

Modelling Mortality with Common Stochastic Long-Run Trends

S everine Gaille^a and Michael Sherris^b

^aFaculty of Business and Economics, Department of Actuarial Science, University of Lausanne, 1015 Lausanne, Switzerland.

E-mail: Severine.Gaille@unil.ch

^bAustralian School of Business, School of Actuarial Studies, University of New South Wales, Sydney NSW 2052, Australia.

E-mail: M.Sherris@unsw.edu.au

Modelling mortality and longevity risk is critical to assessing risk for insurers issuing longevity risk products. It has challenged practitioners and academics alike because of first the existence of common stochastic trends and second the unpredictability of an eventual mortality improvement in some age groups. When considering cause-of-death mortality rates, both aforementioned trends are additionally affected by the cause of death. Longevity trends are usually forecasted using a Lee-Carter model with a single stochastic time series for period improvements, or using an age-based parametric model with univariate time series for the parameters. We assess a multivariate time series model for the parameters of the Heligman-Pollard function, through Vector Error Correction Models which include the common stochastic long-run trends. The model is applied to circulatory disease deaths in U.S. over a 50-year period and is shown to be an improvement over both the Lee-Carter model and the stochastic parameter ARIMA Heligman-Pollard model. *The Geneva Papers* (2011) 36, 595–621. doi:10.1057/gpp.2011.19

Keywords: mortality trends; Heligman-Pollard model; Lee-Carter model; VECM; causes of death; mortality forecasts

Introduction

Longevity and mortality improvement has been driven by many factors over the last century. These factors have impacted different ages to varying extents resulting in common improvements across age groups. As a result, longevity and mortality risk across ages is expected to include common stochastic trends. The Lee-Carter model,¹ and variants of it, is often used for mortality and longevity risk applications by industry practitioners and academics. This model is highly parameterized and includes a single common stochastic trend across ages fitted using a time series model. An alternative approach² is to fit a parametric model to the age structure of mortality and apply time series models to the parameters. This provides for a more parsimonious model but does not capture common stochastic trends across the age structure of mortality. Our first attempt to improve these models was to use a Vector Autoregression (VAR) for the parameters which captures dependence between the

¹ Lee and Carter (1992).

² McNown and Rogers (1989, 1992).

parameters. This was further used in an analysis of mortality trends in a number of developed countries.³ However, VAR models do not include long-run relationships between the parameters that will be important in maintaining the relative levels of mortality for different ages. Our second attempt was then to use Vector Error Correction Models (VECM) that are applied in econometrics to model multivariate dynamic systems. They include time dependency between economic variables, common stochastic trends and long-run equilibrium relationships.

Modelling cause-of-death mortality rates is increasingly important in forecasting trends in aggregate mortality improvement.⁴ Models that better capture trends and that can be applied to cause-of-death mortality rates will provide practitioners with an improved understanding of mortality risks and factors determining future mortality trends. Indeed, causes of death reflect underlying socio-economic factors and provide important insights into future trends and volatility for longevity and mortality modelling.⁵ Cause elimination models as well as cause-delay models have been developed⁶ assuming independent causes of death. Projections of mortality rates by cause of death have also been considered.⁷ The Lee-Carter model has been used for cause of death data.⁸ In general, models used for mortality at a population level are not appropriate for cause-specific rates as demonstrated by an analysis of six causes of death on Italian data.⁹ Different causes of death have different age patterns, different time trends and these vary by country.¹⁰ New approaches are then needed to model mortality by cause of death. A flexible model that can capture different age patterns and maintain consistent patterns while allowing trends to differ by age is required for a cause of death approach.

This paper models the age structure of cause-specific mortality rates with a parametric mortality model, the Heligman-Pollard model.¹¹ The parameters for this model are readily interpreted according to their impact on different age ranges. These parameters are modelled with a VECM allowing for time dependency and long-run trends between the parameters. To demonstrate the model performance, it is estimated over the period 1950–2000 for United States circulatory system deaths. The VECM approach is shown to be an improvement over the Lee-Carter model and approaches using an Autoregressive Integrated Moving Average (ARIMA) process for the parameters and allows a more realistic quantification of risk.

This paper provides practitioners with new methods for modelling mortality, in particular, cause-of-death mortality rates. The paper demonstrates the benefits of incorporating common trends in longevity and mortality risk models to be used by insurers issuing longevity and mortality risk products. The model assessed in the paper

³ Njenga and Sherris (2009).

⁴ Gaille (2010); Gutterman and Vanderhoof (1998); Tabeau *et al.* (1999).

⁵ Hanewald (2011).

⁶ Manton *et al.* (1980) and Olshansky (1987).

⁷ Tabeau *et al.* (2001).

⁸ Wilmoth (1993).

⁹ Barugola and Maccheroni (2007).

¹⁰ Gaille and Sherris (2010).

¹¹ Heligman and Pollard (1980).

improves existing models and this is demonstrated in the improved forecasting performance.

The paper begins with a summary of the theoretical background for a VECM analysis in the next section, following which the data used is described. Long-run equilibriums between age-based risk factors are estimated and discussed in the later section. An improved mortality forecasting approach using a stochastic parameter Heligman-Pollard VECM is provided in the subsequent section. The improved accuracy of this new forecasting approach over the Lee-Carter model and ARIMA processes is then demonstrated for United States circulatory system deaths in the penultimate section. The final section concludes.

VAR and VECM, theoretical background

VAR model dynamic interactions between a set of variables. They capture dependence through time and between variables. This multivariate time series approach has been popular among economists for several decades and is often used to model time series of economic variables. VECM, an extension of VAR, include long-run equilibrium relationships between variables, also referred to as common stochastic trends, using the concept of cointegration.

A p th-order VAR, denoted as VAR(p), explains each variable with p lags of itself and the other variables in the model. Denoting the n variables at time t by the $(n \times 1)$ vector \mathbf{y}_t , a VAR is written as

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \dots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t,$$

where \mathbf{c} is a $(n \times 1)$ vector of constants and Φ_i is a $(n \times n)$ matrix of autoregressive coefficients for $i=1, 2, \dots, p$. The $(n \times 1)$ vector $\boldsymbol{\varepsilon}_t$ is a vector of white noises, with

$$E(\boldsymbol{\varepsilon}_t) = \mathbf{0},$$

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_l) = \begin{cases} \boldsymbol{\Omega} & \text{for } t = l \\ \mathbf{0} & \text{for } t \neq l, \end{cases}$$

where $\boldsymbol{\Omega}$ is a symmetric positive definite covariance matrix.

Estimates of the parameters of a VAR(p) and the associated asymptotic distributions assume stationarity of the process. A VAR(p) is (weakly) stationary if its first and second moments are constant over time, that is $E(\mathbf{y}_t)$ and $E(\mathbf{y}_t \mathbf{y}'_{t-j})$ are independent of time t , although they will usually differ with the time lag j . In this case the process has a constant mean (no trend) and its variance does not change over time. However, many time series do have a trend so that non-stationarity must be considered as well.

If a variable (x_t) is non-stationary and becomes stationary by taking first differences

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

then the variable is referred to as 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. 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 can be captured by a long-run equilibrium relationship. A linear combination of these variables will then exist such that the relation is stationary even if each variable is not.

Consider the n variables in vector \mathbf{y}_t , all $I(1)$, and related by

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

This relationship holds on average in the long run. At a particular point in time, there will be deviations from the equilibrium such that

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

where z_t is a stochastic variable representing that deviation.¹² If a long-run equilibrium exists, then z_t will be stationary. In this case these integrated variables are referred to as cointegrated and the above relation as the cointegration relation.

The cointegration relation can be written in vector and matrix notations as

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

with

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

The vector $\boldsymbol{\beta}$ 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. For example, if there are five variables, the first two may be linked by a cointegration relation and the last three by another. In such a situation, the vector $\boldsymbol{\beta}$ is a matrix with each of its columns being a cointegrating vector. Thus

$$\begin{aligned} \boldsymbol{\beta} &= (\boldsymbol{\beta}_1 \quad \boldsymbol{\beta}_2 \dots \boldsymbol{\beta}_r) \\ &= \begin{pmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1r} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2r} \\ \vdots & & & \vdots \\ \beta_{n1} & \beta_{n2} & \dots & \beta_{nr} \end{pmatrix}, \end{aligned}$$

with $\boldsymbol{\beta}_i$ the i th cointegration relation, for $i=1, 2, \dots, r$. The stationary vector $\boldsymbol{\beta}' \mathbf{y}_t$ contains the r linearly independent cointegrated relations of the n variables

¹² Lütkepohl (2005).

under study.¹³ If the columns of β represent all the linearly independent relations, that is all other cointegrated relations are a linear combination of the columns of β , then there are exactly r cointegrating relations among the elements of y_t and $(\beta_1 \ \beta_2 \ \dots \ \beta_r)$ forms a basis of the space of cointegrating vectors.

The cointegrated relations, given by the matrix β , are not uniquely defined. Each cointegrating vector could be multiplied by a constant and the relationship will remain the same. Broadly speaking, for any non-zero vector v with r elements, $v'\beta'y_t$ also represents the r cointegrated relations. As a result constraints are required for some values of the matrix β .

The cointegration relations are important in VAR modelling. A VAR(p) can equivalently be written as

$$\nabla y_t = c + \xi_1 \nabla y_{t-1} + \xi_2 \nabla y_{t-2} + \dots + \xi_{p-1} \nabla y_{t-p+1} + \Pi y_{t-1} + \varepsilon_t,$$

where

$$\begin{aligned} \Pi &= -(\mathbf{I}_n - \Phi_1 - \dots - \Phi_p); \\ &= \alpha\beta'; \\ &= \text{matrix of rank } r; \end{aligned}$$

α = a $(n \times r)$ loading matrix;

β = a $(n \times r)$ matrix containing the r vectors forming a basis of the space of cointegration;

$$\xi_i = -(\Phi_{i+1} + \dots + \Phi_p) \quad \text{for } i=1, \dots, p-1.$$

which is known as the VECM 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 ($j=1, \dots, r$) on the variable i ($i=1, \dots, n$).

The rank of the matrix Π indicates the number of cointegrated relations among the variables of the process. Three different cases are possible: There is no cointegrated relation ($r=0$). A VAR($p-1$) may be applied on the first differences of the variables. Or all linear combinations are stationary ($r=n$). Thus, all the variables in the process are stationary. Or there are r cointegrated relations ($0 < r < n$), such that $\Pi = \alpha\beta'$. 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 , c and ξ_i for $i=1, 2, \dots, (p-1)$. As a summary, the following steps are used to estimate a VECM:

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

¹³ In this paper, we consider variables that are integrated of order one, in which case, cointegrated relations are necessarily stationary. The general framework is given in Hamilton (1994) and L utkepohl (2005).

2. Unit root tests on the variables: For a process to be stationary, the characteristic polynomial of its VAR should have all its roots outside the complex unit circle.¹⁴ 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 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 in order 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.
4. If the variables are $I(1)$ and if there is no cointegration, a VAR($p-1$) on the first difference is fitted. Otherwise, the appropriate VECM should be found.
5. Model validation: test for residual autocorrelations and non-normality.

Figure 1 provides a summary diagram of the estimation process.

Data

The data used are for five-year age group cause-specific mortality rates collected from the Mortality Database administered by the WHO.¹⁵ Mortality rates were determined as the number of persons for each age group and sex who die in a particular year of a specified cause, divided by the number of persons of that age group and sex alive at the beginning of the year. Mid-year population provided by the WHO database is used as an approximation for the beginning-of-the-year population. Indeed, as the population increases over time, mid-year population is an appropriate approximation, partly including migrations. Besides, diseases of the circulatory system for females in the United States over the period 1950–2005 are used and are known to be reliable. Circulatory diseases are recognised as the most important cause of death in developed economies, especially for ages above 30–50.¹⁰ They are important in modelling future aggregate mortality rates as well as capturing the most recent trends in mortality improvement. This cause represented around 40 per cent of total deaths in United States in 1955, 56 per cent in 1970, 50 per cent in 1985 and 41 per cent in 2000.

Causes of death are defined by the International Classification of Diseases (ICD), which ensures consistencies between countries. The ICD changed three times between 1950 and 2006, from ICD-7 to ICD-10, in order to take into account progresses in science and technology and to refine the classification. In the United States, the ICD changed in 1968, 1979 and 1999. The raw data is then not directly comparable for different periods. To make it comparable, comparability ratios are computed. The aim is to smooth mortality rates across the classifications. The average of the mortality

¹⁴ Hamilton (1994) and Lütkepohl (2005).

¹⁵ World Health Organization (2009).

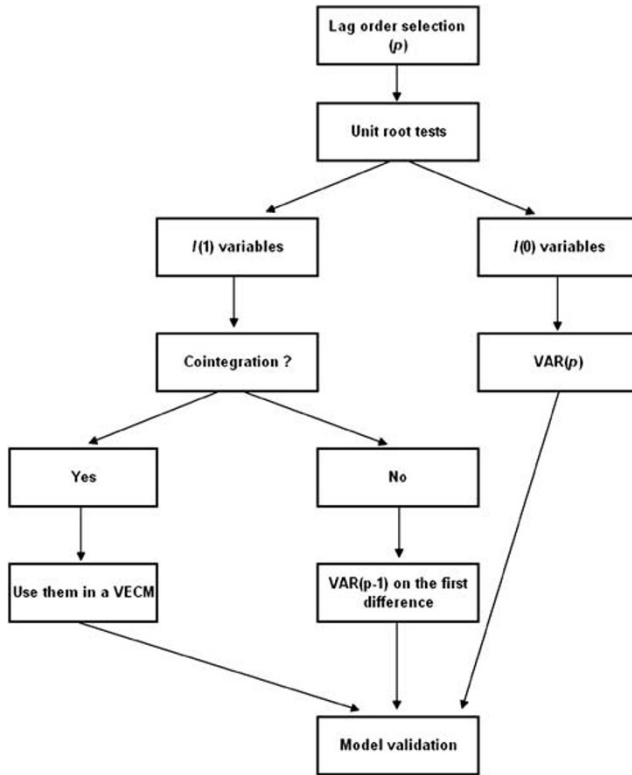


Figure 1. Process for estimating a VECM.

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.

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. Discontinuities in the mortality rates at the junction points between two classifications have been adjusted by these comparability ratios. The data used are the adjusted mortality rates.

Long-run equilibrium among age-based risk factors

The Heligman-Pollard (H-P) model has been used to model mortality rates in the United States over the period 1900–1985¹⁶ and the multi-exponential model for

¹⁶ McNown and Rogers (1989).

cause-specific mortality rates in the United States over the period 1960–1985.¹⁷ In previous research, the resulting time series of the parameters have been modelled with ARIMA processes and these models are used to forecast the value of the parameters and hence the complete age profile of mortality. Although univariate models are assumed for the parameters, the authors recognise the potential improvement from incorporating the covariation between the parameters.

We extend this approach, using multivariate time series allowing for time dependency and long-run trends between the parameters. The H-P model is used for cause-specific mortality rates and a VECM is used to model the parameters as stochastic factors. The H-P model is a concise representation of mortality by age, each parameter having a demographic meaning and is given by

$$q_{x,t} = A_t^{(x+B_t)C_t} + D_t e^{-E_t(\ln(x)-\ln(F_t))^2} + \frac{G_t H_t^x}{1 + K_t G_t H_t^x},$$

where

$q_{x,t}$ = probability of dying at age x , in year t ;

$p_{x,t} = 1 - q_{x,t}$;

Z_t = value of the parameter Z at time t ,

for $Z \in (A, B, C, D, E, F, G, H, K)$.

This model is a sum of three terms: the first represents mortality rates during childhood; the second, mortality at middle ages (the accident hump); and the last, mortality rates at older ages. The parameters have interpretation as factors impacting differing age ranges. Although the H-P model has nine parameters, which means the model has nine stochastic factors, these are required in order to capture the shape of the mortality curve for the whole age range. Table 1 summarises the demographic meaning of the nine parameters and the range of values they may have.

The model estimation process in Figure 1 is applied to the mortality rates due to diseases of the circulatory system for females in the United States. The parameters of the H-P function are first estimated by weighted least squares, the weights used implying a minimisation of the relative error as suggested by Heligman and Pollard.¹¹ Because of issues regarding multicollinearity, irregular changes of the parameter estimates from one year to another may occur even if the mortality rates of adjoining years are similar.¹⁸ This multicollinearity makes it more difficult to fit the H-P function to any given data set. Without setting constraints on the parameters, convergence problems occur during the non-linear minimisation, as parameter B tends to increase to values greater than ten, even if it was expected to be smaller than one based on previous studies,¹⁹ and parameter F tends to reach the upper bound of 150.

¹⁷ McNown and Rogers (1992).

¹⁸ Bell (1997) and Tabeau *et al.* (1999).

¹⁹ Gaille (2010); Heligman and Pollard (1980); McNown and Rogers (1989).

Table 1 Parameters of the Heligman-Pollard model

<i>Age range</i>	<i>Parameter</i>	<i>Demographic meaning</i>
Childhood	$A \in (0, 1)$	Level of infant mortality
	$B \in (0, \infty)$	Mortality change between age zero and age one (the larger the value, the smaller the change)
	$C \in (0, 1)$	Speed at which the mortality rates decline at young ages
Middle ages:	$D \in (0, 1)$	Intensity of the accident hump
Accident hump	$E \in (0, \infty)$	Age range at which the accident hump occurs (varies inversely with the spread of the hump)
	$F \in (0, 150)$	Age at which the hump is the highest
Older ages	$G \in (0, 1)$	General level of mortality rates
	$H \in (0, \infty)$	Steepness of the curve
	$K \in (-\infty, \infty)$	Curvature at the last ages (a positive value indicates that the age pattern of mortality on a semi-logarithmic scale is concave downward; a negative value indicates a concave upward curve)

Fixing the value of some of the parameters, either *a priori* or estimated from the data, has been proposed as a solution,²⁰ thus reducing the number of stochastic factors in the model. Since parameter *B* converges to an optimum value of one for a few years, without setting any constraints during the minimisation process, we fixed it to be equal to one. This assumption does not have an important effect on the fit of the H-P function on mortality as parameter *B* has a negligible impact at ages other than zero. Parameter *F* is also fixed at its median value of 33.85. By fixing these two parameters, the non-linear minimisation converges without any difficulties to an optimum value. Only seven parameters are then required in the VECM and forecasted to give the complete age profile of mortality rates.

Lag order selection

A lag order of one was found to provide a best fit to the parameters for the mortality rates, which is consistent with the autoregressive nature of mortality. This is determined based on AIC, HQ, SC and FPE tests performed on the data. These four criteria have the smallest value at a lag order of one, even with a constant or a trend included in the VAR for the parameters.

Unit root tests

KPSS, ADF and PP tests reveal similar results. The only parameter clearly trend stationary according to the three tests is parameter *C*, even at a 1 and 10 per cent significance levels. KPSS test does not reject the null hypothesis of stationarity at a 10 per cent significance level. ADF and PP tests reject the null hypothesis of non-stationarity at a

²⁰ Gaille (2010); McNown and Rogers (1992); Tabeau *et al.* (1999).

Table 2 Tests for the number of cointegrated relations, 1950–2005, Circulatory system, U.S.A., females (a) Trace test; (b) Maximum-eigenvalue test

<i>r</i>	Trace stat	Critical values			
		10%	5%	2.50%	1%
<i>(a)</i>					
5	0.66	2.69	3.76	4.95	6.65
4	7.80	13.33	15.41	17.52	20.04
3	25.14	26.79	29.68	32.56	35.65
2	48.59	43.95	47.21	50.35	54.46
1	88.91	64.84	68.52	71.8	76.07
0	142.96	89.48	94.15	98.33	103.18
<i>r</i>	Eigen stat	Critical values			
		10%	5%	2.50%	1%
<i>(b)</i>					
5	0.66	2.69	3.76	4.95	6.65
4	7.14	12.07	14.07	16.05	18.63
3	17.34	18.6	20.97	23.09	25.52
2	23.45	24.73	27.07	28.98	32.24
1	40.32	30.9	33.46	35.71	38.77
0	54.06	36.76	39.37	41.86	45.1

Notes: A null hypothesis is accepted at a $\alpha\%$ significance level when the statistic is lower than the corresponding critical value. Thus, these tables indicate that two cointegrated relations are accepted at a 2.5% significance level.

These two tests assess the number of long-run equilibrium relationships among the parameters of the Heligman-Pollard model (A , D , E , G , H and K) that is among age-based risk factors.

1 per cent significance level. Results are less conclusive for parameter D . ADF and PP tests reject the null hypothesis of non-stationarity at a 5 per cent significance level, but do not reject it at a 1 per cent significance level. KPSS test does not reject the null hypothesis of stationarity at a 1 per cent significance level but rejects it at the 5 per cent level. All three tests confirm the five remaining parameters as non-stationary.

In order to allow for this, two models are fitted, one assuming only parameter C is trend stationary and the other assuming parameters C and D are trend stationary. Normality tests on the residuals of both models as well as autocorrelations among the residuals show that the best results are obtained with parameter D modelled as non-stationary. This model is the one used. Cointegrated relations model then common stochastic trends in the six non-stationary parameters A , D , E , G , H and K .

Cointegrated relations

Table 2 presents the results for the trace test and the maximum-eigenvalue test of the Johansen's procedure. These two tests assess the number of long-run equilibrium relationships among the stochastic parameters (factors) of the Heligman-Pollard model (A , D , E , G , H and K). 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 $r < n$. This test indicates that we do not reject the null hypothesis of two, three, four or five cointegrated relations against six cointegrated relations at a 2.5 per cent significance level.

The maximum-eigenvalue statistic tests the null hypothesis of r cointegrated relations against the hypothesis of $r + 1$ cointegrated relations. Table 2(b) shows that the null hypothesis of two cointegrated relations is not rejected, while the null hypothesis of one cointegrated relation is rejected indicating that two cointegrated relations exist. These two long-run equilibrium relationships for the parameters of the Heligman-Pollard model determine how changes in the age-based risk factors move relative to each other. They allow for stochastic trends in the mortality curve, while maintaining long-run relationships between ages.

Fitted VECM

Johansen’s procedure is used to estimate the resulting VECM. The fitted VECM with B and F fixed is

$$\begin{bmatrix} \nabla A_t \\ \nabla C_t \\ \nabla D_t \\ \nabla E_t \\ \nabla G_t \\ \nabla H_t \\ \nabla K_t \end{bmatrix} = \begin{bmatrix} 4.34 \cdot 10^{-4} \\ 0.12 \\ -4.03 \cdot 10^{-3} \\ 5.27 \cdot 10^2 \\ 4.57 \cdot 10^{-5} \\ -0.91 \\ -6.62 \cdot 10^2 \end{bmatrix} + \begin{bmatrix} -1.97 \cdot 10^{-6} & 8.99 \cdot 10^{-8} \\ -2.98 \cdot 10^{-3} & -3.03 \cdot 10^{-4} \\ 3.02 \cdot 10^{-6} & -2.89 \cdot 10^{-6} \\ -0.43 & 0.37 \\ -2.61 \cdot 10^{-7} & 2.24 \cdot 10^{-9} \\ 1.61 \cdot 10^{-3} & -5.26 \cdot 10^{-4} \\ 0.69 & -0.45 \end{bmatrix} \times \begin{bmatrix} A_{t-1} \\ D_{t-1} \\ E_{t-1} \\ G_{t-1} \\ H_{t-1} \\ K_{t-1} \end{bmatrix} .$$

$$\times \begin{bmatrix} -1.02 \cdot 10^4 & 8.50 \cdot 10^4 & -0.39 & -1.73 \cdot 10^6 & -1.41 \cdot 10^2 & 0.15 \\ -6.52 \cdot 10^3 & -1.52 \cdot 10^5 & -0.36 & 2.79 \cdot 10^6 & 1.08 \cdot 10^3 & -1.31 \end{bmatrix} \times$$

The second term on the right-hand side of this equation gives the effect of the two cointegrating relationships between the parameters. They can be written as

$$\begin{aligned} z_{1t} &= -1.02 \cdot 10^4 A_t + 8.50 \cdot 10^4 D_t - 0.39 E_t \\ &\quad - 1.73 \cdot 10^6 G_t - 1.41 \cdot 10^2 H_t + 0.15 K_t, \\ z_{2t} &= -6.52 \cdot 10^3 A_t - 1.52 \cdot 10^5 D_t - 0.36 E_t \\ &\quad + 2.79 \cdot 10^6 G_t + 1.08 \cdot 10^3 H_t - 1.31 K_t, \end{aligned}$$

where z_{1t} and z_{2t} are two stochastic variables representing the deviation from the equilibrium. These two variables are stationary. The sign associated with each

parameter is of significance in these relations.²¹ The first cointegrated relation shows that as mortality around age zero decreases, that is parameter A decreases, then either the accident hump will decrease or impact a smaller age range, so that parameter D will decline or E will increase, or mortality for the elderly will increase, with parameter K declining, parameter G or parameter H increasing, or a combination of these impacts. This relationship reflects the historical data and the relative changes in the mortality curve across all ages. The VECM assumes these estimated long-run relationships will continue when used for forecasting. The curve changes stochastically but is constrained by the long-run equilibrium relationships between the parameters given by the cointegrating vectors.

Increases in mortality at older ages will occur in the Heligman-Pollard model if there is an increase in either or both parameters G and H or a decrease in parameter K . This is captured in both cointegrating vectors as the coefficients for parameters G and H have the same sign, while the coefficient of K is of an opposite sign.

In a comparison of mortality for males and females in Switzerland and in the United States using the Heligman-Pollard model,²² mortality rates at older ages are shown to be lower, but the age pattern steeper, in Switzerland than in the United States, since parameter G is smaller and parameter H higher in Switzerland. Similarly, female mortality at older ages is lower and increases at a faster pace than male mortality in both countries. This link between these two parameters is reflected in the cointegrations, as a decrease in parameter G may produce an increase in parameter H for the relation to stay at the equilibrium.

The relation between these parameters that is reflected in the cointegrated relations can be understood based on recent changes in the leading causes of death. Deaths from circulatory system have been more important in the past. More recently, cancer is becoming the major cause of death for middle ages, while the circulatory system still remains the most important cause for ages above 75.¹⁰ Thus, circulatory mortality at middle ages has decreased more abruptly over the last 50 years than mortality at older ages, which is reflected through a decrease in parameter G and an increase in parameter H . The fitted mortality rates using the model are given in Figure 2 along with the original rates for comparison. The model fits the data well reproducing the age structure and changes through time.

Model validation

The residuals of the model are tested for normality as well as any remaining autocorrelation. The test statistics are summarised in Table 3. 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

²¹ The size of the coefficients is influenced by the value taken by the parameters. For example, as parameters A , D and G are small, their associated coefficient is high.

²² Gaille (2010).

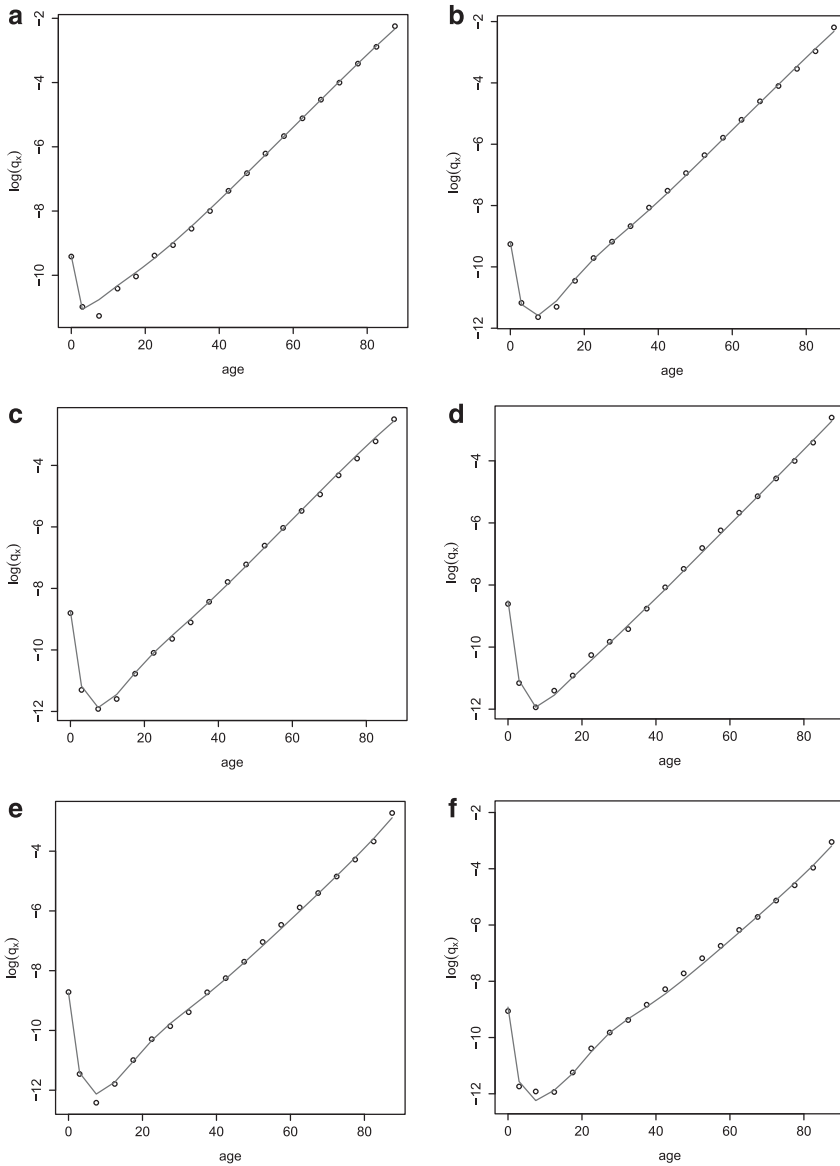


Figure 2. Sample years of the log-mortality rates and the fitted VECM, Circulatory system, U.S.A., females. (a) 1955; (b) 1965; (c) 1975; (d) 1985; (e) 1995; (f) 2005.

Note: The dots represent the observed mortality rates, while the fitted model is depicted by the curve.

small sample size.²³ As shown by the p -values, the null hypothesis is not rejected at a 5 per cent significance level.

²³ As in L  tkepohl (2005).

Table 3 Tests on residuals of the fitted VECM, 1950–2005, Circulatory system, U.S. females

<i>Type of test</i>	<i>Name of the test</i>	<i>Statistic value</i>	<i>p-value</i>
Autocorrelation	Portmanteau (15 lags)	776.68	0.07
	Portmanteau (25 lags)	1224.33	0.39
Normality	Skewness	8.39	0.30
	Kurtosis	8.05	0.33
	Both	16.44	0.29

Notes: 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.

Tests for normality are based on the third and fourth central moments (skewness and kurtosis) of a normal distribution.²⁴ The test statistic labelled *both* in Table 3 is a joint test of skewness and kurtosis. The three tests clearly show that the null hypothesis of normality is not rejected. The fitted model residuals for the parameters of the Heligman-Pollard model satisfy the assumption of normally distributed and no-autocorrelation. The model is a good representation of the data generating process and thus captures the trends and age dependence of mortality rates for deaths due to the circulatory system in the United States.

Projections

The fitted VECM provides a good fit to the data with 33 parameters for 19 age groups over 56 years based on 1,064 observations. Parameters are used to model stochastic trends and to ensure that long-run relationships between different ages are maintained through cointegration. These are very important features of the model approach.

To illustrate the model forecasting performance, the model is fitted to a shorter data period and projections are performed in order to compare the forecasts with the actual data. At least 50 years of observations are usually needed in order to reliably estimate a VECM. The model is fitted to data over the period 1950–2000 and then forecasted for the following 20 years. Parameter *B* is assumed fixed at a value of one and parameter *F* is assumed fixed at the median value over the period 1950–2000 of 25.42. Since actual data is only available until 2005, the first five years of forecasted mortality rates are compared with actual mortality.

To assess the performance of the VECM approach, two other models are fitted over the period 1950–2000 and used to forecast mortality rates until 2005. The first is a variant of the well-known Lee-Carter model, which has been successfully applied at the population level for United States data in the past. The second model fits the Heligman-Pollard model for mortality rates and uses univariate time series, traditional ARIMA processes, to model the parameters.

²⁴ For a detailed description of these tests, see Lütkepohl (2005).

VECM

The VECM fitting process is applied for the time period 1950–2000. AIC, HQ, SC and FPE tests confirm a lag order of one. ADF and PP tests reject the null hypothesis of non-stationarity at a 1 per cent significance level for parameters *C*, *D* and *E*, while KPSS test does not reject the null hypothesis of trend stationarity at a 10 per cent significance level only for parameter *C* and at a 1 per cent significance level for parameter *E*. Parameters *C*, *D* and *E* are assumed trend stationary based on support for this from at least two out of the three tests.

Two cointegrated relations are estimated for the remaining four parameters, as shown in Table 4. These tests determine the number of long-run equilibrium relationships among the parameters of the Heligman-Pollard model (*A*, *G*, *H* and *K*).

The estimated VECM is

$$\begin{bmatrix} \nabla A_t \\ \nabla C_t \\ \nabla D_t \\ \nabla E_t \\ \nabla G_t \\ \nabla H_t \\ \nabla K_t \end{bmatrix} = \begin{bmatrix} 2.13 \cdot 10^{-3} \\ -4.25 \\ 4.79 \cdot 10^{-4} \\ -6.60 \cdot 10^2 \\ -1.59 \cdot 10^{-4} \\ 0.79 \\ -3.65 \cdot 10^2 \end{bmatrix} + \begin{bmatrix} -2.07 \cdot 10^{-6} & 1.61 \cdot 10^{-6} \\ -2.72 \cdot 10^{-3} & -1.22 \cdot 10^{-3} \\ 2.74 \cdot 10^{-7} & 1.46 \cdot 10^{-7} \\ -5.88 \cdot 10^{-2} & -0.30 \\ -2.07 \cdot 10^{-7} & -1.44 \cdot 10^{-8} \\ 9.11 \cdot 10^{-4} & 1.05 \cdot 10^{-4} \\ 0.24 & -0.24 \end{bmatrix} \times \begin{bmatrix} A_{t-1} \\ G_{t-1} \\ H_{t-1} \\ K_{t-1} \end{bmatrix}.$$

Table 4 Tests for the number of cointegrated relations, 1950–2000, Circulatory system, U.S. females (a) Trace test; (b) Maximum-eigenvalue test

<i>r</i>	Trace stat	Critical values			
		10%	5%	2.50%	1%
<i>(a)</i>					
3	4.27	2.69	3.76	4.95	6.65
2	12.71	13.33	15.41	17.52	20.04
1	40.04	26.79	29.68	32.56	35.65
0	88.70	43.95	47.21	50.35	54.46
<i>r</i>	Eigen stat	Critical values			
		10%	5%	2.50%	1%
<i>(b)</i>					
3	4.27	2.69	3.76	4.95	6.65
2	8.44	12.07	14.07	16.05	18.63
1	27.34	18.6	20.97	23.09	25.52
0	48.66	24.73	27.07	28.98	32.24

Notes: A null hypothesis is accepted at a $\alpha\%$ significance level when the statistic is lower than the corresponding critical value. Thus, these tables indicate that two cointegrated relations are accepted at a 2.5% significance level. These two tests assess the number of long-run equilibrium relationships among the parameters of the Heligman-Pollard model (*A*, *G*, *H* and *K*) that is among age-based risk factors.

Table 5 Tests on residuals of the fitted VECM, 1950–2000, Circulatory system, U.S. females

Type of test	Name of the test	Statistic value	p-value
Autocorrelation	Portmanteau (15 lags)	872.36	8.7×10^{-5}
	Portmanteau (25 lags)	1451.54	2.01×10^{-6}
Normality	Skewness	12.19	0.09
	Kurtosis	11.02	0.14
	Both	23.21	0.06

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

The cointegrating relationships are

$$\begin{aligned}
 -8.12 \cdot 10^3 A_t + 3.61 \cdot 10^5 G_t + 5.50 \cdot 10^2 H_t - 0.65 K_t &= z_{1t}, \\
 3.44 \cdot 10^4 A_t + 5.44 \cdot 10^6 G_t + 1.86 \cdot 10^3 H_t - 2.27 K_t &= z_{2t}.
 \end{aligned}$$

These are similar to the cointegrated relations for the VECM fitted over the full period 1950–2005 with parameters G and H having coefficients with similar signs, while the coefficient of K is of opposite sign. The middle age parameters D and E are no longer included.

Normality tests as well as tests on the autocorrelations among the residuals are shown in Table 5. The normality of the residuals is not rejected at a 5 per cent significance level. However, the Portmanteau test indicates some autocorrelation remains in the residuals. A VECM with higher order, such as two or three lags, may then be better. Since the longer period model fit indicated a VECM with higher order was not required and since a model with as few parameters as possible is expected to have a better forecasting performance, a lag of one is used for the shorter period as well.

Figure 3 displays graphs of the parameters for the fitted model and the projected values. Figure 4 shows the observed mortality rates in 2000 (dots) along with forecasts for the VECM.

Lee-Carter model

In order to assess the performance of the VECM a variant of the Lee-Carter model,¹ which has become a standard in mortality modelling, is used for comparison. The model decomposes the logarithm of the force of mortality into two components, one capturing the age pattern of average mortality rates and the other a common time trend with differential impacts by age. The model is written as

$$\ln \mu_{x,t} = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t},$$

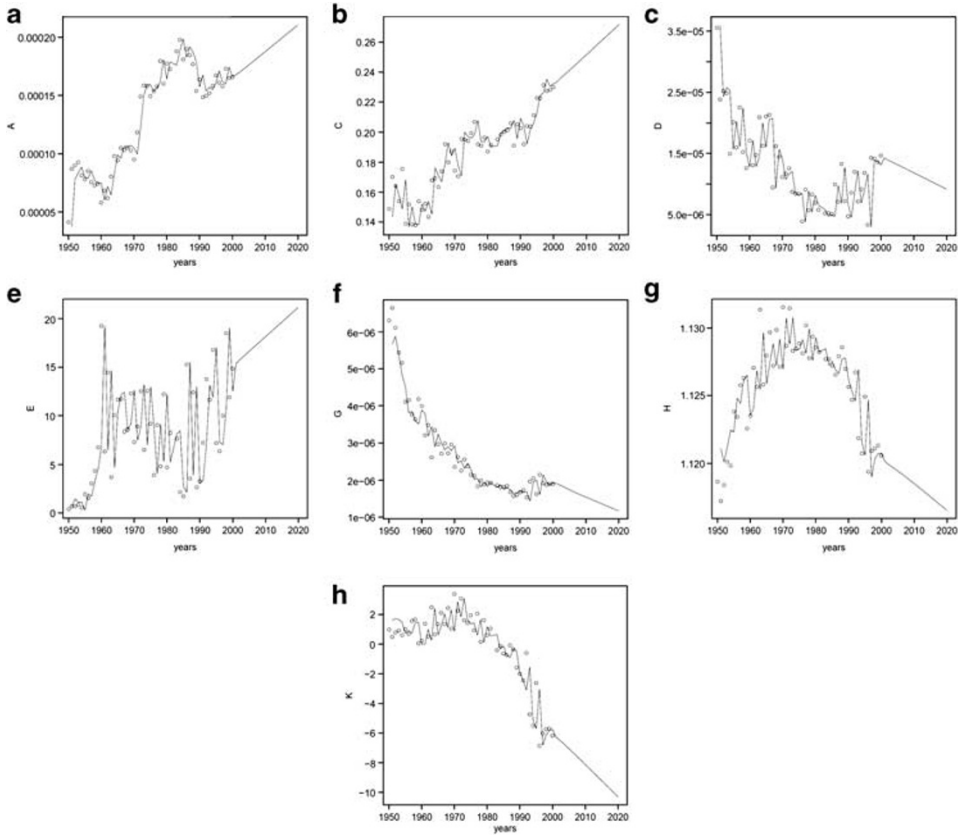


Figure 3. Parameter estimates and forecasts of the Heligman-Pollard model, Circulatory System, U.S. females. (a) Parameter *A*; (b) Parameter *C*; (c) Parameter *D*; (e) Parameter *E*; (f) Parameter *G*; (g) Parameter *H*; (h) Parameter *K*.

Notes: The Heligman-Pollard model is fitted to mortality rates over the period 1950–2000. The dots in the graphs are the parameters based on the data. The curves are the VECM fitted on the parameters along with the VECM forecasted values (from 2001 to 2020).

with

- $\mu_{x,t}$ force of mortality at age x , and year t ;
- α_x mean value over time, at age x , of the logarithm of the force of mortality;
- β_x deviation from the mean value α_x at age x .
Reflects the impact of the time trend represented by κ_t on age.
The higher its absolute value, the larger the changes in mortality at that age;
- κ_t mortality rates trend over time;
- $\varepsilon_{x,t}$ random changes including those not captured by the model;
= errors assumed to have mean zero and variance σ^2 (homoscedasticity).

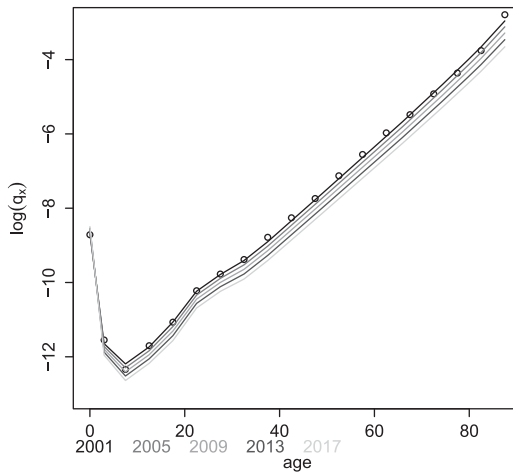


Figure 4. Forecasted mortality rates over the period 2001–2020, Circulatory System, U.S. females.
Notes: The actual mortality rates of 2000 are represented by the dots and the forecasts—using the Heligman-Pollard model and forecasting the parameters with VECM—by the curves.

To ensure this interpretation of the parameter α_x , the sum over the estimator $\hat{\kappa}_t$ is set equal to zero,

$$\sum_{t=t_{\min}}^{t_{\max}} \hat{\kappa}_t = 0.$$

Since if N is the number of years of observations,

$$\begin{aligned} \prod_{t=t_{\min}}^{t_{\max}} \mu_{x,t} &= \exp\left(\hat{\alpha}_x N + \hat{\beta}_x \sum_{t=t_{\min}}^{t_{\max}} \hat{\kappa}_t\right) \\ &= \exp(\hat{\alpha}_x N), \end{aligned}$$

which leads to

$$\hat{\alpha}_x = \frac{1}{N} \sum_{t=t_{\min}}^{t_{\max}} \ln \mu_{x,t}, \quad \forall x.$$

For an identifiable model, another constraint on the parameters is specified, which usually is

$$\sum_{x=x_{\min}}^{x_{\max}} \hat{\beta}_x = 1.$$

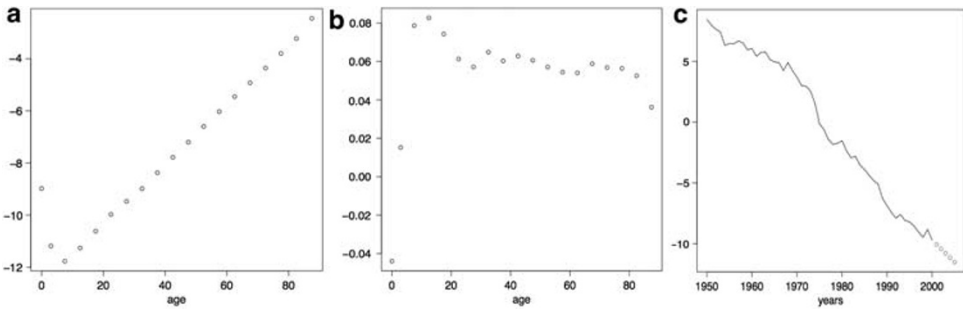


Figure 5. Parameters of the Lee-Carter model, Circulatory System, U.S. females. (a) Parameter α ; (b) Parameter β ; (c) Parameter κ .

Notes: The parameters are estimated by maximum likelihood. Parameter κ is forecasted from 2001 to 2005 (dots) using an ARIMA(0, 1, 0) model with a drift of -0.36 .

The parameters are estimated numerically using maximum likelihood estimation, assuming that the number of deaths at age x follows a Poisson distribution with mean $l_{x,t} \cdot \mu_{x,t}$, $l_{x,t}$ being the population of age x alive at the beginning of year t .²⁵ The parameter estimates are shown in Figure 5.

In the Lee-Carter model, a single common factor across ages determines the general level of mortality improvement over time. This improvement is modelled with a single time series (κ_t) that is projected in order to forecast the complete age profile of mortality. Lee and Carter suggest the use of a simple random walk with drift, that is an ARIMA(0, 1, 0) model. Our analysis confirms an ARIMA(0, 1, 0) process with a drift of -0.36 as the best model.²⁶

ARIMA processes

As another alternative for comparison, the parameters of the Heligman-Pollard model fitted over the period 1950–2000 are modelled with ARIMA processes, instead of a VECM. We follow the approach applied in previous works,²⁷ that is trends are removed from the variables by first or second differencing. According to ADF, PP and KPSS tests, first differencing on each parameter is necessary and sufficient to assure stationarity. ARIMA($k, 1, (k-1)$) models are successively fitted for each parameter of the Heligman-Pollard model, increasing k by one, as suggested by Pandit and Wu.²⁶ The F -criterion is used to decide which model is the most adequate between an ARIMA($k, 1, (k-1)$) and an ARIMA($(k+1), 1, k$). The F -criterion is a test to check for improvement in the residual sum of squares of the error ε_t . It tests the assumption that some of the coefficients in a model are restricted to zero.²⁶

²⁵ Delwarde and Denuit (2006).

²⁶ Pandit and Wu (2001).

²⁷ Bell (1997); McNown and Rogers (1989); McNown and Rogers (1992).

Table 6 Fitted ARIMA($p, 1, q$) processes, Circulatory system, U.S. females

	A: ARIMA (2, 1, 2)			C: ARIMA (2, 1, 1)			D: ARIMA (2, 1, 0)			E: ARIMA (3, 1, 2)		
	Values	CI 95%		Values	CI 95%		Values	CI 95%		Values	CI 95%	
ϕ_1	-0.66	-1.01	-0.31	-1.36	-1.58	-1.13	-0.72	-1.00	-0.44	0.37	0.16	0.59
ϕ_2	-0.55	-0.82	-0.27	-0.58	-0.81	-0.35	-0.36	-0.63	-0.08	-0.24	-0.46	-0.02
ϕ_3										-0.65	-0.86	-0.43
θ_1	0.76	0.57	0.96	1.00	0.87	1.13				-1.27	-1.42	-1.12
θ_2	1.00	0.80	1.20							1.00	0.81	1.19
$\sum \hat{\epsilon}_t^2$	$6.58 \cdot 10^{-9}$			$4.15 \cdot 10^{-3}$			$7.56 \cdot 10^{-10}$			713.81		
	G: ARIMA (2, 1, 1)			H: ARIMA (2, 1, 1)			K: ARIMA (2, 1, 0)					
	Values	CI 95%		Values	CI 95%		Values	CI 95%				
ϕ_1	0.49	0.21	0.77	-1.48	-1.69	-1.28	-0.71	-0.98	-0.45			
ϕ_2	0.48	0.22	0.75	-0.64	-0.85	-0.43	-0.29	-0.55	-0.03			
θ_1	-0.84	-1.04	-0.65	1.00	0.88	1.12						
$\sum \hat{\epsilon}_t^2$	$5.00 \cdot 10^{-12}$			$1.65 \cdot 10^{-4}$			56.71					

Notes: All models are identified and estimated using the 50 observations from 1950 to 2000. First differencing transformation is performed on every parameter.

Once the appropriate ARIMA($k, 1, (k-1)$) process is found, the confidence interval of the coefficients is checked to ensure parameters are significantly different from zero. If some intervals include zero, the *F*-criterion is applied to determine the adequacy of a model without the corresponding coefficients. To check the adequacy of the model found using the procedure described by Pandit and Wu, ARIMA($p, 1, q$) processes for all combinations with $p \leq 3$ and $q \leq 2$ are fitted and compared through the *F*-criterion, resulting in similar choice of the ARIMA model. Finally, residuals of the fitted model are checked to ensure they are uncorrelated. Table 6 shows the best fitting ARIMA models resulting from this procedure. Each model has no significant residual autocorrelation according to the Portmanteau test.

Model comparison

Comparison of the forecasts for the alternative models can be assessed from the plots given in Figures 6–8. Figure 6 shows the forecasted mortality rates from the fitted VECM (curve) compared with the actual data (dots). The forecasts from the Lee-Carter model are shown in Figure 7. Figure 8 shows the forecasted age profiles of mortality from the fitted ARIMA models.

These figures show that the Lee-Carter model tends to produce relatively poor forecasts by over-estimating the decline of mortality at young ages. The Heligman-Pollard model along with ARIMA processes underestimates mortality improvement, particularly at older ages. The VECM provides better forecasts and does not exhibit either of these features.

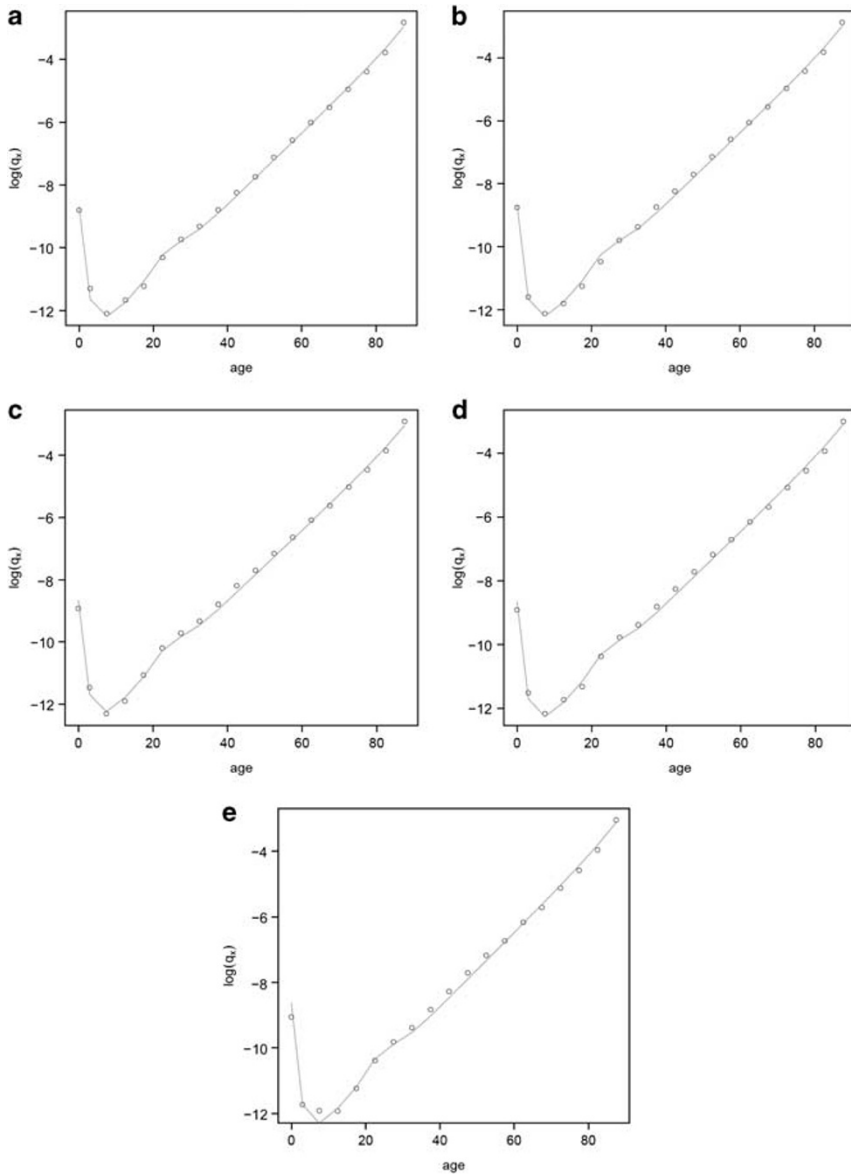


Figure 6. Sample years of the log-mortality rates and the forecasted values according to VECM, Circulatory system, U.S. females. (a) 2001; (b) 2002; (c) 2003; (d) 2004; (e) 2005.

Notes: The dots represent the observed mortality rates, while the forecasted values—using the Heligman-Pollard model and forecasting the parameters with VECM—are depicted by the curve.

The forecasting performance of the three methods is also compared by assessing the match between the actual and the forecasted mortality rates from 2001 to 2005 using a mean absolute percentage error statistic (MAPE), the average of the absolute

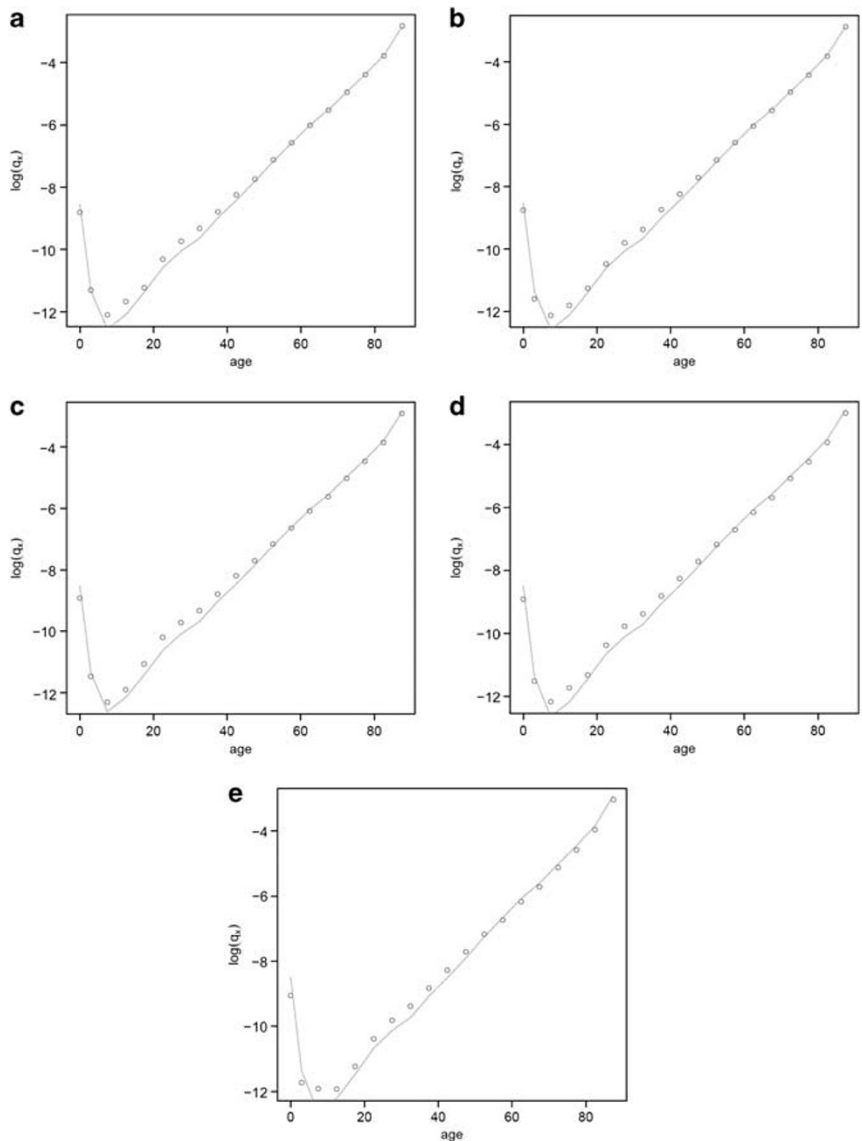


Figure 7. Sample years of the log-mortality rates and the forecasted values according to the Lee-Carter model, Circulatory system, U.S. females. (a) 2001; (b) 2002; (c) 2003; (d) 2004; (e) 2005.

Notes: The dots represent the observed mortality rates, while the forecasted values from the Lee-Carter model are depicted by the curve.

percentage errors between the forecasted and the observed mortality rates computed over the 19 age groups. The projections are also compared with a no-change forecast. The no-change forecast takes the mortality of 2000 as the forecasted mortality rates of the following five years. The ratio of the square root of two mean square errors (MSE)

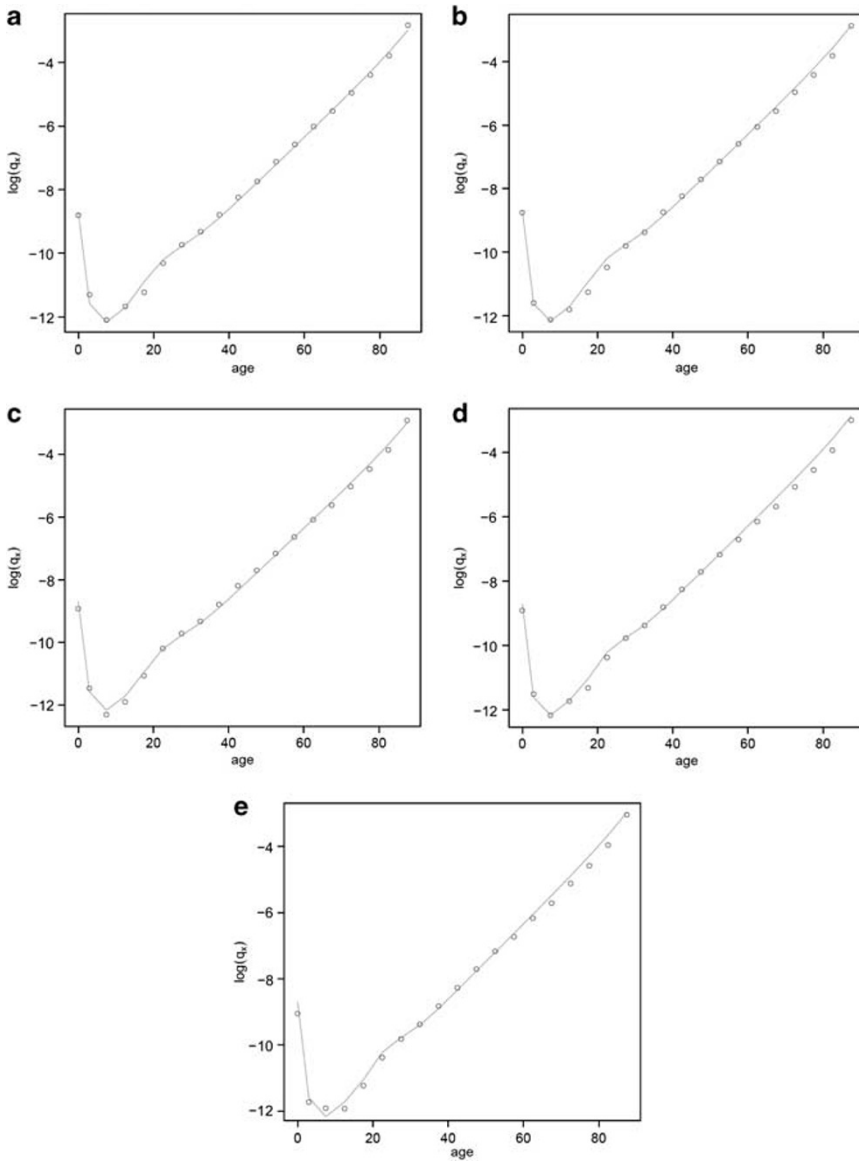


Figure 8. Sample years of the log-mortality rates and the forecasted values according to ARIMA models, Circulatory system, U.S. females. (a) 2001; (b) 2002; (c) 2003; (d) 2004; (e) 2005.

Notes: The dots represent the observed mortality rates, while the forecasted values—using the Heligman-Pollard function and forecasting the parameters with ARIMA processes—are depicted by the curve.

computed over the 19 age groups is used. The first MSE is computed between the forecasted and the observed mortality rates. The second MSE is between the no-change forecast and the observed mortality rates. A statistic with a value smaller than unity indicates that the model forecasts are more accurate than

Table 7 Mean absolute percentage error and ratio of mean square error compared to no-change forecast, Circulatory system, U.S. females

Year	VECM		Lee-Carter		ARIMA	
	MAPE (%)	vs no-change forecast	MAPE (%)	vs no-change forecast	MAPE (%)	vs no-change forecast
2001	9.56	3.21	13.04	0.44	9.90	3.51
2002	9.33	1.27	14.15	0.30	11.08	1.25
2003	11.71	0.87	17.85	0.40	10.71	0.78
2004	11.84	0.43	19.33	0.55	15.37	0.97
2005	13.46	0.33	24.03	0.57	17.06	0.59

Notes: MAPE is the mean absolute percentage error. The *vs no-change forecast* is the ratio of the square root of the mean square error of one of the three considered models with the square root of the mean square error of the no-change forecast. The no-change forecast takes the mortality of 2000 as fixed and as being the forecasted mortality rates of the following five years.

All summary statistics are averages over the 19 age groups and describe the situation at a single point in time.

the base assumption of no change in mortality. Table 7 shows the model comparison results.

The forecasts from the VECM are closer to the actual mortality rates than the forecasts of the Lee-Carter model or the forecasts performed with the Heligman-Pollard model and ARIMA processes. The forecast error (MAPE) with the VECM is lower and increases more slowly with time than the other two models. This is expected to apply for projections over longer time horizons. The VECM performance comes from its ability to capture relationships between the ages better than these alternative models.

The main benefit of using the Heligman-Pollard model over the Lee-Carter model is that it smooths mortality rates across ages and reduces the number of parameters for the age structure mortality curve. The Heligman-Pollard model has sufficient flexibility to allow for the full age range of mortality as well as having parameters that are interpreted for their effect on mortality rates. Treating these parameters as stochastic factors allows for more general dynamics than in the Lee-Carter model, even though there is an increase in the number of parameters for the random factors in the model.

The Lee-Carter model appears to be more accurate at older ages during the early forecasted years than a stochastic parameter model based on the Heligman-Pollard model. As mortality rates at older ages are higher than rates at young and middle ages, the absolute error is higher at older ages. The statistic that compares each model with the no-change forecast uses the squared error. In Table 7 the statistic comparing the no-change forecast for the Lee-Carter model is then lower for the first three years than the statistic of the other two models. For the later years this is reversed with the VECM Heligman-Pollard model having lower comparative forecast error. For longer forecast horizons, the age pattern given by the Heligman-Pollard model produces better fits and relationships between ages are maintained by the VECM taking into account long-run stochastic trends for the parameters of the model. The VECM

Heligman-Pollard model is expected to be more accurate for longer forecast periods than five years because of this.

Conclusion

This paper presents a modelling and forecasting approach for mortality using VECM concepts for a parametric model for the age structure of mortality rates. Because of its ability to capture changing age structures and to model long-run trends, it has the flexibility to model different causes of death as well as aggregate mortality rates. To demonstrate the application of the model, it is fitted to cause-of-death mortality rates for the circulatory system in the United States over the period 1950–2000 and compared with alternative models previously proposed.

The model is a stochastic parameter Heligman-Pollard model, where the time series of the parameters are modelled with a VECM. The VECM incorporates common stochastic trends through long-run equilibrium relationships. The model captures the age dependence of cause-specific mortality rates through dependence between the parameters. Indeed, the parameters of the Heligman-Pollard model have interpretation as factors impacting specific age ranges—young, middle and older. The demographic meaning of the parameters allows for an analysis of the long-run equilibriums or cointegration relations in the model. The expected relationship between the parameters at older ages is shown to be reflected in these cointegration relations.

The analysis demonstrates for circulatory diseases mortality that a forecasting model based on the Heligman-Pollard model with a VECM for the parameters is an improvement over the well-known Lee-Carter model as well as over a model based on the Heligman-Pollard model with ARIMA processes for the parameters. As the forecast horizon is lengthened, the model based on the Heligman-Pollard model, with a VECM for the parameters, increases in forecasting accuracy. The Heligman-Pollard model uses a smaller number of parameters to model the age structure of mortality rates compared with the Lee-Carter model. The stochastic model for the parameters provides an adequate representation of the randomness of the process and the use of a VECM ensures consistent future mortality curves are forecasted, which is not the case if simpler ARIMA time series models are used. A stochastic parameter VECM form of the Heligman-Pollard model has also been shown to perform well at a national population level based on Australian mortality data.²⁸ Finally, new theories on time series models and applications to mortality are being developed.²⁹ Standard unit root tests are potentially unreliable in the presence of bounded time series, as they tend to over-reject the null hypothesis of a unit root, even asymptotically. Conventional cointegration tests may also be misleading in the presence of bounded time series. These new theories have the potential to improve the forecasting performance of the Heligman-Pollard model with a VECM for the parameters.

²⁸ Njenga and Sherris (2011).

²⁹ Cavaliere (2005).

Acknowledgements

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

- Barugola, T. and Maccheroni, C. (2007) '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*, pp. 481–482.
- Bell, W.R. (1997) 'Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates', *Journal of Official Statistics* 13(3): 279–303.
- Cavaliere, G. (2005) 'Limited time series with a unit root', *Econometric Theory* 21: 907–945.
- Delwarde, A. and Denuit, M. (2006) '*Construction de Tables de Mortalité Périodiques et Prospectives*', Paris, France: Economica.
- Gaille, S. (2010) *Improving longevity and mortality risk models*, PhD thesis, University of Lausanne.
- Gaille, S. and Sherris, M. (2010) *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.
- Gutterman, S. and Vanderhoof, I.T. (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.
- Hanewald, K. (2011) 'Explaining mortality dynamics: The role of macroeconomic fluctuations and cause of death trends', *North American Actuarial Journal* 15(2): 290–314.
- Heligman, L. and Pollard, J.H. (1980) 'The age pattern of mortality', *Journal of the Institute of Actuaries* 107(434): 49–80.
- Lee, R.D. and Carter, L. (1992) 'Modeling and forecasting US mortality', *Journal of the American Statistical Association* 87(419): 659–671.
- Lütkepohl, H. (2005) *New Introduction to Multiple Time Series Analysis*, Germany: Springer.
- Manton, K.G., Patrick, C.H. and Stallard, E. (1980) 'Mortality model based on delays in progression of chronic diseases: Alternative to cause elimination model', *Public Health Reports* 95(6): 580–588.
- McNown, R. and Rogers, A. (1989) 'Forecasting mortality: A parameterized time series approach', *Demography* 26(4): 645–660.
- McNown, R. and Rogers, A. (1992) 'Forecasting cause-specific mortality using time series methods', *International Journal of Forecasting* 8(3): 413–432.
- Njenga, C.N. and Sherris, M. (2009) *Longevity risk and the econometric analysis of mortality trends and volatility*, Australian School of Business Research Paper No 2009ACTL08, University of New South Wales, Sydney.
- Njenga, C.N. and Sherris, M. (2011) *Modeling mortality with a Bayesian vector autoregression*, Australian School of Business Research Paper No 2011ACTL04, University of New South Wales, Sydney.
- Olshansky, J.S. (1987) 'Simultaneous/multiple cause-delay (SIMCAD): An epidemiological approach to projecting mortality', *Journal of Gerontology* 42(4): 358–365.
- Pandit, S.M. and Wu, S.-M. (2001) *Time Series and System Analysis with Applications*, Malabar: Krieger, Florida.
- Tabeau, E., Ekamper, P., Huisman, C. and Bosch, A. (1999) 'Improving overall mortality forecasts by analysing cause-of-death, period and cohort effects in trends', *European Journal of Population* 15(2): 153–183.
- Tabeau, E., Van Den Bergh Jeths, A. and Heathcote, C. (2001) *Forecasting Mortality in Developed Countries. Insights from a Statistical, Demographic and Epidemiological Perspective*, Dordrecht: Kluwer Academic Publishers.
- Wilmoth, J.R. (1993) *Computational methods for fitting and extrapolating the Lee*, Technical report, National Institute of Aging.
- World Health Organization (2009) 'Who mortality database', January 2009, www.who.int/whosis/mort/download/en/index.html.

About the Authors

S verine Gaille is an Assistant Professor in Actuarial Science at the University of Lausanne, Switzerland. Besides a PhD in Actuarial Science, she has a Certificate in Population Study from the University of Geneva, Switzerland. Her research focuses on longevity risk and mortality modelling, looking for more efficient models in mortality projections. Along with her work at the university, she has worked for the International Labour Organization as well as for some consulting pension fund firms.

Michael Sherris is Professor of Actuarial Studies at the Australian School of Business, UNSW, Sydney, Australia. He was appointed as Foundation Professor of Actuarial Studies at UNSW in 1998. He was 2009 President of the Asia Pacific Risk and Insurance Association. He is a Fellow of the Institute of Actuaries of Australia, the Institute of Actuaries (U.K.) and the Society of Actuaries (North America). His current research interests focus on insurer risk-based capital, and longevity risk modelling and risk management.