteristics and available observation periods (USA, Japan, France, England and Wales, and Australia). Thanks to this approach, general common patterns are revealed in regard to the interaction existing between the causes of death. At a later point, our analysis could be extended to include other countries as well.

We see multiple ways of how our findings could be used in practice. First, the general patterns revealed by our approach can serve as a theoretical point of comparison for epidemiological studies on the joint development of cause-specific mortality rates due to particular factors, e.g., air pollution impacting not only respiratory, but also circulatory mortality rates (Zmirou et al. 1998); sedentary behavior impacting both circulatory and cancer mortality rates (Matthews et al. 2012); body mass index providing contrasting effects on circulatory and respiratory mortality rates (Breeze et al. 2006); influenza vaccinations reducing all cause-specific mortality rates (Wang et al. 2007); heat waves impacting several cause-specific mortality rates at once (Basagaña et al. 2011; Rey et al. 2007) etc. In a similar way, results of such comprehensive assessments of cause-specific mortality rates, as the Global Burden of Disease Study (GBDS 2013), can be confronted with those delivered by our model.

As previously mentioned, copula-based models are capable of taking into account the dependence between the cause-specific mortality rates. In the same time, copulas are, strictly speaking, not identifiable (Tsiatis 1975). For this reason, research articles usually present several copulas and play with different parameter values, as these choices can have a tremendous impact on the projection results (Dimitrova et al. 2013; Li and Lu 2019). Efforts are made to narrow the set of possible parameters (Li and Lu 2019) and the question of how to estimate the correlations between the causes of death remains open (Dimitrova et al. 2013). Our study provides a new basis that can be used to calibrate copula-based models as it shows explicitly the extent to which cause-specific mortality

rates depend on each other.

Additionally, we contribute to the current discussion regarding whether a cause of death should be considered as endo- or exogenous. In Arnold and Sherris (2016) the authors observed that the results of the cointegration analysis paralleled the classification used by biologists and demographers between the exogenous and endogenous causes of death. Although this classification is not univocal, under the exogenous causes of death most researchers understand diverse external or environmental factors that produce death, while the endogenous causes of death correspond to biological forces that lead to death (Carnes et al. 2006; Arnold and Sherris 2016). As different views exist on this topic (Carnes and Olshansky 1997), we bring the discussion forward by showing that only the External causes can be classified as entirely exogenous, whereas this is not the case for the infectious and parasitic diseases.

We summarize our findings in a comprehensive form with the objective to help practitioners set more informed assumptions when designing scenarios of the possible future evolution of mortality by cause.

The paper is organized as follows: in Section 1.2 we briefly present the data preparation process along with some theoretical notions of the cointegration analysis. Results from the impulse-response analysis, short- and long-term dynamics of the cause-specific mortality rates are then presented in Section 1.3. Section 1.4 concludes.

1.2 Data and the cointegration framework

1.2.1 Preparing the data

We obtained the data for the present study from the WHO Mortality Database (World Health Organization 2016) that contains the midyear population and the death numbers by country, year, sex, age group and cause of death as far back as 1950. Five developed countries were chosen for the analysis: USA, Japan, France, England and Wales, and Australia (further shortened to US, JP, FR, E&W, and AU respectively). We explicitly chose countries from different parts of the world (Americas, Europe, Asia, and Oceania) in order to have geographically representative experiences, and in every part of the world, we chose a developed country with the largest population.

To ensure consistency between the countries, the WHO defines the causes of death according to the International Classification of Diseases (ICD). This classification changed three times since the inception of the database, switching from ICD-7 to ICD-10 in order to account for advances in medical science and to refine the classification. We split the causes of death under each classification into five main groups: infectious and parasitic diseases (I&P), cancer, diseases of the circulatory system, diseases of the respiratory system and external causes (Table 1.1). These groups account for approximately 70-80%

Table 1.1: Five main groups of causes of death according to the versions of the *International Classification of Diseases*

Causes of death	ICD 7	ICD 8	ICD 9	ICD 10
I&P	001-138	001-136	001-139	A00-B99
Cancer	140-239	140-239	140-239	C00-D48
Circulatory	400-468	390-458	390-437	I00-I99
Respiratory	470-527	460-519	460-519	J00-J98
External	E810-E999	E810-E999	E800-E999	V00-Y89

of deaths in recent years and made up approximately 50%-70% of deaths at the onset of the observations.

The WHO database splits the death numbers according to the *primary* cause of death, and for this reason, we will ignore the potential presence of the secondary cause, third cause etc. Moreover, we would have to signigicanlty change our approach in order to incorporate the information on the secondary cause of death, for example. For this reason, our results would not hold in presense of several causes leading to death.

The data are divided into the following age groups: "deaths at 0 years", "at 1", "at 2", "at 3", "at 4", then into five-year age groups "5-9 years", ..., "90-94 years", and finally "deaths at 95 years and above". Having created two new age groups by grouping together the ages from 1 to 4 as well as 85 and above, we obtained the cause-specific mortality rates by following transformations:

1. Grouping the death numbers according to the five causal categories.
2. Distributing the number of deaths at unspecified age proportionally among known age groups.
3. Calculating simple mortality rates as the number of deaths by age, sex and cause divided by the mid-year population by age and sex:

$$m_{x,t,d,s,c} = d_{x,t,d,s,c}/l_{x,t,s,c},$$

with

$d_{x,t,d,s,c}$ = number of deaths at age x, in year t, for cause of death d, gender s and country c;

$l_{x,t,s,c}$ = mid-year population at age x, in year t, gender s and country c;

$m_{x,t,d,s,c}$ = central death rate at age x, in year t, for cause of death d, gender s and country c.

4. Applying the comparability ratios to ensure that the observations under the different

versions of the ICD are comparable. A comparability ratio is defined in such a way that the average of the mortality rates over the last two years of a classification coincides with the average of the mortality rates over the first two years of the next classification. For the whole period under the observation, the mortality rates in a new classification are divided by the comparability ratios linking this classification with the previous one(s). In this way, we can smooth the mortality rates across the classifications and remove the discontinuities.

5. Calculating the age-standardized central death rates, the standard population being equal to 1) the US male population in 2007; 2) the Japanese female population in 2009. In this manner, we ensure that the age structure of the population is the same for all countries and does not change over time. By using one relatively young (USA) and one relatively old (Japan) reference population, we can analyze if the population age structure has an impact on the behavior of the cause-specific mortality rates. In total, we obtain 20 datasets: 5 countries, 2 genders, and 2 population structures.

   The age-standardized death rate $m^{US}_{t,d,s,c}$ in year t for cause d, gender s and country c, assuming that the population age structure is constant over the whole observation period and is equal to the age structure of the US males population in 2007 is calculated as follows:

$$m^{US}_{t,d,s,c} = d^{US}_{t,d,s,c}/l_{2007,males,USA},$$
$$d^{US}_{t,d,s,c} = \sum_x m_{x,t,d,s,c} \times l_{x,2007,males,USA}.$$

   The age-standardized death rate $m^{JP}_{t,d,s,c}$ in year t for cause d, gender s and country c, assuming that the population age structure is constant over the whole observation period and is equal to the age structure of the JP females population in 2009 is calculated as follows:

$$m^{JP}_{t,d,s,c} = d^{JP}_{t,d,s,c}/l_{2009,females,JP},$$
$$d^{JP}_{t,d,s,c} = \sum_x m_{x,t,d,s,c} \times l_{x,2009,females,JP}.$$

   Age-standardized death rates for selected years using the US males population base are shown in table 1.7 of Appendix.

6. Taking the natural logarithm of the death rates. Hereafter we will work with the vector of time series $\mathbf{y}_t$ for each gender s, country c, and population age structure $p \in (US, JP)$:

$$\mathbf{y}^p_{t,s,c} = \begin{pmatrix} log(m^p_{t,I\&P,s,c}) \\ log(m^p_{t,Cancer,s,c}) \\ log(m^p_{t,Circulatory,s,c}) \\ log(m^p_{t,Respiratory,s,c}) \\ log(m^p_{t,External,s,c}) \end{pmatrix}.$$

To ease the notation, we will sometimes omit the indexes c, s and p, and work with a vector of mortality rates $\mathbf{y}_t = (y_{1t}, y_{2t}, y_{3t}, y_{4t}, y_{5t})^T$, keeping in mind that a separate VECM equation is formulated for each country, sex, and population age structure.

We thus use the same database as in Arnold and Sherris (2016) except for the additional years of observations that we added whenever this was possible (Table 1.2).

Table 1.2: Observation periods by country

Country	Arnold and Sherris (2016)	Current study
USA	1950 - 2007	1950 - 2007
Japan	1950 - 2009	1950 - 2013
France	1952 - 2008	1952 - 2011
England and Wales	1950 - 2009	1950 - 2013
Australia	1950 - 2004	1950 - 2004

When we started the current study, the WHO database provided the information on the mid-year population for the USA only until 2007, and for unknown reasons, the data on Australian numbers of deaths for 2005 were also missing. As a consequence, we were obliged to limit the time series for these two countries to years 2007 and 2004 respectively.

As we will see in the following sections, the conclusions stated in Arnold and Sherris (2016) were reconfirmed using the longer time series for Japan, France, and England and Wales.

1.2.2 Cointegration analysis in application to the cause-specific mortality rates

As already mentioned above, the causes of death are not independent. Cointegration analysis is then a tool that can help to understand better and model the dependence between the cause-specific mortality rates. As introduced in Engle and Granger (1987), the time series $\mathbf{y}_t$ that consist of the n non-stationary elements $\{y_{it}\}$, for $i = 1, ..., n$, are said to be cointegrated with a cointegrating vector β if a linear combination $\beta'\mathbf{y}_t$ is stationary:

$$\beta_1 y_{1t} + \beta_2 y_{2t} + ... + \beta_n y_{nt} = z_t, \tag{1.1}$$

where z_t is a stationary variable of stochastic deviations. In other words, while being non-stationary themselves, the cointegrated time series do not drift too far away from each other, i.e., there exists a long-run equilibrium relationship between them. Also, there may be more than one cointegrating vector, so that β becomes a matrix. The variables are then linked to each other by several cointegration relations, and each relation is linearly independent from the others.

In Arnold and Sherris (2015, 2016), the time series of all cause-specific mortality rates were found to be non-stationary and to have stochastic trends. It was also shown that at least one cointegrating relation existed between the causes of death in each country.

Multivariate dynamic systems of the non-stationary but cointegrated variables can then be modelled using a Vector Error Correction Model (VECM), an extension of the Vector AutoRegression (VAR) models, which includes not only the time dependency between the variables up to a lag $p-1$, but also long-run equilibrium relations between them:

$$\Delta \mathbf{y}_t = \mathbf{c} + \mathbf{d}t + \mathbf{\Gamma}_1 \Delta \mathbf{y}_{t-1} + \mathbf{\Gamma}_2 \Delta \mathbf{y}_{t-2} + \cdots + \mathbf{\Gamma}_{p-1} \Delta \mathbf{y}_{t-p+1} + \mathbf{\Pi} \mathbf{y}_{t-1} + \epsilon_t \quad [1] \tag{1.2}$$

where $\Delta \mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$ denote the first differences of the data time series, $\mathbf{c}$ and $\mathbf{d}$ are $(n \times 1)$ vectors of constants, $\mathbf{\Gamma}_i$ is a $(n \times n)$ matrix of autoregressive coefficients for $i = 1, 2, ..., p-1$, and $\mathbf{\Pi} \mathbf{y}_{t-1}$ represents the cointegrated term. The latter provides the model with the information on the long-run equilibrium between the variables that would otherwise be lost if a VAR model were applied to the differenced variables. The rank of the matrix $\mathbf{\Pi}$ corresponds to the number of cointegration relations.

The first differences of the cause-specific mortality rates being stationary[2] (as verified in Arnold and Sherris, 2016), the equation (1.2) will only hold if the term $\mathbf{\Pi} \mathbf{y}_{t-1}$ is also stationary, that is, if the variables are cointegrated. Then the $(n \times 1)$ vector ϵ_t is a vector of white noise terms[3], with

$$E(\epsilon_t) = \mathbf{0}, \tag{1.3}$$

$$E(\epsilon_t \epsilon_l) = \begin{cases} \mathbf{\Omega} & \text{for } t = l \\ \mathbf{0} & \text{for } t \neq l, \end{cases} \tag{1.4}$$

where $\mathbf{\Omega}$ is a symmetric positive definite matrix. More details on the VECM and VAR models can be found in Hamilton (1994) and Lütkepohl (2005).

[1] The corresponding VAR model has p lags: $\mathbf{y}_t = \mathbf{c} + \mathbf{d}t + \xi_1 \mathbf{y}_{t-1} + \xi_2 \mathbf{y}_{t-2} + \cdots + \xi_p \mathbf{y}_{t-p} + \epsilon_t$

[2] This observation applies to the studied times series only and should be verified for other countries and time periods.

[3] We use the Gaussian setting as this allows us to use the cointegration analysis to model the dependence between the cause-specific mortality rates. Other models such as Poisson autoregression are naturally thinkable and would bring a different and complementary perspective to the analysis when modelling death counts.

The number of the cointegrating relations, if any, can then be found using the trace and the maximum eigenvalue tests developed by Johansen (1995). The Johansen approach also allows finding the matrix $\mathbf{\Pi}$ as

$$\mathbf{\Pi} = \alpha\beta', \tag{1.5}$$

where β is a $(n \times r)$ matrix containing r vectors each representing a cointegration relation and α is a $(n \times r)$ loading matrix that indicates how a particular variable is impacted by the cointegration relation. Under the Johansen approach, we can also test for the form of the deterministic elements. Let $\mu_t = \mathbf{c} + \mathbf{d}t$ denote the deterministic part of the model and let $\mathbf{d} = \alpha\rho + \alpha_\perp\gamma$, where $\alpha\alpha_\perp = 0$. As the mortality rates are known to have a trend, we will consider the following forms of the deterministic elements (Johansen, 1995):

- NT: no trend in the VECM, but a linear trend in the levels of the variables: $\mathbf{c} \neq 0, \rho = 0, \gamma = 0,$ hence $\mathbf{d} = 0$,
- TC: linear trend in the cointegration relation combined with a linear trend in the levels of the variables (i.e., no linear trend in the differenced variables): $\mathbf{c} \neq 0, \rho \neq 0, \gamma = 0,$ hence $\mathbf{d} = \alpha\rho$,
- QT: linear trend in the differenced variables, thus the quadratic trend in the levels of the variables (i.e., the VAR model) : $\mathbf{c} \neq 0, \rho \neq 0, \gamma \neq 0,$ hence $\mathbf{d} = \alpha\rho + \alpha_\perp\gamma$.

In the tables that follow we will refer to the abbreviations NT, TC and QT when describing the form of the deterministic elements chosen for a particular dataset.

Once the coefficients of the VECM model (equation 1.2) are estimated, they allow us to assess the short- and long-term dynamics of the system. Indeed, the coefficients of the $\mathbf{\Gamma}_i$ matrices indicate if and to what extent the cause-specific mortality rates interact in the short run. On the other hand, the analysis of the coefficients of the matrices α and β provide us with the information on the long-term relationships in the system.

In particular, the Johansen approach can be used to test if every coefficient in the cointegration relation (i.e., in the matrix β) is significantly different from zero. If this is not the case, we can conclude that a particular variable does not participate in the long-run equilibrium. In Arnold and Sherris (2016) it was found that in all countries and at least for one of the sexes the pair of mortality rates corresponding to the Infectious&Parasitic diseases and the External causes did not appear significantly in the long-run equilibrium. The cointegration analysis hence showed that the long-term equilibrium relationship existed only between the mortality rates that could be classified as endogenous causes of death (Cancer, Circulatory, and Respiratory diseases), exogenous causes (Infectious&Parasitic diseases, External causes) being excluded from it. Interestingly, this result coincides with the distinction used by biologists and demographers

between the exogenous and endogenous causes of death. In this paper, we will complement this study by analyzing first, the short-term component and second, the matrix α, that is, the impact that the cointegration relation performs on a particular mortality rate.

1.2.3 Introducing the lag of 2 to the VECM setup

A usual step when working with a VECM setup is to define the lag order to be used in the VECM or the corresponding VAR model. Although in Arnold and Sherris (2016) the VAR models with the lag order of one were indicated as optimal using Akaike's Information Criterion, Hannan-Quinn Criterion, Schwarz Criterion, and Final Prediction Error [4] for some of the datasets, the corresponding model cannot be used to answer our research question. Indeed, in this case the VECM equation has no lagged values, consists only of the cointegration relation, errors and the eventual deterministic terms and implies that there is no connection between the first differences of the cause-specific mortality rates in the short run:

$$\Delta \mathbf{y}_t = \mathbf{c} + \mathbf{dt} + \mathbf{\Pi} \mathbf{y}_{t-1} + \epsilon_t. \tag{1.6}$$

Hence, in our case we need a VAR model with a lag order of at least two in order to have a full range of parameters in the VECM. So as a preliminary step, we decide to allow for the presence of the $\mathbf{\Gamma}_1 \Delta \mathbf{y}_{t-1}$ term on the right-hand side of the equation (1.2). This gives us the possibility to study the relative importance as well as the significance of the coefficients of the corresponding parameter matrix $\mathbf{\Gamma}_1$. Shall some of the matrix $\mathbf{\Gamma}_1$ coefficients turn out to be significantly different from 0, we will be able to analyze the short-run adjustments of the cause-specific mortality rates.

The models that were chosen as best describing the datasets in Arnold and Sherris (2016) comprised already the VAR(2) models for some of the countries. To be able to make the full analysis of the short-run adjustments, we check if for every dataset we can find models with the lag order of two that would suitably describe the data.

First, we apply the Johansen approach to define the number of cointegration relations and the form of the deterministic elements, then we test the residuals of the fitted VECM. The models suggested by the Johansen approach are shown in the Tables 1.8-1.11 of Appendix (second column). These are the models that will be used in the subsequent analysis of the short- and long-term dynamics of the cause-specific mortality rates. Further columns contain the results of the tests on the residuals of the fitted VECM. The

[4] These criteria are based on the natural logarithm of the determinant of the estimate of the residual covariance matrix $\hat{\Sigma}_\epsilon = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t'$, where T is the number of observations in the time series, to which a penalty for the number of parameters is added. While in the general case it is important to pay attention to the parsimony of the model, the weight of the penalty is to some extent arbitrary and can vary depending on the objectives of the study.

overall fit is similar to that of the models proposed in Arnold and Sherris (2016) except for the lower fit for the Japanese datasets. Also, it was not possible to find any suitable VAR(2) model for the E&W females with the JP females population structure, so in the following sections, we will use 19 datasets instead of 20.

For these new models, we also need to check the significance of the β matrix coefficients. As we can see in Tables 1.3 and 1.4, for 15 out of 19 considered datasets the cause-specific mortality rates corresponding to the causes I&P and External do not appear significantly (at a 1% significance level) in the long-term steady-states, which confirms the conclusion made in Arnold and Sherris (2016).

Table 1.3: p values for the null hypothesis that the I&P and the External causes of death are not significantly different from zero, US males population base, VAR(2) models

Country	Model	Males	Females
US	VAR(2), QT, 1 CR	0.0655	0.0007
JP	VAR(2), TC, 2 CR	0.4878	0.0810
FR	VAR(2), NT, 1 CR	0.1945	-
	VAR(2), QT, 1 CR	-	0.0062
E&W	VAR(2), QT, 1 CR	0.1607	0.0015
AU	VAR(2), NT, 1 CR	-	0.0438
	VAR(2), QT, 1 CR	0.2570	-

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation. A null hypothesis is accepted at a α% significance level when the p value is higher than α%.

As already mentioned, the VAR(2) models indicated in the Tables 1.3 and 1.4 will be used in the analysis that follows in section 1.3.

1.3 Dynamics of the cause-specific mortality rates

In the following sections, we present detailed analysis for the two datasets: US and JP males using the US males population structure. We summarize the most interesting findings and provide the details in the Appendix for the remaining 17 datasets.

1.3.1 Impulse-response analysis

First, to get a high-level overview of interactions between the cause-specific mortality rates (as described by the VECM equations built in the preceding chapter) we apply the

Table 1.4: p values for the null hypothesis that the I&P and the External causes of death are not significantly different from zero, JP females population base, VAR(2) models

Country	Model	Males	Females
US	VAR(2), QT, 1 CR	0.0173	0.0000
JP	VAR(2), QT, 1 CR	0.0530	0.1906
FR	VAR(2), NT, 1 CR	0.0999	0.0696
E&W	VAR(2), QT, 1 CR	0.0788	No model
AU	VAR(2), NT, 1 CR	0.1917	-
	VAR(2), QT, 1 CR	-	0.0691

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation. A null hypothesis is accepted at a α% significance level when the p value is higher than α%.

framework of impulse-response analysis (see, e.g., Lütkepohl, 2005). At this point, we do not differentiate between the short- and long-term elements of the VECM, and analyze the system as a whole. More detailed analysis of the short- and long- term components including the statistical significance of the parameters will follow.

Basically, impulse-response analysis means that we first give a single shock to one cause-specific mortality rate and then analyze and compare the responses to this shock from every other cause-specific mortality rate. In this way, we study the impact of an unexpected change in a particular mortality rate on the dynamics of the system of mortality rates as this was observed in the past. The initial value taken by the variable that receives the shock is equal to its own standard deviation.

Our approach supposes that the event that induces a sudden change in the value of the cause-specific mortality rate does not disappear in the following years and that its influence remains. For example, should the Cancer mortality rate go down due to a new medicine, this medicine remains available and maintains its inhibiting influence on the Cancer mortality rate. A different approach should be taken if one wants to model a temporary impact of an event. For example, the coronavirus pandemics has lead to a rapid increase in the I&P mortality rate, but as we all hope, this impact will not last forever. Should a vaccine or an effective treatment be found, the impact of the coronavirus on the I&P mortality rate will substantially go down or even fully disappear. In order to accommodate such "back-to-normal" return in a VECM setting, one could think of adding a shock of the opposite sign few years later after the event that caused the first shock. As such analysis surpasses the compass of the present thesis, in what follows we will study the impact of the long-lasting events only.

When analyzing the results, we successively adopt two points of view. First, we compare the impacts that a particular cause induces on other cause-specific mortality rates. Then, we compare the responses of a particular cause to the individual shocks received from the rest of mortality rates. In this way, we are able to determine not only if a particular cause influences the others and to what extent, but also if it is influenced by the rest of the causes and to what extent.

Once a shock is given to a particular cause-specific mortality rate, it propagates in the system and confers new values to the rest of the variables. This development can be suitably exposed on a chart. For example, Figure 1.1 shows the responses of every cause-specific mortality rate to the shock given to the Circulatory mortality rate for US males with the US males population structure (standard deviation of the differenced Circulatory mortality rate = 0.0235). Overall, the I&P, as well as the Respiratory mortality rates, show the most important reactions to the shock given to the Circulatory mortality rate. The Cancer and the External mortality rates are insignificantly impacted by the shock given to the Circulatory mortality rate. As for the Circulatory mortality rate itself, after having received the initial shock, it maintains the increased value until the end of the simulation period.

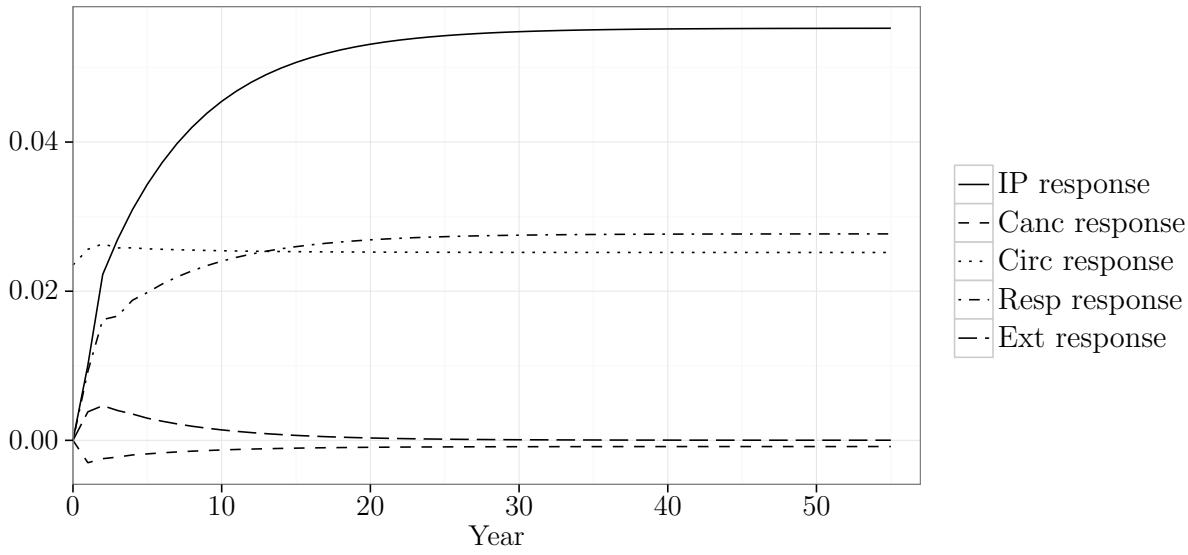


Figure 1.1: Responses to the shock given to the Circulatory cause, US males, US males population base

The responses to the shocks from the Circulatory cause observed in the dataset for the JP males with the US males population structure are shown on Figure 1.2 (standard deviation of the differenced Circulatory mortality rate = 0.0501). Like the US males dataset, the Respiratory cause shows the most important response from the shock given to the Circulatory mortality rate. The response of the I&P mortality rate is slightly less

important than that of the Respiratory rate. Interestingly, both responses have a negative sign whereas in the US males dataset they also have the same sign, but a positive one. One further observation for the JP males dataset is that the External causes also show a non-negligible response to the shock from the Circulatory cause.

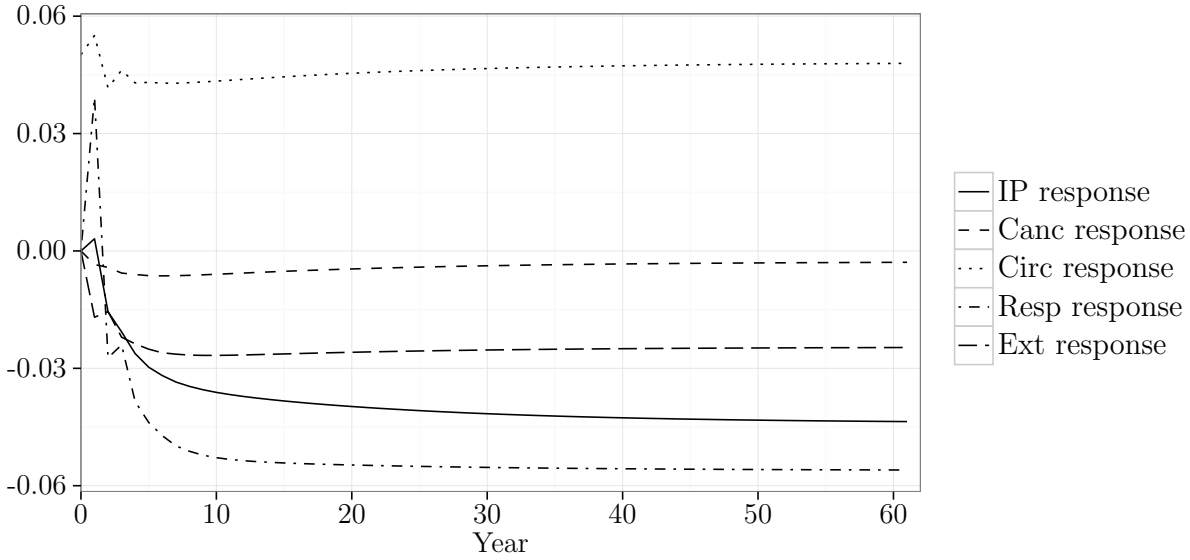


Figure 1.2: Responses to the shock given to the Circulatory cause, JP males, US males population base

We see that in both cases the system comes rather quickly to a new equilibrium. As the same observation holds for the rest of the datasets, we will compare the responses following individual shocks at time t = 20 years.

On Figures 1.1 and 1.2 the responses are shown in absolute values. However, since the cause-specific mortality rates have different standard deviations, each system receives a shock of a different amplitude. As such, the responses are not comparable between the datasets, that is, a response that would be considered high in one dataset can be considered as medium or low in another dataset. To bring the results to the same comparable basis, we will divide the absolute responses by the standard deviation of the cause-specific mortality rate that receives the shock. Then the response of the Respiratory cause to the shock from the Circulatory cause, i.e., the value of the cause-specific Respiratory mortality rate at time t = 20 will be expressed as a proportion of the shock received by the system, i.e., of the standard deviation of the cause-specific Circulatory mortality rate.

The results for every dataset are shown in Tables 1.12 and 1.13 of the Appendix. For the sake of readability, along with a numerical value we provide a label that indicates if the response can be considered as low, medium, or high: low if $|response| < 3/8$, medium if $3/8 \leq |response| < 7/8$, and high if $7/8 \leq |response|$. These labels are indicative only and were chosen to provide a roughly equal number of responses "medium" and

"high" (40% of all responses), the rest being attributed to the category "low" (60% of all responses). The tables are organized as follows: Each row contains responses of all causes to the shock given to the cause X, each column contains responses of the cause Y to the shocks from all causes. In this way, we can judge simultaneously if a particular cause impacts each of the remaining causes and if it reacts to the shocks received from other causes and to what extent.

A synopsis of the observations summarized across the 19 datasets is presented in Table 1.5.

Table 1.5: Impulse-response analysis: response of the mortality rate Y from the shock given to the cause X, high-level summary across all countries, sexes and population structures

X \ Y	I&P	Cancer	Circulatory	Respiratory	External
I&P		Low	Low	Low	Low
Cancer	High		Med	High	High
Circulatory	High	Low		High	Med
Respiratory	Med	Low	Low		Low
External	High	Low	Low	High	

In a nutshell:

- The I&P and the Respiratory causes have virtually no impact on all other causes, but show important responses to the shocks received from them.
- The Cancer and the Circulatory mortality rates have an important impact on other causes, especially on the I&P and the Respiratory mortality rates, but show little response to the shocks from other causes.
- The External causes have an equivocal behavior. On the one hand, they have almost no impact on the Cancer and the Circulatory causes, but importantly impact the I&P and the Respiratory causes. On the other hand, they are not impacted by the I&P and the Respiratory causes, but show important responses to the shocks from the Cancer and the Circulatory causes.

This first analysis shows that the cause-specific mortality rates have different behaviors. In the same time, when a system is analyzed as a whole, many effects are necessarily blended. Therefore, we need to decompose our analysis by separately assessing the short- and long-term dynamics of the system of the mortality rates to understand better how the causes of death are related to each other. In the following subsections, we will see what

drivers in particular lie behind the observed development of the cause-specific mortality rates.

1.3.2 Short-term dynamics

Once the VECM equations are estimated for each dataset, we can use them to separate the short-term adjustments from the long-term dynamics for each cause-specific mortality rate. Indeed, if a particular coefficient γ_{ij} of matrix $\mathbf{\Gamma}_1$ is significant, then cause i is influenced by cause j in the short run. We calculate the standard deviations and the corresponding t-ratios of the estimates as shown in Lütkepohl (2005).

We start with the dataset for US males with the US males population structure. In the preceding section the following model was chosen as best describing this dataset:

$$\Delta \mathbf{y}_t = \mathbf{c} + \mathbf{dt} + \mathbf{\Gamma_1} \Delta \mathbf{y}_{t-1} + \mathbf{\Pi} \mathbf{y}_{t-1} + \epsilon_t =$$

$$= \begin{bmatrix} \mathbf{-2.7716} \\ -0.1500 \\ -0.2428 \\ \mathbf{-2.0573} \\ 0.1558 \end{bmatrix} + \begin{bmatrix} \mathbf{0.0096} \\ 0.0000 \\ 0.0000 \\ \mathbf{0.0050} \\ 0.0000 \end{bmatrix} \mathbf{t} + \begin{bmatrix} -0.1212 & -0.7074 & -0.1879 & 0.1743 & 0.3232 \\ -0.0036 & 0.0150 & \mathbf{-0.1659} & -0.0082 & \mathbf{0.1329} \\ -0.0429 & -0.1211 & 0.0337 & -0.0892 & 0.1963 \\ -0.1345 & -0.2944 & -0.0850 & \mathbf{-0.3810} & \mathbf{1.1189} \\ 0.0416 & -0.3232 & 0.1981 & \mathbf{-0.1455} & 0.2267 \end{bmatrix} \Delta \mathbf{y}_{t-1}$$

$$+ \begin{bmatrix} \mathbf{-0.0331} \\ -0.0020 \\ -0.0030 \\ \mathbf{-0.0257} \\ 0.0019 \end{bmatrix} \begin{bmatrix} 1.7716 & -5.4985 & -18.6015 & 13.2167 & 14.1321 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$$

The significant coefficients are in bold with the selected significance level of 5%. While many of the $\mathbf{\Gamma}_1$ coefficients are not significant, the cause-specific mortality rates from Cancer, Respiratory, and External causes are influenced by the lagged values of Circulatory and External, Respiratory and External, and Respiratory causes respectively. We see that in this dataset three out of five cause-specific mortality rates experience the short-term adjustments from other causes. Hence, it was justified to use the VAR(2) setup and include the lagged values of $\Delta \mathbf{y}_t$ into the model. Otherwise, an essential piece of information on the development of the cause-specific mortality rates would not have been accounted for. Another interesting observation is that only the Respiratory mortality rate shows the autoregressive feature. In other words, the corresponding cause-specific mortality rate is dependent on the lagged value of itself.

As for the dataset of JP males with the US males population structure, the chosen VECM has two cointegration relations with a constant and a trend, the latter being restricted to the cointegration term:

$$\Delta \mathbf{y}_t = \mathbf{c} + \mathbf{\Gamma_1} \Delta \mathbf{y}_{t-1} + \alpha\beta'(\mathbf{y}_{t-1} + (t-1)) + \epsilon_t =$$

$$= \begin{bmatrix} 0.3960 \\ \mathbf{-0.8539} \\ -0.7600 \\ -1.8252 \\ -0.7695 \end{bmatrix} + \begin{bmatrix} -0.1679 & 0.5510 & 0.1982 & 0.0335 & 0.2440 \\ 0.0143 & \mathbf{-0.3818} & -0.0268 & 0.0184 & 0.0406 \\ 0.1343 & 0.7531 & 0.1711 & \mathbf{-0.1692} & 0.0055 \\ 0.1477 & 2.1179 & \mathbf{1.3108} & \mathbf{-0.5537} & -0.1487 \\ 0.0307 & 0.3048 & -0.2898 & 0.1305 & -0.0906 \end{bmatrix} \Delta \mathbf{y}_{t-1}$$

$$+ \begin{bmatrix} \mathbf{0.0261} & \mathbf{0.3600} \\ \mathbf{-0.0186} & \mathbf{-0.0425} \\ -0.0126 & 0.0399 \\ 0.0054 & \mathbf{0.8417} \\ -0.0154 & -0.0149 \end{bmatrix} \left[\begin{bmatrix} 1.5951 & 7.7055 & 0.9876 & -1.5822 & 1.8454 \\ -1.0848 & 9.3172 & -10.1473 & -6.6630 & -3.7839 \end{bmatrix} \mathbf{y}_{t-1} + \begin{bmatrix} 0.1851 \\ -0.3817 \end{bmatrix} (t-1) \right]$$

$$+ \boldsymbol{\epsilon}_t$$

Also for this dataset, many of the $\mathbf{\Gamma}_1$ coefficients are not significant. On the other hand, the cause-specific mortality rates corresponding to the causes Cancer, Circulatory, and Respiratory causes are influenced by the lagged values of the Cancer, Respiratory, Circulatory and Respiratory mortality rates respectively. Again, three out of five cause-specific mortality rates experience the short-term adjustments from other causes. Therefore, it would not be justified to use the VAR(1) setup for JP males with the US males population structure. Like the US males dataset, the Respiratory cause shows the autoregressive feature as well as the Cancer cause.

After the analysis was repeated for the rest of the datasets using both the US males and the JP females population structures, the results can be summarized as follows:

- In every dataset, there is at least one cause-specific mortality rate that is significantly impacted by other causes in the short run. For this reason, it would not be optimal to use a model which does not account for short-term interactions, i.e. has no $\mathbf{\Gamma}_i$ matrix. This case would correspond to a VAR model with the lag order one or, put differently, to a VECM model with a lag order zero. For this reason, we will use VAR models with a lag of two throughout our analysis.

- While in the short run the I&P and Cancer causes are rarely impacted by other causes, they also infrequently impact the rest of the causes, i.e., they show a development mostly independent from other causes in the short run.

- On the other hand, the Circulatory, Respiratory, and the External causes are frequently impacted by one or more causes in the short run and also occasionally

impact other causes. Hence, these cause-specific mortality rates are more linked in their development to other causes than the I&P and Cancer mortality rates are.

- The Respiratory cause consistently shows the autoregressive feature. In other words, in many datasets the corresponding cause-specific mortality rate is dependent on the lagged value of itself.

- For all datasets, the larger part of the significant coefficients are negative, i.e., more often than not the change in the cause-specific mortality rate goes in the opposite direction of the short-term variation of this and/or other cause-specific mortality rates at the previous point in time. More specifically, this means that if the mortality rate of a particular cause of death increases (decreases), the other causes will tend to decrease (increase) in the short run.

The detailed overview of the significant coefficients in $\mathbf{\Gamma}_1$ matrix for each dataset is presented in the tables 1.14 to 1.17 of the Appendix.

1.3.3 Long-term dynamics

The α matrix allows us to estimate how deviations from the steady-states impact the cause-specific mortality rates. For $r = 1$ (which is the case for the majority of the datasets), we can write the long-term component as follows:

$$\mathbf{\Pi y}_{t-1} = \alpha\beta'\mathbf{y_{t-1}} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix} \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \end{bmatrix} \mathbf{y}_{t-1} =$$

$$= \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix} (\beta_1 y_{1t-1} + \beta_2 y_{2t-1} + \beta_3 y_{3t-1} + \beta_4 y_{4t-1} + \beta_5 y_{5t-1}) \tag{1.7}$$

This way, if a particular coefficient α_i is significant, the long-term component on the right-hand side of the equation (1.2) is important in explaining the past variations of the corresponding cause-specific mortality rate on the left-hand side. Moreover, the value of this coefficient shows the extent to which the long-term component contributes to the variation of the cause-specific mortality rate in question. As in the previous subsection, we calculated the standard deviations and the corresponding t-ratios of the estimates of

Table 1.6: p values for the null hypothesis that α_i is not significantly different from zero, US males population structure

Country/Sex	Model		ΔIP_t	$\Delta Canc_t$	$\Delta Circ_t$	$\Delta Resp_t$	ΔExt_t
US Males	VAR(2), QT, 1 CR	α_i	0.000	0.085	0.243	0.001	0.496
JP Males	VAR(2), TC, 2 CR	α_{1i}	0.026	0.000	0.183	0.837	0.182
		α_{2i}	0.000	0.001	0.500	0.000	0.836

α as shown in Lütkepohl (2005). Table 1.6 shows the p-values for the US males and JP males datasets using the US males population structure.

On the one hand, this example shows that for the US males dataset the long-term component enters the equations for the I&P and Respiratory mortality rates with a significant coefficient (at a 5% significance level). On the other hand, the equations for Cancer, Circulatory, and External mortality rates are not significantly impacted by the long-term equilibrium. As for JP males, there are two long-term components that each enter with a significant coefficient the equations for the I&P and Cancer mortality rates; also, the second component enters with a significant coefficient the equation for the Respiratory mortality rate.

We repeated the analysis for the rest of the datasets, and the results can be summarized as follows:

- The I&P and the Respiratory causes seem to be the most impacted by the long-term component: The corresponding α_i coefficients are significant in 15 out of 19 datasets. A similar observation was made using the framework of the impulse-response analysis, as there the I&P and the Respiratory mortality rates showed an important reaction to the shocks from other causes. Hence, these shocks propagate in the system via the cointegration relation(s).

- The External causes seem to be the least impacted: The corresponding α_i coefficients are significant in only 5 out of 19 datasets. Interestingly, the results of the impulse-response analysis for the External causes were equivocal in that there was an important reaction to the shocks from the Cancer and the Circulatory causes, but a low response to the shocks from the I&P and the Respiratory causes. The impact from the Cancer and the Circulatory causes may hence come from the short-term adjustments.

- Results for the Cancer and the Circulatory causes are more difficult to interpret: The corresponding α_i coefficients are significant in respectively 9 and 11 out of 19 datasets. The results of the impulse-response analysis also showed low reactions of the Cancer and the Circulatory mortality rates to the shocks from other causes.

Interestingly, while as was mentioned above, the I&P and the External causes do not participate conjointly in the long-term equilibrium, they show different behaviors when it comes to the impact they experience from this long-term steady-state. Indeed, the cointegration relations often enter the equation for the I&P mortality rate with a significant coefficient, but seldom have an effect on the External causes. Therefore, only the External causes show behavior that is entirely independent from the long-term equilibrium state and, possibly, aging.

The overview of the results for the remaining datasets is shown in the Tables 1.18 and 1.19 of the Appendix.

1.3.4 Long-term vs. short-term dynamics

In the previous sections, we have analyzed the short- and long-term elements separately. Now we want to assess the relative importance of the long- and short-run forces. For this purpose, we break down the expected cause-specific mortality rates at time t, based on the information available up to time t-1, in two elements: the short-term (ST) and the long-term (LT) components. By comparing the behavior of each of these elements with the realized change in the mortality rates, we assess the relative importance of the long- and short-run forces in terms of their contribution to the variation of the cause-specific mortality rates.

For illustrative purposes, we present the results for the Respiratory equation for US males (Figure 1.3) and the I&P equation for JP males (Figure 1.4), both datasets using the US males population structure. In the first case, the actual mortality changes fluctuate primarily with the short-term components (the correlation coefficient between $\Delta Resp_t$ and the LT: 0.268, between $\Delta Resp_t$ and the ST: 0.336):

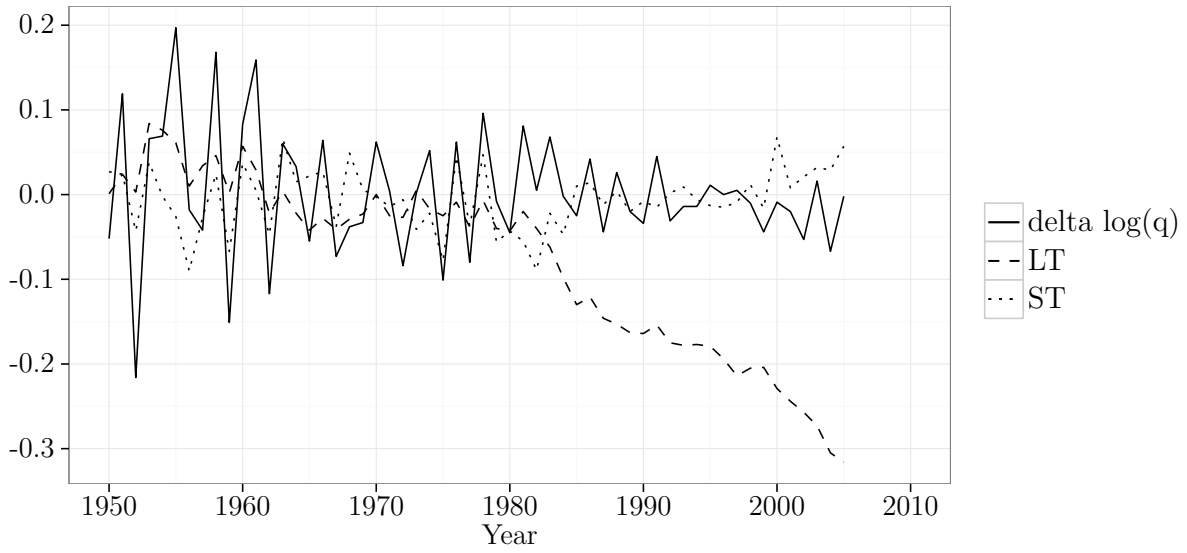


Figure 1.3: Respiratory cause: actual mortality changes, long- and short-term components (US males, US males population structure)

As for JP males, the actual mortality changes fluctuate primarily with the long-term component (the correlation coefficient between ΔIP_t and the LT: 0.570, between ΔIP_t and the ST: -0.177).

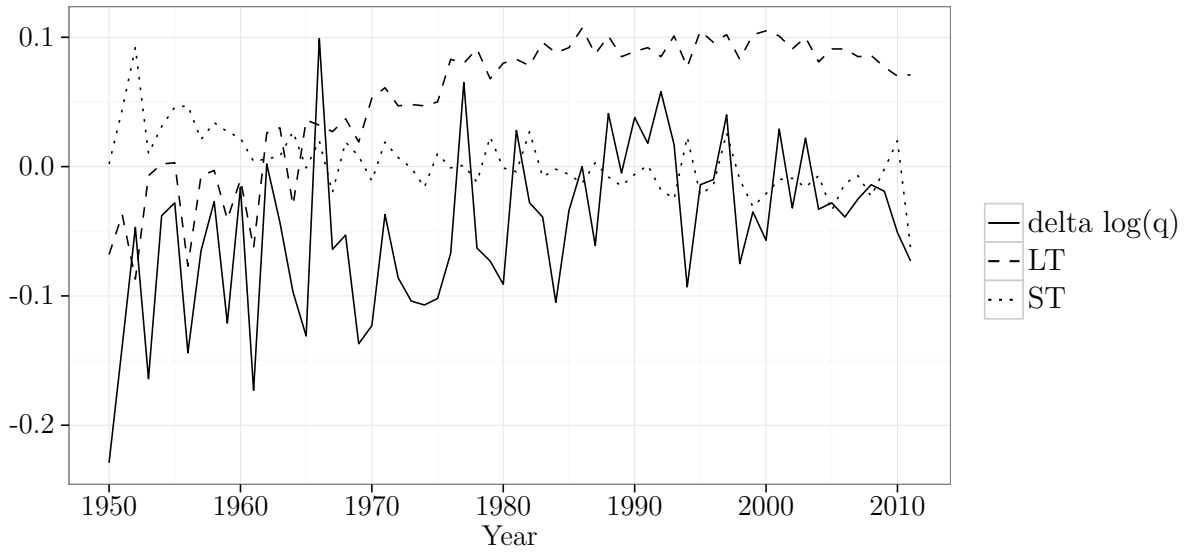


Figure 1.4: I&P cause: actual mortality changes, long- and short-term components (JP males, US males population structure)

As not every equation contains the long-term component, the results for the rest of the datasets are formulated for those cases where the long-term component is present with a significant coefficient α_i.

- Out of 15 datasets for which the I&P mortality rate equation contains the long-term component, in 13 cases the data fluctuates primarily with the long-term component (i.e., the correlation coefficient between the data points and the long-term component is higher than that between the data points and the short-term component).
- Out of 9 datasets for which the Cancer mortality rate equation contains the long-term component, in 8 cases the data fluctuates primarily with the long-term component.
- Out of 11 datasets for which the Circulatory mortality rate equation contains the long-term component, in 4 cases the data fluctuates primarily with the long-term component.
- Out of 15 datasets for which the Respiratory mortality rate equation contains the long-term component, in 7 cases the data fluctuates primarily with the long-term component.
- Out of 5 datasets for which the External mortality rate equation contains the long-term component, only in 1 case does the data fluctuate primarily with the long-term component.

Summarizing the results stated above, we can say that every time the equation contains the long-term component, the cause-specific mortality rate resembles in its behavior the long-term component rather than the short-term one for the causes I&P and Cancer. The opposite is true for the Circulatory and External causes. The Respiratory mortality rate resembles in its behavior the long-term component as often as it resembles the short-term component.

This observation is not surprising for the I&P and the Respiratory causes. As we have seen in previous sections, only these cause-specific mortality rates are often impacted by the long-term equilibrium state. In the same time, the I&P mortality rate is rarely, and the Respiratory is frequently impacted by the short-term component. It could have been expected that the I&P mortality rate data will fluctuate with the long-term component in the majority of cases where the I&P mortality rate contains the cointegration relation(s). In its turn, the Respiratory mortality rate fluctuates either with the short-term component or with the long-term component in roughly similar proportions. Therefore, the correlation analysis reinforces the conclusions of the previous sections for these two rates.

A similar conclusion holds for the Circulatory mortality rate: as it is frequently impacted in the short run and only occasionally in the long, the short-term components play unsurprisingly a more important role in the correlation analysis.

Regarding the Cancer mortality rate, we have seen that it was infrequently impacted by both short- and long-term components. As none of the effects can be called dominant,

the correlation analysis helps to identify the component that plays a more important role in the development of this cause-specific mortality rate, in this case, the long-term.

As for the External causes, the corresponding cause-specific mortality rate is often impacted by the short-term component and virtually never by the long-term one. Even in those rare cases in which the cointegration relation enters the equation with a significant coefficient, the data fluctuate more with a short-term component.

The detailed results for each dataset are shown in Tables 1.20-1.23 of the Appendix.

1.4 Conclusions

The analysis of dynamics of the cause-specific mortality rates shows that they are dependent on each other in both short- and long-run. Although the observed experience will never exactly repeat itself in the future, the following observations can help practitioners set more informed assumptions on the future development of mortality rates:

- The common long-run trend shared by the cause-specific mortality rates is contingent on the evolution of the Cancer, the Circulatory, and the Respiratory mortality rates, as these are the causes that significantly contribute to the cointegration relation between the mortality rates.

- Once the common long-run trend is defined, it more heavily impacts the development of the I&P and the Respiratory mortality rates and to a lesser extent the development of the Cancer and the Circulatory mortality rates. The External causes are exempt from the influence of the common long-term relationship between the causes.

- In the short run, the Respiratory mortality rate consistently shows the autoregressive feature.

- Although the short-run dependencies are more challenging to model, they are significantly pronounced in the development pattern of the Circulatory, Respiratory, and the External mortality rates. In other words, these rates are dependent on each other in the short run.

Coming back to the conclusion made in Arnold and Sherris (2016) that the I&P and the External causes do not participate conjointly in the long-term steady-state, we see that these causes differ in the way they are impacted by the long-term equilibrium. Though the I&P mortality rate is often impacted by the cointegration relation(s) and when it is, fluctuates more with the long-term component, the External causes show the opposite behavior: the corresponding rate is almost never impacted by the cointegration relation(s), and when it is, it fluctuates more with the short-term component.

We see that the development of the External causes mortality rate is completely independent from the long-term equilibrium both in terms of the contribution to, and influence experienced from, the steady state. This is a behavior of what could be called a genuinely exogenous cause of mortality as we observe no long-term impact to or from this cause. It develops in a way that is entirely independent of the observed equilibrium between the rest of the cause-specific mortality rates and is subject to only short-term shocks from other causes. Basically, this observation is not surprising, as under the External causes are grouped such causes as transport and other accidents (falls, poisoning, accidental fire, drowning), suicides, homicides, and war injuries. So, it is rather difficult to imagine a link connecting these mortality rates to the rest of the mortality causes that could be observed over a long time. On the contrary, these causes of mortality can rather be characterized by randomness and "bad luck" rather than by a steady long-term development.

In turn, the I&P mortality rate does not influence the long-term equilibrium observed between the cause-specific mortality rates but is rather sensitive to the impacts received from this equilibrium. Occasionally, it is also subject to short-term shocks from other causes. We can conclude that while the evolution of the I&P mortality rate does not influence the development of other cause-specific mortality rates, i.e., a sudden increase or drop in the I&P mortality rate will not affect the rest of the cause-specific mortality rates, its own development depends to a great extent from other causes of death, especially in the long run. Such behavior cannot be described as fully independent, and so the I&P cause cannot be classified as a truly exogenous cause in the same way the External causes can.

These observations are consistent with the intuition that the biological processes of aging are reflected in the common stochastic trend shared by the cause-specific mortality rates. Indeed, while it can seem that the infectious or parasitic diseases are similar to the external causes in that the origin of the force affecting the human body lies outside, the underlying biological processes are more complicated, as human beings are not equal when they face an infection. Even during severe epidemics, the probability of getting sick and dying depends to a large extent on the internal immune forces of the person, which in turn, depend, among other factors, on age. A well-known example is influenza that is the most dangerous for the elderly. When advancing in age, we are more and more confronted with competing risks such that a decrease in mortality from the circulatory diseases, for example, would leave more vulnerable persons alive who could then die from an infectious disease during an epidemic. It is then understandable that the I&P mortality rate, while not being a part of the long-term equilibrium, is substantially affected by it. Our results then confirm and reinforce the link between the cointegration relations observed within a set of cause-specific mortality rates and the biological processes of aging.

One further possible application of the present study is the calibration of copula-based

models for the cause-specific mortality rates that remains an open question. Indeed, due to the identifiability issue raised by Tsiatis (1975), one usually has to assume that the dependence is represented by a known copula with known parameters. In the same time, copula-based models are highly sensitive with respect to these choices (Dimitrova et al. 2013). As pointed out in Kaishev et al. (2007), the free parameters could be set according to a priori available (medical) information, about the degree of pairwise dependence between the two competing risks, expressed through Kendall's τ and/or Sperman's ρ. In the absence of additional information and to demonstrate the sensitivity of the results, the authors use four different copulas and five different values of Kendall's τ, ranging from -0.91 (extreme negative dependence) to 0.91 (extreme positive dependence). In Li and Lu (2019), the authors go further and by introducing hierarchical Archimedean copulas succeed in building a model that allows for different levels of association between the causes of death. For this, they group the causes in different clusters based on the (assumed) level of dependence between the causes, but also admit that the introduced hierarchical structure is not unique. Although our study cannot provide an exact value of parameters to be used in copula-based models, a certain degree of pairwise association (correlation) between the causes of death can be inferred from the results of the impulse-response analysis (Section 1.3.1). This could help researchers working with copula-based models further reduce the possible range of free parameters that otherwise have to be chosen arbitrarily. Also, the revealed differences in the long- and short-term development of the cause-specific mortality rates can serve as the basis for building clusters of the causes.

In the current study, we limited our analysis to the total cause-specific mortality rates and did not differentiate by age. Yet, it is intuitively clear that when analyzed by age, the mortality rates will present different development patterns. As the cause-specific death numbers are available in the WHO database by five-year age groups, it seems to be a promising path to integrate the age specifics of the mortality rates into the modelling process. However, this remains a challenging task, as, on the one hand, the cointegration tests have been developed for systems with maximum 12 variables (Osterwald-Lenum 1992), and, on the other, the observation horizon, which goes back as far as 1950, is also rather brief. In our opinion, analysis trying to overcome these difficulties while preserving the information on the age profile has the potential to deliver additional insights on the interaction of the cause-specific mortality rates. Moreover, the biological processes of aging may probably be easier to measure once the data on young ages are excluded from the analysis, as by definition, the aging risk factor becomes more important the longer we live. For this reason, the analysis of the cause-specific mortality rates excluding young ages may provide a better measure of the aging process.

Appendix

Table 1.7: Age-standardized central death rates for selected years, US males population base, x10^3

		Females			Males		
		1960	1980	2000	1960	1980	2000
US	I&P	0.0852	0.0630	0.1014	0.1888	0.1215	0.1686
	Cancer	1.6269	1.5857	1.5786	2.1394	2.4409	2.2483
	Circulatory	4.3148	2.9428	1.9744	6.8425	5.2315	3.2263
	Respiratory	0.4595	0.3292	0.5124	0.8670	0.8205	0.8512
	External	0.4516	0.3804	0.3058	1.1217	1.0368	0.7731
JP	I&P	0.4350	0.0842	0.0629	0.8160	0.2099	0.1426
	Cancer	1.4299	1.3156	1.1361	2.0147	2.1748	2.1092
	Circulatory	1.6850	1.3904	0.6263	2.2670	1.8094	0.8788
	Respiratory	0.9022	0.3804	0.3551	1.3673	0.6830	0.7941
	External	0.4644	0.3230	0.2476	1.0982	0.7577	0.7161
FR	I&P	0.1759	0.1085	0.0833	0.4487	0.2171	0.1502
	Cancer	1.6852	1.5080	1.3868	2.5972	3.0608	2.8610
	Circulatory	2.1922	1.6042	0.9415	3.3823	2.7013	1.6697
	Respiratory	0.8405	0.3023	0.3385	1.3661	0.6451	0.5930
	External	0.5020	0.5267	0.3623	1.1683	1.1520	0.8470
EW	I&P	0.0721	0.0387	0.0557	0.1764	0.0636	0.0708
	Cancer	1.7114	1.7764	1.5838	2.6498	2.8131	2.2807
	Circulatory	3.7128	2.7551	1.4769	5.7063	4.9203	2.5757
	Respiratory	0.8536	0.9398	0.8048	1.9318	1.8931	1.2785
	External	0.3823	0.3044	0.1780	0.6787	0.5561	0.4205
AU	I&P	0.0695	0.0296	0.0419	0.1518	0.0535	0.0843
	Cancer	1.4781	1.4868	1.4368	2.0721	2.3765	2.1570
	Circulatory	4.0123	3.0032	1.5502	6.5525	5.1796	2.4928
	Respiratory	0.5108	0.4362	0.6109	1.1511	1.1897	1.0071
	External	0.4747	0.4236	0.2625	1.0533	0.9189	0.6118

Table 1.8: Tests on residuals of the fitted VECM, males, US males population base, VAR(2) models

		p value					
		Autocorrelation			Normality		
Country	Model	15 lags	25 lags	35 lags	Skewness	Kurtosis	Both
US	VAR(2), QT, 1 CR	0.212	0.146	0.181	0.546	0.020	0.066
JP	VAR(2), TC, 2 CR	0.526	0.764	0.810	0.749	0.002	0.015
FR	VAR(2), NT, 1 CR	0.354	0.644	0.824	0.467	0.154	0.244
E&W	VAR(2), QT, 1 CR	0.209	0.146	0.207	0.510	0.558	0.607
AU	VAR(2), QT, 1 CR	0.594	0.297	0.405	0.876	0.002	0.027

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.9: Tests on residuals of the fitted VECM, females, US males population base, VAR(2) models

		p value					
		Autocorrelation			Normality		
Country	Model	15 lags	25 lags	35 lags	Skewness	Kurtosis	Both
US	VAR(2), QT, 1 CR	0.133	0.007	0.021	0.768	0.141	0.369
JP	VAR(2), TC, 2 CR	0.539	0.601	0.769	0.007	0.000	0.000
FR	VAR(2), QT, 1 CR	0.066	0.286	0.491	0.420	0.038	0.080
E&W	VAR(2), QT, 1 CR	0.389	0.307	0.353	0.025	0.000	0.000
AU	VAR(2), NT, 1 CR	0.238	0.284	0.262	0.652	0.059	0.175

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.10: Tests on residuals of the fitted VECM, males, JP females population base, VAR(2) models

		p value					
		Autocorrelation			Normality		
Country	Model	15 lags	25 lags	35 lags	Skewness	Kurtosis	Both
US	VAR(2), QT, 1 CR	0.496	0.202	0.202	0.879	0.189	0.511
JPn	VAR(2), QT, 1 CR	0.168	0.103	0.081	0.225	0.047	0.052
FR	VAR(2), NT, 1 CR	0.398	0.615	0.534	0.139	0.009	0.009
E&W	VAR(2), QT, 1 CR	0.225	0.094	0.099	0.722	0.556	0.743
AU	VAR(2), NT, 1 CR	0.521	0.397	0.437	0.407	0.013	0.034

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.11: Tests on residuals of the fitted VECM, females, JP females population base, VAR(2) models

		p value					
		Autocorrelation			Normality		
Country	Model	15 lags	25 lags	35 lags	Skewness	Kurtosis	Both
US	VAR(2), QT, 1 CR	0.168	0.005	0.021	0.123	0.010	0.009
JP	VAR(2), QT, 1 CR	0.469	0.625	0.664	0.000	0.000	0.000
FR	VAR(2), NT, 1 CR	0.096	0.426	0.381	0.324	0.097	0.127
E&W	No suitable VAR(2) model						
AU	VAR(2), QT, 1 CR	0.548	0.688	0.353	0.488	0.046	0.108

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.12: Impulse-response analysis: response of the mortality rate Y from the shock given to the cause X, US males population base

Cause		Males					Females				
X \ Y	Country	I&P	Canc	Circ	Resp	Ext	I&P	Canc	Circ	Resp	Ext
I&P	US	0.53 (Med)	-0.01 (Low)	-0.02 (Low)	-0.2 (Low)	0.1 (Low)	-0.05 (Low)	-0.12 (Low)	-0.11 (Low)	-0.47 (Med)	0.24 (Low)
I&P	JP	0.97 (High)	-0.13 (Low)	-0.09 (Low)	-0.12 (Low)	-0.14 (Low)	0.58 (Med)	0.03 (Low)	0.13 (Low)	0.01 (Low)	0.01 (Low)
I&P	FR	1.15 (High)	-0.17 (Low)	-0.07 (Low)	0.38 (Med)	-0.21 (Low)	1.2 (High)	-0.03 (Low)	0.07 (Low)	0.22 (Low)	-0.18 (Low)
I&P	E&W	1.59 (High)	-0.14 (Low)	-0.09 (Low)	-0.33 (Low)	-0.38 (Med)	0.37 (Low)	-0.03 (Low)	-0.08 (Low)	-0.33 (Low)	0.17 (Low)
I&P	AU	0.73 (Med)	0.01 (Low)	-0.02 (Low)	-0.12 (Low)	-0.03 (Low)	-0.1 (Low)	0.01 (Low)	0.16 (Low)	-0.1 (Low)	0.2 (Low)
Canc	US	2.05 (High)	1.13 (High)	-0.15 (Low)	0.75 (Med)	-0.78 (Med)	3.39 (High)	1.64 (High)	-0.61 (Med)	0.8 (Med)	-1.78 (High)
Canc	JP	1.06 (High)	0.24 (Low)	-0.27 (Low)	0.75 (Med)	-0.26 (Low)	4.37 (High)	0.5 (Med)	0.03 (Low)	3.38 (High)	0.98 (High)
Canc	FR	1.46 (High)	0.16 (Low)	-0.92 (High)	0.64 (Med)	-0.86 (Med)	2.27 (High)	0.36 (Low)	0.15 (Low)	4.56 (High)	-0.45 (Med)
Canc	E&W	1.61 (High)	0.58 (Med)	-0.1 (Low)	-2.57 (High)	-1.42 (High)	-4.09 (High)	0.7 (Med)	-0.69 (Med)	-2.96 (High)	0.59 (Med)
Canc	AU	-0.18 (Low)	0.68 (Med)	-0.22 (Low)	-1.33 (High)	-0.37 (Low)	0.97 (High)	0.77 (Med)	-0.01 (Low)	0.57 (Med)	0.03 (Low)
Circ	US	2.26 (High)	-0.04 (Low)	1.07 (High)	1.14 (High)	0.01 (Low)	1.34 (High)	0.05 (Low)	1.42 (High)	1.19 (High)	0.22 (Low)
Circ	JP	-0.79 (Med)	-0.09 (Low)	0.91 (High)	-1.09 (High)	-0.52 (Med)	0.1 (Low)	-0.09 (Low)	0.22 (Low)	-1.03 (High)	-0.61 (Med)
Circ	FR	-1.82 (High)	1.16 (High)	1.96 (High)	-1.23 (High)	1.29 (High)	-0.23 (Low)	0.13 (Low)	0.86 (Med)	-0.51 (Med)	0.17 (Low)
Circ	E&W	0.67 (Med)	-0.11 (Low)	0.69 (Med)	-0.03 (Low)	-0.33 (Low)	0.92 (High)	0.06 (Low)	1.01 (High)	1.04 (High)	-0.2 (Low)
Circ	AU	0.4 (Med)	0.05 (Low)	0.83 (Med)	0.67 (Med)	0.16 (Low)	-0.2 (Low)	0.01 (Low)	0.71 (Med)	-0.7 (Med)	0.01 (Low)
Resp	US	-1.37 (High)	-0.08 (Low)	-0.07 (Low)	0.01 (Low)	0.06 (Low)	-0.77 (Med)	-0.11 (Low)	-0.12 (Low)	0.19 (Low)	-0.01 (Low)
Resp	JP	-0.27 (Low)	0.05 (Low)	-0.05 (Low)	0.15 (Low)	0.07 (Low)	-0.02 (Low)	0.01 (Low)	-0.14 (Low)	0.35 (Low)	0.13 (Low)
Resp	FR	-0.41 (Med)	0.2 (Low)	0.05 (Low)	0.05 (Low)	0.13 (Low)	-0.19 (Low)	0.01 (Low)	-0.08 (Low)	0.19 (Low)	-0.02 (Low)
Resp	E&W	-0.05 (Low)	-0.02 (Low)	-0.05 (Low)	0.53 (Med)	-0.04 (Low)	-0.56 (Med)	-0.03 (Low)	-0.13 (Low)	0.31 (Low)	0.07 (Low)
Resp	AU	-0.34 (Low)	-0.05 (Low)	-0.11 (Low)	-0.02 (Low)	-0.07 (Low)	0.65 (Med)	-0.02 (Low)	-0.16 (Low)	0.84 (Med)	-0.17 (Low)
Ext	US	-4.06 (High)	-0.08 (Low)	0.16 (Low)	-1.01 (High)	1.71 (High)	-3.39 (High)	-0.38 (Med)	-0.16 (Low)	-1.47 (High)	1.38 (High)
Ext	JP	0.42 (Med)	-0.19 (Low)	-0.26 (Low)	-0.32 (Low)	0.68 (Med)	-1.35 (High)	0.16 (Low)	0.58 (Med)	-0.13 (Low)	0.79 (Med)
Ext	FR	0.44 (Med)	-0.28 (Low)	-0.08 (Low)	1.08 (High)	0.63 (Med)	-0.6 (Med)	0.07 (Low)	-0.01 (Low)	0.16 (Low)	1.07 (High)
Ext	E&W	-0.38 (Med)	-0.06 (Low)	-0.05 (Low)	-0.56 (Med)	0.53 (Med)	0.01 (Low)	-0.04 (Low)	-0.05 (Low)	-0.35 (Low)	0.98 (High)
Ext	AU	0.07 (Low)	-0.04 (Low)	0.04 (Low)	0.42 (Med)	0.84 (Med)	-1.1 (High)	0.02 (Low)	0.32 (Low)	-0.16 (Low)	1.05 (High)

Table 1.13: Impulse-response analysis: response of the mortality rate Y from the shock given to the cause X, JP females population base

Cause		Males					Females				
X \ Y	Country	I&P	Canc	Circ	Resp	Ext	I&P	Canc	Circ	Resp	Ext
I&P	US	0.41 (Med)	-0.02 (Low)	-0.04 (Low)	-0.21 (Low)	0.09 (Low)	1.25 (High)	0.04 (Low)	-0.07 (Low)	0.16 (Low)	0.03 (Low)
I&P	JP	0.92 (High)	-0.02 (Low)	0.18 (Low)	0.64 (Med)	0.11 (Low)	0.81 (Med)	0 (Low)	0.09 (Low)	0.1 (Low)	0.1 (Low)
I&P	FR	1.07 (High)	-0.17 (Low)	0 (Low)	0.37 (Low)	-0.14 (Low)	1.26 (High)	-0.04 (Low)	-0.14 (Low)	0.37 (Low)	-0.48 (Med)
I&P	E&W	0.78 (Med)	-0.03 (Low)	-0.08 (Low)	0.05 (Low)	0.23 (Low)	No suitable VAR(2) model				
I&P	AU	0.3 (Low)	0.03 (Low)	0.08 (Low)	-0.08 (Low)	0.04 (Low)	0.71 (Med)	0.03 (Low)	0 (Low)	0.15 (Low)	-0.02 (Low)
Canc	US	1.22 (High)	1.26 (High)	-0.02 (Low)	0.78 (Med)	-0.48 (Med)	-2.25 (High)	0.94 (High)	-0.8 (Med)	-2.37 (High)	-1.13 (High)
Canc	JP	0.9 (High)	0.95 (High)	0.73 (Med)	2 (High)	0.79 (Med)	0.46 (Med)	0.99 (High)	1.3 (High)	2.44 (High)	0.76 (Med)
Canc	FR	0.69 (Med)	0.5 (Med)	-0.88 (High)	-0.39 (Med)	-0.75 (Med)	2.95 (High)	0.35 (Low)	-0.8 (Med)	5.79 (High)	-2.01 (High)
Canc	E&W	-0.57 (Med)	1.15 (High)	0.4 (Med)	-0.87 (Med)	0.08 (Low)	No suitable VAR(2) model				
Canc	AU	1.15 (High)	0.65 (Med)	-0.24 (Low)	-0.19 (Low)	-0.27 (Low)	0.1 (Low)	0.84 (Med)	0 (Low)	-0.92 (High)	0.29 (Low)
Circ	US	2.65 (High)	-0.06 (Low)	1.16 (High)	1.52 (High)	0.02 (Low)	3.3 (High)	0.44 (Med)	1.53 (High)	2.48 (High)	0.02 (Low)
Circ	JP	-0.41 (Med)	-0.06 (Low)	0.91 (High)	-0.67 (Med)	-0.48 (Med)	-0.46 (Med)	-0.03 (Low)	0.6 (Med)	-1.05 (High)	-0.52 (Med)
Circ	FR	-1.39 (High)	1.28 (High)	2.15 (High)	-1.13 (High)	1.5 (High)	0.03 (Low)	0.09 (Low)	0.97 (High)	0.08 (Low)	0.09 (Low)
Circ	E&W	1.01 (High)	0.03 (Low)	1.04 (High)	0.51 (Med)	-0.55 (Med)	No suitable VAR(2) model				
Circ	AU	-0.33 (Low)	0.13 (Low)	0.91 (High)	-0.11 (Low)	0.07 (Low)	0.27 (Low)	0.03 (Low)	0.59 (Med)	-0.63 (Med)	-0.25 (Low)
Resp	US	-1.12 (High)	-0.05 (Low)	-0.12 (Low)	-0.03 (Low)	-0.06 (Low)	-0.75 (Med)	-0.14 (Low)	-0.11 (Low)	0.16 (Low)	-0.03 (Low)
Resp	JP	-0.07 (Low)	0.02 (Low)	-0.21 (Low)	0.13 (Low)	-0.01 (Low)	-0.03 (Low)	0.02 (Low)	-0.15 (Low)	0.25 (Low)	0.06 (Low)
Resp	FR	-0.3 (Low)	0.21 (Low)	0.06 (Low)	0.14 (Low)	0.16 (Low)	-0.23 (Low)	0.02 (Low)	-0.03 (Low)	0.13 (Low)	0.06 (Low)
Resp	E&W	-0.36 (Low)	-0.02 (Low)	-0.11 (Low)	0.61 (Med)	0.19 (Low)	No suitable VAR(2) model				
Resp	AU	-1.36 (High)	0.13 (Low)	0.27 (Low)	0.67 (Med)	0.14 (Low)	-0.15 (Low)	-0.03 (Low)	-0.08 (Low)	0.14 (Low)	0.09 (Low)
Ext	US	-4.06 (High)	-0.11 (Low)	0.11 (Low)	-1.4 (High)	1.47 (High)	-5.92 (High)	-0.93 (High)	-0.27 (Low)	-3.36 (High)	1.56 (High)
Ext	JP	0.14 (Low)	0.01 (Low)	-0.04 (Low)	-0.53 (Med)	0.75 (Med)	0.08 (Low)	0.01 (Low)	0.07 (Low)	0.13 (Low)	0.68 (Med)
Ext	FR	1.11 (High)	-1.05 (High)	-0.73 (Med)	2.26 (High)	-0.09 (Low)	-1.49 (High)	0.18 (Low)	0.52 (Med)	-1.41 (High)	2.39 (High)
Ext	E&W	1.17 (High)	0.09 (Low)	0.39 (Med)	-0.18 (Low)	-0.06 (Low)	No suitable VAR(2) model				
Ext	AU	0.72 (Med)	-0.15 (Low)	-0.18 (Low)	0.08 (Low)	0.78 (Med)	0.16 (Low)	0.03 (Low)	0.17 (Low)	0.69 (Med)	0.94 (High)

Table 1.14: $\mathbf{\Gamma}_1$ coefficients that are significantly different from zero, significance level of 0.05, males, US males population base

Country	Model	ΔIP_t	$\Delta Canc_t$	$\Delta Circ_t$	$\Delta Resp_t$	ΔExt_t
US	VAR(2), QT, 1 CR	-	Circ,Ext	-	Resp,Ext	Resp
JP	VAR(2), TC, 2 CR	-	Canc	Resp	Circ,Resp	-
FR	VAR(2), NT, 1 CR	-	-	Resp	Circ,Resp	Circ,Resp
E&W	VAR(2), QT, 1 CR	Ext	-	-	IP,Circ Resp	Ext
AU	VAR(2), QT, 1 CR	IP	Canc, Ext	Circ	-	Ext

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.15: $\mathbf{\Gamma}_1$ coefficients that are significantly different from zero, significance level of 0.05, females, US males population base

Country	Model	ΔIP_t	$\Delta Canc_t$	$\Delta Circ_t$	$\Delta Resp_t$	ΔExt_t
US	VAR(2), QT, 1 CR	Canc,Ext	Ext	Canc	Resp	Canc,Resp
JP	VAR(2), TC, 2 CR	IP,Resp	-	Resp	-	Circ,Resp Ext
FR	VAR(2), QT, 1 CR	Canc,Circ	-	-	Canc,Circ Resp	IP
E&W	VAR(2), QT, 1 CR	-	-	Resp	Resp	Resp
AU	VAR(2), NT, 1 CR	IP,Resp	Canc	Circ	Circ,Resp	-

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.16: $\mathbf{\Gamma}_1$ coefficients that are significantly different from zero, significance level of 0.05, males, JP females population base

Country	Model	ΔIP_t	$\Delta Canc_t$	$\Delta Circ_t$	$\Delta Resp_t$	ΔExt_t
US	VAR(2), QT, 1 CR	Ext	-	-	Resp,Ext	Resp
JP	VAR(2), QT, 1 CR	Ext	-	-	-	Resp
FR	VAR(2), NT, 1 CR	-	-	Canc,Resp	Circ,Resp	Resp
E&W	VAR(2), QT, 1 CR	-	-	Circ	Resp	Ext
AU	VAR(2), NT, 1 CR	Circ	Canc	Circ	Resp	Ext

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.17: $\mathbf{\Gamma}_1$ coefficients that are significantly different from zero, significance level of 0.05, females, JP females population base

Country	Model	ΔIP_t	$\Delta Canc_t$	$\Delta Circ_t$	$\Delta Resp_t$	ΔExt_t
US	VAR(2), QT, 1 CR	Canc,Ext	Ext	-	Resp,Ext	Canc,Resp
JP	VAR(2), QT, 1 CR	-	-	-	-	Circ,Ext
FR	VAR(2), NT, 1 CR	-	-	Resp	Canc,Circ Resp	IP
E&W	No suitable VAR(2) model					
AU	VAR(2), QT, 1 CR	-	Canc	Circ	-	Circ

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.18: Equations to which the long-term component enters with a statistically significant coefficient α_i, significance level of 0.05, US males population base

Country	Model		Males	Females
US	VAR(2), QT, 1 CR	α_i	ΔIP_t, $\Delta Resp_t$	ΔIP_t, $\Delta Canc_t$ $\Delta Circ_t$, $\Delta Resp_t$
JP	VAR(2), TC, 2 CR	α_{1i}	ΔIP_t, $\Delta Canc_t$	$\Delta Circ_t$, $\Delta Resp_t$
		α_{2i}	ΔIP_t, $\Delta Canc_t$, $\Delta Resp_t$	ΔIP_t, $\Delta Canc_t$ $\Delta Circ_t$, $\Delta Resp_t$
FR	VAR(2), NT, 1 CR	α_i	ΔIP_t, $\Delta Canc_t$, $\Delta Resp_t$	-
	VAR(2), QT, 1 CR	α_i	-	ΔIP_t, $\Delta Canc_t$, $\Delta Resp_t$
E&W	VAR(2), QT, 1 CR	α_i	$\Delta Canc_t$, $\Delta Resp_t$ ΔExt_t	ΔIP_t, $\Delta Circ_t$ ΔExt_t
AU	VAR(2), NT, 1 CR	α_i	-	ΔIP_t, $\Delta Circ_t$, ΔExt_t
	VAR(2), QT, 1 CR	α_i	ΔIP_t, $\Delta Circ_t$, $\Delta Resp_t$	-

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.19: Equations to which the long-term component enters with a statistically significant coefficient α_i, significance level of 0.05, JP females population base

Country	Model		Males	Females
US	VAR(2), QT, 1 CR	α_i	ΔIP_t, $\Delta Resp_t$	ΔIP_t, $\Delta Canc_t$ $\Delta Circ_t$, $\Delta Resp_t$
JP	VAR(2), QT, 1 CR	α_i	$\Delta Circ_t$, $\Delta Resp_t$	ΔIP_t, $\Delta Circ_t$, $\Delta Resp_t$
FR	VAR(2), NT, 1 CR	α_i	ΔIP_t, $\Delta Canc_t$, $\Delta Circ_t$ $\Delta Resp_t$, ΔExt_t	ΔIP_t, $\Delta Resp_t$
E&W	VAR(2), QT, 1 CR	α_i	ΔExt_t	No suitable VAR(2) model
AU	VAR(2), NT, 1 CR	α_i	ΔIP_t, $\Delta Circ_t$, $\Delta Canc_t$	-
	VAR(2), QT, 1 CR	α_i	-	$\Delta Circ_t$, $\Delta Resp_t$

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.20: Correlation coefficients between the actual changes in mortality rates and the long- and short-term components, males, US males population base

Country	Model		ΔIP_t	$\Delta Canc_t$	$\Delta Circ_t$	$\Delta Resp_t$	ΔExt_t
US	VAR(2), QT, 1 CR	LT	-0.271			0.268	
		ST	-0.164			0.336	
JP	VAR(2), TC, 2 CR	LT	0.570	0.797		0.530	
		ST	-0.177	-0.434		0.325	
FR	VAR(2), NT, 1 CR	LT	0.478	0.597		0.368	
		ST	0.023	-0.311		0.613	
E&W	VAR(2), QT, 1 CR	LT		-0.589		0.069	0.018
		ST		0.023		0.575	0.155
AU	VAR(2), QT, 1 CR	LT	-0.389		0.450	0.239	
		ST	-0.058		0.203	-0.059	

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.21: Correlation coefficients between the actual changes in mortality rates and the long- and short-term components, females, US males population base

Country	Model		ΔIP_t	$\Delta Canc_t$	$\Delta Circ_t$	$\Delta Resp_t$	ΔExt_t
US	VAR(2), QT, 1 CR	LT	0.076	0.477	0.290	0.192	
		ST	-0.172	0.012	0.343	0.451	
JP	VAR(2), TC, 2 CR	LT	0.667	0.511	0.763	0.744	
		ST	-0.181	-0.146	-0.315	-0.386	
FR	VAR(2), QT, 1 CR	LT	0.415	0.272		0.582	
		ST	0.083	0.226		0.259	
E&W	VAR(2), QT, 1 CR	LT	0.131		0.410		0.069
		ST	0.286		0.425		0.240
AU	VAR(2), NT, 1 CR	LT	0.529		0.291		0.192
		ST	0.183		0.457		0.311

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.22: Correlation coefficients between the actual changes in mortality rates and the long- and short-term components, males, JP females population base

Country	Model		ΔIP_t	$\Delta Canc_t$	$\Delta Circ_t$	$\Delta Resp_t$	ΔExt_t
US	VAR(2), QT, 1 CR	LT	-0.266			0.275	
		ST	-0.200			0.372	
JP	VAR(2), QT, 1 CR	LT			0.538	0.729	
		ST			0.522	0.069	
FR	VAR(2), NT, 1 CR	LT	0.355	0.553	0.033	0.366	0.271
		ST	0.259	-0.276	0.405	0.621	0.189
E&W	VAR(2), QT, 1 CR	LT					0.130
		ST					0.272
AU	VAR(2), NT, 1 CR	LT	0.600	0.191	0.081		
		ST	-0.017	0.432	0.487		

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Table 1.23: Correlation coefficients between the actual changes in mortality rates and the long- and short-term components, females, JP females population base

Country	Model		ΔIP_t	$\Delta Canc_t$	$\Delta Circ_t$	$\Delta Resp_t$	ΔExt_t
US	VAR(2), QT, 1 CR	LT	0.617	0.589	0.088	0.301	
		ST	-0.160	-0.006	0.365	0.438	
JP	VAR(2), QT, 1 CR	LT	0.533		-0.054	0.378	
		ST	0.142		-0.236	-0.411	
FR	VAR(2), NT, 1 CR	LT	0.406			0.623	
		ST	0.084			0.339	
E&W	No suitable VAR(2) model						
AU	VAR(2), QT, 1 CR	LT			0.514	0.576	
		ST			0.450	0.383	

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

Bibliography

Arnold, S. and M. Sherris, 2013, Forecasting Mortality Trends allowing for Cause-of-Death Mortality Dependence, *North American Actuarial Journal*, 17(4):273–282.

Arnold, S. and M. Sherris, 2015, Causes-of-Death Mortality: What Do We Know on their Dependence?, *North American Actuarial Journal*, 19(2):116–128.

Arnold, S. and M. Sherris, 2016, International Cause-Specific Mortality Rates: New Insights from a Cointegration Analysis., *ASTIN Bulletin*, 46(1):9–38.

Baillie, R. and D. Selover, 1987, Cointegration and models of exchange rate determination, *International Journal of Forecasting*, 3:43–51.

Basagaña, X., C. Sartini, J. Barrera-Gómez, P. Dadvand, J. Cunillera, B. Ostro, J. Sunyer, and M. Medina-Ramón, 2011, Heat waves and cause-specific mortality at all ages, *Epidemiology*, 22(6):765–772.

Booth, H. and L. Tickle, 2008, Mortality modelling and forecasting: A review of methods, *Annals of Actuarial Science*, 3:3–43.

Breeze, E., R. Clarke, M. J. Shipley, M. G. Marmot, and A. E. Fletcher, 2006, Cause-specific mortality in old age in relation to body mass index in middle age and in old age: follow-up of the Whitehall cohort of male civil servants, *International Journal of Epidemiology*, 35(1):169–178.

Cairns, A., 2013, Modeling and management of longevity risk, In P. B. Hammond, R. Maurer, and O. S. Mitchell, editors, *Recreating Sustainable Retirement: Resilience, Solvency, and Tail Risk*, pages 71–88. Oxford, UK: Oxford University Press.

Carnes, B. A., L. R. Holden, S. J. Olshansky, T. M. Witten, and J. S. Siegel, 2006, Mortality partitions and their relevance to research on senescence, *Biogerontology*, 7:183–198.

Carnes, B. A. and S. J. Olshansky, 1997, A biologically motivated partitioning of mortality, *Experimental Gerontology*, 32(6):615–631.

Chiang, C. L., 1968, *Introduction to Stochastic Process in Biostatistics*. New York: John Wiley and Sons.

Clarida, R., 1992, Co-Integration, Aggregate Consumption, and the Demand for Imports: A Structural Economic Investigation, Research Paper 9213, Federal Reserve Bank of New York.

Costa, D. L., 2005, Causes of improving health and longevity at older ages: a review of the explanations, *Genus*, 61(1):21–38.

Cutler, D., A. Deaton, and A. Lleras-Muney, 2006, The Determinants of Mortality, *The Journal of Economic Perspectives*, 20(3):97–120.

Debón, A., F. Montes, and R. Sala, 2006, A Comparison of Models for Dynamic Life Tables. Application to Mortality Data from the Valencia Region (Spain), *Lifetime Data Analysis*, 12:223–244.

Dimitrova, D. S., S. Haberman, and V. Kaishev, 2013, Dependent competing risks: Cause elimination and its impact on survival, *Insurance: Mathematics and Economics*, 53(2):464–477.

Engle, R. and C. Granger, 1987, Co-integration and Error Correction: Representation, Estimation and Testing., *Econometrica*, 55:251–276.

GBDS, 2013, Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013, *The Lancet*, 385:117–171.

Hamilton, J. D., 1994, *Time Series Analysis*. Princeton: Princeton University Press.

Himes, C. L., 1994, Age Patterns of Mortality and Cause-of-Death Structures in Sweden, Japan, and the United States, *Demography*, 31(4):633–650.

Horiuchi, S. and J. R. Wilmoth, 1997, Age Patterns of the Life Table Aging Rate for Major Causes of Death in Japan, 1951-1990, *Journal of Gerontology: Biological Sciences*, 52A(1):B67–B77.

Johansen, S., 1988, Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control*, 12:231–254.

Johansen, S., 1995, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. New York: Oxford University Press.

Johansen, S. and K. Juselius, 1992, Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK, *Journal of Econometrics*, 53:211–244.

Kaishev, V. K., D. S. Dimitrova, and S. Haberman, 2007, Modelling the joint distribution of competing risks survival times using copula functions., *Insurance: Mathematics and Economics*, 41:336–361.

Li, H. and Y. Lu, 2019, Modeling cause-of-death mortality using hierarchical Archimedean copula, *Scandinavian Actuarial Journal*, 2019(3):247–272.

Lo, S. and R. A. Wilke, 2010, A copula model for dependent competing risks, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(2):359–376.

Lütkepohl, H., 2005, *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.

Mackenbach, J.-P., A. Kunst, H. Lautenbach, Y. Oei, and F. Bijlsma, 1999, Gains in life expectancy after elimination of major causes of death: revised estimates taking into account the effect of competing causes, *Journal of Epidemiology & Community Health*, 53(1):32–37.

Manton, K. G. and G. C. Myers, 1987, Recent trends in multiple-caused mortality 1968 to 1982: Age and cohort components, *Population Research and Policy Review*, 6:161–176.

Manton, K. G. and S. S. Poss, 1979, Effects of dependency among causes of death for cause elimination life table strategies, *Demography*, 16(2):313–327.

Manton, K. G., E. Stallard, and H. Trolley, 1991, Limits to human life expectancy: Evidence, prospects, and implications, *Population and Development Review*, 17(4):603–637.

Matthews, C., S. George, S. Moore, H. Bowles, A. Blair, Y. Park, R. Troiano, A. Hollenbeck, and A. Schatzkin, 2012, Amount of time spent in sedentary behaviors and cause-specific mortality in US adults, *The American Journal of Clinical Nutrition*, 95(2):437–445.

Osterwald-Lenum, M., 1992, A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics, *Oxford Bulletin of Economics and Statistics*, 54(3):461–472.

Rey, G., E. Jougla, A. Fouillet, G. Pavillon, P. Bessemoulin, P. Frayssinet, J. Clavel, and D. Hémon, 2007, The impact of major heat waves on all-cause and cause-specific mortality in France from 1971 to 2003, *International Archives of Occupational and Environmental Health*, 80(7):615–626.

Tsiatis, A., 1975, A nonidentifiability aspect of the problem of competing risks., *Proceedings of the National Academy of Sciences*, 72(1):20–22.

Wang, C.-S., S.-T. Wang, C.-T. Lai, L.-J. Lin, and P. Choua, 2007, Impact of influenza vaccination on major cause-specific mortality, *Vaccine*, 25:1196–1203.

World Health Organization, 2016, WHO Mortality Database. `http://www.who.int/whosis/mort/download/en/index.html`.

Zmirou, D., J. Schwartz, M. Saez, A. Zanobetti, B. Wojtyniak, G. Touloumi, C. Spix, A. Ponce de León, Y. Le Moullec, L. Bacharova, J. Schouten, A. Pönkä, and K. Katsouyanni, 1998, Time-series analysis of air pollution and cause-specific mortality, *Epidemiology*, 9(5):495–503.

Chapter 2

Cause-Specific Mortality Rates: Common Trends and Differences

2.1 Introduction

The ever-decreasing mortality rates represent one of the biggest challenges that the insurance industry has ever faced. It is hence very important to understand well the past developments of mortality and be able to build sustainable forecasts for the future. This colossal task has been occupying many researchers as well as practitioners for several decades and resulted in innumerable number of models, approaches, and practices. An interested reader can find a review thereof in Booth and Tickle (2008), Cairns (2013) and Debón et al. (2006), including their references. While most of the existing models deal with the all-cause mortality, we believe that integrating the information on different causes of death into the model can bring additional insight and improve the model's fit. In the same time, cause-specific mortality rates are dependent, but the dependency between them is unobservable and so, highly difficult to model.

Cointegration analysis represents a tool that allows us to take into account the dependency between the non-stationary variables. Two or more variables are said to be cointegrated if there exists a linear combination of them that is stationary. Such linear combination describes then the link (or the dependency) between the variables in the long run and can be included into a vector autoregressive model. This approach was initially developed to model the econometric time series, but later also gained popularity in the field of mortality modelling. As a vector of age-specific mortality rates usually has more elements than a cointegration relation can incorporate, some authors overcame this difficulty by concentrating on the pairwise cointegration between the age-specific mortality rates (Darkiewicz and Hoedemakers, 2004) or cointegration in a subset of mortality rates, e.g. higher ages (Lazar and Denuit, 2009). In Gaille and Sherris (2011) the authors reduced the age dimension by applying cointegration analysis to the parameters

of the Heligman-Pollard model. Using total mortality rates, Njenga and Sherris (2011) were able to formulate a cointegrated model incorporating five country-specific mortality rates, whereas Arnold and Sherris (2013) applied a model allowing for cointegration relations between five cause-specific mortality rates. Also, very productive was the idea to apply the cointegration analysis to the time trends extracted from the mortality rates by the means of the Lee-Carter model. In this way, cointegration relations between the population-specific time trends were used in modelling and forecasting mortality rates by Li and Hardy (2011), Yang and Wang (2013), Zhou et al. (2014), and Hunt and Blake (2015) in a multi-population setting. Salhi and Loisel (2017) also studied the cointegration relations between two populations, but used the pairwise cointegration between the age-specific mortality rates for higher ages. A further example is the work by Li and Lu (2017) who enrich the vector-autoregression model for age-specific mortality rates with a cointegration element that ensures the non-divergence of the mortality rates at different ages.

For our part, we want to concentrate on the different causes of death with the help of the cointegration analysis. In this relation, Arnold and Sherris (2013) showed that the cause-specific mortality rates were cointegrated and confirmed this finding for both sexes in ten different countries. Using the setup of the vector error correction models, Arnold and Sherris (2016) identified the optimal model structure for the cause-specific mortality rates that allowed the authors to spot the difference in behavior between the exo- and endogenous causes of death in the long run. Using a slightly different set of the optimal VECMs, Arnold and Glushko (2021) analyzed the short- and long-term interactions between the cause-specific mortality rates, namely how each cause-specific mortality rate impacts and reacts to the shocks received from the rest of the causes. To achieve this goal, Arnold and Glushko (2021) analyzed the *stationary* part of the model by studying the short run (autoregressive) and the long run (cointegration) terms.

Now we want to take a different angle: if certain variables are cointegrated, they must share common stochastic trends that are eliminated in the cointegration relation. In other words, as shown by Stock and Watson (1988), for any cointegrated variables x_t and y_t there must exist a common factor representation of the following form:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} A \\ 1 \end{bmatrix} f_t + \begin{bmatrix} \tilde{y}_t \\ \tilde{x}_t \end{bmatrix}, \tag{2.1}$$

where $\tilde{y}_t$ and $\tilde{x}_t$ are both I(0), $(1, -A)$ is the cointegrating vector and f_t represent the common stochastic factors impacting the system. In the present paper, we will study the common stochastic trends shared by the cause-specific mortality rates, i.e. the *non-stationary* part of the model which is usually ignored when only cointegration relations are analyzed (as this was done in Arnold and Sherris (2016) and Arnold and Glushko

(2021)). To achieve this, we will, first, recover these common factors, and second, see if any similarities or common patterns can be found between different countries. For the sake of comparability and consistence, we use the same set of models as in Arnold and Glushko (2021).

In Arnold and Sherris (2016) the authors found that the cointegration techniques applied to the cause-specific mortality data provided a first bridge between econometrics and biology, two areas of study essential for actuaries, and that cointegration relations reflected the biological theory on aging. This became possible once the distinction between the exo- and endogenous causes, developed by biologists and demographers, was considered. Although this classification is not clear-cut, to the first group of causes of death most researchers attribute various external or environmental factors that produce death, while the endogenous causes of death correspond to biological forces that lead to death (Carnes et al., 2006). From the distinction between the exo- and endogenous causes follows the idea that endogenous mortality reflects fundamental processes of the human body referred to as the biological processes of aging. Since the authors found that only the endogenous causes appeared in the cointegration relations, *these relations may have the potential to capture the statistical characteristics of the biological processes of aging* (Arnold and Sherris, 2016). Also, common stochastic trends shared by the cause-specific mortality rates could represent the aging processes, because the aging process is known to be stochastic (Hayflick, 2004) and a potential mixture of several stochastic processes (Holliday, 2004). Then, the biological aging of the body is indeed the underlying risk factor influencing the causes of death (Olshansky et al., 2002) and is captured by the common stochastic trends of the cointegrating system. However, Arnold and Sherris (2016) only analysed the cointegration relations and did not try to find an expression for these common stochastic trends.

In the present work, we want to go further and investigate the first intuition of Arnold and Sherris (2016) by recovering and studying the common stochastic factors f_t as they are shown in (2.1). Gonzalo and Granger (1995) mention several reasons why it may be interesting to recover f_t, namely to 1) simplify a complex model; 2) decompose (y_t, x_t) into two components $(f_t, (\tilde{y}_t, \tilde{x}_t))$ that transmit different kinds of information (while the permanent component f_t corresponds to the trends present in the data, the transitory component $(\tilde{y}_t, \tilde{x}_t)$ conveys the information on the short-term shocks and cycles); 3) study the subdivisions of a large system by first finding the common factors in every subdivision and then studying the cointegration among them.

Cause-specific mortality rates reflect numerous impacts and processes that range from medical advances, changes in lifestyles and nutrition to epidemics and aging. In spite of the evident differences between the past experience of different countries, it is still reasonable to expect that some of these processes will be present in all countries due to their universal character, e.g., aging. So, our objective is to recover the common stochastic

factors shared by the cause-specific mortality rates from one country and propose an approach allowing us to make comparisons with a goal to find similarities across five countries. Should a certain pattern be found in all tested countries, this would allow us to expect that we are dealing with some fundamental process common to human species as such. So, in this way, we are able to shed light on the processes that underlie the development of mortality in every tested country. Although we are not yet able to identify these processes with a certainty, we believe that the possibility to give them a mathematical expression can help to improve our understanding of the past development of the mortality rates.

To achieve this, we, first, construct the set of common factors in every country that has a lower number of dimensions than the initial variables. As it turns out that this number is still too high to allow direct comparison between the countries, we further concentrate the information available in the set by using the principal component analysis. When comparing the charts of the principal components, we noticed that the form of the first elements was similar on all charts. To study further the observed resemblance, we test for cointegration using the Johansen maximum likelihood tests (Johansen, 1988). This allows us to examine cointegration in a large system of data variables (5 cause-specific mortality rates for 5 countries and 2 sexes) that would not have been possible without the initial reduction in its dimensionality. At the next step, we find that once we put together the first components extracted in every country, they are indeed cointegrated.

We believe that, although the cause-specific mortality rates show different development patterns across countries, this observation could mean that the common factors reflect some similar intrinsic stochastic processes which occur in every dataset (country). As we work with the cause-specific mortality rates, these processes could point to some fundamental mechanisms, typical for human species, such as biological aging.

The paper is organized as follows: in Section 2.2 we briefly present the data that we used in the study, then continue with some theoretical notions of the cointegration analysis and lay out the methodology used to extract and condense the common stochastic factor in Section 2.3. The application of these tools to the data is presented in Section 2.4, while Section 2.5 concludes.

2.2 Data

For this study, we used the same data as in Arnold and Glushko (2021) and refer the interested reader to this article for the details on the data preparation process. We mention here the main points.

- Data were retrieved from the WHO Mortality Database (World Health Organization, 2016) that contains the mid-year population and the death counts by country,

year, sex, age group, and cause of death. The earliest observations in this database go as far back as 1950.

- The WHO database splits the death numbers according to the *primary* cause of death, and for this reason, we will ignore the potential presence of the secondary cause, third cause etc. Moreover, we would have to signigicanlty change our approach in order to incorporate the information on the secondary cause of death, for example. For this reason, our results would not hold in presense of several causes leading to death.

- We considered the following countries and observation periods: USA (1950-2007), Japan (1950-2013), France (1952-2011), England and Wales (1950-2013), and Australia (1950-2004), subsequently shortened to US, JP, FR, E&W, and AU respectively. These countries were chosen as they participate in the database from the onset, belong to the developed countries with important population sites and are located in different parts of the world (North America, Asia, Europe, and Oceania). This choice ensures that we have at our disposal the longest possible series of rich and reliable observations, also accounting for the variety of possible geographic impacts.

- WHO defines the causes of death according to the International Classification of Diseases (ICD). By applying the comparability ratios [1] we ensured that observations were comparable across the different versions of ICD that switched from ICD-7 to ICD-10 since the inception of the database.

- Causes of deaths were split in groups of infectious and parasitic diseases, cancer, diseases of the circulatory system, diseases of the respiratory system, and external causes. These are the most important groups of causes of deaths. They account for approximately 70-80% of deaths in recent years and made up approximately 50%-70% of deaths at the onset of the observations.

- Central death (or mortality) rates were calculated as the number of deaths by age, sex, and cause divided by the mid-year population by age and sex.

- Central death rates were age-standardized using the US male population of 2007 as the standard population (more details on this procedure are given in Appendix A). We will work with the total rates and will not differentiate by age, otherwise there would be more variables than the cointegration analysis can accommodate.

[1] A comparability ratio serves to remove the discontinuities between the observation periods: it makes the average of the mortality rates over the last two years of a classification coincide with the average of the mortality rates over the first two years of the next classification. So, the mortality rates in every classification are divided by the comparability ratio(s) linking this classification to the previous one(s).

Age-standardized death rates for selected years using the US males population base as well as the charts showing their evolution over the entire observation period are shown in Table 2.5 and on Figure 2.6 in the Appendix A.

- Equations were estimated for the time series of ordered $(n \times 1)$ vectors of the cause-specific mortality rates after the application of the natural logarithm:

$$\mathbf{y}_{t,s,c} = \begin{pmatrix} log(m_{t,s,c}^{I\&P}) \\ log(m_{t,s,c}^{Cancer}) \\ log(m_{t,s,c}^{Circulatory}) \\ log(m_{t,s,c}^{Respiratory}) \\ log(m_{t,s,c}^{External}) \end{pmatrix},$$

where n is the number of the analyzed cause-specific mortality rates, here $n = 5$, t denotes the time, s the gender, and c the country.

2.3 Theoretical framework

2.3.1 VECM and the common stochastic factors

Arnold and Sherris (2015, 2016) showed that the cause-specific mortality rates were non-stationary and so, contained stochastic trends. It was also demonstrated that at least one cointegration relation existed between the variables and for this reason, it was possible to build a Vector Error Correction Model (VECM) describing the development of the cause-specific mortality rates. VECMs represent an extension of the Vector AutoRegression (VAR) models and allow modelling the dependency between the lagged values of the differenced variables and the variables in levels through the so-called cointegration term $\alpha\beta'\mathbf{y}_{t-1}$. Supposing that there are r cointegration relations, i.e. that there exists a matrix β of rank r such that $\beta'\mathbf{y}_t$ is I(0), the corresponding VECM has the following form:

$$\Delta\mathbf{y}_t = \mathbf{c} + \mathbf{d}t + \alpha\beta'\mathbf{y}_{t-1} + \sum_{n=1}^{p-1} \xi_i \Delta\mathbf{y}_{t-i} + \epsilon_t, \ t = 1...T \tag{2.2}$$

where

- $\mathbf{c}$ and $\mathbf{d}$ are $(n \times 1)$ vectors of constants;
- ξ_i is a $(n \times n)$ matrix of autoregressive coefficients for $i = 1, 2, ..., p-1$;
- β is a $(n \times r)$ matrix containing r vectors each representing a cointegration relation;
- α is a $(n \times r)$ loading matrix that indicates how a particular variable is impacted by the cointegration relation;

- ϵ_t is a $(n \times 1)$ vector of white noise errors.

Hamilton (1994) and Lütkepohl (2005) are the extensive references on the VECM and VAR models.

Further, as mentioned in Stock and Watson (1988), cointegrated multivariate time series comprise at least one common trend and so, can be expressed as a sum of a reduced number of common stochastic trends, plus transitory, or stationary, components. In other words, $\mathbf{y}_t$ can be explained in terms of a smaller number $(n-r)$ of I(1) variables, f_t, called common factors or common long-memory components, plus some I(0) components $\tilde{\mathbf{y}}_t$:

$$\underset{n\times 1}{\mathbf{y_t}} = \underset{n\times k}{A_1}\ \underset{k\times 1}{f_t} + \underset{n\times 1}{\tilde{\mathbf{y}}_\mathbf{t}}\,, \tag{2.3}$$

where $k = n - r$.

The objective is then to estimate f_t as a linear combination of the original variables using the methodology developed by Gonzalo and Granger (1995) which we briefly present in the Appendix C. Although the f_t have fewer number of dimensions than the original data, further reduction of dimensionality may be needed to allow comparing of the common factors across countries.

2.3.2 Principal component analysis

One of the most popular dimension-reduction techniques is the principal component analysis (PCA) and the book by Jolliffe (2002) is an extensive reference on the subject. Simply put, the idea of the PCA is to reduce the dimensionality of a dataset and, in the same time, preserve as much as possible the variation present in the data.

Suppose that $\mathbf{y}$ is a vector of n random variables with a known covariance matrix Σ. If Σ is a positive definite, it can be decomposed as $\Sigma = \Gamma'\Lambda\Gamma$, where columns $v_1, v_2, ..., v_n$ of Γ are the eigenvectors corresponding to the ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ which form the main diagonal of the matrix Λ. It can be shown that the vector $\mathbf{x} = \Gamma\mathbf{y}$ will have the same total variance as the vector $\mathbf{y}$ and the first component of $\mathbf{x}$ given by $x_1 = v_1'\mathbf{y}$ will have the maximum variance of any linear combination $a'\mathbf{y}$ such that $||a|| = 1$ (Hogg et al., 2014). For this reason, x_1 is called the first principal component (PC) of $\mathbf{y}$ and can be used as a proxy for the information contained in $\mathbf{y}$.

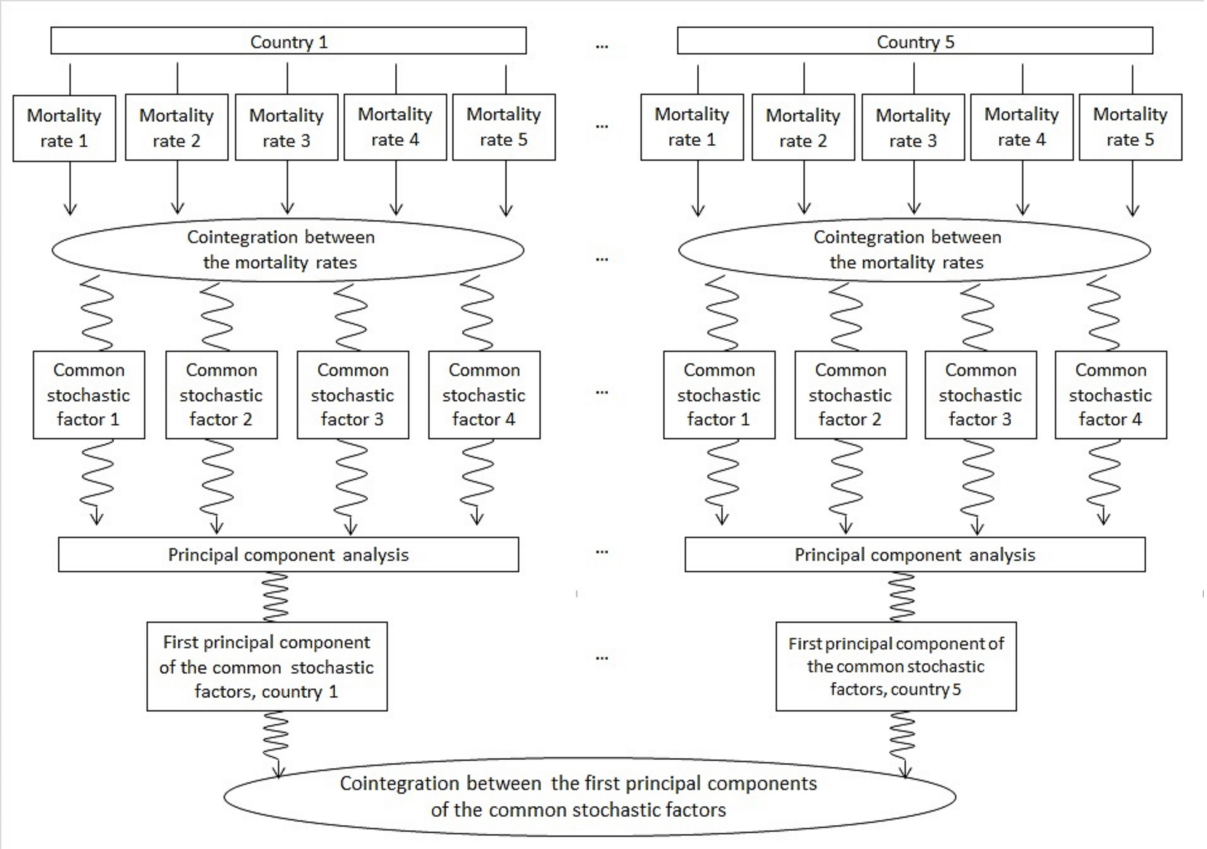
In our data, some elements of the vector $\mathbf{y}$ substantially outweigh the rest of the causes (e.g. circulatory and cancer death rates). In this case it is recommended to use the correlation matrix Σ^* of $\mathbf{y}$ when calculating the PCs (Jolliffe, 2002).

So, we apply the PCA in order to express the information contained in the common stochastic factors in one principal component. Such a reduction in the dimensionality of the data will allow us to compare the first principal components for different countries and sexes using the cointegration tests. This process is schematically shown in Figure

2.1.

Added to above, the PCA is a well-known method in the mortality modelling field. One of the most popular mortality models, the Lee-Carter model, basically extracts and projects a unique time trend from a matrix of mortality rates by assuming that the vector of the mortality reductions is time-invariant. Yang et al. (2010) go further and account for the variant mortality improvements at different ages.

Figure 2.1: Two-level cointegration analysis of the cause-specific mortality rates.



2.4 Application to the cause-specific mortality rates

2.4.1 Estimation of the common stochastic factors

To estimate the common stochastic factors, one has first to find the VECM equations that best describe the datasets. Here, not only the number of cointegrating relations, but also the form of the deterministic part of the model play an important role. Let $\mu_t = \mathbf{c} + \mathbf{d}t$ denote the deterministic part of the model (2.2) and suppose that the parameter d can be decomposed in the directions of $\delta_\perp$ and δ such that $\delta\delta_\perp = 0$. Then $\mathbf{d} = \delta\rho + \delta_\perp\gamma$, where ρ and γ are the decomposition parameters. As the mortality rates are known to have

Table 2.1: Vector Error Correction Models chosen for the analysis.

Country	Males	Females
US	VAR(2), QT, 1 CR	VAR(2), QT, 1 CR
JP	VAR(2), TC, 2 CR	VAR(2), TC, 2 CR
FR	VAR(2), NT, 1 CR	VAR(2), QT, 1 CR
E&W	VAR(2), QT, 1 CR	VAR(2), QT, 1 CR
AU	VAR(2), QT, 1 CR	VAR(2), NT, 1 CR

Note: QT = quadratic trend in the VAR; TC = linear trend in the cointegration relation; NT = no trend; CR = cointegration relation.

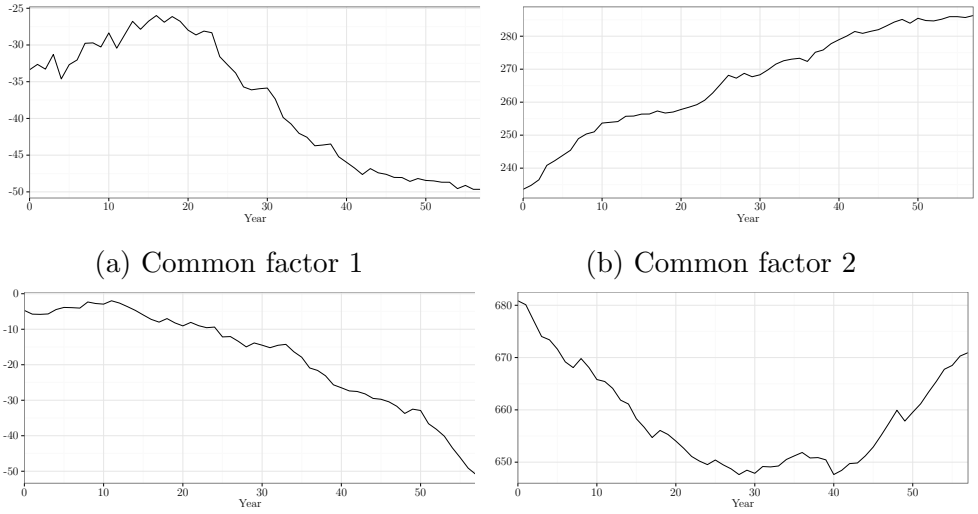
a trend, we will consider the following forms of the deterministic elements (Johansen, 1995):

- NT: no trend in the VECM, but a linear trend in the levels of the variables: $\mathbf{c} \neq 0, \rho = 0, \gamma = 0,$ hence $\mathbf{d} = 0,$
- TC: linear trend in the cointegration relation combined with a linear trend in the levels of the variables (i.e., no linear trend in the differenced variables): $\mathbf{c} \neq 0, \rho \neq 0, \gamma = 0,$ hence $\mathbf{d} = \delta\rho,$
- QT: linear trend in the differenced variables, thus the quadratic trend in the levels of the variables (i.e., the VAR model) : $\mathbf{c} \neq 0, \rho \neq 0, \gamma \neq 0,$ hence $\mathbf{d} = \delta\rho + \delta_\perp\gamma.$

We will use the same VECMs as in Arnold and Glushko (2021) that we reproduce here for the sake of completeness (Table 2.1).

Once the VECM coefficients are calculated for every dataset, we estimate the common factors $f_t = \alpha_\perp \mathbf{y_t}$ as described in the previous section. The maximum likelihood estimates of $\hat{\alpha}_\perp = (\hat{m}_{r+1}, ..., \hat{m}_n)$, where r is a number of cointegration relations, that is 1 or 2 for the datasets used in the study, suggest that the number of dimensions of the common factor component f_t will be 4 or 3 respectively. Hence, by estimating common stochastic factors we reduced the number of dimensions in our system, but comparing common factors across countries remains complicated. Figure 2.2 shows the common factors estimated for the dataset of US males, common factors for the rest of the datasets are shown on figures 2.7 - 2.15 of the Appendix B.

Figure 2.2: Common stochastic factors f_t, US males.



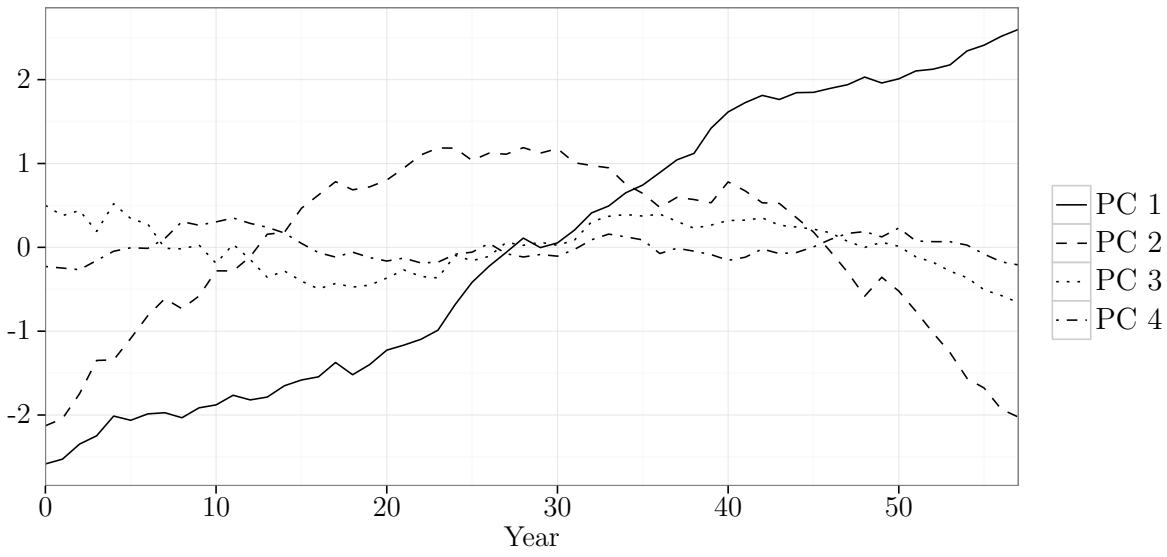
(a) Common factor 1 (b) Common factor 2

(c) Common factor 3 (d) Common factor 4

At present, we would like to compare the common stochastic factors across five countries, but as we have 3 to 4 factors for each country, we cannot apply the cointegration analysis. For this, we need to further reduce the number of dimensions to be able to compare the common trends across the different datasets, and for this, we use the principal component analysis.

The principal components for the US males dataset are shown in Figure 2.3 (see Figures 2.16 - 2.24 of the Appendix B for the rest of the countries and sexes). As expected, the first PC has the maximum variance, whereas the rest of the components represent fluctuations around the zero line.

Figure 2.3: PCA applied to common factors, US males.



For US males, the first principal component explains approximately 72% of the to-

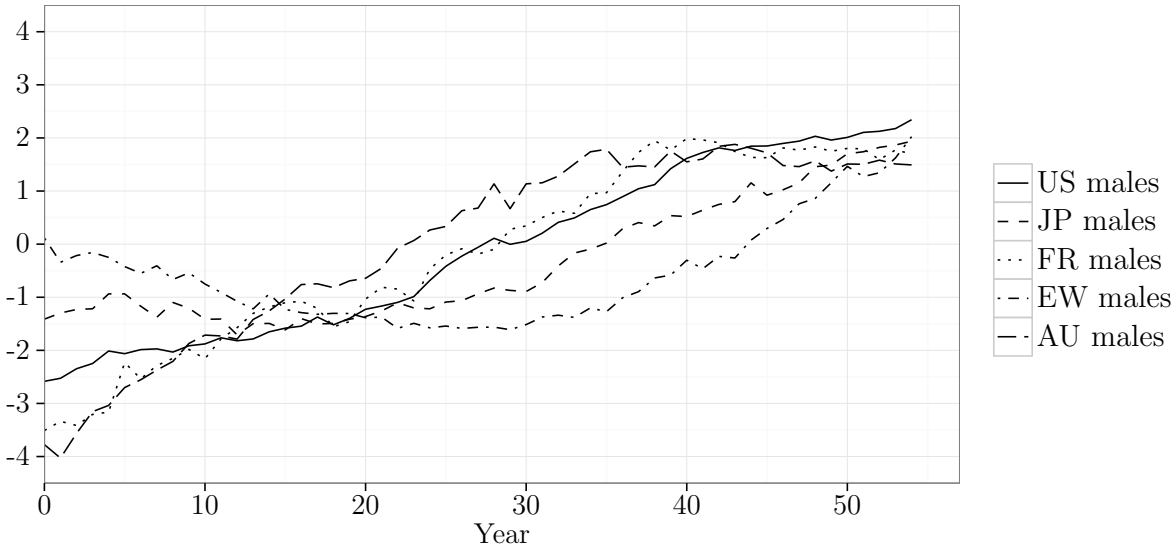
tal variance. Table 2.2 shows proportions for the remaining datasets: for all datasets except Japanese males the first principal component accounts for at least 70% of the total variance. For the sake of comparability, we will keep and compare the first PCs in every country and for every sex and will consider the second and the rest of the principal components as negligible.

Table 2.2: Proportion of the variance explained by the first principal component.

	US	JP	FR	EW	AU
Males	72%	64%	79%	71%	79%
Females	79%	91%	76%	74%	90%

When comparing the charts of the principal components, it is interesting to notice that the forms of the first components on each chart have a high degree of resemblance. The similarity becomes even more striking once the first PCs from every dataset are put on the same chart (Figures 2.4 and 2.5). To improve comparability, we multiplied some of the PCs by -1 as we know that such operations have no impact on the orthogonality or the variance accounted for by a given principal component.

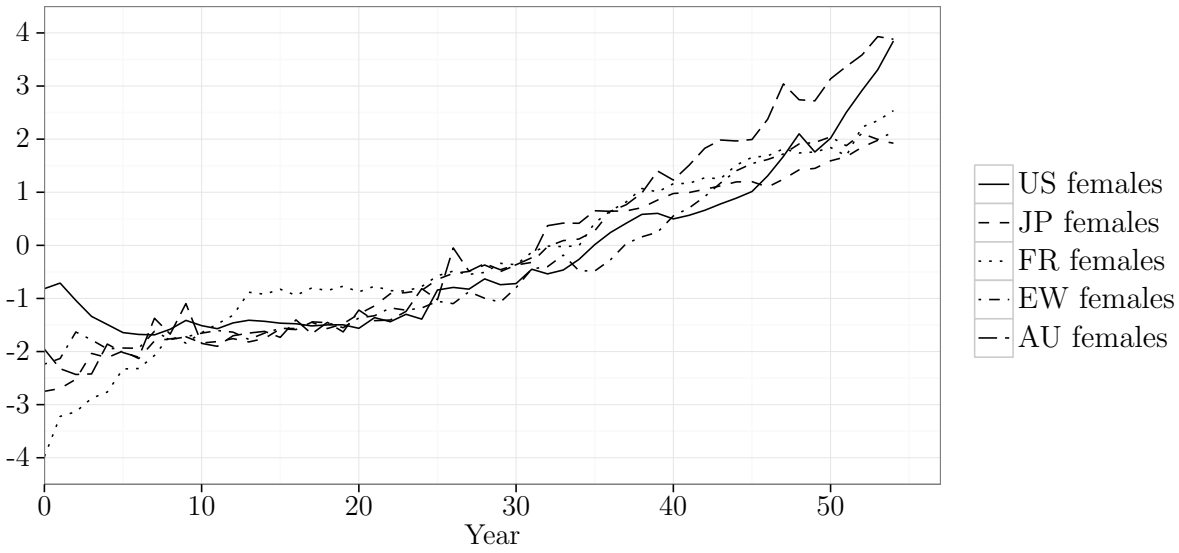
Figure 2.4: First PCs for each country, males.



Although the cause-specific mortality rates showed rather different development profiles depending on the country, we see that the patterns of the common stochastic factors, condensed using the principal component analysis, share a lot of similarities across the datasets. The resemblance is even more pronounced for the female datasets. So, we would like to measure the closeness between the first principal components of the common stochastic factors in a formal way, using the tools of the cointegration analysis. If some non-stationary variables are found to be cointegrated, then there exists their linear combination that is stationary even if each variable is not. Also, cointegrated variables

move together over the long term and are subject to the influence of the same common trends. If the first PCs from every dataset are cointegrated, this will mean that they are linked to each other in their long-term development.

Figure 2.5: First PCs for each country, females.



2.4.2 Cointegration between the first principal components of the common stochastic factors

To test if the first principal components of the common stochastic factors are cointegrated, we apply the Johansen maximum likelihood procedure (Johansen, 1988). This is possible only for the time series with the same number of observations. For this reason, we are obliged to cut the first PCs at the length of the shortest among them.

Under the null hypothesis H_0 of the *trace test* there are exactly r cointegrating relations among the data vector, in our case five first principal components from five countries, against the alternative hypothesis H_A that there are $n = 5$ cointegration relations. From the upper part of the Table 2.3 we see that H_0 is rejected for the $r = 0$ and is accepted for $r = 1$ at 5%, 2.5% and 1% significance level. So according to the trace test there is 1 cointegration relation for the male dataset. Under the null hypothesis H_0 of the *maximum eigenvalue test* there are exactly r cointegrating relations among the data vector against the alternative hypothesis H_A that there are $r+1$ cointegration relations. From the lower part of the Table 2.3 we see that H_0 is rejected for the $r = 0$ and is accepted for $r = 1$ at all shown significance levels. We can conclude that according to both tests there is 1 cointegration relation for the male dataset. Thus, although the common factors evolve stochastically, their development is driven by one long-run equilibrium relationship that remains stationary over the years.

Similar observations hold for the female datasets with the number of cointegrating relations being equal to 2 in the NT case and to 1 in the QT case (numerical results shown in Appendix B, Tables 2.6 and 2.7).

Table 2.3: Tests for the number of cointegration relations, male datasets

Trace statistics		Critical values (Case NT)			
r	Males	10%	5%	2.5%	1%
4	3.50	2.69	3.76	4.95	6.65
3	11.75	13.33	15.41	17.52	20.04
2	25.78	26.79	29.68	32.56	35.65
1	44.80	43.95	47.21	50.35	54.56
0	83.14	64.84	68.52	71.80	76.07
Max eigenvalue statistics		Critical values (Case NT)			
r	Males	10%	5%	2.5%	1%
4	3.50	2.69	3.76	4.95	6.65
3	8.25	12.07	14.07	16.05	18.63
2	14.03	18.60	20.97	23.09	25.52
1	19.02	24.73	27.07	28.98	32.24
0	38.34	30.90	33.46	35.71	38.77

A null hypothesis is accepted at α% significance level when the statistic is lower than the corresponding critical value. Hence, the hypothesis of r equal to 1 is accepted by both tests at 5%, 2.5% and 1%.

For both male and female datasets taken separately, Johansen tests for the form of the deterministic term indicate that either no trend (case NT) or a quadratic trend (case QT) should be included in the process. For male datasets, the results of cointegration test with a quadratic trend are inconclusive, and for this reason, we present the results for the NT case only. For female datasets, both forms of the deterministic term provide good model fit. The null hypotheses of no autocorrelation and normality of the residuals are not rejected for the male, as well as the female dataset (Table 2.4).

Table 2.4: Tests on residuals of the fitted VECM

Type of the test	Name of the test	p value		
		Males NT	Females NT	Females QT
		$r = 1$	$r = 2$	$r = 1$
Autocorrelation	Portmanteau (15 lags)	0.348	0.970	0.991
	Portmanteau (25 lags)	0.162	0.975	0.982
	Portmanteau (35 lags)	0.159	0.985	0.991
Normality	Skewness	0.335	0.653	0.180
	Kurtosis	0.311	0.967	0.603
	Both	0.309	0.935	0.339

We also tested the male and female datasets together for the cointegration between the first principal components of the common stochastic factors. As the shape of the first PCs is quite different between the male and female datasets, it is not surprising that we do not find any cointegration relation when we test the first principal components of the common stochastic factors for both sexes simultaneously.

As for the first principal components of the common stochastic factors tested separately for each sex, they are indeed cointegrated. This means that first, the cause-specific mortality rates, being cointegrated themselves, share some stochastic trends that are common to all causes. Then, once the corresponding common stochastic factors are explicitly extracted from the cause-specific mortality rates and condensed, they, in turn, also share some common stochastic trends, but on the next level, i.e. across different countries.

2.5 Discussion and conclusion

Although cause-specific mortality rates show different development patterns across countries, sexes and historical periods, as we deal with the death rates of the human species, it is reasonable to expect that similarities and common features also exist between these patterns. The steady decrease of the mortality rates that has been observed in many parts of the world for more than a century is also due to factors and effects that are universally present, albeit to a different extent. Among such factors one can cite medical advances, changes in lifestyles and nutrition, epidemics, and aging. Cointegration analysis in conjunction with the identification of the common stochastic factors as proposed by Gonzalo and Granger (1995) is then a practical tool that can efficiently help to elicit possible common long-term regularities and trends.

Among the factors influencing the development of mortality rates, biological aging has probably the most universal character as it increases vulnerability to all common causes of death (Olshansky et al., 2002; Hayflick, 2004). Although there is no single generally

accepted definition of aging, in simple terms it is usually defined as a progressive loss of normal body functions that leads to death (Holliday, 2004). In biological systems, aging is believed to be a stochastic process that is caused by the increasing loss of molecular fidelity that, with time, becomes superior to the repair capacity of the organisms. Also, age changes represent the greatest risk factor for age-associated diseases (Hayflick, 2004).

In spite of its paramount importance to virtually every aspect of human life, the process of aging is not yet fully understood and many different theories of aging exist, each having its merits and weaknesses (for a review thereof see Holliday (2004) and Jin (2010) including the references). Moreover, it is still not clear if aging can be measured since a reliable biomarker of aging is yet to be found (Butler et al., 2004). Should such measure be discovered, we could then study the development of mortality rates in function of advancement in aging. Instead, we presently find ourselves in a situation when a set of mortality rates evolving in time *must* reflect the effect of the aging processes, but it is not clear if and how this effect can be made explicit on the basis of the observed mortality rates only.

The intuition behind the application of the cointegration analysis to the death rates is to model simultaneously the development of several main cause-specific mortality rates and repeating the analysis for a number of populations. Should any patterns or trends be revealed that are common to all or the majority of the datasets included in the study, this could point to some fundamental processes that are proper to the human species. In Arnold and Sherris (2016) the authors found that only the endogenous mortality rates participated in the cointegration relations, i.e. in the long-term equilibrium states between the causes. It was assumed that this could indicate the link between the common stochastic trends shared by every cause-specific mortality rate and the processes of biological aging.

In this study, we explicitly measured the common stochastic factors as well as proposed an approach allowing us to make comparisons across countries. For this, we first, extracted the common stochastic factors from the sets of the cause-specific mortality rates in every country and for each sex. These are the factors that impact every cause-specific mortality rate in a particular country. Then, using the principal component analysis we condensed these factors and used the first principal components in the subsequent cross-country analysis. We have found that there exists at least one cointegration relation between the first principal components of the common stochastic factors from different countries, which means that they are also subject to some universal stochastic trends that deploy their impact in all countries. Consequently, these universal stochastic trends might reflect some intrinsic processes that occur in every country. Our results, combined with those of Arnold and Sherris (2016) tend to indicate that these universal stochastic trends are describing some features of the processes of biological aging, although at this point, we cannot state with certainty to what exactly these trends correspond. As we

now know that they exist and can be made explicit from the data, further research is needed in order to identify the mechanisms, likely biological, that are behind the observed behavior of the cause-specific mortality rates. Because of its universal and predominant character, aging is one of the possibilities that should not be omitted.

We believe that our results will bring forward the discussion on how to measure the biological processes of aging and the related research in the fields of demography, economics, biology or epidemiology. In addition, our study shows that as the same stochastic trends are present in all datasets, similar assumptions for the intra-cause dependence can be used across different countries. On the other hand, we saw that the first principal components of the common stochastic factors for males and females when put together were not cointegrated. This could indicate that there is an important divergence in the biology of aging for men and women and for this reason, different assumptions should be used for each sex when modelling the development of the cause-specific mortality rates. Our next steps will consist in using the results of the cointegration analysis to improve the cause-specific mortality forecasts.

Appendices

Appendix A. Age-standardization of the cause-specific mortality rates

To calculate the age-standardized death rates, we first, calculate the simple mortality rates for each country as the number of deaths by age, sex and cause divided by the mid-year population by age and sex. Next, we assume that the population age structure is constant over the whole observation period and is equal to the age structure of the US males population in 2007:

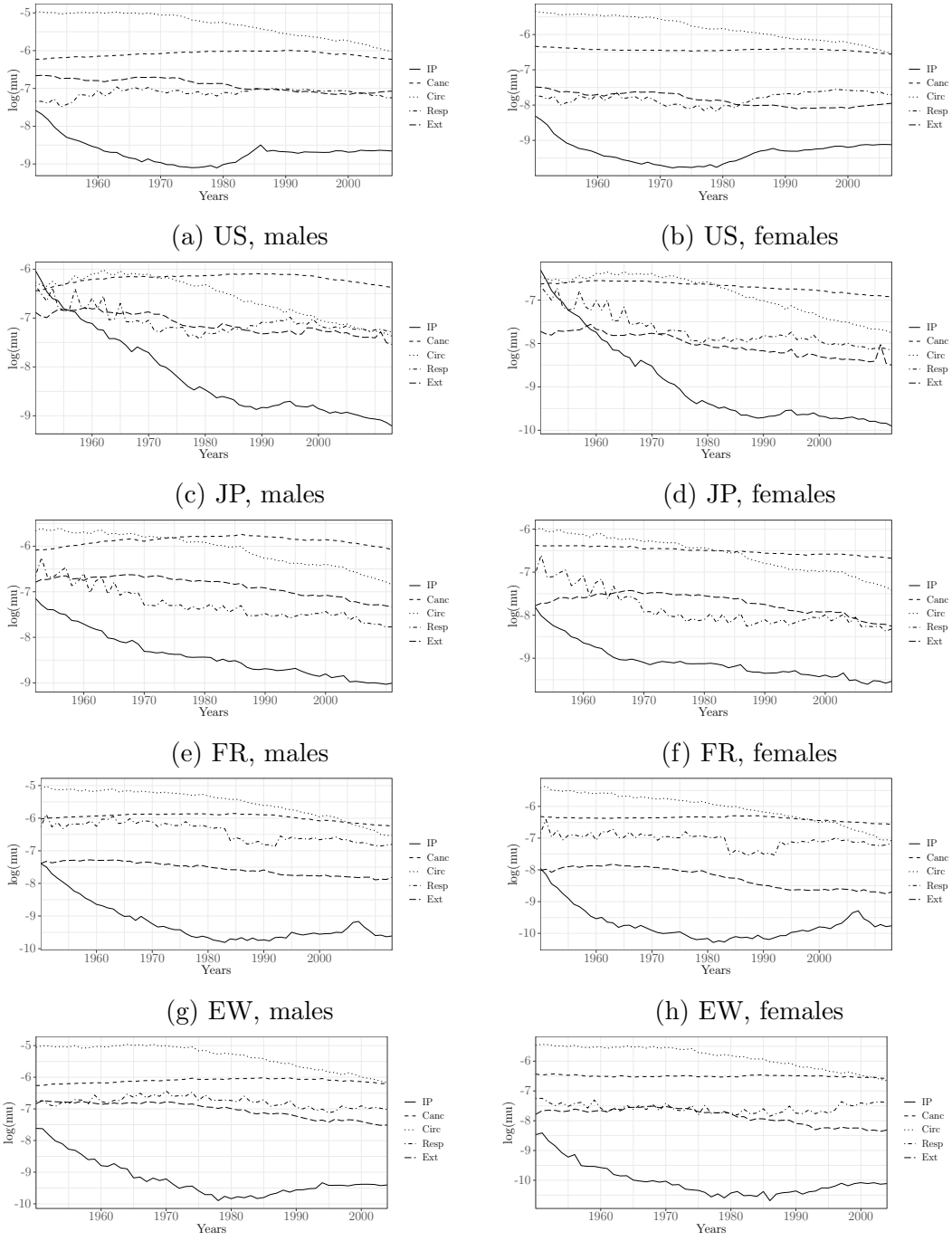
$$\begin{aligned} m_{x,t,d,s,c} &= d_{x,t,d,s,c}/l_{x,t,s,c} \\ d^*_{t,d,s,c} &= \sum_x m_{x,t,d,s,c} \times l_{x,2007,males,USA} \\ m^*_{t,d,s,c} &= d^*_{t,d,s,c}/l_{2007,males,USA} \end{aligned}$$

where

$$\begin{aligned} d_{x,t,d,s,c} &= \text{number of deaths at age } x\text{, in year } t\text{, for cause of death } d\text{,} \\ &\quad \text{gender } s \text{ and country } c\text{;} \\ l_{x,t,s,c} &= \text{mid-year population at age } x\text{, in year } t\text{, gender } s \text{ and country } c\text{;} \\ m_{x,t,d,s,c} &= \text{central death rate at age } x\text{, in year } t\text{, for cause of death } d\text{,} \\ &\quad \text{gender } s \text{ and country } c\text{.} \end{aligned}$$

Table 2.5: Age-standardized central death rates for selected years, x10^3

		Females			Males		
		1960	1980	2000	1960	1980	2000
US	I&P	0.0852	0.0630	0.1014	0.1888	0.1215	0.1686
	Cancer	1.6269	1.5857	1.5786	2.1394	2.4409	2.2483
	Circulatory	4.3148	2.9428	1.9744	6.8425	5.2315	3.2263
	Respiratory	0.4595	0.3292	0.5124	0.8670	0.8205	0.8512
	External	0.4516	0.3804	0.3058	1.1217	1.0368	0.7731
JP	I&P	0.4350	0.0842	0.0629	0.8160	0.2099	0.1426
	Cancer	1.4299	1.3156	1.1361	2.0147	2.1748	2.1092
	Circulatory	1.6850	1.3904	0.6263	2.2670	1.8094	0.8788
	Respiratory	0.9022	0.3804	0.3551	1.3673	0.6830	0.7941
	External	0.4644	0.3230	0.2476	1.0982	0.7577	0.7161
FR	I&P	0.1759	0.1085	0.0833	0.4487	0.2171	0.1502
	Cancer	1.6852	1.5080	1.3868	2.5972	3.0608	2.8610
	Circulatory	2.1922	1.6042	0.9415	3.3823	2.7013	1.6697
	Respiratory	0.8405	0.3023	0.3385	1.3661	0.6451	0.5930
	External	0.5020	0.5267	0.3623	1.1683	1.1520	0.8470
EW	I&P	0.0721	0.0387	0.0557	0.1764	0.0636	0.0708
	Cancer	1.7114	1.7764	1.5838	2.6498	2.8131	2.2807
	Circulatory	3.7128	2.7551	1.4769	5.7063	4.9203	2.5757
	Respiratory	0.8536	0.9398	0.8048	1.9318	1.8931	1.2785
	External	0.3823	0.3044	0.1780	0.6787	0.5561	0.4205
AU	I&P	0.0695	0.0296	0.0419	0.1518	0.0535	0.0843
	Cancer	1.4781	1.4868	1.4368	2.0721	2.3765	2.1570
	Circulatory	4.0123	3.0032	1.5502	6.5525	5.1796	2.4928
	Respiratory	0.5108	0.4362	0.6109	1.1511	1.1897	1.0071
	External	0.4747	0.4236	0.2625	1.0533	0.9189	0.6118



(a) US, males (b) US, females

(c) JP, males (d) JP, females

(e) FR, males (f) FR, females

(g) EW, males (h) EW, females

(i) AU, males (j) AU, females

Figure 2.6: Age-standardized central log-death rates by cause

Appendix B

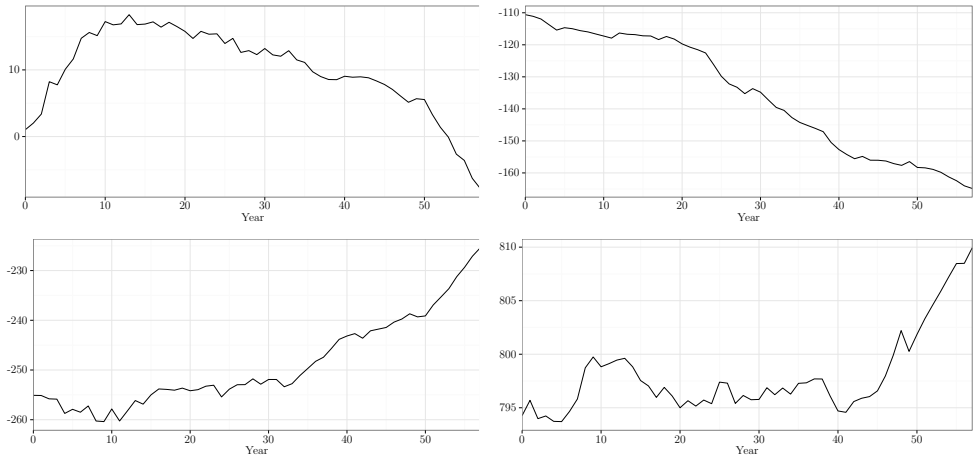


Figure 2.7: Common stochastic factors, US females.

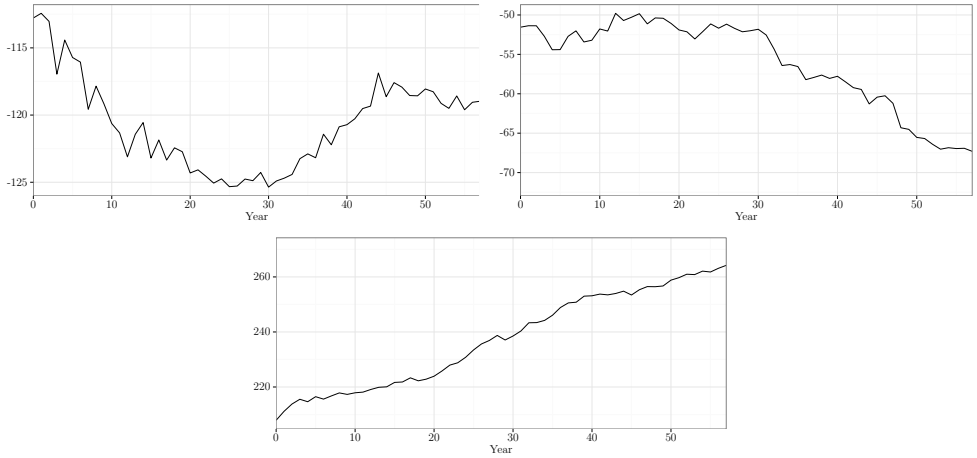


Figure 2.8: Common stochastic factors, JP males.

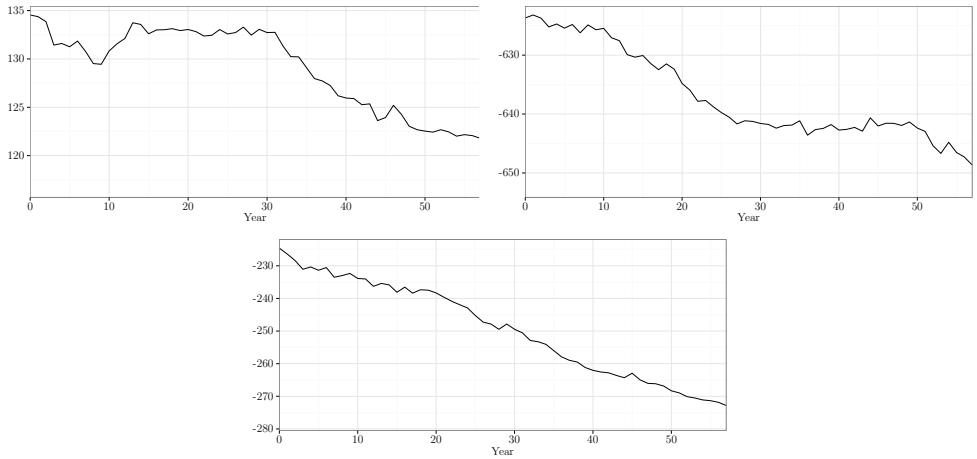
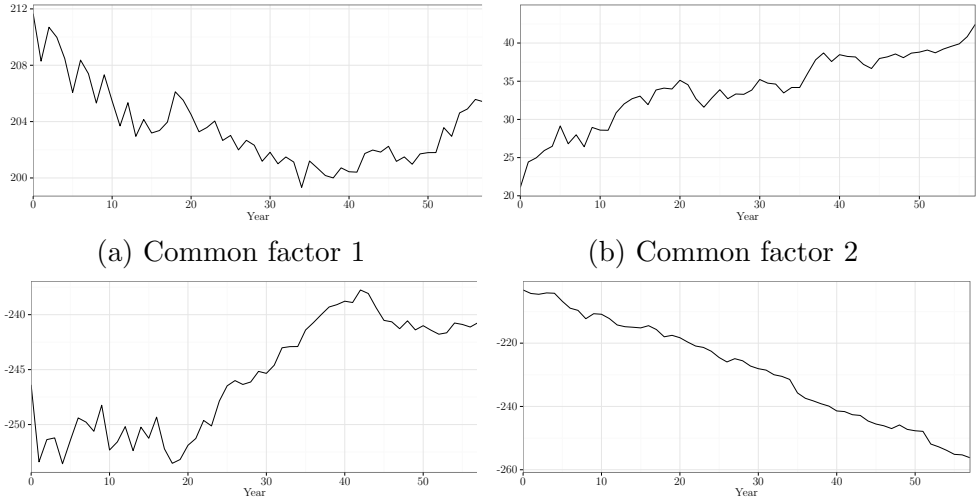


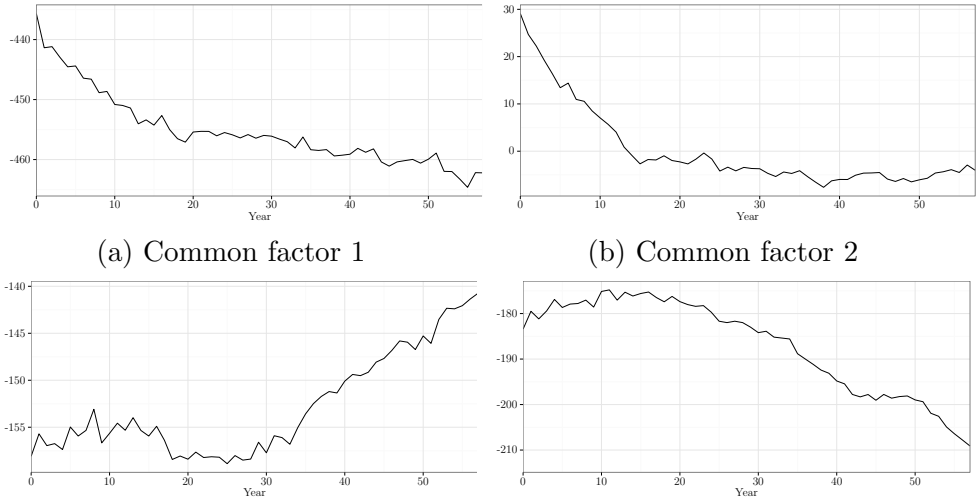
Figure 2.9: Common stochastic factors, JP females.



(a) Common factor 1 (b) Common factor 2

(c) Common factor 3 (d) Common factor 4

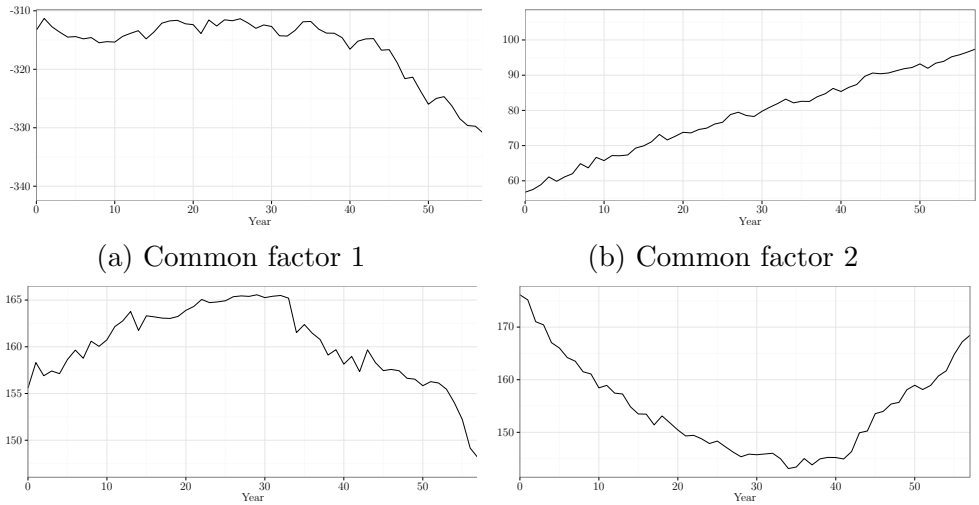
Figure 2.10: Common stochastic factors, FR males.



(a) Common factor 1 (b) Common factor 2

(c) Common factor 3 (d) Common factor 4

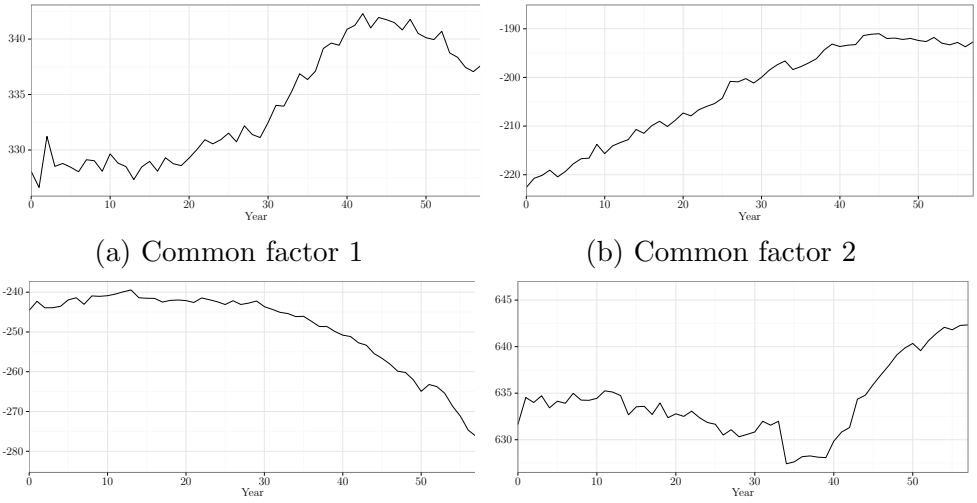
Figure 2.11: Common stochastic factors, FR females.



(a) Common factor 1 (b) Common factor 2

(c) Common factor 3 (d) Common factor 4

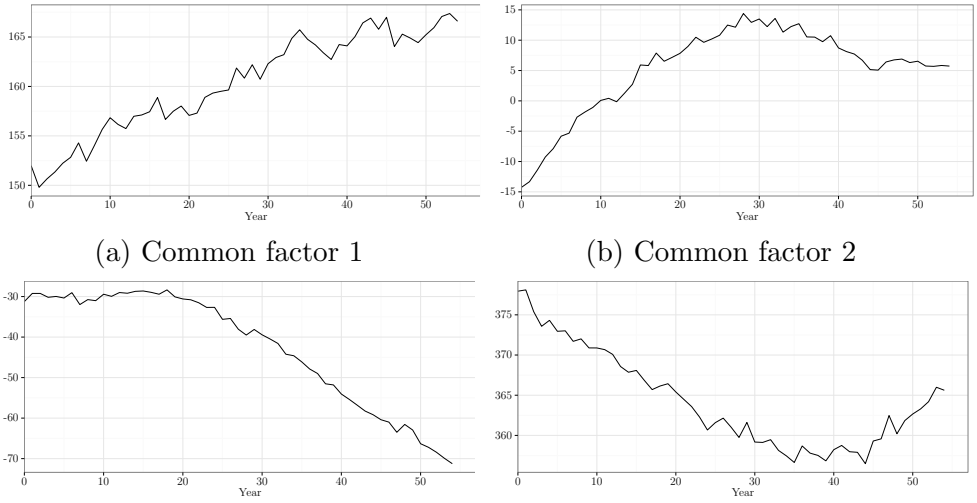
Figure 2.12: Common stochastic factors, EW males.



(a) Common factor 1 (b) Common factor 2

(c) Common factor 3 (d) Common factor 4

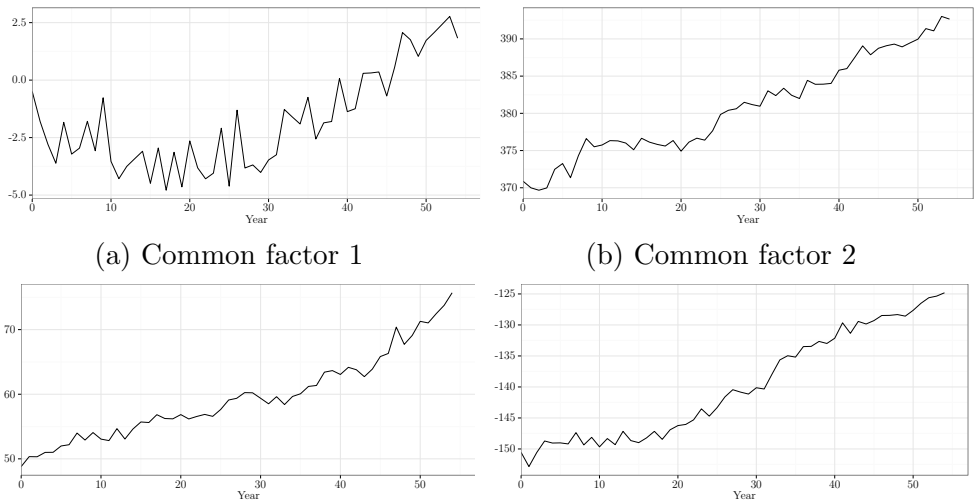
Figure 2.13: Common stochastic factors, EW females.



(a) Common factor 1 (b) Common factor 2

(c) Common factor 3 (d) Common factor 4

Figure 2.14: Common stochastic factors, AU males.



(a) Common factor 1 (b) Common factor 2

(c) Common factor 3 (d) Common factor 4

Figure 2.15: Common stochastic factors, AU females.

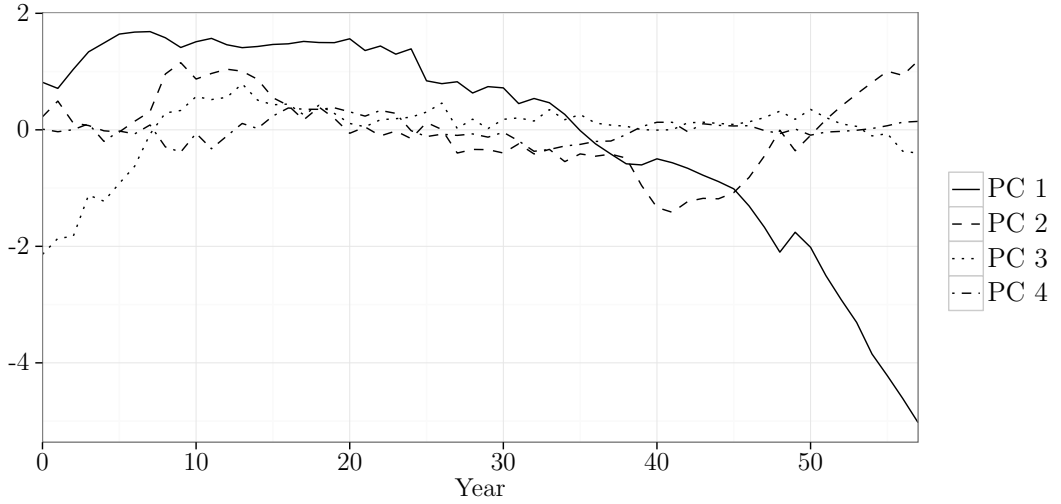


Figure 2.16: PCA applied to common factors, US females.

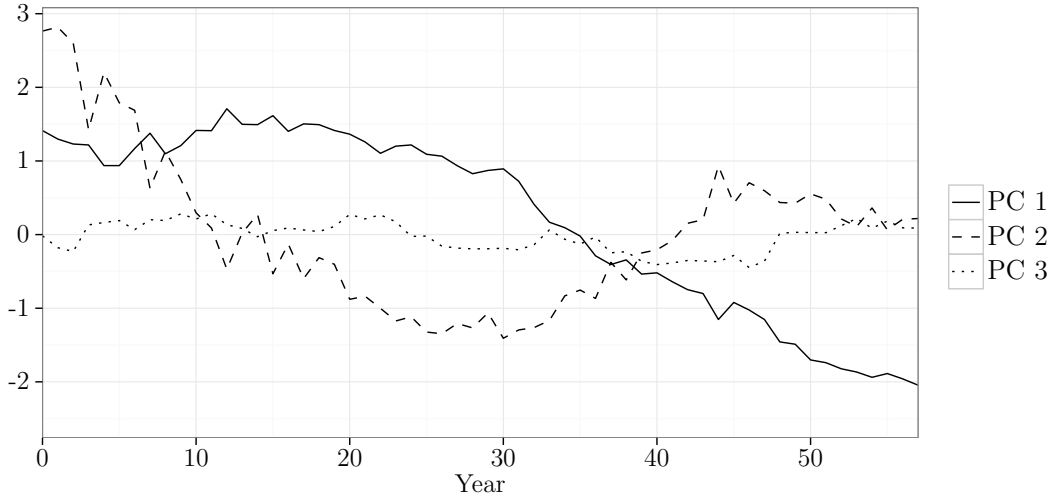


Figure 2.17: PCA applied to common factors, JP males.

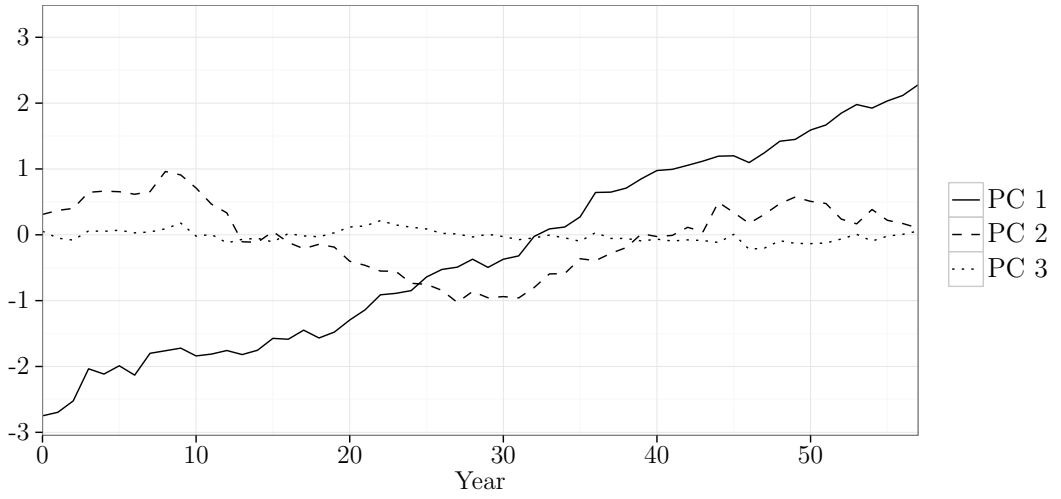


Figure 2.18: PCA applied to common factors, JP females.

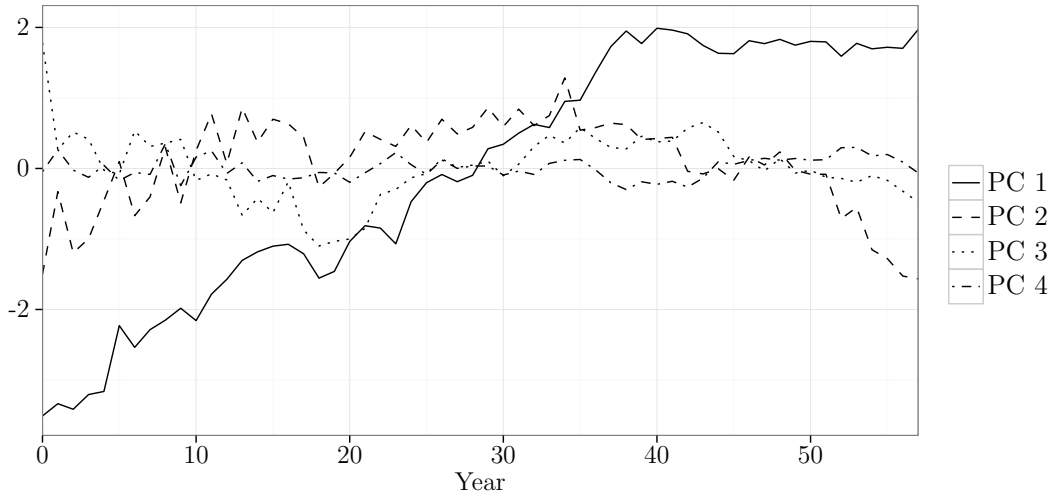


Figure 2.19: PCA applied to common factors, FR males.

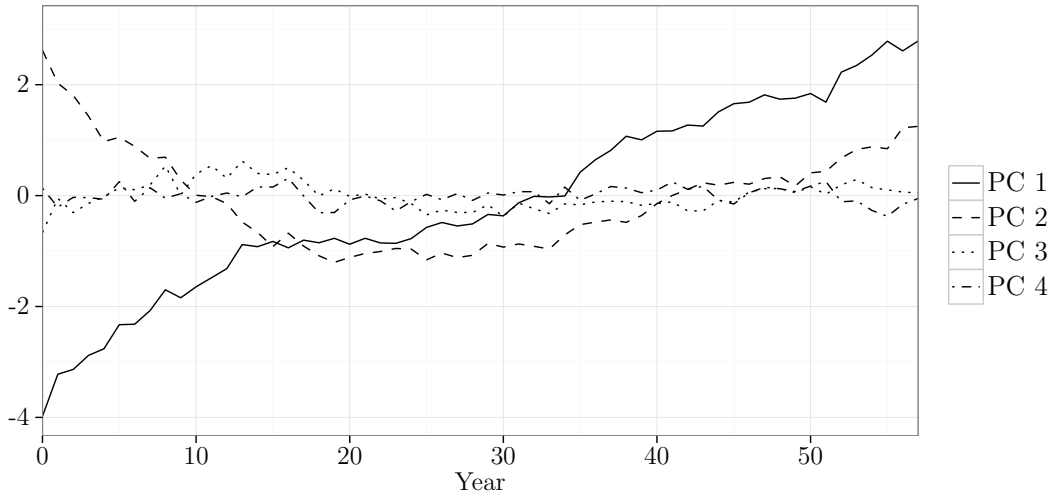


Figure 2.20: PCA applied to common factors, FR females

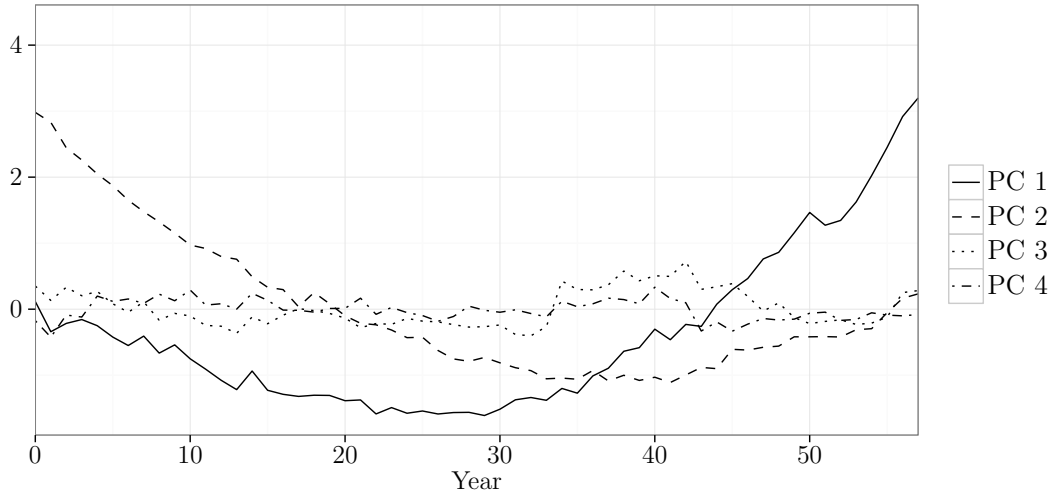


Figure 2.21: PCA applied to common factors, EW males.

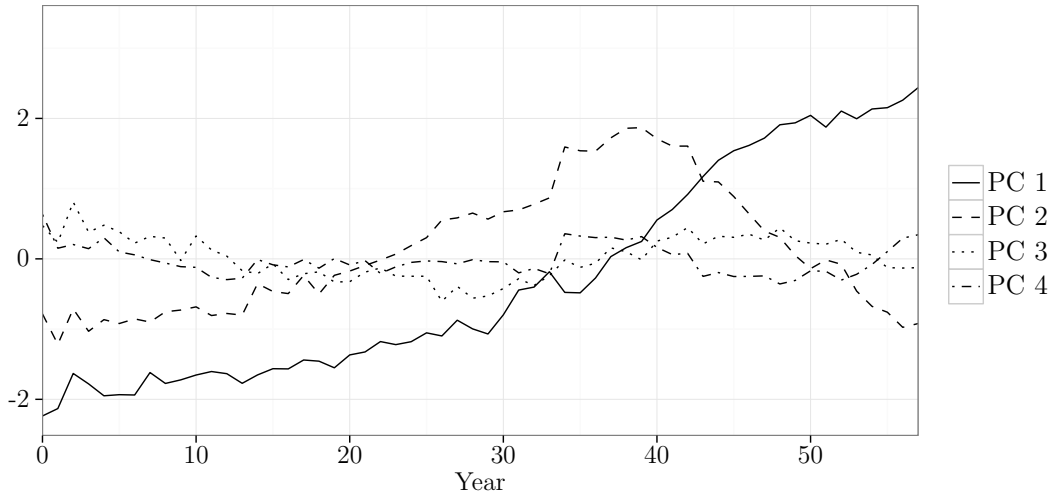


Figure 2.22: PCA applied to common factors, EW females.

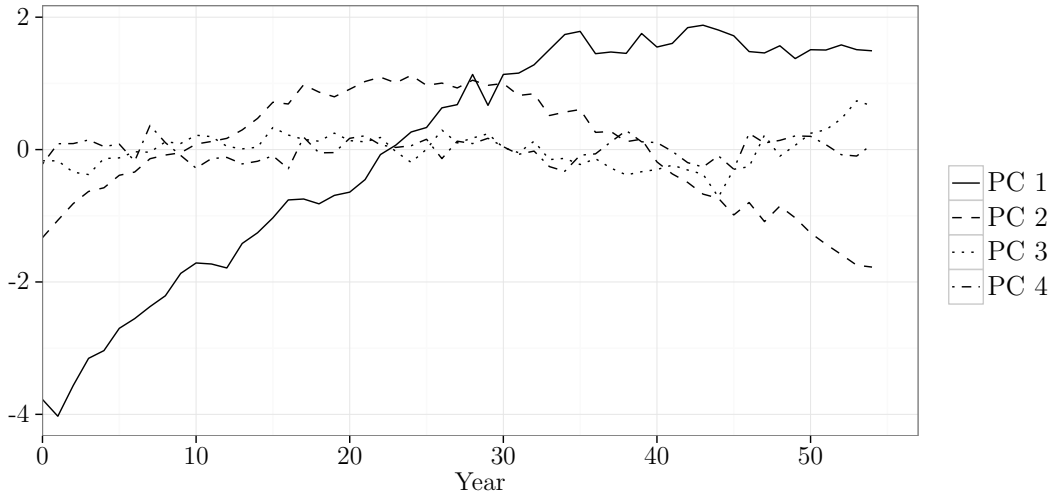


Figure 2.23: PCA applied to common factors, AU males.

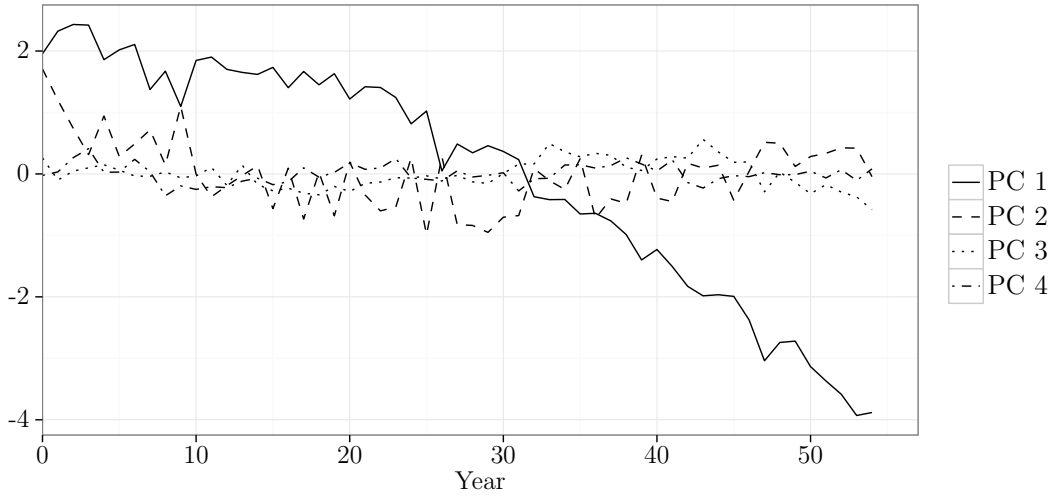


Figure 2.24: PCA applied to common factors, AU females.

Table 2.6: Tests for the number of cointegration relations, female datasets, case NT

Trace statistics		Critical values (NT)			
r	Females	10%	5%	2.5%	1%
4	0.79	2.69	3.76	4.95	6.65
3	6.16	13.33	15.41	17.52	20.04
2	18.18	26.79	29.68	32.56	35.65
1	52.39	43.95	47.21	50.35	54.56
0	101.41	64.84	68.52	71.80	76.07
Max eigenvalue statistics		Critical values (NT)			
r	Females	10%	5%	2.5%	1%
4	0.79	2.69	3.76	4.95	6.65
3	5.37	12.07	14.07	16.05	18.63
2	12.02	18.60	20.97	23.09	25.52
1	34.20	24.73	27.07	28.98	32.24
0	49.02	30.90	33.46	35.71	38.77

A null hypothesis is accepted at α% significance level when the statistic is lower than the corresponding critical value. Hence, the hypotheses of r equal to 2 is accepted at all significance levels by both tests.

Table 2.7: Tests for the number of cointegration relations, female datasets, case QT

Trace statistics		Critical values (QT)			
r	Females	10%	5%	2.5%	1%
4	0.05	2.57	3.74	4.85	6.40
3	5.37	16.06	18.17	20.13	23.46
2	15.75	31.42	34.55	36.94	40.49
1	35.22	50.74	54.64	57.79	61.24
0	83.85	73.40	77.74	80.74	85.78
Max eigenvalue statistics		Critical values (QT)			
r	Females	10%	5%	2.5%	1%
4	0.05	2.57	3.74	4.85	6.40
3	5.32	14.84	16.87	18.57	21.47
2	10.38	21.53	23.78	26.07	28.83
1	19.47	27.76	30.33	32.56	35.68
0	48.64	33.74	36.41	38.68	41.58

A null hypothesis is accepted at α% significance level when the statistic is lower than the corresponding critical value. Hence, the hypotheses of r equal to 1 is accepted at all significance levels by both tests.

Appendix C. Estimation of the common stochastic factors

The methodology developed by Gonzalo and Granger (1995) allows to estimate f_t as defined in (2.3) by imposing that

1. f_t be linear combinations of the original variables, in our case the cause-specific mortality rates:
$$\underset{k\times 1}{f_t} = \underset{k\times n}{B_1}\ \underset{n\times 1}{\mathbf{y_t}}\,, \tag{2.4}$$
2. the remaining stationary part $\tilde{\mathbf{y}}_\mathbf{t}$ does not have any permanent effect on $\mathbf{y_t}$.

Substituting (2.4) in (2.3), we obtain that $\tilde{\mathbf{y}}_\mathbf{t} = (I - A_1B_1)\mathbf{y_t}$. In other words, the stationary component $\tilde{\mathbf{y}}_\mathbf{t}$ is also a linear combination of the non-stationary variables $\mathbf{y_t}$ which is only possible for $\tilde{\mathbf{y}}_\mathbf{t} = A_2\beta'\mathbf{y_t} = A_2z_t$, where $z_t = \beta'\mathbf{y_t}$ is the cointegration relation. The authors show that the only linear combination of $\mathbf{y_t}$ such that $\tilde{\mathbf{y}}_\mathbf{t}$ has no long-run effect on $\mathbf{y_t}$ is

$$f_t = \underset{k\times n}{\alpha_\perp}\ \underset{n\times 1}{\mathbf{y_t}}\,, \tag{2.5}$$

where $\alpha'_\perp\alpha = 0$ and $k = n - r$.

The condition imposed in (2.4) not only helps to identify f_t, but also makes them observable by linking f_t to the original variables. Both conditions make f_t "a good candidate to summarize the long-run behavior of the original variables" (Gonzalo and Granger, 1995). The authors also show that these conditions allow identifying f_t up to a non-singular matrix multiplication to the left. The resulting factor model is:

$$\mathbf{y_t} = A_1f_t + A_2z_t, \tag{2.6}$$

where $f_t = \alpha_\perp\mathbf{y_t}$ and $z_t = \beta'\mathbf{y_t}$, and satisfies the following properties:

- The common factors f_t are not cointegrated.
- $Cov(\Delta f^*_{it}, z^*_{j,t-s}) = 0 \quad (i = 1, ..., k; j = 1, ..., n-k; s \geq 0)$,
where $\Delta f^*_{it} = \Delta f_{it} - E(\Delta f_{it} \mid \text{lags}(\Delta\mathbf{y_{t-1}}))$ and $\Delta z^*_{it} = \Delta z_{it} - E(\Delta z_{it} \mid \text{lags}(\Delta\mathbf{y_{t-1}}))$

The second property is another way of saying that z_t does not cause f_t in the long run. It also follows that any alternative definition of f_t will vary only by I(0) components and therefore will be cointegrated.

To solve for the coefficients of (2.2), following Johansen (1988) we concentrate the model by regressing $\Delta\mathbf{y}_t$ and $\mathbf{y}_{t-1}$ on $(\Delta\mathbf{y}_{t-1}, ..., \Delta\mathbf{y}_{t-p+1})$ which gives the residuals R_{0t} and R_{1t} respectively, as well as the residual product matrices S_{ij}

$$S_{ij} = T^{-1}\sum_{j=1}^{T} R_{0t}R_{1t},\ \ i, j = 0, 1. \tag{2.7}$$

The concentrated model is then

$$R_{0t} = \alpha\beta' R_{1t} + \epsilon_t, \tag{2.8}$$

and β is estimated using the reduced-rank regression from the following eigenvalues problem

$$\mid \lambda S_{11} - S_{10} S_{00}^{-1} S_{01} \mid = 0. \tag{2.9}$$

After ordering the eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_n$ and corresponding eigenvectors $\hat{V} = (\hat{v}_1, \hat{v}_2, ..., \hat{v}_n)$, the maximum likelihood estimates of the cointegration term of the VECM are obtained as $\hat{\beta} = (\hat{v}_1, \hat{v}_2, ..., \hat{v}_r)$ and $\hat{\alpha} = S_{01}\hat{\beta}$.

We proceed in a similar way to estimate $\alpha_\perp$ by solving the equation

$$\mid \lambda S_{00} - S_{01} S_{11}^{-1} S_{10} \mid = 0, \tag{2.10}$$

which gives the eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_n$ and corresponding eigenvectors $\hat{M} = (\hat{m}_1, \hat{m}_2, ..., \hat{m}_n)$, normalized such that $\hat{M}' S_{00} \hat{M} = I$. The $\alpha_\perp$ that defines f_t is then

$$\alpha_\perp = (\hat{m}_{r+1}, ..., \hat{m}_n). \tag{2.11}$$

We can see that the set of the common factors f_t has indeed a lower number of dimensions than the initial data, but if r is equal to 1 or 2, further reduction of dimensionality may be needed.

Bibliography

Arnold, S. and V. Glushko, 2021, Short- and long-term dynamics of cause-specific mortality rates using cointegration analysis, *North American Actuarial Journal*.

Arnold, S. and M. Sherris, 2013, Forecasting Mortality Trends allowing for Cause-of-Death Mortality Dependence, *North American Actuarial Journal*, 17(4):273–282.

Arnold, S. and M. Sherris, 2015, Causes-of-Death Mortality: What Do We Know on their Dependence?, *North American Actuarial Journal*, 19(2):116–128.

Arnold, S. and M. Sherris, 2016, International Cause-Specific Mortality Rates: New Insights from a Cointegration Analysis., *ASTIN Bulletin*, 46(1):9–38.

Booth, H. and L. Tickle, 2008, Mortality modelling and forecasting: A review of methods, *Annals of Actuarial Science*, 3:3–43.

Butler, R. N., R. Sprott, H. Warner, J. Bland, R. Feuers, M. Forster, H. Fillit, S. M. Harman, M. Hewitt, M. Hyman, K. Johnson, E. Kligman, G. McClearn, J. Nelson, A. Richardson, W. Sonntag, R. Weindruch, and N. Wolf, 2004, Biomarkers of Aging: From Primitive Organisms to Humans, *Journal of Gerontology: BIOLOGICAL SCIENCES*, 59A(6):560–567.

Cairns, A., 2013, Modeling and management of longevity risk, In P. B. Hammond, R. Maurer, and O. S. Mitchell, editors, *Recreating Sustainable Retirement: Resilience, Solvency, and Tail Risk*, pages 71–88. Oxford: Oxford University Press.

Carnes, B. A., L. R. Holden, S. J. Olshansky, T. M. Witten, and J. S. Siegel, 2006, Mortality partitions and their relevance to research on senescence, *Biogerontology*, 7:183–198.

Darkiewicz, G. and T. Hoedemakers, 2004, How the Co-Integration Analysis Can Help in Mortality Forecasting, Dtew research report or 0406, K.U. Leuven.

Debón, A., F. Montes, and R. Sala, 2006, A Comparison of Models for Dynamic Life Tables. Application to Mortality Data from the Valencia Region (Spain), *Lifetime Data Analysis*, 12:223–244.

Gaille, S. and M. Sherris, 2011, Modeling Mortality with Common Stochastic Long-Run Trends, *The Geneva Papers on Risk and Insurance - Issues and Practice*, 36(4):595–621.

Gonzalo, J. and C. Granger, 1995, Estimation of common long-memory components in cointegrated systems, *Journal of Business & Economic Statistics*, 13(1):27–35.

Hamilton, J. D., 1994, *Time Series Analysis*. Princeton: Princeton University Press.

Hayflick, L., 2004, "Anti-Aging" Is an Oxymoron, *Journal of Gerontology: BIOLOGICAL SCIENCES*, 59A(6):573–578.

Hogg, R. V., J. McKean, and A. T. Craig, 2014, *Introduction to Mathematical Statistics*. Boston, Mass.: Pearson Education.

Holliday, R., 2004, The Multiple and Irreversible Causes of Aging, *Journal of Gerontology: BIOLOGICAL SCIENCES*, 59A(6):568–572.

Hunt, A. and D. Blake, 2015, Modelling longevity bonds: Analysing the swiss re kortis bond, *Insurance: Mathematics and Economics*, 63:12–29.

Jin, K., 2010, Modern Biological Theories of Aging, *Aging and Disease*, 1(2):72–74.

Johansen, S., 1988, Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control*, 12:231–254.

Johansen, S., 1995, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. New York: Oxford University Press.

Jolliffe, I. T., 2002, *Principal Component Analysis*. New York: Springer.

Lazar, D. and M. Denuit, 2009, A multivariate time series approach to projected life tables, *Applied Stochastic Models in Business and Industry*, 25:806–823.

Li, H. and Y. Lu, 2017, Coherent forecasting of mortality rates: a sparse vector-autoregression approach, *ASTIN Bulletin*, 47(2):563–600.

Li, J. S.-H. and M. R. Hardy, 2011, Measuring Basis Risk in Longevity Hedges, *North American Actuarial Journal*, 15(2):177–200.

Lütkepohl, H., 2005, *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.

Njenga, C. and M. Sherris, 2011, Longevity Risk and the Econometric Analysis of Mortality Trends and Volatility, *Asia-Pacific Journal of Risk and Insurance*, 5(2).

Olshansky, S. J., L. Hayflick, and B. Carnes, 2002, Position Statement on Human Aging , *Journal of Gerontology: Biological sciences*, 57A(8):B292–B297.

Salhi, Y. and S. Loisel, 2017, Basis risk modelling: a co-integration based approach, *A Journal of Theoretical and Applied Statistics*, 51(1):205–221.

Stock, J. H. and M. W. Watson, 1988, Testing for common trends, *American Statistical Association*, 83(404):1097–1107.

World Health Organization, 2016, WHO Mortality Database. `http://www.who.int/whosis/mort/download/en/index.html`.

Yang, S. and C.-W. Wang, 2013, Pricing and securitization of multi-country longevity risk with mortality dependence, *Insurance: Mathematics and Economics*, 52:157–169.

Yang, S., J. C. Yue, and H.-C. Huang, 2010, Modeling longevity risks using a principal component approach: A comparison with existing stochastic mortality models, *Insurance: Mathematics and Economics*, 46:254–270.

Zhou, R., Y. Wang, K. Kaufhold, J. Siu-Hang Li, and K. Seng Tan, 2014, Modeling Period Effects in Multi-Population Mortality Models: Applications to Solvency II, *North American Actuarial Journal*, 18(1):150–167.

Chapter 3

Forecasting Cause-Specific Mortality Rates Using the Insights from the Cointegration Analysis

3.1 Introduction

The difficulty to project the mortality rates due to the ongoing increases of life expectancy and the uncertainty related to these increases is a well-known topic in actuarial science. Due to its importance for pension providers and social security systems, this question has attracted a lot of attention from researchers and practitioners alike. A review of the existing models can be found, for example, in Booth and Tickle (2008) including their references. Still, new approaches continue to be developed with an objective to improve our understanding of the mortality rates' evolution as well as the quality of their forecasts.

Among the stochastic mortality models, the Lee-Carter model (Lee and Carter, 1992) together with its various extensions is possibly one of the most widely known and used. Initially applied to the all-cause mortality in a single population context, it was enhanced by different authors to take into account the cohort effects as well as to incorporate multiple populations: an impressive genealogy of the models is provided in Cairns (2013) and an overview of the multipopulational extensions in Villegas et al. (2017).

Originally, modelling efforts concerned the all-cause mortality for the simple reason that these were the only data at hand. As the amount of available statistics grew with time, modelling the cause-specific mortality rates became possible. In this regard, there are two aspects to consider. On the one hand, it seems a natural step to disaggregate the total mortality by causes of death when one knows that the cause-specific mortality rates had vastly different development patterns in the past. Also, as mentioned in Tabeau et al. (1999), while it is practically impossible to make empirically valid assumptions for the total mortality due to the extremely large number of the mortality determinants,

trends in the cause-specific mortality can be linked with the risk factors of diseases. Further, as the prevalence of diseases heavily depends on age, aggregate mortality rates forecasts for a specific age group may not be complete. Overall, as Gutterman and Vanderhoof (1998) put it, "we must be able to decompose past trends and recognize their causes to help us feel our way to the future" which is not possible if one works with the aggregate mortality rates. On the other hand, the disaggregated approach has its practical and theoretical drawbacks: the effects of misclassification of deaths by cause, limited length of the available time series, inferior data quality and the dependence structure between the causes are cited among reasons that did not permit to obtain forecasts superior to those obtained for the aggregated mortality rates (Wilmoth, 1995; Booth and Tickle, 2008).

However, the forecasts of the cause-specific mortality rates are needed not only as a path leading to the aggregate mortality, but also in their own value for many purposes such as estimation of the health care and disability costs in the ageing populations (Tabeau et al., 1999). While remaining conscious of the difficulties the cause-specific approach entails, we align with the view that analyzing and modelling the cause-specific mortality rates can improve our understanding of the past and improve our forecasts for the future.

To overcome the problem of the limited and sometimes volatile number of observations by sex, age, and cause, we propose to apply the multipopulational approach in a context of the cause-specific mortality rates. Lyu et al. (2020) note that the cause-specific mortality rates have not yet been modelled in a multipopulational setup to the same extent as this has been the case for the all-cause mortality rates. We believe that the multipopulational modelling for the cause-specific mortality rates could be justified for two reasons. First, one could expect that similar to the all-cause mortality, there will be resemblance in the development patterns of the cause-specific mortality rates between countries having comparable socioeconomic conditions, medical advances quickly spreading across the developed countries. Second, the multipopulation approach ensures that the forecasts built for the countries included into the group are converging. As there is no reason to expect that the future mortality rates from cancer will be substantially different between France and Spain, for example, the country- and cause-specific mortality rates should not be diverging in the long term, and the multipopulational modelling is a way to ensure the coherence of the forecasts.

The question is then how to choose the countries to be included into the group or, in other words, what can serve as a measure of the sufficient coherence between the experiences of two countries? If certain countries have similar experience in the past, then they should be modelled together in a multipopulational context. Li and Lee (2005) define an explanation ratio for the augmented common factor model. They suggest including countries in the group if the corresponding ratios are "large enough" while stressing that this criteria is left intentionally vague and should be tempered by judgment. In Lyu

et al. (2020) the authors are confronted with a similar task of comparing cause-specific experiences of three countries that they propose to solve by using a beta-convergence test from the growth literature. This test verifies whether the cause-specific mortality rates in different countries tend to the same level and improve at the same speed. Both approaches arrive at simple "yes/no" answers that summarize several decades of age-specific observations, and as such, are necessarily simplistic. We propose a complementary approach that consists in using the cointegration analysis to assess if the cause-specific mortality rates in different countries exhibited coherent development in the past. Indeed, should the cancer mortality rates from France and Spain, for example, be cointegrated, this would mean that they were linked in their development in the past and are not expected to wander from each other or diverge in the long run. Should a common development pattern be revealed between these two countries, this would justify using the cause-specific mortality rates in a multipopulational model that explicitly imposes a common trend on the country-specific rates, such as the one proposed by Li and Lee (2005). Specifically, we will analyze the cause-specific mortality experience of several countries by applying the cointegration analysis to the country-specific mortality time trends by cause of death. We believe that this new angle will provide additional evidence whether countries should be modelled together on the cause-specific level or not. For the future, finding a more nuanced answer to the difficult question of comparing country-specific mortality experiences expressed in vast matrices of observations seems to be an interesting and promising research topic.

At the second stage, we verify if the Li-Lee model built for the countries having the cointegrated cause-specific mortality rates allows improving the forecasts in comparison with a benchmark approach, that is the Lee-Carter model. Indeed, we observe that for the male as well as for the female cause-specific mortality rates the Li-Lee model helps to improve the forecasts for most of the countries included in our study. The cointegration analysis can hence deliver a helpful answer to the question regarding the countries to be included into a multipopulational model for the cause-specific mortality rates.

The paper is organized as follows: in Section 3.2 we briefly present the data used in the study regarding causes and countries chosen for the study. A brief theoretical review of the mortality models that will be used as well as of the cointegration analysis is exposed in Section 3.3. The results of the cointegration tests along with the comparison of the forecasts are presented in Section 3.4. Section 3.5 concludes.

3.2 Data

Below we briefly present the main steps of the data preparation process that follows a path similar to the one described in Arnold and Sherris (2016):

- The data comes from the WHO Mortality Database (World Health Organization, 2020) that collects the mid-year population and the death numbers by country, year, sex, age group, and cause of death since its creation in 1950.

- As the WHO database splits the death numbers according to the *primary* cause of death, we will ignore the potential presence of the secondary cause, third cause etc. Also, we would have to signigicanlty change our approach in order to incorporate the information on the secondary cause of death, for example. For this reason, our results would not hold in presense of several causes leading to death.

- In order to limit the extent to which countries' experiences differ due to the social-economic factors, we chose the five most populated Western European countries participating in the database from its onset: France, Italy, Netherlands, Spain, and England and Wales, subsequently shortened to FR (1), IT (2), NL (3), SP (4), and EW (5) respectively. In contrast to the preceeding chapters where we wanted to have a variety of experiences, here we want the countries' conditions be as close as possible.

- We are going to build a model involving the data from different countries and for this reason, we are obliged to cut the observations for all countries at the shortest available observation window. This corresponded to the time period 1952-2014 at the moment when the data were retrieved (October 2020).

- WHO Mortality database provides the data for the age groups: "deaths at 0 years", "at 1", "at 2", "at 3", "at 4", "5-9 years", ..., "90-94 years", and finally "deaths at 95 years and above". To deal with the age groups, we, first, created two new age groups by grouping together the ages from 1 to 4 and 85 and above. Second, we distributed the number of deaths at unspecified age proportionally among the all known age groups.

- Causes of death are clustered into five main groups: infectious and parasitic diseases, cancer, diseases of the circulatory system, diseases of the respiratory system, and external causes. We define these groups of causes of death under the different versions of the International Classification of Diseases (ICD) as shown in Table 3.1. Naturally, there is more than one way to perform such grouping, and for the sake of comparability with earlier studies (Arnold and Sherris, 2016; Arnold and Glushko, 2021a), we keep these five groups of causes of death. Cause-specific mortality rates for selected years are shown on Figure 3.2 in the Appendix.

Table 3.1: Main groups of causes of death according to the versions of the *International Classification of Diseases*.

Causes of death	ICD 7	ICD 8	ICD 9	ICD 10
IP	001-138	001-136	001-139	A00-B99
Cancer	140-239	140-239	140-239	C00-D48
Circulatory	400-468	390-458	390-437	I00-I99
Respiratory	470-527	460-519	460-519	J00-J98
External	E810-E999	E810-E999	E800-E999	V00-Y89

- We calculate the cause-specific central death (mortality) rates as the number of deaths by age, sex, and cause divided by the mid-year population by age and sex:

$$m_{x,t}^{d,s,c} = d_{x,t}^{d,s,c} / l_{x,t}^{s,c},$$

with

$d_{x,t}^{d,s,c}$ = number of deaths at age x, in year t, for cause of death d, gender s and country c;

$l_{x,t}^{s,c}$ = mid-year population at age x, in year t, gender s and country c;

$m_{x,t}^{d,s,c}$ = central death rate at age x, in year t, for cause of death d, gender s and country c.

- We apply the comparability ratios to ensure the comparability between the observations under the different versions of the ICD. In this way, the discontinuities between the observation periods are removed. Indeed, a comparability ratio makes the average mortality rates of the last two years of a classification equal to the average mortality rates of the first two years of the following classification. Once the mortality rates in every classification are divided by the comparability ratio(s) linking this classification to the previous one(s), observations become comparable across the different versions of the ICD.[1]

- For our analysis, we will use the data for the age groups 20 years and older as the cause-specific data for the younger age groups are known to be sparse. Similar approach was taken in Lyu et al. (2020).

- All equations were estimated for the natural logarithms of the cause-specific mortality rates:

$$ln(m_{x,t}^{d,s,c}),$$

[1]Further details on the data preparation process involving comparability ratios can be found in Arnold and Sherris (2015).

where
$x \in \{20-24, 25-29, ..., 80-84, 85+\}, t \in \{1952-2014\}, d \in \{IP, Canc, Circ, Resp, Ext\}$,
$s \in \{Males, Females\}, c \in \{FR, IT, NL, SP, EW\}$.

3.3 Theoretical framework

3.3.1 Lee-Carter model for the cause-specific mortality rates

As already mentioned, the Lee-Carter model or simply LC (Lee and Carter, 1992) is possibly the most widely used model for the mortality rates and for this reason, it often serves as a comparison point in studies aiming to improve the quality of a forecast. With this objective in mind, we apply the Lee-Carter model to the mortality rates separately for each cause, sex, and country :

$$ln(m_{x,t}^{d,s,c}) = \alpha_x^{d,s,c} + \beta_x^{d,s,c} k_t^{d,s,c} + \epsilon_x^{d,s,c}(t) \tag{3.1}$$

To ease the notation in what follows, we will sometimes omit the indexes d, s, c. Following the approach proposed by Brouhns et al. (2002), we estimate the parameters α_x, β_x and k_t by maximizing the log-likelihood based on the Poisson model for the number of deaths:

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{k}) = \sum_{x,t} (d_{x,t}(\alpha_x + \beta_x k_t) - l_{x,t} exp(\alpha_x + \beta_x k_t)) + \text{constant}, \tag{3.2}$$

and applying the constrains $\sum_x \hat{\beta}_x = 1$, $\sum_t \hat{k}_t = 0$.

The Box-Jenkins methodology is used to generate the appropriate ARIMA time series model and project k_t.

3.3.2 Li-Lee model for the cause-specific mortality rates

As an extension of the Lee-Carter model, Li and Lee (2005) proposed the augmented common factor model for the multi-population context:

$$ln(m_{x,t}^{d,s,c}) = \alpha_x^{d,s,c} + B_x^{d,s} K_t^{d,s} + \beta_x^{d,s,c} k_t^{d,s,c} + \epsilon_x^{d,s,c}(t), \tag{3.3}$$

where $B_x^{d,s} K_t^{d,s}$ is the common factor and $\beta_x^{d,s,c} k_t^{d,s,c}$ is the population-specific factor.

Like the Lee-Carter model, the $\alpha_x^{d,s,c}$ are obtained as the average mortality rates, in this case, by cause:

$$\alpha_x^{d,s,c} = \frac{\sum_t ln(m_{x,t}^{d,s,c})}{T} \tag{3.4}$$

The remaining model parameters are defined in two steps. First, the $B_x^{d,s}$ and $K_t^{d,s}$ are obtained from applying the ordinary LC model to the aggregate group mortality rates. In this way, the common trend of mortality change is identified. Second, the population-specific factor is obtained from the residual matrix $ln(m_{x,t}^{d,s,c}) - \alpha_x^{d,s,c} - B_x^{d,s}K_t^{d,s}$ to which the strategy of the ordinary LC model is applied. Li and Lee (2005) suggest to assess the performance of this model for a particular population by constructing the explanation ratio as follows:

$$R(c) = 1 - \frac{\sum_{x,t}(ln(m_{x,t}^{d,s,c}) - \alpha_x^{d,s,c} - B_x^{d,s}K_t^{d,s} - \beta_x^{d,s,c}k_t^{d,s,c})^2}{\sum_{x,t}(ln(m_{x,t}^{d,s,c}) - \alpha_x^{d,s,c})^2}. \tag{3.5}$$

The authors propose to include the population c to the studied group if the explanation ratio $R(c)$ is "large enough", leaving the criteria intentionally vague as other considerations may play a role. For example, a country may not be a part of the group in the past, but its mortality can be expected to follow a similar path in the future. In the current study, we propose to use the cointegration analysis which we briefly present below to assess if two countries should be modelled together using the Li-Lee approach.

Like the Lee-Carter model, we use the Box-Jenkins methodology to generate the appropriate ARIMA time series model and build the forecasts for the K_t and k_t^c.

3.3.3 Cointegration analysis as a measure of coherence

According to Engle and Granger (1987), the time series $\mathbf{y}_t$ that consist of the n non-stationary variables $(y_{1t}, y_{2t}, ...y_{nt})'$ with $t = 1, ..., T$ are said to be *cointegrated of order 1* or I(1) when there exists a linear combination of its elements $\beta'\mathbf{y}_t$ that is stationary or I(0):

$$\beta_1 y_{1t} + \beta_2 y_{2t} + ... + \beta_2 y_{nt} = z_t, \tag{3.6}$$

where z_t is a stationary variable of stochastic deviations. Then $\beta' = (\beta_1, \beta_2, ..., \beta_n)$ is said to be a cointegrating vector and $\beta'\mathbf{y}_t$ is a cointegration relation.

Should such a linear combination exist, this means that non-stationary variables remain linked to each other in their long-term development. It is also possible that there is more than one cointegrating vector, so that β becomes a matrix. Each cointegration relation is then linearly independent from the others.

Arnold and Sherris (2015, 2016) studied the age-standardized cause-specific mortality

rates within different countries:

$$\mathbf{y}_{t,s,c} = \begin{pmatrix} ln(m^{IP}_{t,s,c}) \\ ln(m^{Canc}_{t,s,c}) \\ ln(m^{Circ}_{t,s,c}) \\ ln(m^{Resp}_{t,s,c}) \\ ln(m^{Ext}_{t,s,c}) \end{pmatrix}$$

and showed that they were non-stationary. They also demonstrated that at least one cointegration relation existed between the variables. This means that the long-term equilibrium relation(s) existed between mortality rates corresponding to different causes inside of a particular country. For this reason, it was possible to build a Vector Error Correction Model (VECM) describing the joint development of the cause-specific mortality rates within every country included into the study. Supposing that there are r cointegration relations, i.e. that there exists a matrix β of rank r such that $\beta'\mathbf{y}_t$ is I(0), the corresponding VECM has then the following form:

$$\Delta\mathbf{y}_t = \mathbf{c} + \mathbf{d}t + \alpha\beta'\mathbf{y}_{t-1} + \sum_{i=1}^{l} \xi_i \Delta\mathbf{y}_{t-i} + \epsilon_t,\ t = 1...T \tag{3.7}$$

where

- $\mathbf{c}$ and $\mathbf{d}$ are $(n \times 1)$ vectors of constants;
- ξ_i is a $(n \times n)$ matrix of autoregressive coefficients for $i = 1, 2, ..., l$;
- l is number of lags;
- β is a $(n \times r)$ matrix containing r vectors each representing a cointegration relation;
- α is a $(n \times r)$ loading matrix that indicates how a particular variable is impacted by the cointegration relation;
- ϵ_t is a $(n \times 1)$ vector of white noise errors.

More details on the VECM can be found in such extensive references on the subject as Hamilton (1994) and Lütkepohl (2005).

For our part, we would like to apply the cointegration analysis from a different perspective by studying the possible cointegration relations between the cause-specific mortality rates corresponding to the same causes, but coming from different countries. For this, we will study all possible pairwise combination of countries. At the same time, age-specific mortality time series present a challenge from a modelling perspective, because to the best of our knowledge, the cointegration tests have been developed for the time series with dimension n less than 12 (Osterwald-Lenum 1992). For this reason, the cointegration tests

cannot be applied to the age-specific mortality time series directly. In Arnold and Sherris (2015, 2016) the authors overcame this difficulty by using the age-standardized mortality rates. In the present study we want to apply an alternative approach and study the cointegration between the time trends extracted from the cause-specific mortality rates, $k_t^{d,s,c}$ as defined in (3.1). So, we will test if the following time series are cointegrated:

$$\mathbf{y}_{t,s,d} = \begin{pmatrix} k_t^{d,s,c_1} \\ k_t^{d,s,c_2} \end{pmatrix},$$

where $c_1 \neq c_2$ and $c_1, c_2 \in \{FR, IT, NL, SP, EW\}$.

To achieve this, we use the trace and the maximum eigenvalue tests developed by Johansen (1995) and test for the existence of the cointegration relation for $\mathbf{y}_{t,s,d}$, i.e. that $\mathbf{\Pi y}_{t,s,d} = \alpha\beta'\mathbf{y}_{t,s,d}$ is stationary. Should this be the case, we proceed with testing for the form of the deterministic terms in (3.7), also developed by Johansen, and at the later stage, with assessing the quality of the fit for every identified VECM using the usual residuals tests. In this way, we verify, first, if the cointegration relation exists, and, second, that the resulting VECM has a good fit.

As suggested by Johansen (1995), we will consider the following cases where $\mathbf{d} = \alpha\rho + \alpha_\perp\gamma$ and $\alpha\alpha_\perp = 0$ to distinguish between the possible forms of the deterministic elements in the VECM:

- NT: no trend in the VECM, but a linear trend in the levels of the variables: $\mathbf{c} \neq 0, \rho = 0, \gamma = 0,$ hence $\mathbf{d} = 0$,
- TC: linear trend in the cointegration relation combined with a linear trend in the levels of the variables (i.e., no linear trend in the differenced variables): $\mathbf{c} \neq 0, \rho \neq 0, \gamma = 0,$ hence $\mathbf{d} = \alpha\rho$,
- QT: linear trend in the differenced variables, thus a quadratic trend in the levels of the variables : $\mathbf{c} \neq 0, \rho \neq 0, \gamma \neq 0,$ hence $\mathbf{d} = \alpha\rho + \alpha_\perp\gamma$.

In what follows, we will use the abbreviations NT, TC and QT to describe the VECM that was chosen, if any, for every tested pair of countries.

Should it be possible to identify a cointegration relation (here at most 1) as well as a VECM having normally distributed and non-correlated residuals, then this would mean that the particular cause-specific mortality rates from two countries experienced a similar development in the past. This observation may then justify the creation of the corresponding group of the countries to be included into the Li-Lee model. By comparing with the historical mortality rates (backtesting) we will be able to see if the existence of the cointegration relation between the time trends extracted from the cause-specific mortality rates of two countries can improve the forecasts of the corresponding cause-specific mortality rates.

3.4 Application

3.4.1 Cointegration relations in cause-specific mortality experiences

To decide if cause-specific experiences of two countries are close enough, as a first step, we extract the country-, cause- and sex-specific time trends $k_t^{d,s,c}$ from the model (3.1). The time trends corresponding to the Infectious&Parasitic diseases for males in five countries are shown on Figure 3.1. The charts for the rest of the causes can be found on Figures 3.3-3.11 in the Appendix. As we can see, on the one hand, there is a general pattern to which all countries tend to. On the other hand, certain differences can be observed between the countries. Hence, we need a formal procedure that could allow us to judge whether the experiences of two countries are close enough to justify the application of the Li-Lee model, i.e. a measure of coherence between the country-specific experiences.

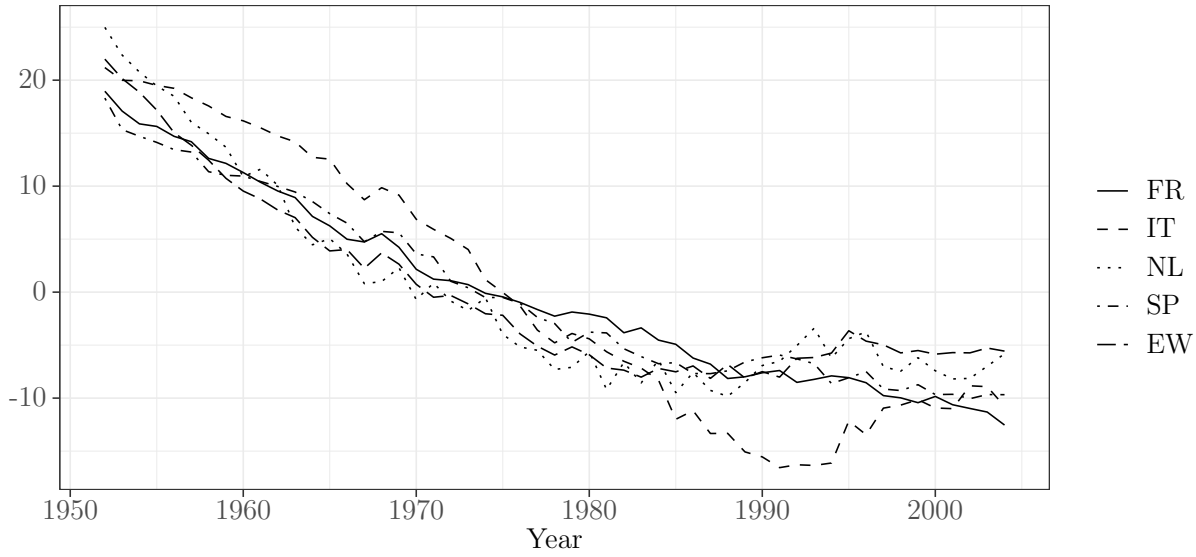


Figure 3.1: Time trends by country for the IP diseases, males.

For this, we check if the time trends corresponding to a particular cause are cointegrated between a pair of countries. If yes, this is an indication that the information contained in the cause-specific mortality rates of one country can enrich the model and improve the forecast of the second country from the pair and vice versa.

To achieve this, we apply the Johansen test and show the number of cointegration relations and the resulting VECM, if any, in the Table 3.2 below for males. We see that for the Infectious&Parasitic diseases the country-specific time trends are cointegrated in 7 pairs (out of 10). Similar observations hold for the Cancer and the Respiratory diseases (6 out of 10 pairs). The External causes happen slightly less often to be cointegrated (in 5 out 10 pairs), but the cointegration is observed more frequently for the Circulatory

diseases (in 9 out of 10 pairs).

The corresponding results for the female country-specific time trends are shown in Table 3.7 in the Appendix. Similar to the male time series, the female time trends for the Infectious&Parasitic diseases are cointegrated in 8 pairs out of 10, in 5 out of 10 pairs for the Cancer, the Circulatory and the Respiratory diseases, and only in 3 pairs out of 10 for the External causes. The detailed results of the maximum eigenvalue and trace tests as well as of the tests for the form of the deterministic elements are available from the authors upon request. Then, the quality of the model fit was assessed using the autocorrelation and the normality tests and the results are shown in Tables 3.8 and 3.9 of the Appendix.

Table 3.2: Number of cointegration relation and the form of the VECM, if any, describing the relation between the country-specific time trends, males.

Countries	IP	Canc	Circ	Resp	Ext
1 & 2	1 CR, QT l=1	1 CR, NT l=1	0 CR	1 CR, TC l=0	1 CR, NT l=0
1 & 3	1 CR, QT l=0	0 CR	1 CR, TC l=0	1 CR, TC l=1	0 CR
1 & 4	0 CR	1 CR, NT l=1	1 CR, NT l=0	0 CR	0 CR
1 & 5	1 CR, NT l=1	1 CR, QT l=1	1 CR, NT l=1	0 CR	1 CR, QT l=0
2 & 3	1 CR, QT l=1	0 CR	1 CR, QT l=0	1 CR, TC l=0	1 CR, TC l=1
2 & 4	0 CR	0 CR	1 CR, NT l=0	1 CR, TC l=1	0 CR
2 & 5	0 CR	1 CR, NT l=1	1 CR, QT l=1	0 CR	1 CR, NT l=0
3 & 4	1 CR, NT l=0	0 CR	1 CR, QT l=0	1 CR, TC l=0	0 CR
3 & 5	1 CR, QT l=1	1 CR, QT l=1	1 CR, NT l=1	1 CR, NT l=0	0 CR
4 & 5	1 CR, NT l=0	1 CR, NT l=1	1 CR, QT l=0	0 CR	1 CR TC, l=0

Note: CR = cointegration relation; QT = quadratic trend in the levels of the variables; TC = linear trend in the cointegration relation; NT = no trend; l = number of lags.

As mentioned in the introduction, our approach that consists in measuring the similarity between country-specific experiences using cointegration analysis is complementary to those proposed by Lyu et al. (2020) and Li and Lee (2005). In the former study the authors analyze the cause-specific experiences in France, Netherlands and Belgium. The

first two countries are included in our study as well (countries 1 and 3 in the Table 3.2). Lyu et al. (2020) arrive at the conclusion that there was no diverging pattern in the cause-specific mortality among all countries and for all causes analysed in their study. Our results cannot be directly compared with those in Lyu et al. (2020) due to a diverging definition of causes of death and a different observation period. Still, we see that the experiences of France and Netherlands can be called similar in terms of the cointegration analysis for three causes of death (IP, Circulatory and Respiratory) for male as well for female datasets (Tables 3.2 and 3.7). Our study thus reveals that there are causes for which the experiences of these two countries have not been as close as one could think.

To follow the approach proposed by Li and Lee (2005), we calculated the explanation ratios for mortality rates by cause and country as per the Li-Lee model (3.3) that we applied to every pair of countries mentioned in the Table 3.2. The explanation ratios for males are shown in the Table 3.3 and for the females in the Table 3.10 in the Appendix. We can see that apart from the External cause in some cases (e.g., the country 4 in the pairs 1&4, 2&4 and 3&4 and 4&5 for males), the application of the Li-Lee model to the rest of the cause- and country-specific mortality rates results in a "large enough" explanation ratios. This observation suggests that all pairs of countries should be modelled together in a multipopulational setting according to the Li and Lee (2005) approach whereas the cointegration analysis delivers a more nuanced answer.

Cointegration in the set of three countries

It should be also noted that the cointegration analysis can be applied to the systems having three or more variables. In the case of the cause-specific mortality rates, once the countries have been analyzed in a pairwise manner, we can conduct an additional scenario of putting together observation coming from three countries. To illustrate the idea, we will use the results for the Cancer mortality rates as shown in the Table 3.2. We see that for males, the countries 1 (FR), 2 (IT) and 5 (EW) as well as 1 (FR), 4 (SP) and 5 (EW) can built two groups of three countries each in which every two countries have cointegrated country-specific time trends: 1&2&5 and 1&4&5. Simultaneous modelling of the cause-specific mortality rates for these countries is then justified by the fact that every pair of the country-specific time trends shares some stochastic trends, and so, there may exist a trend shared together by all three countries. We do not analyze the combination 2&3&5, for example, because the countries 2 and 3 do not have cointegrated time trends, and so, there is less reason to believe that modelling three countries together can bring an additional benefit in comparison with the two-country model already built. It is even more so for a combination like 1&2&3 in which only the countries 1 and 2 have cointegrated time trends. The cointegration relation present between them will still exist in the three-variable system, but the three-country modelling will hardly bring

Table 3.3: Country-specific explanation ratios by cause, males

Countries	I&P	Canc	Circ	Resp	Ext
1 & 2	0.9663	0.7383	0.9612	0.9502	0.9242
	0.9547	0.8188	0.9701	0.9552	0.9271
1 & 3	0.9762	0.8702	0.9610	0.9588	0.9386
	0.7939	0.7271	0.7795	0.7051	0.8530
1 & 4	0.9645	0.8602	0.9623	0.9506	0.8687
	0.9567	0.8492	0.9550	0.9256	0.5690
1 & 5	0.9587	0.8418	0.9545	0.9518	0.8613
	0.9181	0.8739	0.9647	0.9322	0.9495
2 & 3	0.9709	0.9343	0.9737	0.9617	0.9507
	0.7617	0.7737	0.7945	0.7052	0.8404
2 & 4	0.9568	0.9328	0.9623	0.9220	0.8942
	0.9493	0.8561	0.9422	0.9204	0.5597
2 & 5	0.9522	0.9379	0.9628	0.9308	0.9305
	0.9463	0.9208	0.9580	0.9189	0.9407
3 & 4	0.7944	0.7930	0.7249	0.7055	0.8720
	0.9665	0.8759	0.9554	0.9401	0.5469
3 & 5	0.8002	0.7801	0.8126	0.7285	0.8320
	0.9534	0.9469	0.9717	0.9509	0.9622
4 & 5	0.9391	0.8839	0.9236	0.9196	0.4792
	0.9180	0.8945	0.9479	0.9276	0.9568

Upper number in the cell corresponds to the explanation ratio for the left country in the pair.

any additional benefit in comparison with the two-country case. For the females, there is only one three-country combination for Cancer in which every two countries have cointegrated country-specific time trends (2&4&5). For the sake of completeness, we apply the cointegration analysis to the identified three-country combinations and show the results in the Table 3.4. Unsurprisingly, the three-country combinations of the country-specific time trends remain cointegrated.

Table 3.4: p values for the null hypotheses of no autocorrelation and normality of the residuals of the VECM fitted to the country-specific time trends, Cancer.

			Autocorrelation			Normality		
Sex	Countries	Model	15 lags	25 lags	35 lags	Skewness	Kurtosis	Both
Males	1 & 2 & 5	l=1, NT, 1 CR	0.8139	0.4927	0.3739	0.2952	0.5610	0.4506
Males	1 & 4 & 5	l=0, NT, 2 CR	0.4338	0.3434	0.1086	0.9902	0.7094	0.9597
Females	2 & 4 & 5	l=1, TC, 1 CR	0.3614	0.8633	0.9841	0.0453	0.1735	0.0428

A null hypothesis is accepted at a α% significance level when the p value is higher than α%.

3.4.2 Cause-specific forecasts for countries having similar experiences

In cases where the country- and cause-specific time trends are cointegrated and assuming that the observed coherence between the experiences of two countries continues in the future, we expect that the Li-Lee model built for this pair of countries will deliver improved forecasts of the cause-specific mortality rates in comparison with the basic Lee-Carter approach. To check this, we use the data for 1952-2004 to estimate the parameters of the Li-Lee and the Lee-Carter models, project the time trends using the ARIMA framework for the 2005-2014, retrieve the projected cause-specific mortality rates and compare with values observed in 2005-2014 on the basis of the mean absolute average percentage error (MAPE). First, the country-, sex- and cause-specific absolute percentage error values (APE) were calculated for each age group x and the projection year t :

$$APE(x,t) = \frac{abs(ln(m_{x,t}^{observed}) - ln(m_{x,t}^{projected}))}{ln(m_{x,t}^{observed})} \tag{3.8}$$

At the second stage, the individual $APE(x,t)$ corresponding to the countries included in the pair were averaged across the pair taking into account the population numbers of each country and then again averaged for all x and t. In this way, we obtained the MAPE values averaged over two countries. The results of these calculations are shown in the Table 3.5. We observe that indeed, for the male rates, the Li-Lee model allows obtaining better forecasts for five pairs of countries out of seven pairs that have cointegrated time

trends for the Infectious&Parasitic diseases, for every pair of countries that has cointegrated time trends for the Cancer diseases, for six out of nine pairs of countries for the Circulatory diseases, for five out of six pairs of countries for the Respiratory diseases, and for two out of five pairs of countries for the External causes.

Similar observations hold for the female mortality rates: for the Infectious&Parasitic diseases, the Li-Lee model permits to obtain more precise forecasts for seven out of eight pairs of countries that have the cointegrated time trends; for four out of five pairs of countries for the Cancer diseases; for every pair of countries that has the cointegrated time trend for the Circulatory diseases; for three out of five pairs of countries for the Respiratory diseases, and for two out of three pairs of countries for the External causes.

Both the Lee-Carter and Li-Lee model give age-specific forecasts for each year in 2005-2014. These forecasts are in fact point estimates. The incertainty related to the estimates is best described using the confidence intervals, but comparing the intervals is a more challenging task that would probably not deliver clear-cut results. For this reason, we will limit our analysis to comparing the point estimates produced by both models.

Forecasts for the set of three countries

The comparison of the forecasts for the Lee-Carter and the Li-Lee model built for three countries is shown in the Table 3.6. We can see that the Li-Lee model substantially improves the quality of the cause-specific forecast for the males whereas the improvement is less pronounced for the females.

Table 3.5: Cause-specific MAPE averaged over two countries.

		Males					Females		
Cause	Countries	LC	Li-Lee	diff	Cause	Countries	LC	Li-Lee	diff
IP	1 & 2	0.0281	0.0297	-0.0016	IP	1 & 2	0.0411	0.0409	0.0002
IP	1 & 3	0.0251	0.0216	0.0034	IP	1 & 3	0.0249	0.0207	0.0042
IP	1 & 5	0.0258	0.0383	-0.0125	IP	1 & 4	0.0360	0.0336	0.0024
IP	2 & 3	0.0383	0.0350	0.0033	IP	1 & 5	0.0297	0.0326	-0.0029
IP	3 & 4	0.0644	0.0437	0.0207	IP	2 & 4	0.0506	0.0453	0.0053
IP	3 & 5	0.0369	0.0299	0.0070	IP	3 & 4	0.0542	0.0396	0.0146
IP	4 & 5	0.0407	0.0209	0.0198	IP	3 & 5	0.0454	0.0376	0.0078
					IP	4 & 5	0.0381	0.0175	0.0205
Canc	1 & 2	0.0440	0.0159	0.0281	Canc	1 & 5	0.0121	0.0132	-0.0010
Canc	1 & 4	0.0397	0.0233	0.0165	Canc	2 & 4	0.0205	0.0158	0.0047
Canc	1 & 5	0.0264	0.0156	0.0108	Canc	2 & 5	0.0145	0.0117	0.0028
Canc	2 & 5	0.0339	0.0117	0.0222	Canc	3 & 5	0.0113	0.0106	0.0007
Canc	3 & 5	0.0220	0.0126	0.0094	Canc	4 & 5	0.0218	0.0169	0.0050
Canc	4 & 5	0.0279	0.0151	0.0128					
Circ	1 & 3	0.0126	0.0110	0.0016	Circ	1 & 3	0.0174	0.0124	0.0050
Circ	1 & 4	0.0202	0.0168	0.0034	Circ	2 & 3	0.0339	0.0189	0.0149
Circ	1 & 5	0.0219	0.0194	0.0025	Circ	2 & 4	0.0361	0.0186	0.0175
Circ	2 & 3	0.0147	0.0158	-0.0011	Circ	2 & 5	0.0344	0.0222	0.0122
Circ	2 & 4	0.0164	0.0131	0.0033	Circ	4 & 5	0.0382	0.0286	0.0096
Circ	2 & 5	0.0181	0.0170	0.0011					
Circ	3 & 4	0.0214	0.0269	-0.0055					
Circ	3 & 5	0.0245	0.0275	-0.0029					
Circ	4 & 5	0.0234	0.0185	0.0049					
Resp	1 & 2	0.0276	0.0231	0.0046	Resp	1 & 2	0.0245	0.0273	-0.0027
Resp	1 & 3	0.0129	0.0144	-0.0015	Resp	1 & 3	0.0162	0.0201	-0.0039
Resp	2 & 3	0.0391	0.0295	0.0096	Resp	2 & 4	0.0288	0.0250	0.0038
Resp	2 & 4	0.0336	0.0254	0.0082	Resp	3 & 4	0.0325	0.0243	0.0083
Resp	3 & 4	0.0282	0.0264	0.0018	Resp	4 & 5	0.0359	0.0197	0.0161
Resp	3 & 5	0.0282	0.0273	0.0009					
Ext	1 & 2	0.0205	0.0160	0.0045	Ext	1 & 2	0.0172	0.0132	0.0040
Ext	1 & 5	0.0216	0.0285	-0.0069	Ext	2 & 5	0.0165	0.0182	-0.0018
Ext	2 & 3	0.0182	0.0163	0.0019	Ext	4 & 5	0.0336	0.0307	0.0029
Ext	2 & 5	0.0178	0.0244	-0.0066					
Ext	4 & 5	0.0335	0.0375	-0.0040					

3.5 Discussion and conclusion

We live in a world that becomes more and more interconnected, globalized, and in many regards less diversified. For some time now, this trend found its reflection in the converg-

Table 3.6: Cause-specific MAPE averaged over three countries, Cancer.

Males				Females			
Countries	LC model	Li-Lee model	diff	Countries	LC model	Li-Lee model	diff
1 & 2 & 5	0.0350	0.0184	0.0166	2 & 4 & 5	0.0189	0.0143	0.0046
1 & 4 & 5	0.0313	0.0156	0.0158				

ing mortality levels around the world (Wilson, 2011). So it seems less and less adequate to forecast mortality rates for individual countries without considering their future development in a larger picture. Also, it has been noted that individual application of the Lee-Carter model to the G7 countries leads to an increase of the largest gap in the life expectancy from about 4 to 8 years over a 50 year forecast horizon (Tuljapurkar et al., 2000). Such results enter in contradiction with the converging pattern of the mortality rates around the world. These considerations have lead Li and Lee (2005) to propose a model that takes into account the membership of the countries in a group by identifying the central tendencies proper to all countries and letting the weight of each country's particularities diminish in the long run.

There is no reason why what is true for the all-cause mortality would not be true for the cause-specific mortality rates. Even more so: as it may be easier to identify the driving factors of the cause-specific mortality than those of the all-cause mortality, it may also be easier to establish the coherence on the cause-specific level (Lyu et al., 2020).

Then the question arises: how to "measure" the coherence of the experiences of several countries, when each experience is contained in a large-scale matrix of observations by age and year? We suggest using the cointegration analysis that allows us precisely to say if two (or more) nonstationary vectors remain close enough to each other over a long period of time to build a stationary linear combination. To reduce the dimensionality of the mortality data we propose to apply the cointegration analysis to the mortality trends extracted by the Lee-Carter model.

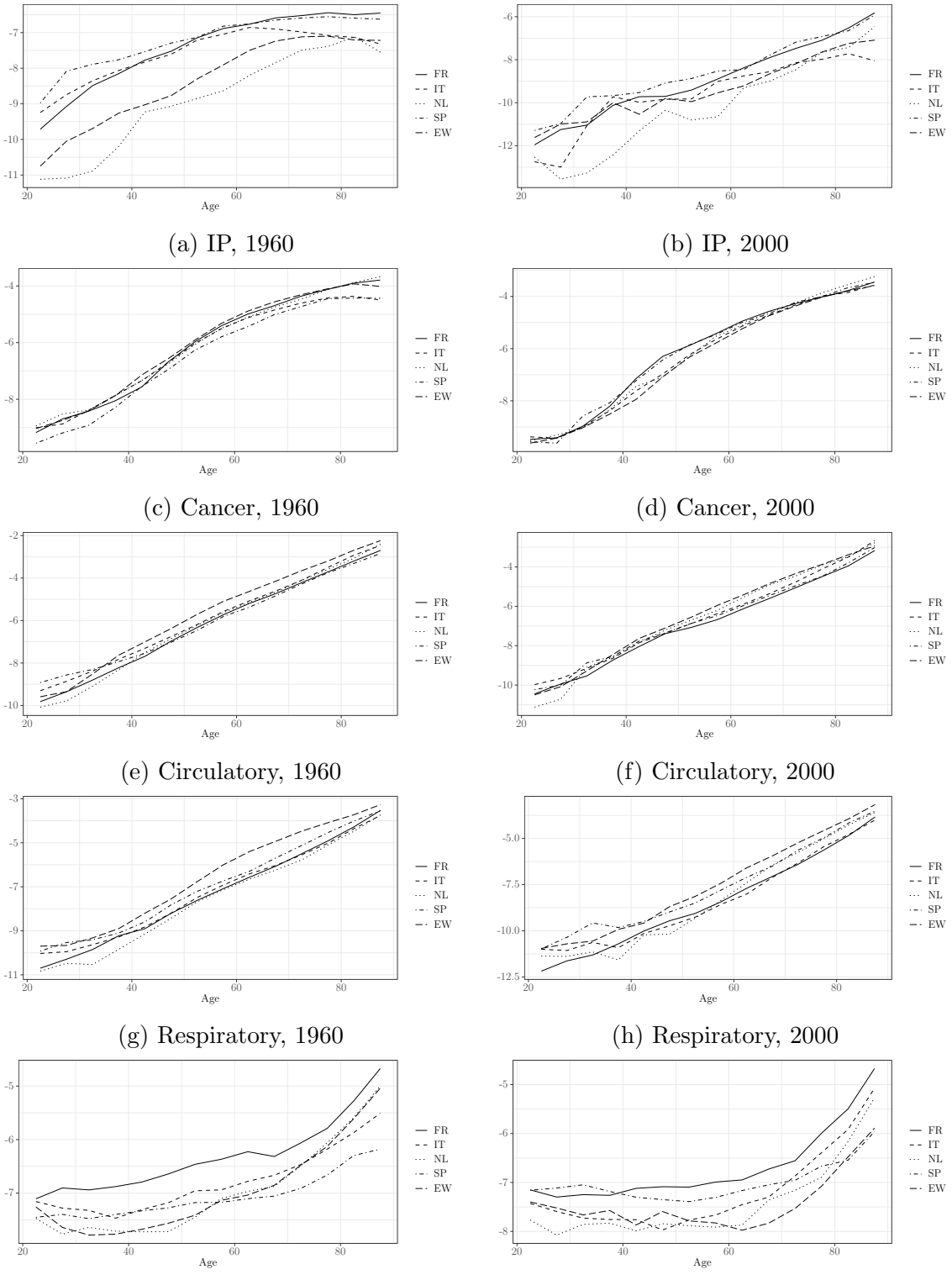
We chose five most populated Western European countries to increase our chances to find the coherence between their respective cause-specific experiences. And indeed, looking at the countries in a pairwise manner, we see that very often their cause-specific time trends are cointegrated. At the same time, one needs to be cautious because not in all cases the cointegration was found. This means that even such similar countries may not have coherent experience for all considered causes of death. At the same time, should one apply the approaches proposed by Lyu et al. (2020) and Li and Lee (2005), this would lead to the conclusion that the county- and cause-specific mortality experiences are to a large extent comparable and so, the corresponding countries should be modelled together. Hence, the cointegration analysis delivers a more nuanced answer.

Once the countries having the cointegrated cause-specific time trends were included

together in a augmented common factor model proposed by Li and Lee (2005), in many cases this allowed improving the forecasting results in comparison with the basic cause-specific Lee-Carter approach. Additionally to ensuring the convergence of the forecasts, the Li-Lee model helps to enrich the experience of one country with the observations from another which can be beneficial in case of limited or volatile data as the country-specific noise is levelled out by the information from the similar countries. In cases when for some causes and combinations of countries no improvement was found, this can probably be explained by the fact that the coherence stated in the past did not continue during the forecast horizon. This is particularly true for such an independent cause as the External causes of death. Indeed, as this cause represents such random events as transport and other accidents (falls, poisoning, accidental fire, drowning), suicides, homicides, and war injuries, there are less reasons to expect that the experiences of any two countries have been following a similar path in the past. At the same time, should this have been the case, it is less probable that the observed similarity of experiences will be stable enough to continue into the future.

If one takes into account the proper character of each cause, the cointegration analysis proves to be a useful tool to assess the similarities between the experiences of two countries and so, helps building more accurate forecasts for the cause-specific mortality rates.

Appendix



(a) IP, 1960

(b) IP, 2000

(c) Cancer, 1960

(d) Cancer, 2000

(e) Circulatory, 1960

(f) Circulatory, 2000

(g) Respiratory, 1960

(h) Respiratory, 2000

(i) External, 1960

(j) External, 2000

Figure 3.2: Log-death cause-specific rates by cause and year, males.

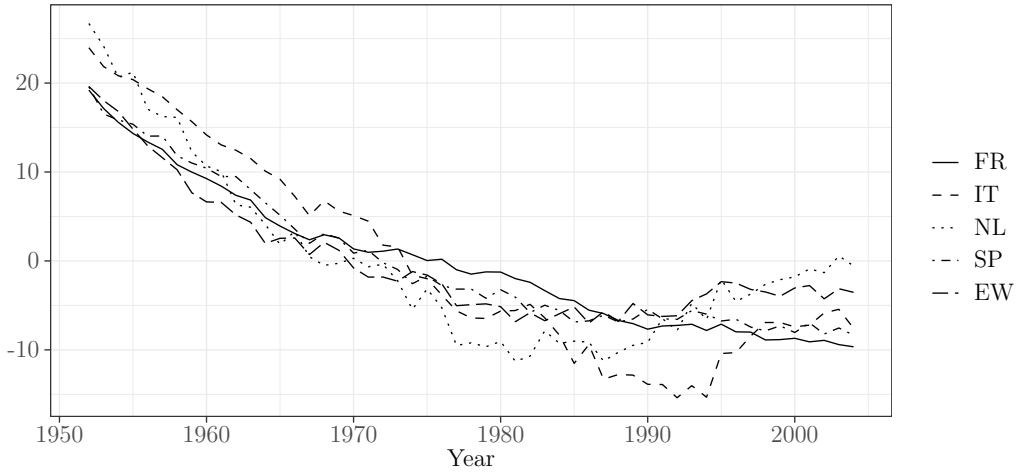


Figure 3.3: Time trends by country for the IP diseases, females.

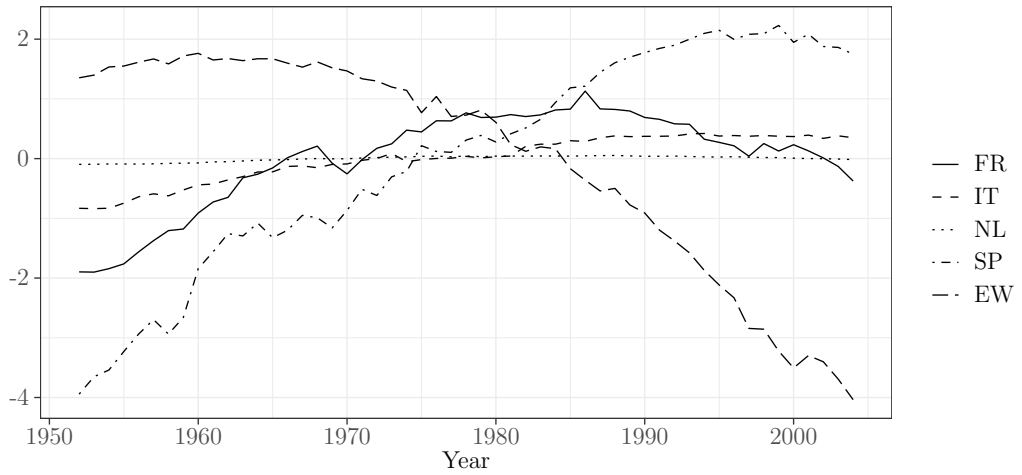


Figure 3.4: Time trends by country for the Cancer diseases, males.

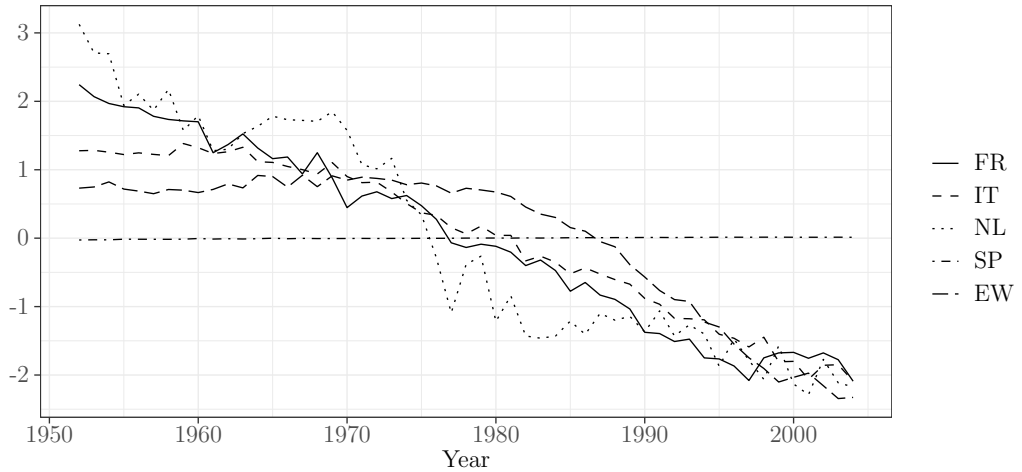


Figure 3.5: Time trends by country for the Cancer diseases, females.

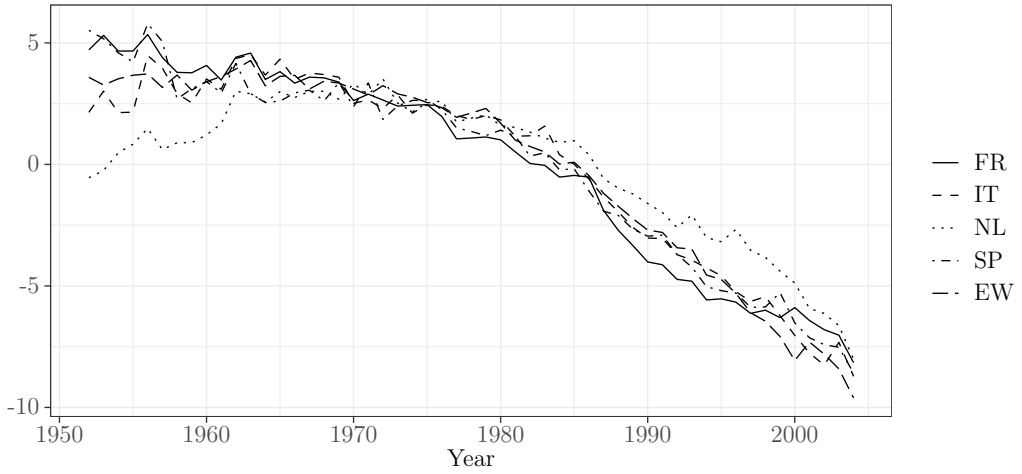


Figure 3.6: Time trends by country for the Circulatory diseases, males.

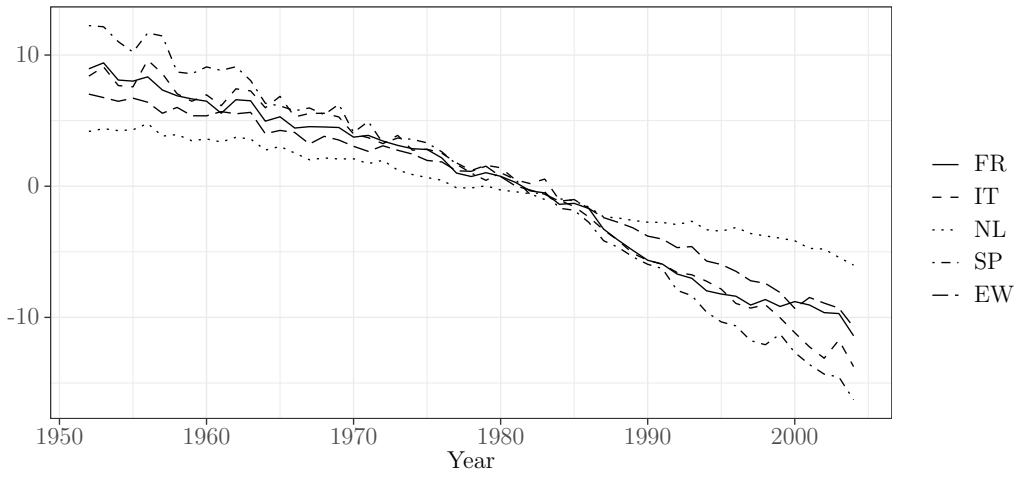


Figure 3.7: Time trends by country for the Circulatory diseases, females.

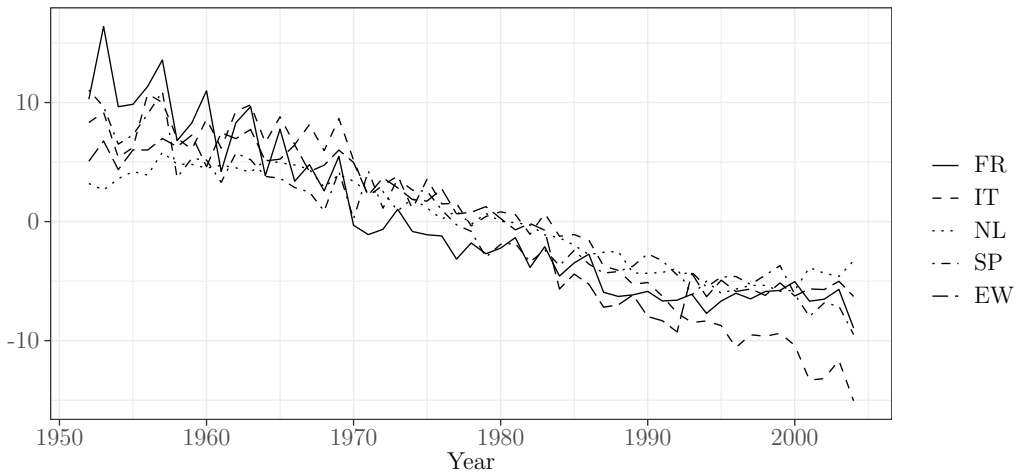


Figure 3.8: Time trends by country for the Respiratory diseases, males.

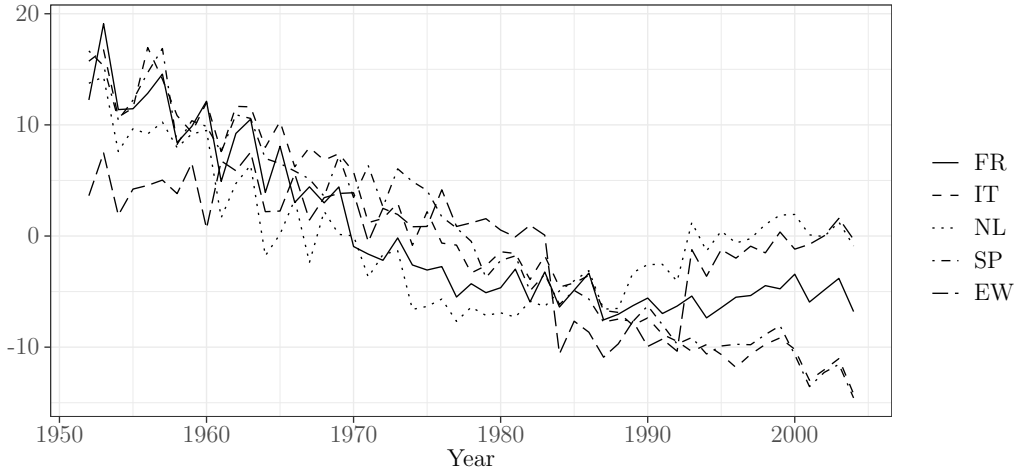


Figure 3.9: Time trends by country for the Respiratory diseases, females.

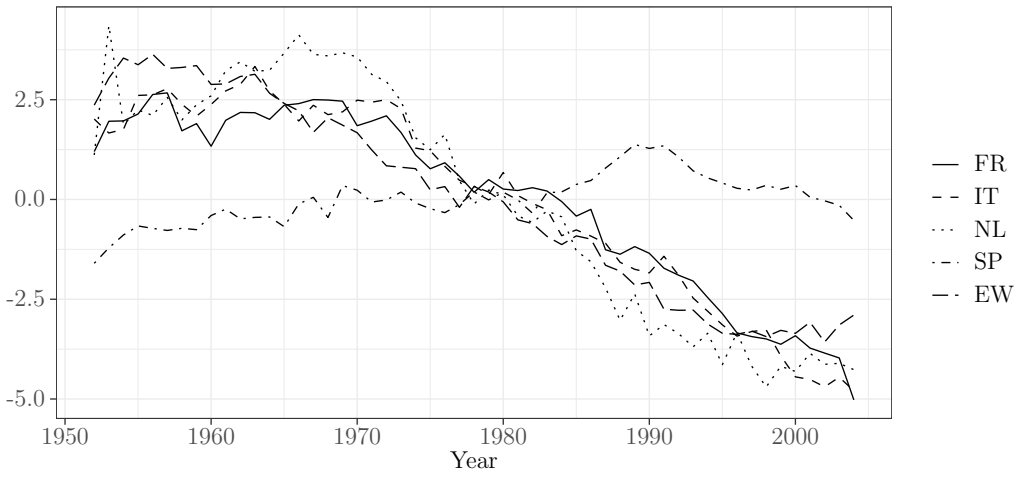


Figure 3.10: Time trends by country for the External causes, males.

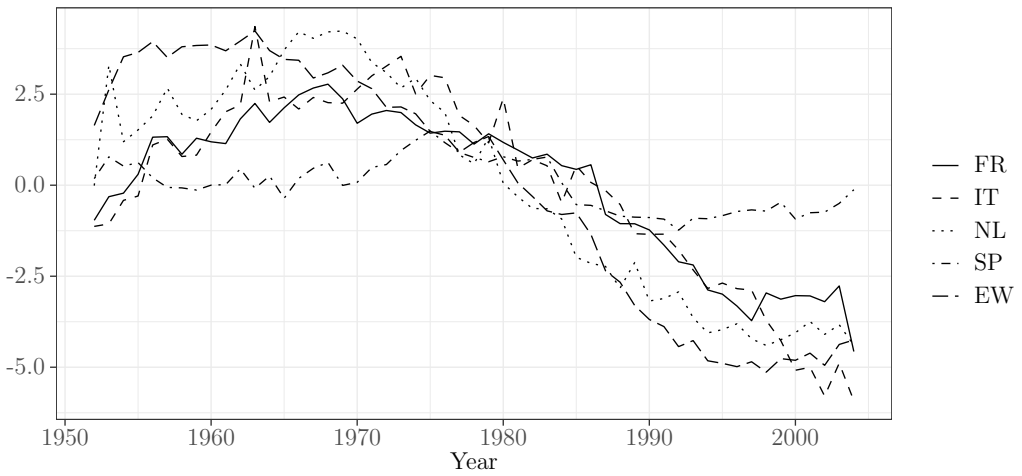


Figure 3.11: Time trends by country for the External causes, females.

Table 3.7: Number of cointegration relations and the form of the VECM, if any, describing the relation between the country-specific time trends, females.

Countries	IP	Canc	Circ	Resp	Ext
1 & 2	1 CR, NT l=1	0 CR	0 CR	1 CR, QT l=1	1 CR, TC l=1
1 & 3	1 CR, NT l=1	0 CR	1 CR, TC l=1	1 CR, TC l=0	0 CR
1 & 4	1 CR, NT l=1	0 CR	0 CR	0 CR	0 CR
1 & 5	1 CR, NT l=0	1 CR, TC l=0	0 CR	0 CR	0 CR
2 & 3	0 CR	0 CR	1 CR, TC l=0	0 CR	0 CR
2 & 4	1 CR, TC l=1	1 CR, TC l=0	1 CR, NT l=0	1 CR, TC l=1	0 CR
2 & 5	0 CR	1 CR, NT l=1	1 CR, NT l=0	0 CR	1 CR, TC l=0
3 & 4	1 CR, QT l=0	0 CR	0 CR	1 CR, TC l=0	0 CR
3 & 5	1 CR, NT l=1	1 CR, NT l=1	0 CR	0 CR	0 CR
4 & 5	1 CR, NT l=1	1 CR, NT l=1	1 CR, QT l=1	1 CR, TC l=0	1 CR, TC l=0

Note: CR = cointegration relation; QT = quadratic trend in the levels of the variables; TC = linear trend in the cointegration relation; NT = no trend; l = number of lags.

Table 3.8: p values for the null hypotheses of no autocorrelation and normality of the residuals of the VECM fitted to the country-specific time trends, males.

			Autocorrelation			Normality		
Cause	Countries	Model	15 lags	25 lags	35 lags	Skewness	Kurtosis	Both
IP	1 & 2	l=1, QT, 1 CR	0.2052	0.3892	0.3008	0.8439	0.2416	0.5281
IP	1 & 3	l=0, QT, 1 CR	0.0669	0.1547	0.1042	0.7046	0.3638	0.6053
IP	1 & 5	l=1, NT, 1 CR	0.4658	0.2906	0.3003	0.4232	0.6015	0.6028
IP	2 & 3	l=1, QT, 1 CR	0.0473	0.3001	0.1290	0.4920	0.1830	0.3068
IP	3 & 4	l=0, NT, 1 CR	0.3759	0.3559	0.4575	0.2122	0.1155	0.1154
IP	3 & 5	l=1, QT, 1 CR	0.3760	0.4781	0.4088	0.0228	0.0987	0.0159
IP	4 & 5	l=0, NT, 1 CR	0.1868	0.3197	0.5076	0.0547	0.6239	0.1495
Canc	1 & 2	l=1, NT, 1 CR	0.2768	0.2467	0.2858	0.5839	0.1612	0.3165
Canc	1 & 4	l=1, NT, 1 CR	0.1564	0.1765	0.6300	0.9285	0.5085	0.8264
Canc	1 & 5	l=1, QT, 1 CR	0.9500	0.8273	0.6646	0.4136	0.7251	0.6611
Canc	2 & 5	l=1, NT, 1 CR	0.7074	0.8131	0.6609	0.2328	0.1005	0.1113
Canc	3 & 5	l=1, QT, 1 CR	0.5439	0.8769	0.8575	0.9339	0.1563	0.4269
Canc	4 & 5	l=1, NT, 1 CR	0.8490	0.7211	0.0938	0.3727	0.5019	0.5007
Circ	1 & 3	l=0, TC, 1 CR	0.1088	0.2020	0.2628	0.4921	0.3433	0.4693
Circ	1 & 4	l=0, NT, 1 CR	0.5027	0.7622	0.6161	0.1131	0.7013	0.2804
Circ	1 & 5	l=1, NT, 1 CR	0.1926	0.4667	0.3768	0.9295	0.5091	0.8272
Circ	2 & 3	l=0, QT, 1 CR	0.7329	0.8532	0.8850	0.1252	0.9719	0.3781
Circ	2 & 4	l=0, NT, 1 CR	0.3959	0.6746	0.8418	0.1476	0.7491	0.3541
Circ	2 & 5	l=1, QT, 1 CR	0.4580	0.8213	0.6247	0.3805	0.7720	0.6536
Circ	3 & 4	l=0, QT, 1 CR	0.6637	0.6956	0.5042	0.8488	0.7402	0.9203
Circ	3 & 5	l=1, NT, 1 CR	0.7050	0.8905	0.6391	0.0412	0.0586	0.0170
Circ	4 & 5	l=0, QT, 1 CR	0.7879	0.9135	0.9558	0.6435	0.7234	0.8214
Resp	1 & 2	l=0, TC, 1 CR	0.2234	0.3956	0.8512	0.8326	0.7200	0.9062
Resp	1 & 3	l=1, TC, 1 CR	0.4658	0.4705	0.1625	0.9059	0.3891	0.7201
Resp	2 & 3	l=0, TC, 1 CR	0.0928	0.1988	0.1980	0.7266	0.6841	0.8445
Resp	2 & 4	l=1, TC, 1 CR	0.6321	0.7747	0.8087	0.1622	0.1049	0.0864
Resp	3 & 4	l=0, TC, 1 CR	0.5709	0.1923	0.2564	0.4238	0.8656	0.7348
Resp	3 & 5	l=0, NT, 1 CR	0.1114	0.0228	0.0163	0.4340	0.5995	0.6105
Ext	1 & 2	l=0, NT, 1 CR	0.6169	0.9758	0.9817	0.1146	0.7351	0.2926
Ext	1 & 5	l=0, QT, 1 CR	0.4720	0.4658	0.7119	0.0480	0.5109	0.1154
Ext	2 & 3	l=1, TC, 1 CR	0.2407	0.2370	0.2027	0.6203	0.1282	0.2809
Ext	2 & 5	l=0, NT, 1 CR	0.1402	0.2740	0.4255	0.7726	0.2176	0.4679
Ext	4 & 5	l=0, TC, 1 CR	0.3539	0.3563	0.6735	0.1618	0.3120	0.2012

A null hypothesis is accepted at a α% significance level when the p value is higher than α%.

Table 3.9: p values for the null hypotheses of no autocorrelation and normality of the residuals of the VECM fitted to the country-specific time trends, females.

			Autocorrelation			Normality		
Cause	Countries	Model	15 lags	25 lags	35 lags	Skewness	Kurtosis	Both
IP	1 & 2	l=1, NT, 1 CR	0.5333	0.6791	0.5972	0.9612	0.5131	0.8418
IP	1 & 3	l=1, NT, 1 CR	0.0399	0.3974	0.4287	0.8135	0.5242	0.7899
IP	1 & 4	l=1, NT, 1 CR	0.3530	0.1626	0.3574	0.7525	0.1424	0.3465
IP	1 & 5	l=0, NT, 1 CR	0.4699	0.4436	0.6921	0.9372	0.1842	0.4758
IP	2 & 4	l=1, TC, 1 CR	0.1839	0.1525	0.3068	0.7851	0.1887	0.4311
IP	3 & 4	l=0, QT, 1 CR	0.2559	0.1389	0.1941	0.7811	0.2086	0.4586
IP	3 & 5	l=1, NT, 1 CR	0.7933	0.8437	0.8806	0.9556	0.2385	0.5649
IP	4 & 5	l=1, NT, 1 CR	0.5775	0.2677	0.3877	0.9402	0.0798	0.2694
Canc	1 & 5	l=0, TC, 1 CR	0.5170	0.7723	0.6811	0.9433	0.2615	0.5919
Canc	2 & 4	l=0, TC, 1 CR	0.5663	0.4804	0.5546	0.0180	0.1735	0.0212
Canc	2 & 5	l=1, NT, 1 CR	0.6308	0.6110	0.8195	0.5690	0.2874	0.4596
Canc	3 & 5	l=1, NT, 1 CR	0.6758	0.3139	0.0809	0.1784	0.3895	0.2548
Canc	4 & 5	l=1, NT, 1 CR	0.1455	0.5899	0.8862	0.0148	0.0225	0.0030
Circ	1 & 3	l=1, TC, 1 CR	0.3867	0.5658	0.7723	0.8762	0.7271	0.9243
Circ	2 & 3	l=0, TC, 1 CR	0.4389	0.3220	0.2340	0.1342	0.6330	0.2944
Circ	2 & 4	l=0, NT, 1 CR	0.9101	0.9757	0.9749	0.4050	0.2932	0.3718
Circ	2 & 5	l=0, NT, 1 CR	0.2280	0.7746	0.7418	0.2993	0.6775	0.5264
Circ	4 & 5	l=1, QT, 1 CR	0.7526	0.8769	0.9453	0.8221	0.7101	0.8980
Resp	1 & 2	l=1, QT, 1 CR	0.0376	0.1869	0.5414	0.8165	0.8554	0.9491
Resp	1 & 3	l=0, TC, 1 CR	0.0011	0.0140	0.0411	0.6019	0.7309	0.8012
Resp	2 & 4	l=1, TC, 1 CR	0.8012	0.4841	0.6196	0.1026	0.2862	0.1330
Resp	3 & 4	l=0, TC, 1 CR	0.1099	0.0362	0.1729	0.0198	0.5736	0.0623
Resp	4 & 5	l=0, TC, 1 CR	0.1805	0.0316	0.0353	0.9974	0.3537	0.7203
Ext	1 & 2	l=1, TC, 1 CR	0.1989	0.3484	0.0206	0.1574	0.6173	0.3237
Ext	2 & 5	l=0, TC, 1 CR	0.1562	0.1251	0.1811	0.6831	0.2083	0.4197
Ext	4 & 5	l=0, TC, 1 CR	0.3593	0.4089	0.4226	0.5324	0.4532	0.5844

A null hypothesis is accepted at a α% significance level when the p value is higher than α%.

Table 3.10: Country-specific explanation ratios by cause, females.

Countries	I&P	Canc	Circ	Resp	Ext
1 & 2	0.9487	0.8176	0.9710	0.9386	0.8788
	0.9378	0.9146	0.9851	0.9653	0.8726
1 & 3	0.9632	0.8817	0.9615	0.9447	0.8901
	0.6074	0.8630	0.7723	0.6747	0.7933
1 & 4	0.9550	0.8387	0.9670	0.9340	0.8767
	0.9415	0.7440	0.9781	0.9579	0.5017
1 & 5	0.9542	0.8081	0.9529	0.9033	0.7996
	0.9151	0.8402	0.9621	0.9103	0.9130
2 & 3	0.9640	0.9239	0.9830	0.9692	0.9071
	0.6371	0.8349	0.7865	0.6785	0.7656
2 & 4	0.9355	0.8977	0.9794	0.9649	0.8372
	0.9387	0.6621	0.9816	0.9607	0.5117
2 & 5	0.9307	0.9123	0.9684	0.9647	0.8820
	0.9003	0.8427	0.9273	0.9047	0.9427
3 & 4	0.6275	0.8430	0.7928	0.6544	0.7034
	0.9523	0.7760	0.9760	0.9617	0.6082
3 & 5	0.6701	0.8381	0.7887	0.6471	0.8156
	0.9328	0.8815	0.9706	0.9269	0.9645
4 & 5	0.9347	0.6697	0.9776	0.9416	0.5685
	0.9155	0.8422	0.9246	0.9097	0.9378

Upper number in the cell corresponds to the explanation ratio for the left country in the pair.

Bibliography

Arnold, S. and V. Glushko, 2021a, Short- and long-term dynamics of cause-specific mortality rates using cointegration analysis, *North American Actuarial Journal*.

Arnold, S. and M. Sherris, 2015, Causes-of-Death Mortality: What Do We Know on their Dependence?, *North American Actuarial Journal*, 19(2):116–128.

Arnold, S. and M. Sherris, 2016, International Cause-Specific Mortality Rates: New Insights from a Cointegration Analysis., *ASTIN Bulletin*, 46(1):9–38.

Booth, H. and L. Tickle, 2008, Mortality modelling and forecasting: A review of methods, *Annals of Actuarial Science*, 3:3–43.

Brouhns, N., M. Denuit, and J. K. Vermunt, 2002, A Poisson log-bilinear regression approach to the construction of projected lifetables., *Insurance: Mathematics and Economics*, 31(3):373–393.

Cairns, A., 2013, Modeling and management of longevity risk, In P. B. Hammond, R. Maurer, and O. S. Mitchell, editors, *Recreating Sustainable Retirement: Resilience, Solvency, and Tail Risk*, pages 71–88. Oxford: Oxford University Press.

Engle, R. and C. Granger, 1987, Co-integration and Error Correction: Representation, Estimation and Testing., *Econometrica*, 55:251–276.

Gutterman, S. and I. T. Vanderhoof, 1998, Forecasting changes in mortality: A search for a law of causes and effects, *North American Actuarial Journal*, 2(4):135–138.

Hamilton, J. D., 1994, *Time Series Analysis*. Princeton: Princeton University Press.

Johansen, S., 1995, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. New York: Oxford University Press.

Lee, R. D. and L. R. Carter, 1992, Modeling and forecasting U.S. mortality, *J. American Statistical Association*, 87:659–671.

Li, N. and R. Lee, 2005, Coherent mortality forecasts for a group of populations: An extension of the lee-carter method., *Demography*, 42(3):575–594.

Lütkepohl, H., 2005, *New Introduction to Multiple Time Series Analysis*. Crown Publishing Group.

Lyu, P., A. De Waegenaere, and B. Melenberg, 2020, A multi-population approach to forecasting all-cause mortality using cause-of-death mortality data, *North American Actuarial Journal*, pages 1–36.

Osterwald-Lenum, M., 1992, A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics, *Oxford Bulletin of Economics and Statistics*, 54(3):461–472.

Tabeau, E., P. Ekamper, C. Huisman, and A. Bosch, 1999, Improving overall mortality forecasts by analysing cause-of-death, period and cohort effects in trends, *European Journal of Population*, 15:153–183. Available at `http://actuaries.asn.au/Library/FSF10_Paper_burgess.pdf`.

Tuljapurkar, S., N. Li, and C. Boe, 2000, A Universal Pattern of Mortality Decline in the G7 Countries, *Nature*, 405:789–792.

Villegas, A., S. Haberman, V. K. Kaishev, and P. Millossovich, 2017, A comparative study of two population models for the assessment of basis risk in longevity hedges, *ASTIN Bulletin*, 47(3):631–679.

Wilmoth, J. R., 1995, Are mortality projections always more pessimistic when disaggregated by cause of death?, *Mathematical Population Studies*, 5(4):293–319.

Wilson, C., 2011, Understanding global demographic convergence since 1950, *Population and Development Review*, 37(2):375–388.

World Health Organization, 2020, WHO Mortality Database. `http://www.who.int/whosis/mort/download/en/index.html`.

General conclusions

The aim of this thesis was to investigate the relations existing between the cause-specific mortality rates through the lens of the cointegration techniques. Because of the presence of competing risks, modelling cause-specific mortality rates requires modelling of the dependence between them. As this dependence is not observable, it is impossible to gather the corresponding statistics. Cointegration analysis offers then a possibility to include the relationship between the non-stationary and dependent cause-specific mortality rates into a model in such a way that the information on the dependency between the rates is preserved. Throughout the thesis, we used the data coming from different countries in order to compare the country-specific results and find common patterns emerging across different datasets, so that we could make a general observation independent of the country-specific experiences.

In the first chapter *Short- and Long-Term Dynamics of Cause-specific Mortality Rates Using Cointegration Analysis*, we applied the cointegration analysis and vector error correction models to model the short- and long-run relationships between the cause-specific mortality rates from USA, Japan, France, England and Wales, and Australia. We analyzed the development of every cause, how it was dependent from other causes and also what impact a particular cause had on the development of the rest of the causes.

In the second chapter *Cause-Specific Mortality Rates: Common Trends and Differences*, we studied the non-stationary part of the system of the cause-specific mortality rates. We showed that once the common stochastic trends were explicitly extracted from the original variables and summarized using the principal component analysis, the first principal components turned out to be cointegrated for the male and female datasets coming from different countries. This observation implies that not only cause-specific mortality rates were cointegrated across different causes inside a particular country, but that there was also a cointegration relation between the first principal components coming from different countries. This means that these country-specific first principal components also shared some common stochastic trends. Although we could not identify to what these common stochastic trends correponsded, we believe that they may indicate a link to some fundamental biological processes, common to the human species, such as aging.

In the third chapter, *Forecasting Cause-Specific Mortality Rates Using the Insights*

from the Cointegration Analysis, we proposed to apply the cointegration analysis as a measure of the coherence between the country-specific experiences. For this, we extracted the country- and cause-specific time trends using the Lee-Carter model and looked for a cointegration relation between two trends corresponding to the same cause and coming from different countries. Should these two time trends be cointegrated, we can consider that the corresponding countries have had coherent development patterns in the past regarding a particular cause of death. Then, this coherence justifies modelling the cause-specific mortality rates of these two countries in a multipopulation model. Indeed, in cases where the two cause-specific time trends turned out to be cointegrated, the multipopulation model showed better forecasting results than the base scenario (Lee-Carter model).

From a practitioner's perspective, life tables represent one of the most important actuarial instruments. Any life table taking into account the information on different causes of death will have to deal with the dependency between the causes which is a challenging task. The present thesis contributes to a better unterstanding of the dependency between cause-specific mortality rates corresponding to different causes and coming from a particular country, as well as between cause-specific mortality rates corresponding to a particular cause and coming from different countries. It paves then the road to future research necessary for building the life tables enriched with the information on different causes of death.