Modeling Long-Run Cause of Death Mortality Trends

Séverine Gaille
Faculty of Business and Economics
Institute of Actuarial Science
University of Lausanne, 1015 Lausanne
Switzerland
Severine.Gaille@unil.ch
+41 21 692 33 76

Michael Sherris
Australian School of Business
School of Actuarial Studies
University of New South Wales, Sydney NSW 2052
Australia
M.Sherris@unsw.edu.au
+61 2 9385 2333

November 17, 2010

Abstract

This paper models relationships between trends in cause of death mortality rates for five main causes of death (circulatory system, cancer, respiratory system, external causes, infectious and parasitic diseases) across nine major countries (USA, Australia, Switzerland, Japan, Singapore, Italy, Norway, Sweden, UK). Trends and relationships between mortality rates for causes of death are important since these trends are hidden in aggregate data. Vector Error Correction Models (VECM) are used to model the common trends in causes of death by country. A VECM is a multivariate dynamic system allowing for long-run relationships between variables and common stochastic trends. The paper demonstrates that mortality rates by causes of death have common stochastic trends in many countries but these also differ across countries highlighting the potential for geographical diversification of mortality trends. The results confirm long-run relationships exist between the five main causes of death, indicating dependence between these competing risks. Cause of death analysis provides valuable information that can improve the estimation of aggregate mortality trends.

Keywords: causes of death, mortality trends, VECM

JEL Classifications: J11, C32, N30, G22, G23

1 Introduction

Models for trends in mortality rates for different ages and sexes as well as for different countries are often based on the assumption that past trends in historical data will continue in the future. Mortality trends and variability reflect many factors and these include changes in the causes of deaths. These causes have differing age patterns and have shown different trends over recent years. At the same time, systematic changes in causes of death have been common across the developing economies. Gaille and Sherris [2010] discuss the factors driving mortality changes based on causes of death. Tuljapurkar et al. [2000] shows how mortality declines have had common trends in the G7 countries although there is evidence of variability in those trends. Booth et al. [2006] also demonstrate common improvement trends based on the Lee-Carter model and variants of the model. Wilmoth [1995] shows how taking into account causes of death can influence projected trends and effectively highlights how cause of death trends are hidden in aggregate data. McNown and Rogers [1992] forecast cause specific rates and Barugola and Maccheroni [2007] also examine cause of death trends.

Vector Autoregressions as well as Vector Error Correction Models (VECM) have been developed in econometrics to model multivariate dynamic systems including time dependency between economic variables and allowing for stochastic trends. VECM include common stochastic trends and long-run equilibrium relationships. These models should provide a better understanding of trends in cause of death mortality rates across countries and implications for modeling aggregate mortality rates. They provide information about estimated long-run relationships between causes based on historical data.

As a result, the application of these models to cause of death mortality rates will provide valuable information about the dependence between causes of death. Indeed, dependence between competing risks are important in constructing aggregate mortality rates. Usually an assumption is made that causes of death are independent. Cause elimination models as well as cause-delay models developed by Manton et al. [1980] and Jay Olshansky [1987] are two well-known examples. Tabeau et al. [1999] as well as McNown and Rogers [1992] have considered the impact on projections of modeling mortality rates by cause of death, assuming independent causes. Estimating the common trends and relationships between the five main causes of death will improve understanding of this dependence for use in competing risk models and constructing aggregate mortality rate trends. This will better inform estimates of future mortality trends and variability.

The paper shows that although many countries have similar trends in cause of death mortality rates, there are differences in groups of countries and in the form of the long-run common stochastic trends. The paper begins with a brief description of VAR and VECM in Section 2. Section 3 summarizes the data source and cause of death rates used to estimate the models. Results from the model fitting and implications for modeling mortality trends are then discussed in Section 4. Section 5 concludes.

2 VAR and VECM Models

Vector AutoRegressive (VAR) models are used to model vectors of variables that are assumed stationary. They model expected changes allowing for lagged relationships between the variables and also for the correlations between the variables (Ndigwako Njenga and Sherris [2009]). For mortality modeling, a vector of age-based mortality rates transformed to stationary variables can be effectively modeled with a VAR. A pth-order vector

autoregression, denoted as VAR(p), based on p lags of the variables in the model is written as

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \epsilon_t, \tag{1}$$

where the n variables at time t are denoted by the $(n \times 1)$ vector $\mathbf{y}_t$, $\mathbf{c}$ is a $(n \times 1)$ vector of constants and $\mathbf{\Phi}_i$ is a $(n \times n)$ matrix of autoregressive coefficients for $i = 1, 2, \ldots, p$. The $(n \times 1)$ vector ϵ_t is a vector of white noise terms, with

$$E(\epsilon_t) = \mathbf{0}, \tag{2}$$

$$E(\epsilon_t \epsilon_l) = \begin{cases} \mathbf{\Omega} & \text{for } t = l \\ \mathbf{0} & \text{for } t \neq l, \end{cases} \tag{3}$$

where $\mathbf{\Omega}$ is a symmetric positive definite matrix. Hamilton [1994] and Lütkepohl [2005] are comprehensive references on these models.

A VAR(p) is suitable for (weakly) stationary processes with constant mean and variance. More generally, $E(\mathbf{y}_t)$ and $E(\mathbf{y}_t \mathbf{y}'_{t-j})$ are assumed independent of time t, but may depend on the time difference j.

Often variables are non-stationary and may have a trend that can be removed by taking differences. A variable (x_t) that is non-stationary can be made stationary by taking first differences

$$\nabla x_t = x_t - x_{t-1}.$$

A variable that becomes stationary by taking differences has stochastic trends. Such a variable is referred to as being integrated of order one, denoted $I(1)$. If the process is integrated of order one, differencing removes the non-stationarity and a VAR(p) can then be fitted to the differenced data. However, differencing will lose any information about long-run trends in the levels of the data. Even if the variables are non-stationary, they may move together with common stochastic trends. These common trends are modeled based on a long-run equilibrium relationship. A linear combination of these variables may then exist such that the relation is stationary even if each variable is not.

Vector Error Correction Models (VECM) include common stochastic trends using cointegration. If the n variables in the vector y_t are all $I(1)$ then, if they are cointegrated, a long-run relationship given by

$$\beta_1 y_{1t} + \beta_2 y_{2t} + \cdots + \beta_n y_{nt} = 0$$

will hold on average in the long-run. Allowing for deviations from the long-run equilibrium relationship this becomes

$$\beta_1 y_{1t} + \beta_2 y_{2t} + \cdots + \beta_n y_{nt} = z_t, \tag{4}$$

where z_t is a stochastic variable representing that deviation. If a long-run equilibrium exists, z_t will be stationary. In this case these integrated variables are referred to as cointegrated.

Equation (4) is written in vector and matrix notations as

$$\beta' \mathbf{y}_t = z_t, \tag{5}$$

with

$$\begin{aligned} \beta &= (\beta_1 \quad \beta_2 \ldots \beta_n)', \\ \mathbf{y}_t &= (y_{1t} \quad y_{2t} \ldots y_{nt})'. \end{aligned} \tag{6}$$

The vector β is referred to as a cointegrating vector. More than one cointegration relation may exist, and thus there might be more than one cointegrating vector, each being linearly independent from the others. In such a situation, the vector β of Equation (6) is a matrix with each of its columns being a cointegrating vector. Thus

$$\begin{aligned} \beta &= (\beta_{\mathbf{1}} \quad \beta_{\mathbf{2}} \ldots \beta_{\mathbf{r}}), \\ &= \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1r} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2r} \\ \vdots & & & \vdots \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nr} \end{pmatrix}, \end{aligned} \tag{7}$$

with $\beta_{\mathbf{i}}$ the ith cointegration relation, for $i = 1, 2, \ldots, r$. The stationary vector $\beta' \mathbf{y}_t$ contains the r linearly independent cointegrated relations of the n variables under study.[1]

The cointegration relations are incorporated in VAR modeling using an alternative VAR(p) representation (see, for example, Hamilton [1994] for a proof)

$$\nabla \mathbf{y}_t = \mathbf{c} + \xi_1 \nabla \mathbf{y}_{t-1} + \xi_2 \nabla \mathbf{y}_{t-2} + \cdots + \xi_{p-1} \nabla \mathbf{y}_{t-p+1} + \mathbf{\Pi} \mathbf{y}_{t-1} + \epsilon_t, \tag{8}$$

where

$$\begin{aligned} \mathbf{\Pi} &= -(\mathbf{I}_n - \mathbf{\Phi_1} - \cdots - \mathbf{\Phi_p}); \\ &= \alpha \beta'; \\ &= \text{matrix of rank } r; \\ \alpha &= \text{a } (n \times r) \text{ loading matrix ;} \\ \beta &= \text{a } (n \times r) \text{ matrix containing the } r \text{ vectors} \\ &\quad \text{forming a basis of the space of cointegration;} \\ \xi_i &= -(\mathbf{\Phi_{i+1}} + \cdots + \mathbf{\Phi_p}) \quad \text{for } i = 1, \ldots, p-1. \end{aligned}$$

Equation (8) is the Vector Error Correction Model of the cointegrated system. Each element is stationary as the first difference of an $I(1)$ process is stationary as are the cointegration relations. The loading matrix α indicates which cointegrated relation has an impact on which variable and to what extent. For example, the element α_{ij} measures the effect of the cointegrated relation j ($j = 1, \ldots, r$) on the variable i ($i = 1, \ldots, n$).

The rank of the matrix $\mathbf{\Pi}$ gives the number of cointegrated relations among the variables of the process. Three different cases are possible:

Case 1: $r = 0$ There is no cointegrated relation. A VAR($p-1$) may be applied on the first difference of the variables.

Case 2: $r = n$ All linear combinations are stationary. Thus, all the variables in the process are stationary.

Case 3: $0 < r < n$ There are r cointegrated relations, such that $\mathbf{\Pi} = \alpha \beta'$. In this case, the cointegrated relations are included in the error correction term.

Johansen's approach is used to estimate the number of cointegrated relations in a process as well as the parameters in the matrices α, β, $\mathbf{c}$ and ξ_i for $i = 1, 2, \ldots, (p-1)$ in Equation (8) (Hamilton [1994] and Lütkepohl [2005]). The following steps are used to estimate a VECM (Figure (1)):

[1] In this paper, we consider variables that are integrated of order one. In that special case, cointegrated relations are necessarily stationary. For a more general framework, see Hamilton [1994] and Lütkepohl [2005].

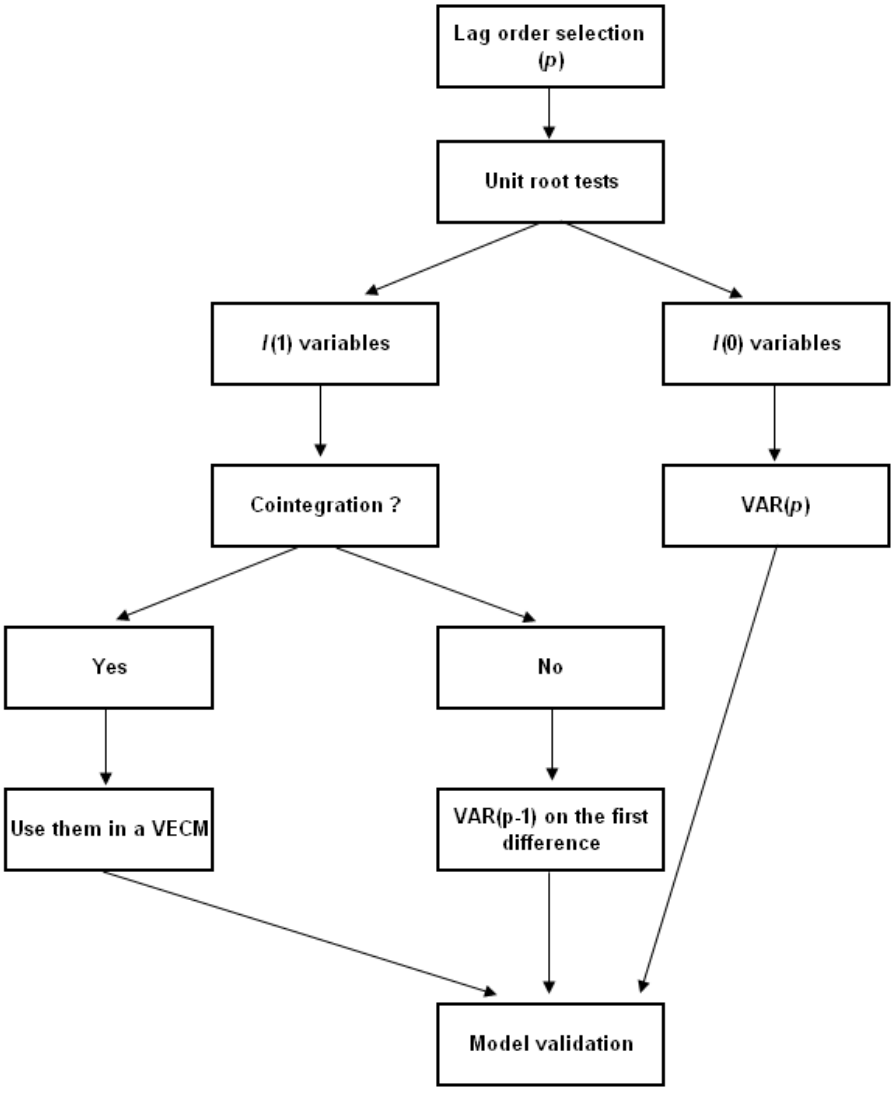

Figure 1: Steps to follow in a VECM analysis

1. Lag order of the VAR, p: Using selection criteria, such as Akaike's Information Criteria (AIC), Hannan-Quinn Criterion (HQ), Schwarz Criterion (SC), Final Prediction Error (FPE), the lag order of the VAR is selected.

2. Unit root tests on the variables considered: For a process to be stationary, the characteristic polynomial of its VAR should have all its roots outside the complex unit circle (Hamilton [1994] and Lütkepohl [2005]). Therefore, if this polynomial has a root equal to unity, some or all the variables are integrated of order one and there might be cointegrated relations among them. Unit root tests, such as the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS), the Augmented Dickey-Fuller test (ADF) or the Phillips-Perron test (PP), are useful tools in order to check for the stationarity of the variables. KPSS tests the null hypothesis that the variable is level or trend stationary, while ADF and PP test the null hypothesis of a unit root, and thus, the null hypothesis of non-stationarity.

3. If the variables are stationary, denoted $I(0)$, a VAR(p) is suitable. If the variables are $I(1)$, the Johansen's procedure is applied to find the number of cointegrated relations. Two test statistics are commonly used in order to find the number of cointegrated relations: the trace test and the maximum-eigenvalue test. The trace test compares the null hypothesis that there are r cointegrated relations against the alternative of n cointegrated relations, where n corresponds to the number of variables under observation and $r < n$. The maximum-eigenvalue statistic tests the null hypothesis of r cointegrated relations against the hypothesis of $r+1$ cointegrated relations.

4. If the variables are $I(1)$ and if there is no cointegration, a VAR($p-1$) on the first difference is estimated. Otherwise, the appropriate VECM should be found.

5. Model validation: test for residual autocorrelations and non-normality.

3 Data

Mortality rates were determined as the number of persons for each age, sex, and country who die in a particular year of a specified cause, divided by the number of persons of that age and sex in the country alive at the beginning of the year. Data were obtained from the Mortality Database administered by the World Health Organization [2009] (WHO) which contains demographic information, including the number of deaths according to the underlying cause of death, for many countries over the last 50 years for five-year age groups. Nine countries were chosen representing different countries in the developed world – North America, Europe, Asia and Oceania. Developing countries were not included since the trends in these countries are expected to be different to the developed economies and the data less reliable. The nine major countries are USA (1950–2005), Australia (1950–2003), Switzerland (1951–2005), Japan (1950–2006), Singapore (1963–2006), Italy (1951–2002), Norway (1951–2005), Sweden (1951–2005), and UK (1950–2006). The five main causes of death are diseases of the circulatory system, cancer, diseases of the respiratory system, external causes, infectious and parasitic diseases.

Causes of death are defined by the International Classification of Diseases (ICD), which ensures consistencies between countries (Table (1(a))). In this study, only the primary causes of death are considered. The ICD changed three times between 1950

and 2006, from ICD-7 to ICD-10, in order to take into account changes in science and technology and to refine the classification. The raw data are then not directly comparable for different periods. To make them comparable, comparability ratios are computed in order to smooth mortality rates across the classifications. The average of the mortality rates over the last two years of a classification is required to coincide with the average of mortality rates over the first two years of the next classification. A comparability ratio is defined as the sum of the probabilities of dying in the first two years of a new classification divided by the sum of the probabilities of dying in the last two years of the previous classification. The dates at which the countries adopted a new classification are presented in Table (1(b)). In order to obtain data comparable over the complete period under observation, the number of deaths in a new classification is divided by the comparability ratio linking this classification with the previous one and previous comparability ratios where appropriate. Most of these ratios take a value between 0.7 and 1.3. They are extremely close to one for cancer and the external causes of death. The higher and smaller values are usually at young and older ages. Discontinuities in the mortality rates at the junction points between two classifications have been removed using these comparability ratios. The analysis in this paper is applied to these adjusted mortality rates.

Table 1: International Classification of Diseases

(a) Coding system

Cause of death	ICD 7	ICD 8	ICD 9	ICD 10	
				Switzerland	Other countries
Circulatory system	A079-A086	A080-A088	B25-B30	1064	I00-I99
Cancer	A044-A060	A045-A061	B08-B17	1026	C00-D48
Respiratpry system	A087-A097	A089-A096	B31-B32	1072	J00-J99
External causes	A138-A150	A138-A150	B47-B56	1095	V00-Y89
Infectious and parasitic diseases	A001-A043	A001-A044	B01-B07	1001	A00-B99

(b) Adoption of new classifications

Country	ICD change	Year
USA	ICD7-8	1968
	ICD8-9	1979
	ICD9-10	1999
Australia	ICD7-8	1968
	ICD8-9	1979
	ICD9-10	1998
Switzerland	ICD7-8	1969
	ICD8-10	1995
Japan	ICD7-8	1968
	ICD8-9	1979
	ICD9-10	1995
Singapore	ICD7-8	1969
	ICD8-9	1979

Country	ICD change	Year
Italy	ICD7-8	1968
	ICD8-9	1979
Norway	ICD7-8	1969
	ICD8-9	1986
	ICD9-10	1996
Sweden	ICD7-8	1969
	ICD8-9	1987
	ICD9-10	1997
United Kingdom	ICD7-8	1968
	ICD8-9	1979
	ICD9-10	2001

The International Classification of Diseases changed three times between 1950 and 2006. The aim of such changes was to take into account progresses in science and technology as well as to refine the categories of the diseases in order to have a more detailed description. With ICD-7, the death numbers were classified in 150 different categories. In ICD-10, 11'468 categories and subcategories exist.

4 Long-Run Trends for Causes of Death

To examine trends by cause of death, standardized aggregate country specific mortality rate is used. To allow for changes in the age structure of the population, the aggregate country specific mortality rate is denoted by $q^*_{c,t,d,s}$, where

$$\begin{aligned} q^*_{c,t,d,s} &= d^*_{c,t,d,s}/l_{c,LY_c,s}, \\ d^*_{c,t,d,s} &= \sum_x (q_{x,c,t,d,s} \times l_{x,c,LY_c,s}), \end{aligned} \tag{9}$$

and

$$\begin{aligned} q_{x,c,t,d,s} &= \text{probability of dying in country } c, \text{ at time } t, \text{ from cause of death } d, \\ &\quad \text{and for a person of sex } s, \text{ and age } x; \\ l_{x,c,LY_c,s} &= \text{number of persons of sex } s, \text{ and age } x, \text{ alive in country } c, \\ &\quad \text{at the beginning of year } LY_c; \\ l_{c,LY_c,s} &= \sum_x (l_{x,c,LY_c,d,s}); \\ &= \text{number of persons of sex } s, \text{ alive in country } c, \\ &\quad \text{at the beginning of year } LY_c; \\ LY_c &= \text{last year under observation for country } c. \end{aligned}$$

The population of the last year under observation is used as a base. Total number of deaths in a particular year t is determined as if the population alive at the beginning of that year was the same as the population of the last year of the data period. For each country and cause $q^*_{c,t,d,s}$ refers to the country cause-specific mortality rate in year t, assuming that the population is constant during the complete period under observation and fixed at the level of the last observed year.

The VECM analysis is applied across the nine major countries for males and females. Long-run equilibrium relationships are estimated between the five main causes of death. The analysis is applied to each country separately and to the logarithm of $q^*_{c,t,d,s}$.

4.1 Lag Order Selection

Out of the four tests performed, at least two of them, if not all of them, indicate a lag order of one as optimal. A VAR(1) is the most suitable model for the aggregate standardized log-mortality rates for causes of death in each of the nine analyzed countries.

4.2 Unit Root Tests

KPSS, ADF and PP tests are performed on the data. A cause of death is said stationary when at least two out of the three tests accept it at a five percent significance level. When some doubts still remain, several models are tested and the one with non-autocorrelated and normally distributed residuals is chosen. Table (2) summarizes the causes of death that are stationary according to these tests. Across the countries most of the causes of death log-mortality rates show evidence of non-stationarity and have stochastic trends. The major exception is the diseases of the respiratory system. In the United States, Australia, Italy (females only), Sweden (females) and United Kingdom (males), the five main causes of death are non-stationary. In Switzerland (males only), Japan, Italy (males),

Norway, Sweden (males) and United Kingdom (females), log-mortality rates for diseases of the respiratory system are the only rates that are stationary. Singapore is different with log-mortality rates for infectious and parasitic diseases as the only stationary cause of death. The shorter period under observation as well as the climate of this country may explain this. Indeed, Singapore is the only country for which less than 50 years are observed and is also the only country with a tropical weather.

Table 2: Stationarity and non-stationarity of the five main causes of death in nine countries

	Males	Females
USA	All causes: UR	All causes: UR
Australia	All causes: UR	All causes: UR
Switzerland	Respiratory: S Other causes: UR	Cancer, Respiratory: S Other causes: UR
Japan	Respiratory: S Other causes: UR	Respiratory: S Other causes: UR
Singapore	I&P: S Other causes: UR	I&P: S Other causes: UR
Italy	Respiratory: S Other causes: UR	All causes: UR
Norway	Respiratory: S Other causes: UR	Respiratory: S Other causes: UR
Sweden	Respiratory: S Other causes: UR	All causes: UR
United Kingdom	All causes: UR	Respiratory: S Other causes: UR

UR = Unit root, that is a non-stationary variable; S = Stationary variable; I&P = Infectious and parasitic diseases.

This table describes the stationarity of the log-mortality rate $\log q^*_{c,t,d,s}$. A variable is said to be stationary when at least two out of the three tests (that is KPSS, ADF and PP) do not reject it at five percent significance level or when it provides the best model according to the model validation criteria.

4.3 Long-Run Equilibrium Relationships

The number of estimated cointegration relations is summarized in Table (3) based on trace and maximum-eigenvalue tests of the Johansen's procedure. These two tests assess the number of long-run equilibrium relationships among the non-stationary causes of death. Several model assumptions are tested and the most efficient one according to the model validation criteria (non-autocorrelated and normally distributed residuals) is shown. In general there is at least one cointegrating relationship between the cause of death log-mortality rates in each country showing that these rates have changed with common stochastic trends. These long-run equilibrium relationships determine how changes in causes of death move relative to each other.

Table 3: Number of cointegrated relations among the five main causes of death in nine countries

	Males	Females
USA	2	1
Australia	3	1
Switzerland	1	0
Japan	1	3
Singapore	0	1
Italy	1	1
Norway	1	1
Sweden	1	1
United Kingdom	2	1

Number of cointegrated relations according to the trace and maximum-eigenvalue tests of the Johansen's procedure at a five percent significance level, except for females in Australia and in United Kingdom as one cointegrated relation is accepted at a 2.5% significance level. For females in Singapore, Norway and Sweden, several models are tested and the table reports the best model according to the model validation criteria.

4.4 Results: Fitted VECM for Causes of Death

Parameters for the fitted VECM for each country and both sexes, based on the stationarity assumptions in Table (2) and the number of cointegrated relations shown in Table (3) are given in Tables (4), (5) and (6). For all these Tables the VECM is estimated for each country, using Johansen's procedure. The variables used in the VECM are the log-mortality rates for the five main causes of death of the country. Males in Singapore as well as females in Switzerland are not represented in these tables since no cointegration relation is found to exist. VAR models were estimated and are discussed later.

To illustrate the application of these tables, the estimated VECM for log-mortality rates by cause of death for males in the United States can be written

$$\begin{bmatrix} \Delta CircSyst_t \\ \Delta Cancer_t \\ \Delta RespSyst_t \\ \Delta ExtCauses_t \\ \Delta I\&P_t \end{bmatrix} = \begin{bmatrix} 0.43874 \\ 0.46001 \\ -0.48234 \\ -0.15794 \\ -4.20466 \end{bmatrix} + \begin{bmatrix} 0.00736 & 0.00606 \\ 0.00700 & 0.00454 \\ 0.00124 & 0.02903 \\ -0.00317 & -0.00497 \\ -0.04774 & 0.02163 \end{bmatrix} \times \begin{bmatrix} 1.03933 & -2.34554 & -0.41691 & -6.95797 & -2.15630 \\ -4.37272 & -11.39015 & 8.37977 & 5.60970 & 1.64404 \end{bmatrix} \times \begin{bmatrix} CircSyst_{t-1} \\ Cancer_{t-1} \\ RespSyst_{t-1} \\ ExtCauses_{t-1} \\ I\&P_{t-1} \end{bmatrix}. \quad (10)$$

Common features between countries are described in Tables (7) and (8).

For countries where the log-mortality rates for diseases of the respiratory system are stationary, there are two patterns in the relationships for the common stochastic trends (Table (7)). Females in Japan, males in Italy, Norway and Sweden all show similar relative changes. Diseases of the circulatory system, cancer and infectious and parasitic diseases have coefficients with the same sign, with the coefficient for external causes of death having an opposite sign. In these four cases, the long-run stochastic trends are such that decreases (increases) in the log-mortality rates of the circulatory system are associated with either increases (decreases) in log-mortality rates for cancer or the infectious and parasitic diseases, or decreases (increases) in log-mortality rates for external causes of death, or a combination of these impacts, so that overall changes are stationary. For males in Switzerland and Japan as well as females in United Kingdom, diseases of the circulatory system and infectious and parasitic diseases have a coefficient with the same sign, while the coefficient for cancer and external causes of death is of opposite sign. As log-mortality rates for diseases of the circulatory system decrease (increase), either log-mortality rates for cancer or external causes of death decrease (increase), or log-mortality rates for infectious and parasitic diseases increase (decrease) for the stochastic trends to remain in equilibrium.

Countries where all causes of death are non-stationary show two patterns in the relationships for the common stochastic trends for males shown in Table (8). There is no common relationship for females. The two long-run equilibrium patterns for males in United Kingdom are similar to one of the two relations for Australia. A decrease (increase) in the log-mortality rates of diseases of the circulatory system in these two countries implies a decrease (increase) in log-mortality rates in cancer, in the diseases of

Table 4: Constants included in a VECM based on the five main causes of death

	Males					Females				
	Circulatory system	Cancer	Respiratory system	External causes	I&P	Circulatory system	Cancer	Respiratory system	External causes	I&P
USA	0.43874	0.46001	-0.48234	-0.15794	-4.20466	0.73665	-0.34753	-2.32401	-1.39652	-7.53256
Australia	-3.89968	-2.26731	-14.24828	-5.56537	-8.19684	-1.08802	-0.42182	-6.84809	-0.51955	-2.34092
Switzerland	-0.65890	-0.95909	4.13417	-2.03313	6.07545	-	-	-	-	-
Japan	-0.58047	-0.39123	0.65844	-0.31950	0.93573	4.43095	-1.99409	18.65621	2.16433	4.17431
Singapore	-	-	-	-	-	-0.18357	-0.16679	0.45846	-0.62157	0.73467
Italy	-0.55202	-0.18801	-0.50883	-0.12029	0.02746	1.06598	-0.06050	5.19679	0.67211	0.39017
Norway	1.63403	-0.13604	-0.62175	0.74013	-6.38061	-2.48457	-3.01543	-9.70387	-1.50241	2.16388
Sweden	-0.12201	-0.20726	-1.45263	1.49949	-7.54633	-1.98343	-0.41957	-8.14290	0.42770	5.82425
UK	-1.32293	-0.41679	-5.06989	0.07603	-0.62117	-0.04302	-0.00512	-0.21213	0.05591	-0.39890

I&P = Infectious and parasitic diseases.

The constants included in the models are given in the line corresponding to the country. For example, 0.44 is the constant for log-mortality rates for diseases of the circulatory system for males in the United States, while -4.2 is for infectious and parasitic diseases.

Table 5: Cointegrated relations between the five main causes of death

		Males					Females				
		Circulatory system	Cancer	Respiratory system	External causes	I&P	Circulatory system	Cancer	Respiratory system	External causes	I&P
USA	1st	1.03933	-2.34554	-0.41691	-6.95797	-2.15630	-2.46855	18.79722	-0.11446	6.42895	-0.23762
	2nd	-4.37272	-11.39015	8.37977	5.60970	1.64404					
Australia	1st	-2.56301	-15.06544	-9.82005	11.56377	-3.71314	-11.78058	-6.42785	-8.05339	10.96734	2.66325
	2nd	-1.91375	11.44725	1.61785	-1.11874	0.50319					
	3rd	18.49540	-20.37665	-0.69072	-28.33015	-1.36802					
Switzerland	1st	-14.89733	17.51768	-	17.78248	-0.64395	-	-	-	-	-
Japan	1st	0.99443	-4.61224	-	-3.33358	1.19568	5.41306	-36.97359	-	5.49919	-0.46654
	2nd						-10.42972	56.45083	-	-1.10891	2.11263
	3rd						-2.41726	-27.85372	-	21.30805	-2.97511
Singapore	1st	-	-	-	-	-	2.10568	0.35377	-7.93496	6.45040	-
Italy	1st	8.41799	15.32077	-	-18.74788	2.39752	15.32377	11.19038	-13.14499	-4.89043	4.19417
Norway	1st	-6.79880	-8.80674	-	4.30630	-2.52123	5.90035	42.09067	-	-3.49216	-2.77631
Sweden	1st	12.07646	16.98234	-	-13.20739	3.88387	-4.60714	30.86239	6.80098	-3.54985	-3.97700
UK	1st	4.55206	-11.07908	-1.43379	3.57710	-0.75337	-2.70738	2.08495	-	1.31605	-1.53318
	2nd	1.31480	-25.36101	-6.91611	18.38995	-4.22144					

I&P = Infectious and parasitic diseases.
These results show, for example that the VECM for females in the United States has an estimated long-run equilibrium relationship given by

$$-2.47 \times CircSyst_t + 18.80 \times Cancer_t - 0.11 \times RespSyst_t + 6.43 \times ExtCauses_t - 0.24 \times I\&P_t = z_t,$$

where z_t is a stationary variable.

Table 6: Loadings of a VECM based on the five main causes of death

		Males					Females				
		Circulatory system	Cancer	Respiratory system	External causes	I&P	Circulatory system	Cancer	Respiratory system	External causes	I&P
USA	**1st**	0.00736	0.00700	0.00124	-0.00317	-0.04774	-0.00503	0.00230	0.01550	0.00922	0.05008
	2nd	0.00606	0.00454	0.02903	-0.00497	0.02163					
Australia	**1st**	-0.01737	-0.00549	-0.08965	-0.00188	-0.04695	-0.01942	-0.00765	-0.12463	-0.00921	-0.04206
	2nd	0.01969	0.00929	0.02488	0.00867	-0.04378					
	3rd	-0.00149	-0.00393	-0.00527	-0.01955	-0.02153					
Switzerland	**1st**	0.00472	0.00705	-0.03053	0.01489	-0.04505	-	-	-	-	-
Japan	**1st**	-0.01638	-0.01147	0.01947	-0.00910	0.02846	-0.01016	-0.00833	0.04035	-0.00232	0.04410
	2nd						-0.02138	0.00219	-0.04205	-0.00557	0.00866
	3rd						-0.00472	-0.00106	0.00108	0.01873	-0.00274
Singapore		-	-	-	-	-	0.01230	0.01217	-0.03460	0.04579	-0.05760
Italy	**1st**	0.02576	0.00927	0.02326	0.00539	-0.00318	-0.01927	0.00110	-0.09332	-0.01192	-0.00762
Norway	**1st**	0.01965	-0.00165	-0.00750	0.00892	-0.07600	0.01034	0.01262	0.04062	0.00627	-0.00915
Sweden	**1st**	0.00108	0.00196	0.01378	-0.01425	0.07121	0.01366	0.00289	0.05640	-0.00302	-0.04043
UK	**1st**	-0.00866	-0.00895	-0.00027	-0.00463	0.04552	-0.00352	-0.00072	-0.03381	0.01202	-0.06217
	2nd	-0.01264	-0.00185	-0.06068	0.00265	-0.02299					

I&P = Infectious and parasitic diseases.

To explain, log-mortality rates of diseases of the circulatory system for males in the United States are affected by the first cointegrated relation with a factor of 0.00736, while the second cointegrated relation has an impact of 0.00606.

Table 7: Long-run equilibrium relationships in countries with similar experience, the diseases of the respiratory system being stationary

		Circulatory system	Cancer	External causes	Infectious and parasitic diseases
Switzerland - Males		-14.90	17.52	17.78	-0.64
Japan	Males	0.99	-4.61	-3.33	1.20
	Females	-2.42	-27.85	21.31	-2.98
Italy - Males		8.42	15.32	-18.75	2.40
Norway - Males		-6.80	-8.81	4.31	-2.52
Sweden - Males		12.08	16.98	-13.21	3.88
UK - Females		-2.71	2.08	1.32	-1.53

Cointegrated relations for specified countries under study (all the cointegrated relations are presented in Table (5)). To illustrate the meaning of the table, a VECM for log-mortality rates by cause of males in Switzerland has one long-run equilibrium relationship (cointegrated relation), written as

$$-14.90 \times CircSyst_t + 17.52 \times Cancer_t + 17.78 \times ExtCauses_t - 0.64 \times I\&P_t = z_t,$$

where z_t is a stationary variable.

Table 8: Long-run equilibrium relationships in countries with similar experience, all causes of death being non-stationary, males

	Circulatory system	Cancer	Respiratory system	External causes	Infectious and parasitic diseases
USA	1.04	-2.35	-0.42	-6.96	-2.16
Australia	-1.91	11.45	1.62	-1.12	0.50
	18.50	-20.38	-0.69	-28.33	-1.37
UK	4.55	-11.08	-1.43	3.58	-0.75
	1.31	-25.36	-6.92	18.39	-4.22

Cointegrated relations for specified countries under study (all the cointegrated relations are presented in Table (5)). To illustrate the meaning of the table, a VECM for log-mortality rates by cause of males in the United States has one long-run equilibrium relationship (cointegrated relation), written as

$$1.04 \times CircSyst_t - 2.35 \times Cancer_t - 0.42 \times RespSyst - 6.96 \times ExtCauses_t - 2.16 \times I\&P_t = z_t,$$

where z_t is a stationary variable.

the respiratory system or in the infectious and parasitic diseases, or an increase (decrease) in log-mortality rates of external causes of death. The other cointegrating relation in Table (8) for Australia is similar to that for the United States. A decrease (increase) in the log-mortality rate of diseases of the circulatory system is associated with a decrease (increase) in log-mortality rates of either or a combination of the four remaining causes.

These relationships reflect the historical data and the relative changes in cause-specific mortality. Despite these similarities, there is significant variation in trends between these causes of death mortality rates across these countries.

4.5 Singapore and Switzerland: VAR Models

For males in Singapore and females in Switzerland there are no common stochastic trends found in the log-mortality rates for the causes of death. In both cases VAR models are fitted. For males in Singapore, Table (9) shows the estimated VAR fitted to the first difference of the non-stationary variables, that is on the first difference of log-mortality rates of diseases of the circulatory system, cancer, diseases of the respiratory system and external causes of death. Infectious and parasitic diseases are stationary and thus, no differencing is required. Table (10) shows the VAR model for log-mortality rates for females in Switzerland. The VAR is fitted to the first difference of the non-stationary variables, that is on the first difference of log-mortality rates for diseases of the circulatory system, external causes of death and infectious and parasitic diseases. Cancer and diseases of the respiratory system are stationary.

4.6 Model Validation

The residuals of the model are tested for normality as well as any remaining autocorrelation. Tables (11) and (12) summarize the significance of the tests for males and females respectively. The Portmanteau test is a test for the overall significance of the residual autocorrelations up to lag l. The Portmanteau statistic has an approximate asymptotic Chi-square distribution for large values of l. The test has a null hypothesis of no-autocorrelation among the residuals up to $l = 15$ and $l = 25$ lags. The statistic used is the Portmanteau statistic adjusted for small sample.[2] Tests for normality are based on the third and fourth central moments (skewness and kurtosis) of a normal distribution.[3] The test statistic labeled *both* in both tables is a joint test of skewness and kurtosis.

The null hypothesis of normality as well as the null hypothesis of no-autocorrelation up to 15 or 25 lags are, in most cases, accepted at a five percent significance level. For males in Italy as well as females in Singapore and United Kingdom, the kurtosis test and the joint test of the kurtosis and skewness reject the null hypothesis of normality. Despite this, the estimated VECM capture the trends in the causes of death data and provide a good fit based on the model assumptions.

5 Conclusion

Mortality rates of many countries show similar trends by age and by cause of death, even if these causes of death have shown differing patterns of improvement and have differential

[2] As in Lütkepohl [2005]

[3] For a detailed description of these tests, see Lütkepohl [2005].

Table 9: Autoregressive coefficients as well as the trend used in the VAR estimated for males in Singapore

	Δ Circulatory system (t)	Δ Cancer (t)	Δ Respiratory system (t)	Δ External causes (t)	Infectious and parasitic diseases (t)
Δ Circulatory system (t-1)	-0.46283	-0.02341	-0.21527	-0.09648	-0.30539
Δ Cancer (t-1)	0.14898	-0.49717	-0.05213	-0.04751	0.76079
Δ Respiratory system (t-1)	0.13283	0.00244	-0.40142	0.06997	-0.11157
Δ External causes (t-1)	0.13739	0.17410	-0.09101	-0.51364	-0.38760
Infectious and parasitic diseases (t-1)	-0.00562	-0.00500	-0.00946	-0.00362	1.00439
Trend	-0.00262	-0.00212	-0.00392	-0.00228	-0.00153

The table reads as follows: The first difference of log-mortality due to cancer at time $t-1$ impacts the first difference of log-mortality due do the diseases of the circulatory system at time t with coefficient 0.149. The diseases of the circulatory system are affected by the five causes as follows

$$\begin{aligned}\nabla CircSyst_t = \quad & - \quad 0.46283 \times \nabla CircSyst_{t-1} + 0.14898 \times \nabla Cancer_{t-1} + 0.13283 \times \nabla RespSyst_{t-1} \\ & + \quad 0.13739 \times \nabla ExtCauses_{t-1} - 0.00562 \times I\&P_{t-1} - 0.00262 \times t.\end{aligned}$$

Table 10: Autoregressive coefficients, the constant as well as the trend for the VAR estimated for females in Switzerland

	Δ Circulatory system (t)	Cancer (t)	Respiratory system (t)	Δ External causes (t)	Δ Infectious and parasitic diseases (t)
Δ Circulatory system (t-1)	0.14614	0.03040	-0.16896	0.21790	0.30522
Cancer (t-1)	-1.42405	0.16423	-1.10560	-0.60341	-0.12140
Respiratory system (t-1)	-0.09331	-0.02942	0.48664	-0.04918	-0.14489
Δ External causes (t-1)	0.17695	0.17611	0.44943	-0.18828	0.07370
Δ Infectious and parasitic diseases (t-1)	0.02529	0.05968	-0.00555	-0.03199	-0.17226
Trend	-0.00817	-0.00435	-0.01076	-0.00384	-0.00043
Constant	-8.96603	-5.08569	-10.01289	-3.86021	-1.78953

The table reads as follows: Log-mortality due to cancer at time $t-1$ impacts the first difference of log-mortality due do the diseases of the circulatory system at time t with coefficient -1.42. Diseases of the circulatory system are related to the five causes as follows

$$\begin{aligned}\nabla CircSyst_t = &+ 0.14614 \times \nabla CircSyst_{t-1} - 1.42405 \times Cancer_{t-1} - 0.09331 \times RespSyst_{t-1} \\ &+ 0.17695 \times \nabla ExtCauses_{t-1} + 0.02529 \times \nabla I\&P_{t-1} - 0.00817 \times t - 8.96603.\end{aligned}$$

Table 11: Tests on residuals of the fitted VECM on causes of death, males

	Portmanteau test		Normality tests		
	15 lags	25 lags	skewness	kurtosis	both
USA	***	***	***	***	***
Australia	***	**	***	***	***
Switzerland	***	***	***	***	***
Japan	***	***	***	***	**
Singapore	***	***	***	***	***
Italy	***	***	***	—	—
Norway	*	*	***	***	***
Sweden	***	***	***	***	***
UK	***	***	***	***	***

* The null hypothesis is accepted at a one percent significance level.
** The null hypothesis is accepted at a 2.5% significance level.
*** The null hypothesis is accepted at a five percent significance level.
– The null hypothesis is rejected.

The Portmanteau statistic tests the null hypothesis of no-autocorrelation among the residuals up to 15 or 25 lags. The normality tests for the residuals are based on the skewness statistic, the kurtosis statistic and a combination of these.

Table 12: Tests on residuals of the fitted VECM on causes of death, females

	Portmanteau test		Normality tests		
	15 lags	25 lags	skewness	kurtosis	both
USA	***	**	***	***	***
Australia	***	***	***	**	***
Switzerland	***	***	***	***	***
Japan	***	***	***	***	***
Singapore	***	***	***	—	—
Italy	***	***	***	***	***
Norway	***	***	***	***	***
Sweden	***	***	***	***	***
UK	***	***	***	—	—

* The null hypothesis is accepted at a one percent significance level.
** The null hypothesis is accepted at a 2.5% significance level.
*** The null hypothesis is accepted at a five percent significance level.
– The null hypothesis is rejected.
The Portmanteau statistic tests the null hypothesis of no-autocorrelation among the residuals up to 15 or 25 lags. The normality tests for the residuals are based on the skewness statistic, the kurtosis statistic and a combination of these.

impacts by age. Common international changes such as development of national health care systems, launch of smoking control measures and other similar health policy changes are impacting mortality rates leading to common trends across countries. As a result, longevity and mortality risk across countries and within a country across causes of death contains common stochastic trends. It is important to incorporate these common trends in longevity and mortality risk models.

By considering aggregate cause of death mortality rates and using models with long-run common stochastic trends, it is possible to estimate equilibrium relationships arising from different causes of death. Comparing these trends across countries allows to identify countries with similar trends. This study uses a multivariate dynamic systems to model log-mortality rates for causes across nine countries. VECM are found to fit accurately the historical data and the dynamics of cause-specific mortality rates.

The results show that long-run equilibrium relationships exist between the mortality rates for the five main causes of death. This confirms the nature of dependence between these competing risks. The often made assumption of independence between mortality rates for causes of deaths is shown not to hold as these rates have common stochastic trends at a country level. Lon-run equilibrium relationships should not be disregarded in any analysis considering the causes of death and should be included in new forecasting mortality models.

The study also demonstrates that groups of countries have similar experience. Females in Japan, males in Italy, Norway and Sweden show similar relative past changes. Males in Switzerland and Japan have similar long-run equilibrium relationships as females in United Kingdom. Males in Australia share similar pattern with males in United States as well as with males in United Kingdom. This information is of primary importance as it highlights the potential for geographical diversification of mortality risk.

6 Acknowledgement

The authors acknowledge the support of ARC Linkage Grant Project LP0883398 Managing Risk with Insurance and Superannuation as Individuals Age with industry partners PwC and APRA. Gaille acknowledges scholarship support from the Swiss National Science Foundation for the project Managing Risk as Individuals Age with Insurance and Superannuation, number PBLAP1-124258.

References

T. Barugola and C. Maccheroni. Sensitivity Analysis of the Lee-Carter Model Fitting Mortality by Causes of Death. *Società Italiana di Statistica, Rischio e Previsione, Atti della riunione intermedia, Padova*, pages 481–482, 2007.

H. Booth, Rob J. Hyndman, L. Tickle, and P. de Jong. Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research*, 15: 289–310, 2006.

S. Gaille. *Improving Longevity and Mortality Risk Models* . PhD thesis, University of Lausanne, 2010.

S. Gaille and M. Sherris. Age Patterns and Trends in Mortality by Cause of Death and Implications for Modeling Longevity Risk . Technical report, The Social Science Research Network Electronic Paper Collection, 2010.

A. M. Gittelsohn. On the Distribution of Underlying Causes of Death. *American Journal of Public Health*, 72(2):133–140, 1982.

S. Gutterman and Irwin T. Vanderhoof. Forecasting Changes in Mortality: A Search for a Law of Causes and Effects. *North American Actuarial Journal*, 2(4):135–138, 1998.

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

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

S. Jay Olshansky. Simultaneous/Multiple Cause-Delay (SIMCAD): An Epidemiological Approach to Projecting Mortality. *Journal of Gerontology*, 42(4):358–365, 1987.

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

Kenneth G. Manton, Clifford H. Patrick, and Eric Stallard. Mortality Model Based on Delays in Progression of Chronic Diseases: Alternative To Cause Elimination Model. *Public Health Reports*, 95:580–588, 1980.

R. McNown and A. Rogers. Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting*, 8:413–432, 1992.

Ministry of Health. Ministry of health, singapore, November 2009. `http://www.moh.gov.sg`.

C. Ndigwako Njenga and M. Sherris. Longevity Risk and the Econometric Analysis of Mortality Trends and Volatility. Technical report, The Social Science Research Network Electronic Paper Collection, 2009.

E. Tabeau, P. Ekamper, C. Huisman, and A. Bosch. Improving Overall Mortality Forecasts by Analysing Cause-of-Death, Period and Cohort Effects in Trends. *European Journal of Population*, 15:153–183, 1999.

S. Tuljapurkar and C. Boe. Mortality Change and Forecasting: How Much and How Little Do We Know? *North American Actuarial Journal*, 2(4):13–47, 1998.

S. Tuljapurkar, Nan Li, and C. Boe. A universal pattern of mortality decline in the G7 countries. *Nature*, 405:789–792, 2000.

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

World Health Organization. Who mortality database, January 2009. `http://www.who.int/whosis/mort/download/en/index.html`.