Modelling trends in cohort survival probabilities

P. Hatzopoulos[a], S. Haberman[b]

[a] Department of Mathematics, Division of Statistics and Actuarial-Financial Mathematics, University of the Aegean, Samos, 83200, Greece
[b] Faculty of Actuarial Science and Insurance, Cass Business School, City University London, 106 Bunhill Row, London EC1Y8TZ, UK

Abstract

A new dynamic parametric model is proposed for analyzing the cohort survival function. A one-factor parameterized polynomial in age effects, complementary log-log link and multinomial cohort responses are utilized, within the generalized linear models (GLM) framework. Sparse Principal component analysis (SPCA) is then applied to cohort dependent parameter estimates and provides (marginal) estimates for a two-factor structure. Modeling the two-factor residuals in a similar way, in age-time effects, provides estimates for the three-factor age-cohort-period model. An application is presented for Sweden, Norway, England & Wales and Denmark mortality experience.

Keywords: Cohort mortality; Multinomial responses; Generalized Linear Models; Mortality forecasting; Sparse Principal Component analysis; Dynamic Linear Regression;

1. Introduction

In statistics, *cohort studies* typically refer to a group of individuals with a specific shared experience within a certain time period. A birth cohort is the most common example, i.e. a group of people who are born during a particular period (e.g., year, decade).

Cohorts and periods are two different ways to analyze demographic events. Cohort life expectancy at birth is the observed average age at death for the group of people who are born during a particular period. In contrast, period life expectancy at birth is the expected average age at death for a *synthetic cohort*, i.e. an imaginary group of people who experience, hypothetically, the demographic conditions of that period throughout life. As mentioned by Benjamin and Pollard (1993), "the period life table based on the deaths over a limited period, is not likely to be reproduced in the future; it mixes the experience of different generations—the lives who contribute to the death rates at advanced ages were born many decades before those who contribute to the rates at young ages".

In order to identify the determinants of mortality trends, it is important to distinguish between period and cohort effects. Period effects indicate immediate effects on mortality arising from contemporaneous events (e.g. wars, accidents). In contrast, cohort effects may reflect long-lasting effects located earlier in the life course (e.g. poor nutrition in infancy, lifetime health behaviors like smoking, poor living conditions), the effects of which the generation carries throughout life. With a cohort effect, the length of time elapsed since some earlier event may be important – for example, time since the cessation of smoking. This distinction between a period effect and a cohort effect is of crucial importance in making projections of mortality. If cohort effects prove to be important, then cohort-based projections may be useful in exploring possible future mortality trends.

A select cohort is a birth cohort characterized by greater rates of mortality improvement than previous and following generations, i.e. *high relative mortality improvement*. A recognized example is in the U.K. population where there are clear signs of birth cohort effects. The generations born between 1925 and 1945 have experienced more rapid improvement than earlier and later generations. Richards (2008) discusses those cohorts born in Britain in the 1930s, often referred to as the "golden cohorts", because they have shown high relative rates of mortality improvement for many years. Possible explanations for this golden cohort include a healthy diet in the 1940s and early 1950s (during a period of national rationing of food) and the introduction of the National Health Service in 1948.

In addition, more insight may be obtained by investigating also cohorts who appear to have particularly *poor* rather than good patterns. For example, Murphy (2009) comments that "although people born in the 1950s in Britain were the first beneficiaries of the welfare state, including the National Health Service, and were raised in an economic boom period with unprecedented family life stability, these postwar cohorts have shown only *low relative mortality improvement* or even mortality increases in some cases".

The debate about the relative importance of period and cohort effects on mortality has been long-standing. Derrick (1927), presented a generation influence for England and Wales in the period 1841–1925 by year of birth. Kermack, McKendrick, and McKinlay (1934) presented a table with hand-drawn contours of age-specific mortality rates for England and Wales for the period 1841–50. They showed a straightforward and easily interpreted cohort effect that alters the relative risk of mortality by the same proportion at most ages. Finch and Crimmins (2004) characterized these results as among studies that clearly documented the existence of cohort effects. As discussed by Murphy (2009), "although the analyses of Derrick and of Kermack and his colleagues are now extensively cited by those who view cohort effects as dominant, Derrick and others who held similar views actually lost the argument at the time, and for the next half-century attention was concentrated almost entirely on period effects. With the resurgence of interest in cohort patterns noted by Davey Smith and Lynch (2004), however, interest in this topic has been renewed, not only for interpreting current mortality but also as a tool for forecasting future mortality".

Barker et al. (1989) argue that "there are plausible causal mechanisms and empirical evidence that suggest that events around the time of birth (both pre-natal and post-natal) lead to different chances of cardiovascular disease starting at late middle-ages". Finch & Crimmins (2004) argue that "inflammation in early childhood may influence mortality at even higher ages", and discuss the idea of a `cohort morbidity phenotype'. Under this hypothesis, "early-life exposure to infectious agents causes an inammatory reaction which can lead to chronic disease in middle age". According to Janssen & Kunst (2005), "mortality from infectious diseases at older ages is related to the disease load experienced during the first year of life, especially exposure to airborne infectious diseases. Infant mortality often used as a proxy of living conditions in infancy and of the disease load during the birth year. The fact that infant mortality started to decline for about the same cohorts as adult mortality suggests that improvements in living conditions have left an imprint on the mortality experience of cohorts up to old age". Their results indicate that factors occurring in adulthood (predominantly due to smoking) and in infancy or childhood (such as poor living conditions) have left an imprint on the mortality experience of birth cohorts up to old age and conclude that "cohort effects appear to be important in determining secular trends in all-cause and cause-specific mortality among the elderly". A similar argument is proposed by Richards et. al. (2005). Beltra´n-Sa´nchez (2012) has developed a multilevel model that accounts for cohort and period effects in later-life mortality. They show that "early-life mortality, which is linked to exposure to infection and poor nutrition, predicts both the estimated cohort mortality level at age 40 and the subsequent Gompertz rate of mortality acceleration during aging. These findings apply to cohorts born in both the 19th century and the early 20th century". Also, Meza et. al. (2010) comment that "Dramatic improvements in life expectancy during the 20th century are commonly attributed to improvements in either health care services or the social and economic environment". They evaluate the hypothesis that improving infant survival produces improvements in adult mortality rates, and argue that "year of birth correlates better with adult mortality than year of death". The analysis of the relationship between month of birth and survival to age 100 carried out by Gavrilova and Gavrilov (2011a) supports "the idea of early-life programming of human aging and longevity".

In contrast, Richards et. al. (2005) state that "mortality is also strongly influenced by lifetime health behaviors. A great deal is now known about mechanisms linking various health behaviors to mortality outcomes, especially smoking and diet. Where there are pronounced and consistent generational differences in such life-course health behaviors, these will also be picked up in a year-of-birth analysis. An example of this is smoking and rates of lung-cancer mortality. Year of birth is therefore a potentially powerful combined proxy for fatal and gestational conditions, early-life disease load and lifetime patterns of health behaviors". Also, Charlton & Murphy (1997) argue that "other causal factors occurring later in life, such as the well-established effects of smoking on mortality, especially lung cancer, are largely manifested at older ages".

In many developed countries there is evidence of a cohort effect. Many authors in recent studies include a

cohort effect in their mortality modelling and analysis. Renshaw and Haberman (2006) incorporate this effect by developing an age-period-cohort version of the Lee-Carter (LC) model which provides an improved fit to the data than the basic LC model. Similarly, Yue et al (2008) apply the Bell (1997) approach to several countries, and propose a jump model to the mortality rates, including two second-order interaction terms (age-period and age-cohort). They suggest that the reduction shift of ages for different time periods can be treated as a "cohort" effect, introducing an age-period-cohort (APC) model. Alai & Sherris (2011), using national population mortality data from Norway and Australia, discuss the relative significance of cohort effects as compared to period effects. Following Cairns et. al. (2008), a key criterion, in order to assess whether any stochastic mortality model is a good model or not, is that the model should incorporate a stochastic cohort effect.

The remainder of the paper is organised as follows. In section 2, we analyse the methodology proposed for analyzing the cohort mortality survival function and the extraction of the age-cohort related terms. In section 3, we explore the procedure proposed for forecast purposes. In section 4, we illustrate the method proposed to implement the age-time related residual components, and finally, in section 5, we provide some concluding remarks. In order to illustrate the methodology, we present a study based on Norway, England & Wales, Sweden and Denmark, for both sexes, total population mortality experience, for calendar years 1846-2008 and individual ages 1,…,100. We choose these countries to illustrate the methodology due to their long history of high quality mortality data. The data are freely provided by the "Human Mortality Database" (www.mortality.org).

2. Model set up

According to Wilmoth et. al. (2007), death rates and probabilities of dying are simpler conceptually for cohorts than for periods. These are found using death counts in Lexis triangles and population estimates by single years of age and time (Figure 1). Cohort rates and probabilities are measured over the parallelogram that follows the lives of individuals who turn age x in one calendar year until their next birthday, at age x +1, in the following calendar year. We assume that there is no migration and deal with the case of a closed population, since the effects of migration on calculated rates and probabilities are negligible so long as migratory flows have the same direction and a similar magnitude over the interval (Wilmoth et. al., 2007).

For cohorts, the (conditional) probability of survival, p_x, between ages x and x +1, given that is alive age x, is found by the product of two components:

$$p_x = \frac{P}{P + D_L} \cdot \frac{P - D_U}{P} = \frac{P - D_U}{P + D_L}$$

where $P=P(x, t+1)$ is the size of the cohort in between the two calendar years (i.e., on January 1st of the latter year), $D_L = D_L(x,t)$ refers to lower triangle deaths for the year the cohort turns age x, and $D_U = D_U(x, t+1)$ refers to upper triangle deaths in the following year. Thus, the probability of surviving from age x to age x +1 is a product of the fraction surviving from age x to the end of the calendar year and the fraction surviving from the beginning of the next calendar year to age x +1.

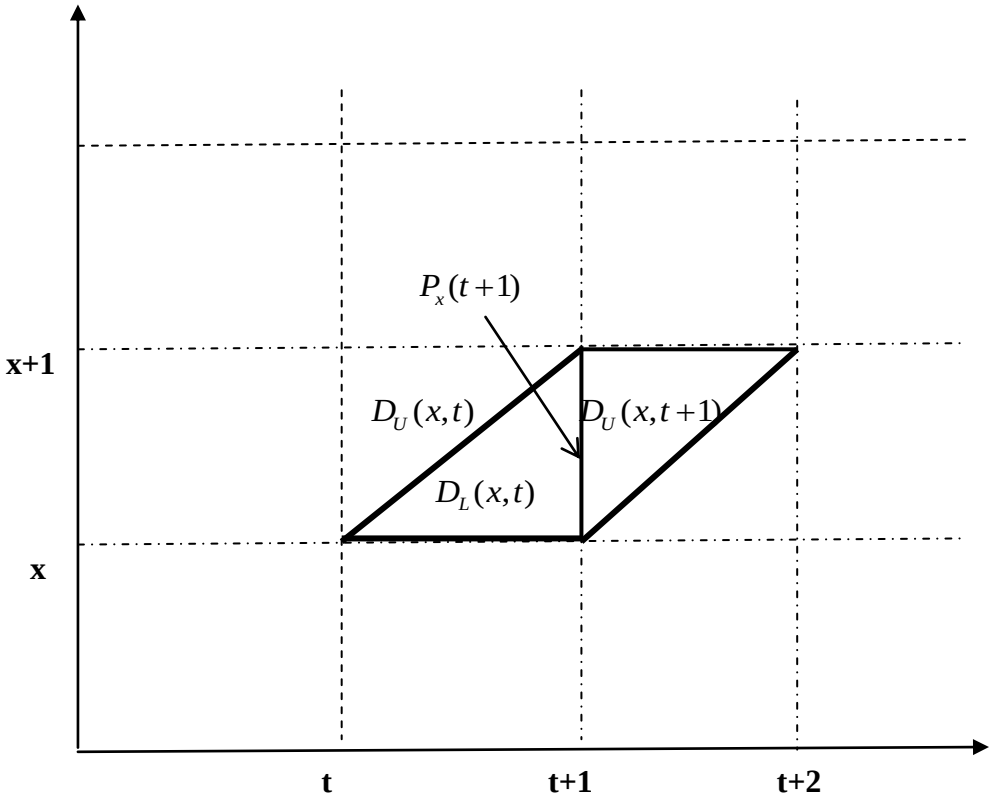


Figure 1: Data for cohort death probabilities.

For a particular cohort (people that have born in the same calendar year), the survival distribution function of the future lifetime, *T*, is defined as ${}_x p_0 := S_T(x) = P(T > x) = \prod_{i=0}^{x-1} p_i$ with $S_T(0)=1$, and denotes the probability that a newborn survives up to age *x*. The corresponding cumulative hazard function is $\Lambda_x = \int_0^x \mu_{u,c} \cdot du = -\log {}_x p_0$, where $\mu_{u,c}$ denotes the force of mortality at age *x* for a particular cohort *c*.

Let $D_x = D_L + D_U$ the total number of deaths for specific age *x*, in the cohort manner as described above. Then, the unconditional probability a newborn to die between ages $[x, x+1)$ is denoted by ${}_{x|}q_0 := P(x \le T < x+1)$, and the corresponding random variable, Δ_x, for the number of deaths, follows a binomial distribution, $\Delta_x \sim Binomial(m, {}_{x|}q_0)$, where *m* is the sum of all deaths across all ages, with estimator ${}_{x|}\hat{q}_0 = \frac{\Delta_x}{m}$.

If we treat the numbers of deaths as polytomous (dependent) responses, for each age *x=0,1,2,…, ω-1*, (where ω denotes the limiting age), they follow the multinomial distribution, with distribution function

$$P(\Delta_0 = D_0, \Delta_1 = D_1, ..., \Delta_{\omega-1} = D_{\omega-1}) = \binom{m}{D_0\ D_1 \ldots\ D_{\omega-1}} \cdot \prod_{x=0}^{\omega-1} ({}_{x|}q_0)^{D_x}$$

where $\binom{m}{D_0\ D_1\ \ldots\ D_{\omega-1}} = \frac{m!}{D_0!D_1!\ldots D_{\omega-1}!}$ with mean $E(\Delta_x) = m \cdot {}_{x|}q_0$, variance $Var(\Delta_x) = m \cdot {}_{x|}q_0 \cdot (1 - {}_{x|}q_0)$ and covariance $Cov(\Delta_x, \Delta_y) = -m \cdot {}_{x|}q_0 \cdot {}_{y|}q_0$.

In this setting, the interval scale is ordered in distinct age ranges $[x, x+1)$, and according to McCullagh & Nelder (1989, page 151), it is preferable to consider models based on the cumulative response probabilities ${}_xq_0 := F_T(x) = P(T \le x) = 1 - {}_xp_0$, rather than the category probabilities ${}_{x|}q_0$, because simple models for the cumulative probabilities are likely to have better properties for ordinal response scales than equally simple models based on the category probabilities. They suggest using the logistic scale or the complementary log-log scale. In general, it is acknowledged that probit and logit models are reasonable choices when the changes in the cumulative probabilities are gradual. The complementary log-log link may be a good model when the cumulative probabilities increase from 0 fairly slowly and then rapidly approach 1.

In view of the above, working with the vector of cumulative (dependent) totals, $Z_x = \sum_{i=0}^{x} \Delta_i$, rather than with the cell counts Δ_x, the cumulative responses have mean $m_x = {}_xq_0$ and covariance matrix $V_{xy} = {}_xq_0 \cdot (1 - {}_yq_0)$, for $x \le y$ and $m = 1$. Utilizing the complementary log-log link function, in age effects, we propose the following model structure for the graduated survival function

$$\eta_x = \log(-\log(1 - {}_x\hat{q}_0)) = \log(-\log({}_x\hat{p}_0)) = \mathbf{L} \cdot \underline{\hat{\beta}} \qquad (1)$$

where the design matrix $\mathbf{L}$ is an orthonormal polynomial in age affects, and $\underline{\hat{\beta}}$ is the vector of the estimated parameters. The model structure (1) subsequently can be applied to each successive cohort independently.

We note that the model structure (1) is equivalent to targeting the cumulative hazard function, as responses, under the log link function:

$$\eta_x = \log(\hat{\Lambda}_x) = \mathbf{L} \cdot \underline{\hat{\beta}}.$$

For the estimation of the unknown parameter vectors $\underline{\beta}$ for each successive cohort, in GLM framework, we minimize the quasi-deviance function (McCullagh & Nelder, 1989, page 335)

$$D(\mathbf{m};\mathbf{Z}) = \frac{2}{3} \cdot (\mathbf{Z} - \mathbf{m})^T \cdot V^{-1}(\mathbf{m}) \cdot (\mathbf{Z} - \mathbf{m}) + \frac{1}{3} \cdot (\mathbf{Z} - \mathbf{m})^T \cdot V^{-1}(\mathbf{Z}) \cdot (\mathbf{Z} - \mathbf{m})$$

where $\mathbf{m}$ denotes the expected cumulative totals and V^{-1} denotes the inverse covariance matrix.

In practice, the minimization of the above quasi-deviance function cannot be applied since the covariance matrix of the expected cumulative totals, $V(\mathbf{m})$, is unknown. For the implementation of the above estimation, in the first step we minimize the linear predictor's mean squared error (MSE), i.e. we minimize the function $Q(\underline{\beta}) = \sum_x \left(\overset{o}{\eta}_x - \eta_x(\underline{\hat{\beta}}) \right)^2$, where $\overset{o}{\eta}_x = \log(-\log(1 - {}_x\overset{o}{q}_0))$ the observed values and $\eta_x(\underline{\hat{\beta}}) = \mathbf{L} \cdot \underline{\hat{\beta}}$ are the expected ones, in order to get an estimation of the covariance matrix $V(\mathbf{m})$.

The minimization of the linear predictor's MSE and the minimization of the quasi-deviance function are subject to particular constraints:

I) Firstly, the inequalities $\frac{d\eta_x}{dx} > 0$ ensure that the survival function is monotonically increasing in all the age range.

II) Moreover, we ensure that the η_x function is strictly concave for the first 10 ages, by imposing $\frac{d^2\eta_x}{dx^2} < 0$. We note that this constraint, is equivalent to requiring that the first derivative of the force of mortality for the first 10 ages is negative, since $\eta'_x = \frac{d}{dx}\log(\Lambda_x) = \frac{\mu_x}{\Lambda_x} \Rightarrow \eta''_x = \frac{\mu'_x \cdot \Lambda_x - \mu_x^2}{\Lambda_x^2} = \frac{1}{\Lambda_x} \cdot \left(\mu'_x - \frac{\mu_x^2}{\Lambda_x} \right) \Rightarrow$ $\eta''_x = \frac{1}{\Lambda_x} \cdot (\mu'_x - \eta'_x \cdot \mu_x)$. Thus, if $\mu'_x < 0 \Rightarrow \eta''_x < 0$ because $\eta'_x \cdot \mu_x > 0$.

III) A further constraint is that the linear predictor is convex after the age of 50, i.e. $\frac{d^2\eta_x}{dx^2} \geq 0$ for *x*>50. This condition is fulfilled by all the mortality experiences under study.

IV) An additional important constraint ensures that the η_x function for the old ages (which is convex under the previous constraint) tends to be linear by imposing that $\left(0 \leq\right) \frac{d^2\eta_{x+1}}{dx^2} < \frac{d^2\eta_x}{dx^2}$ for successive ages (specifically the observed values from various mortality experiences satisfy this condition for the ages above 65 for females and above 70 for males, see graph 5.2). The analysis by Gavrilova and Gavrilov (2011b) of several single-year extinct U.S. birth cohorts shows that the mortality trajectory at advanced ages follows the Gompertz law (i.e. linearity for the force of mortality in log-scale) up to the ages 106 years without significant mortality deceleration. These findings are supported by another study of independent data on siblings of centenarians drawn from verified and accurate U.S. family histories. We note that the asymptotical linearity of the linear predictor η_x is comparable with the linearity of the force of mortality on the logarithmic scale. Indeed, $\mu_x = -\frac{{}_x p'_0}{{}_x p_0} = -\frac{d}{dx}\left(-e^{\eta_x}\right) = e^{\eta_x} \cdot \eta'_x \Rightarrow \log(\mu_x) = \eta_x + \log(\eta'_x)$. If the mortality trajectory at advanced ages follows the Gompertz Law then $\eta_x + \log(\eta'_x) = a + b \cdot x \Rightarrow \eta'_x + \frac{\eta''_x}{\eta'_x} = b \Rightarrow \eta''_x = \eta'_x \cdot (b - \eta'_x)$. As a result, because $\eta''_x \geq 0$, we have that $\eta'_x \leq b$, and also that $\lim_{x\to\infty} \eta'_x = b$ since η'_x is an increasing function of *x*.

Under the above 4 constraints, the minimum number of polynomial parameters allowed in model (1) is 5 (i.e. must be $k_1 \geq 5$).

The minimization of the quasi-deviance function, $D(\mathbf{m};\mathbf{Z})$, under the above 4 constraints is generally referred to as a constrained nonlinear optimization or nonlinear programming problem. Specifically, the minimization method utilizes a quasi-Newton approximation to the Hessian of the Lagrangian at $\underline{\beta}$, using a sequential quadratic programming method. In this method, the function solves a quadratic programming subproblem at each iteration (Fletcher and Powell, 1963).

The optimum degree k_1 for the orthonormal polynomials, in model (1), can be determined by monitoring the quasi-deviance profile. Specifically, we examine the total differences between successive quasi-deviance functions, for all the *n*-cohorts combined (Figure 2). That is, analogous with the analysis of variance, we work out the first differences of deviances for a sequence of models, each including one more parameter more than

the previous one and we sum up these differences for all the *n*-cohorts. In the analysis of quasi-deviance, possible maximum values of the total differences will support the choice the particular model. As is clear from Figure 2, models with k_1=9 number of parameters are the optimum choice for all the mortality experiences (Sweden, Norway, England & Wales and Denmark for both sexes).

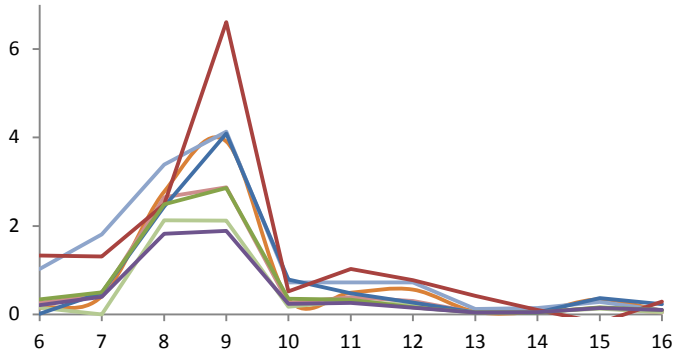


Figure 2: Analysis of quasi-deviance, for the mortality experiences (Sweden, Norway, England & Wales and Denmark for both sexes), under model structure (1): each value represents the first differences of deviances for a sequence of models, with number of parameters k_1*-1* and k_1. High values support the choice of the particular model. The maximum values for all the mortality experiences is k_1 =*9*.

3. Extraction of the age-cohort effects

3.1 Forecasting inside the parallelogram EBDH

According to the proposed method, we have utilized only the data inside the parallelogram EBCF, in Figure 3. The recently available data from the triangle FCH have not yet been used; with only full cohorts have been included in the model structure (1). We need to extrapolate the data in the triangle HCD in order to get estimates for cohorts *c(n+1)* up to *c(n+ω)*, i.e. inside the parallelogram EBDH. Initially, we start with the model (1), defined in the parallelogram EBCF. As a result, we produce a random matrix of estimated parameters, $\mathbf{B}$, of order *n* by k_1. In an analogous way with the SVD approach, we apply SVD, or equivalently Principal Component analysis (PCA), to the matrix $\mathbf{B}$ to extract the k_1-mortality dynamics in age-cohort effects (see Hatzopoulos and Haberman (2009)), leading to the age-cohort (full) model structure

$$\log(-\log({}_x\hat{p}_0(c))) = A(x) + \sum_{i=1}^{k_1} g_i(x)\cdot Y_i(c) \qquad (2)$$

$A(x)$ is a set of age-specific constants, describing the relative pattern of mortality by age. The $g_i(x)$ values, describe the relative importance of the complementary log-log survival probability, at age *x*, to variations and trends in the cohort index c, as described by the PC dynamics ($Y_i(c)$ zero centred trends).

For projection purposes, we propose the following procedure:

i) Forecast the $Y_i(n+1)$ values, for *i=1,2,...,* k_1.

ii) Obtain forecast value ${}_x\hat{p}_0(n+1)$ and let the "observed" value $z_x(n+1) = {}_x\hat{p}_0(n+1)$, for $x=\omega$

iii) Fit model (1) for the (*n+1*)-cohort, using as observed values $z_x(n+1) = {}_xp_0(n+1)$ if $x<\omega$ and $z_x(n+1) = {}_x\hat{p}_0(n+1)$ if $x=\omega$, and get estimated parameter vector $\underline{\hat{\beta}}(n+1)$

iv) Derive model (2) adding the (*n+1*)-cohort

The above process can be repeated for successive cohorts up to the $(n+\omega)$-cohort, in order to derive the model structure (2). The only difference in step ii) & iii) is that the "observed" values $z_x(n+j) = {}_x\hat{p}_0(n+j)$ are defined for ages $x= \omega-j+1,...,\omega$ (for $j=1,...,\omega$).

We note that, according to the above iterative forecasting procedure, we finally utilize all the observed experience inside the triangle FCH, using a succession of re-fits. This kind of method seems to be more reasonable in comparison with a pure extrapolated technique from parallelogram EBCF to parallelogram EBDH. Especially, for the first iterative re-fits, the proposed method is based mostly on the observed survival values.

For the later iterative re-fits (where the forecast method is based on both the observed and projected survival values, and in particular for the last cohorts where the above method is based mostly on the forecast survival values), we note the following points. According to the above method, from the projected values of the $\hat{Y}_i(n+j)$ PCs in step i) and forecast values ${}_x\hat{p}_0(n+j)$ for ages $x= \omega-j+1,...,\omega$ in step ii), we obtain in step iii) the 'predicted values' of the parameter estimates $\hat{\underline{\beta}}(n+j)$, and finally in step iv) we derive the PC scores $Y_i(c)$ and age-specific scores $g_i(x)$ respectively, adding the (n+j)-cohort. Hence, in step iv) we apply PCA to the matrix $\mathbf{B} = \{\hat{\underline{b}}(1), \hat{\underline{b}}(2),..., \hat{\underline{b}}(n+j)\}'$, with associated covariance matrix $\mathbf{\Sigma}$, or equivalently we apply eigenvalue decomposition, to the covariance matrix $\mathbf{\Sigma}$, with associated matrix of eigenvectors $\mathbf{P} = [\underline{e}_1, \underline{e}_2, \cdots, \underline{e}_{k_1}]$ and vector of eigenvalues $\mathbf{\Lambda} = \left[\lambda_1, \lambda_2, ..., \lambda_{k_1}\right]'$. We note that the matrix of eigenvectors $\mathbf{P}$ is a projection matrix which transforms the matrices $\mathbf{B}^r$ and $\mathbf{L}$ into PC scores and age-specific scores respectively, since $\mathbf{Y} = \mathbf{B}^r \cdot \mathbf{P}$ is the matrix of PC scores and $\mathbf{G} = \mathbf{L} \cdot \mathbf{P}$ is the matrix of age-specific scores (where $\mathbf{B}^r$ denotes the rescaled matrix: $\mathbf{B}^r = \mathbf{B} - \underline{\bar{b}}'$, if $\underline{\bar{b}} = (\bar{b}_0, \bar{b}_1, ..., \bar{b}_{k_1})'$ is the vector of the mean values). From the Cayley-Hamilton Theorem we have, for the relationship $\mathbf{\Sigma} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$, that, if a real function $f(x)$ is given by a power series: $f(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + ...$ then $f(\mathbf{\Sigma}) = \mathbf{P} f(\mathbf{\Lambda}) \mathbf{P}'$. This property implies that, any changes in the covariance matrix $\mathbf{\Sigma}$, during the forecast period, of the form $f(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + ...$, will not affect the matrix of eigenvectors $\mathbf{P}$. It seems, under this property, that the predicted $\hat{\mathbf{Y}}$ values inherit a degree of robustness in terms of structural changes of the covariance matrix $\mathbf{\Sigma}$ during the forecast period.

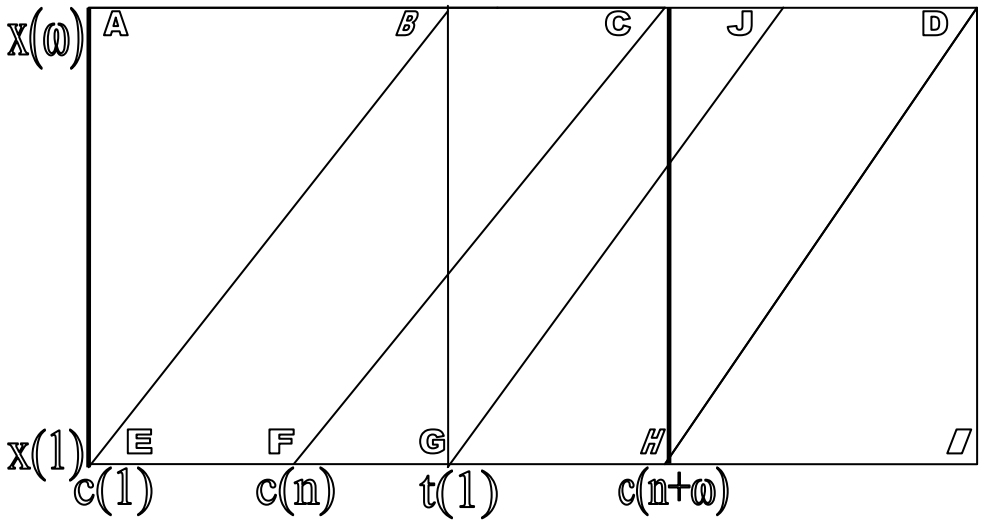


Figure 3: Diagram of observed and extrapolated data.

In step i) of the above procedure, the first basic component, $\mathbf{Y}_1(c)$, is assumed to evolve as an Integrated Random Walk (IRW) trend:

$$Y_1(c) = Y_1(c-1) + \beta_{c-1}$$

$$\text{where } \beta_c = \beta_{c-1} + \varsigma_c \text{ with } \varsigma_c \sim N(0, \sigma_\varsigma^2)$$

for $c=n+1,...,n+\omega$, which when estimated tends to be relative smooth.

For the remaining cohort related principal components, $Y_i(c)$ for $i=2,...,\ k_1$, we advocate a specific class of Dynamic Linear Regression (DLR) models:

$$Y_i\ (c) = \alpha_i + b_{i,c} \cdot c + e_{i,c} \text{ with } e_{i,c} \sim N(0, \sigma_{e_i}^2)$$

for each cohort $c=n+1,...,n+\omega$ (the so-called regressor), with the slope being a stochastic time variable parameter that follows a first order autoregressive process, for stationary PCs, or a random walk process, for non-stationary PCs:

$$b_{i,c} = \varphi_i \cdot b_{i,c-1} + \eta_{i,c} \text{ with } \eta_{i,c} \sim N(0, \sigma_{\eta_i}^2)$$

The innovations $e_{i,c}$ and $\eta_{i,c}$ are assumed to be white noise mutually uncorrelated random variables, representing the system disturbances and measurement noise respectively (Harvey, 1991). If the correlation coefficients φ are less than 1 in absolute value, then the model is called a first-order autoregressive model, where the strength of the correlation between two values of the time series is decreasing as their distance apart in time increases. If φ is equal to 1 the model is called a 'random walk' model (the "border line non-stationary" case). A crucial feature of the model is the signal-noise ratio, or the so-called the noise-variance ratio hyper-parameters (NVR): $NVR_i = \sigma_{\eta_i}^2 / \sigma_{e_i}^2$. If necessary, a pre-determined NVR value can be used to control the required level of smoothness: thus, small values enhance the smoothness.

Experiments with various mortality experiences have shown that these PCs can be represented adequately under the DLR model structures under a first-order autoregressive model. Besides, adopting this particular DLR modelling, gives us the required flexibility, in case where the remaining PC behaves as a non-stationary process (even though the stationary condition is rejected only for some period during the above proposed iterative re-fitting).

Although, we could use the same DLR structure for the first basic component, $\mathbf{Y}_1(c)$ (with φ equal to 1), experiments with the abovementioned mortality experiences gave evidence in favour of the IRW model trend. Under the IRW structure, the projections ${}_x\hat{p}_0(n+j)$, as described in the previous iterative procedure, give us more "reasonable" forecast values. That is, at each iteration, we observe more smoothed transitions, under the IRW model structure, between the observed ${}_x p_0(n+j)$ survival values and the forecast ${}_x\hat{p}_0(n+j)$ survival values (in particular in the neighbourhood of ages where the "observed" values $z_x(n+j)$ change from ${}_x p_0(n+j)$ to ${}_x\hat{p}_0(n+j)$).

The DLR time series models are simply regression models in which the explanatory variables are functions of time and the parameters are time-varying. State space models employ the Kalman filter technique to provide a computationally efficient framework through which we can derive estimates of the stochastic parameters and predicted future values. Predictions are made by extrapolating the estimated components into the future, while

smoothing algorithms give the best estimate of the state at any point within the sample (Harvey, 1991). The computations have been implemented in Matlab using the Captain Toolbox (Taylor, 2007).

In model structure (2), the vectors of PC scores, Y_i (for *i=1,2,...,*k_1), are linear combinations of the k_1-dimensional vectors of the estimated parameters $\hat{\underline{\beta}}$ in model structure (1) (see Model 1, Hatzopoulos and Haberman (2009)). Under the assumption that these vectors of $\hat{\underline{\beta}}$ are asymptotically normally distributed and using the fact that PC analysis, in step iv) of the above procedure, ensures that $Cov(Y_i, Y_j)$ =0, for $i \neq j$, we have that the stochastic processes $Y_i(c)$, for *i=1,2,...,* k_1, are independent of each other. Therefore, in step i) of the above procedure, the vectors of PC scores, Y_i (for *i=1,2,...,*k_1), may be forecasted independently of each other.

3.2 SPCA and central components

Having completed the above forecast procedure we end up with $n+\omega$ cohorts, inside the parallelogram EBDH, with associated random matrix of estimated parameters, $\mathbf{B}$, of order $n+\omega$ by k_1, identical to model structure (1). Then, we apply Sparse Principal Component Analysis (SPCA) to the matrix of the estimated parameters, or equivalently eigenvalue decomposition to the sparse covariance matrix of the estimated parameters (Hatzopoulos & Haberman (2011)), leading to the *age-cohort (sparse) full model* structure which is identical to the model structure (2), inside the parallelogram EBDH.

Like ordinary PCA, SPCA can be used to reduce the dimensionality of a data set by retaining those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones (Mardia et. al. 1997). Such low-order components often contain the "most important" aspects of the data and we keep a "small" subset of them, say $p(<\kappa_1)$, which explains the "majority" of the variances. In SPCA, we seek a trade-off between the two goals of explanatory power (explaining most of the variance or information in the data) and interpretability (making sure that the factors involve only a few variables) (Luss and Aspremont, 2006). As has been analyzed in Hatzopoulos & Haberman (2011), the SPCA approach avoids the high interdependence structure for different range of ages, with possible spurious interpretations and forecasting and improves this problematic dependent structure by giving a better clustering of significant high factor loading values.

Thus, keeping a "small" subset $p(<\kappa_1)$ of the SPCs, which explains the key characteristics in cohort effects, leads to the *age-cohort (sparse) model* structure:

$$\log(-\log({}_x\hat{p}_0(c))) = A(x) + \sum_{i=1}^{p} g_i(x) \cdot Y_i(c) + \varepsilon_x(c) \tag{3}$$

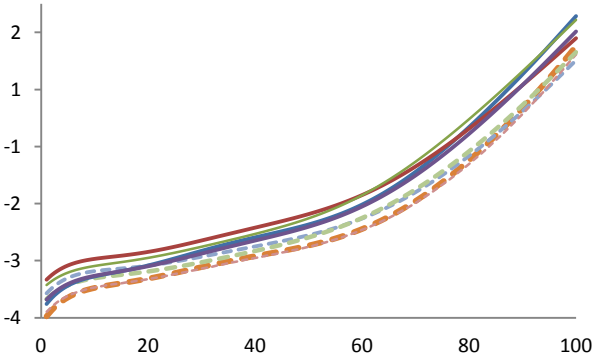
The disturbance term $\varepsilon_x(c)$, is the error component at age *x* in cohort *c* and denotes the deviation of the model represented by the excluded SPCs. Table 1 shows the number of components included (*p*-values) in age-cohort effects, for each mortality experience.

Because the SPC dynamics $Y_i(c)$ are not uncorrelated, the total variance explained by the correlated SPCs cannot be represented by $tr(Y' \cdot Y)$, where *Y* denotes the SPCs. According to Zou et al (2006), using QR decomposition, the adjusted variance can be easily computed. Thus, if $Y = Q \cdot R$, where *Q* is orthonormal and *R* is upper triangular, then the explained total variance is equal to $\sum_{j=1}^{k} R_{jj}^2$, and the percentage variance explained

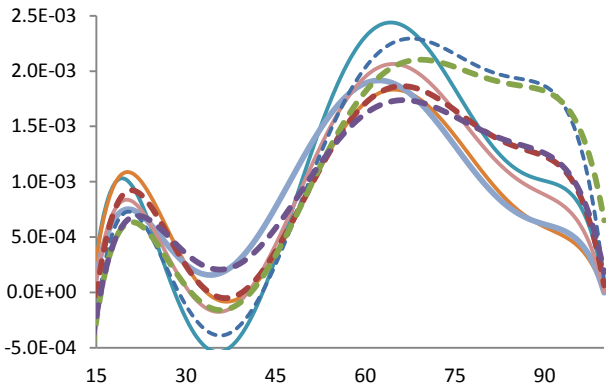
by the *j*-component is $R_{jj}^2 / \sum_{j=1}^{k} R_{jj}^2$. For all the mortality experiences under investigation, the first SPC explains the vast majority of the sparse eigenvalues, in the complementary log-log scale, usually above the 99,7% (Table 1, $VE(Y_1(c))$ values).

In Figure 5, graph 5.1 ($A(x)$ values) shows the main age profile for the survival function on the log-log scale, based on the *age- cohort (sparse) model* (model structure 3) in the parallelogram EBDH (Figure 3). The dashed lines refer to females' experience, with clear lower values, that indicate lower mortality than for males, and graph 5.2 shows the second differences of the main age profile, $A(x)$, for ages 15+, which is in accordance with the particular constraints described previously in section 2.

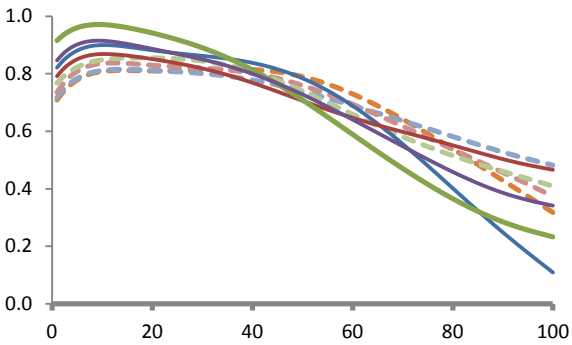
Graph 5.3 shows the $g_1(x)$ age scores for the core $Y_1(c)$ cohort trend, based on the *age- cohort (sparse) model* (model structure 3) in the parallelogram EBDH (Figure 3). We note higher values (almost double) for young ages relative to old ages. Graph 5.4 demonstrates the prominent $Y_1(c)$ cohort trend, where it is clear that from the beginning of the 20th century that there has been a structural change in mortality improvements. After the year 1900, we observe (and forecast) steady rates of mortality improvements. This is in contrast to the situation for cohorts before year 1900, where we observe, irregular, lower rates of mortality improvements with poorer rates at the start of the study (1846).



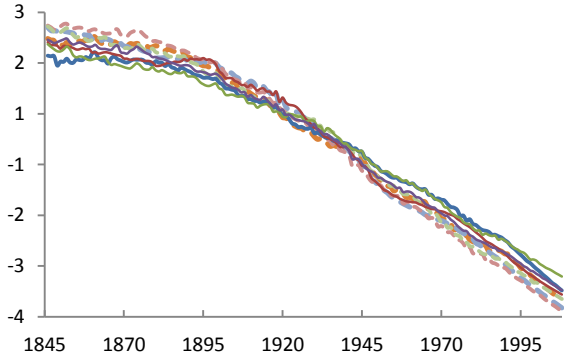
5.1 $A(x)$ for all mortality experiences



5.2 $A''(x)$ for all mortality experiences



5.3 $g_1(x)$ for all mortality experiences



5.4 $Y_1(c)$ for all mortality experiences

Figure 5: Main components, for all mortality experiences (Sweden, Norway, England & Wales and Denmark for both sexes - dashed lines refer to females' experience), based on the *age-cohort model* (model structure 3) in the parallelogram EBDH (Figure 3). $A(x)$ is the main age profile (the main additive effect), $A''(x)$ is the second differences of the main age profile for ages 15+, $g_1(x)$ is the age scores for the central $Y_1(c)$ cohort trend.

3.3 Cohort dynamics

In order to describe the cohort dynamics of the combined survival probabilities (under the log-log transformation), we exploit the first main SPC ($Y_1(c)$ values), for cohorts born after 1900. We fit and smooth them under a specific class of Dynamic Linear Regression (DLR) models:

$$Y_1(c) = \alpha + b_c \cdot c + e_c \quad \text{with} \quad e_c \sim N(0, \sigma_e^2)$$

with the slope being a stochastic time variable parameter that follows a random walk process:

$$b_c = b_{c-1} + \eta_c \quad \text{with} \quad \eta_c \sim N(0, \sigma_\eta^2)$$

The innovations e_c and η_c are assumed to be white noise random variables and mutually uncorrelated. The Noise-Variance Ratio hyper-parameters (NVR) was fixed to the value $NVR = 10^{-7}$.

Table 1, gives the *p*-values for the Jarque-Bera test of the null hypothesis that the residuals comes from a normal distribution, and the *p*-values for the Ljung-Box Q-test to assess the presence of autocorrelation for the residuals at multiple lags jointly.

	Female				***Male***			
	Norway	***E&W***	***Sweden***	***Denmark***	***Norway***	***E&W***	***Sweden***	***Denmark***
k_1	9	9	9	9	9	9	9	9
p	3	3	5	4	4	3	5	3
$VE(Y_1(c))$	99,80%	99,83%	99,91%	99,77%	99,78%	99,89%	99,90%	99,71%
p-value (Jarque-Bera test)	48,0%	0,1%	66,7%	66,5%	43,6%	0,1%	85,1%	0,4%
p-value (Ljung-Box test)	1,8%	2%	51,3%	51,7%	4,8%	32,7%	22,4%	11,6%
k_2	8	8	8	8	8	5	5	8
q	3	3	3	2	3	3	2	2
$VE(Y_1^t(t))$	68%	71%	63%	61%	60%	72%	65%	57%
$VE(Y_2^t(t))$	27%	22%	34%	25%	31%	26%	33%	35%

Table 1: Number of parameters (k_1) used in model (1), number of components used (*p*) in model (3) and associated variance explained ($VE(Y_1(c))$ values) for the first SPC, *p*-values for the Jarque-Bera test of the null hypothesis that the residuals comes from a normal distribution under the DLR model structure since year 1900, *p*-values for the Ljung-Box Q-test to assess the presence of autocorrelation for the residuals under the DLR model structure since year 1900, number of parameters (k_2) used in model (4), number of components used (*q*) in model (5) and associated variance explained ($VE(Y_1^t(t))$ and $VE(Y_2^t(t))$ values) for the first 2 SPCs.

Figure 6, displays the smoothed values for the slopes b_c (the left graphs refer to females' experience). All the graphs are on the same scale for comparison purposes. These slopes describe the rates of mortality improvements in a smooth way.

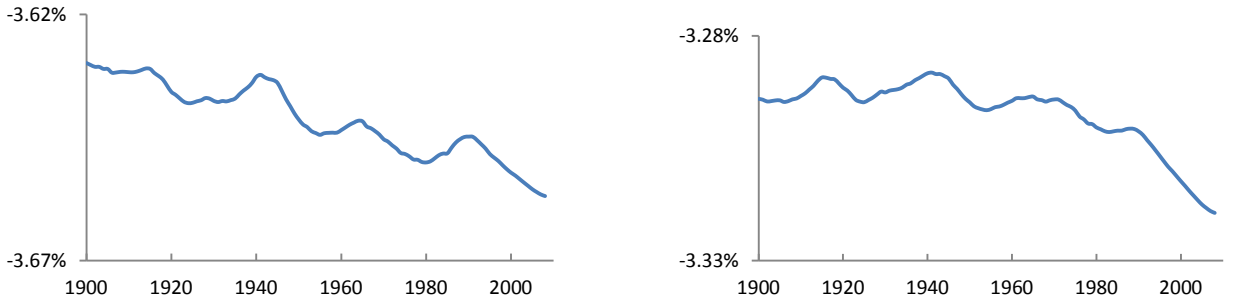
As discussed earlier, according to the above iterative forecasting procedure, we finally exploit all of the observed mortality experience inside the trapezium EBCH (Figure 3), using a succession of re-fits. For the first *n* iterative re-fits, the iterative procedure is based only on the observed survival values (parallelogram EBCF). For the later iterative re-fits, the forecast method is based on both the observed and projected survival values (parallelogram FCDH). In particular, for the last cohorts, the above method is based mostly on the forecast survival values inside the triangle HCD. As a result, in Figure 6, the rates of mortality improvements are the

results of both observed and forecasted mortality rates, especially for cohorts born after 1950 where the number of forecast values exceeds the observed values.

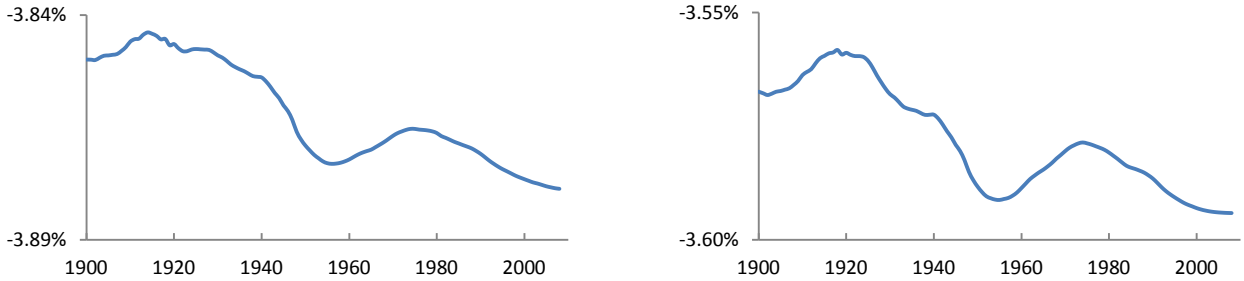
A common feature is the stagnation, or even deterioration, of the mortality improvements at the decades of 1900s-1920s, and a relative faster rates of improvements after the decade of the 1930s (except for the Norway experience, where we observe faster rates of improvements after the decade of the1940s). An important decade, for relative mortality improvements, seems to be the 1950s. According to Andreev & Vaupel (2006) "in the 1950s age specific pattern of mortality improvement was remarkably similar among countries. The highest rates of improvement are observed in infancy and childhood with progressive slackening of rates of improvement at the higher ages. This pattern underwent substantial country-specific transformations in its shape over time. The transformation took place mostly in the 1960s when death rates for many countries were stagnant or increasing".

In case of E&W, the shapes of the slope, for both males-females, indicate a select cohort effect for people born during decades 30's up to 50's. This well documented finding is consistent with Richards et. al. (2005), who comment that "The cohort effect is the observed phenomenon that people born in the U.K. between 1925 and 1945 have experienced more rapid improvement in mortality than generations born on either side of this period".

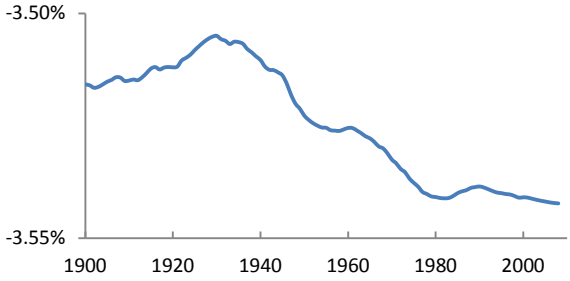
Figure 6.2 also supports the findings of other researchers regarding the cohorts born after the 1940s. Thus, Murphy (2009) argues that "More attention should be given to those born after the 'golden cohorts' in the period 1945 to 1965 in Britain who will dominate mortality trends in the future, since they appear to have worse mortality than might be expected with little or no mortality improvement when compared with their immediate predecessors to date. This is particularly anomalous since they were the first products of the Welfare State, including the National Health Service, and brought up in a boom economic period with unprecedented family life stability". Similarly, Andreev & Vaupel (2006) comment that "in the 1950s and in the 1960s pattern of mortality improvement is similar to that found in many other countries: general decline in the 1950s with especially rapid improvements at younger ages and deceleration of progress in the 1960s with increases in mortality among ages 15-20 and those in the forties".

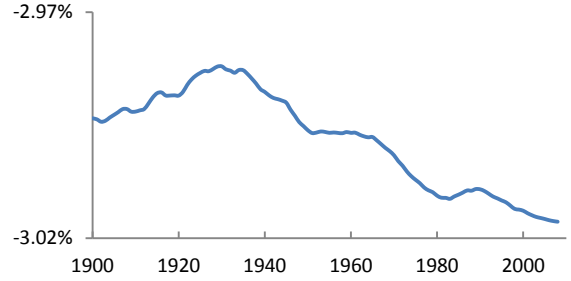


6.1 Norway - smoothed b_c values under the DLR model structure (Females left – Males right)

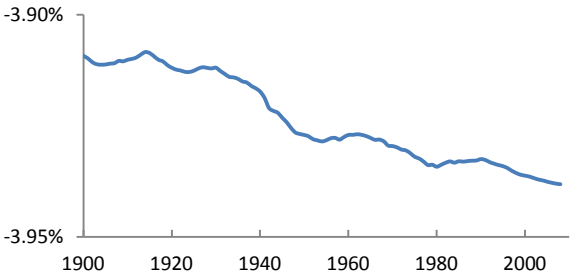


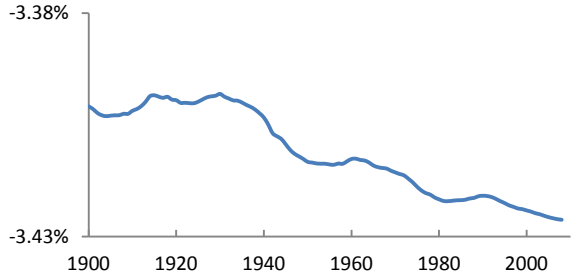
6.2 E&W - smoothed b_c values under the DLR model structure (Females left – Males right)





6.3 Sweden - smoothed b_c values under the DLR model structure (Females left – Males right)





6.4 Denmark - smoothed b_c values under the DLR model structure (Females left – Males right)

Figure 6: Cohort dynamics of the combined survival probabilities - rates of overall mortality improvements in cohort effects (smoothed b_c values under the DLR model structure), based on the *age-cohort (sparse) model* (model structure 3) in the parallelogram EBDH (Figure 3). Females left graphs – Males right graphs.

3.4 Unexplained Variance

In order to derive the most important cohort trends-effects, in the model structure (3), we define for each added component the Unexplained Variance (UV) for each age, comparable to that one considered in the Lee-Carter model (Lee and Carter, 1992). For each age, we find the variance of the time series of survival rates and the variance of the error between the model's predicted survival rates and the actual survival rates, inside the parallelogram EBDH. The ratio of these two variances then defines the UV:

$$UV_x(j) = \frac{\text{var}\left(error(x, j)\right)}{\text{var}\left({}_x\overset{o}{p}_0\right)}$$

where ${}_x\overset{o}{p}_0$ denote the observed survival rates and $error(x, j)$ denote the error between the observed survival rates and the survival rates under model (3) with *j*-components, for *j=1,2,…,p*, i.e.

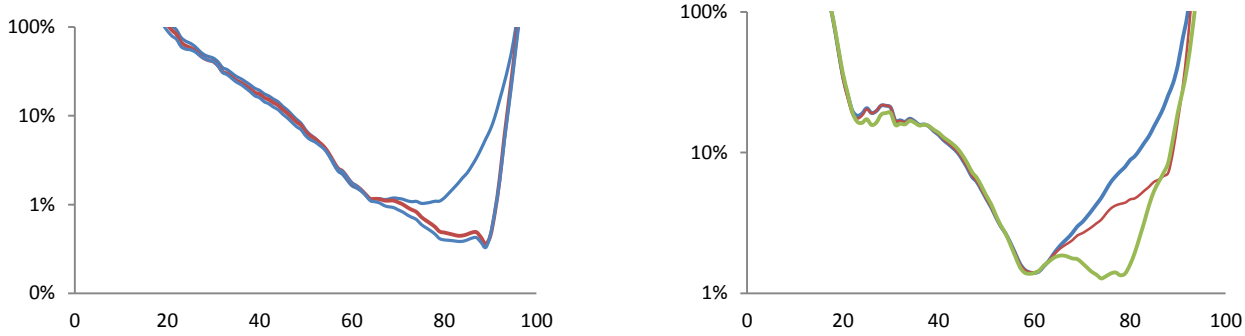
$$error(x, j) = \exp\left(-\exp\left(\log(-\log({}_x\overset{o}{p}_0)) - A(x) + \sum_{i=1}^{j} g_i(x)\cdot Y_i(c)\right)\right)$$

This measure of fit will reveal, for each $Y_i(c)$ cohort trend, the contribution by age to the UV. Figure 7 shows the UVs, for each added component (in order of significance – lower curves correspond to the least significant components, left graphs refer to females' experience). The graphs are on the log-scale, because there is considerably more variation for the very old and the very young ages (this may be due to the relative small exposures at the higher ages or relative small numbers of deaths at the lower ages and also recording issues). The upper lines in these graphs correspond to $UV_x(1)$ values, i.e. the UV when using the first interaction term, which captures the majority of the total variance explained under the model structure (3).

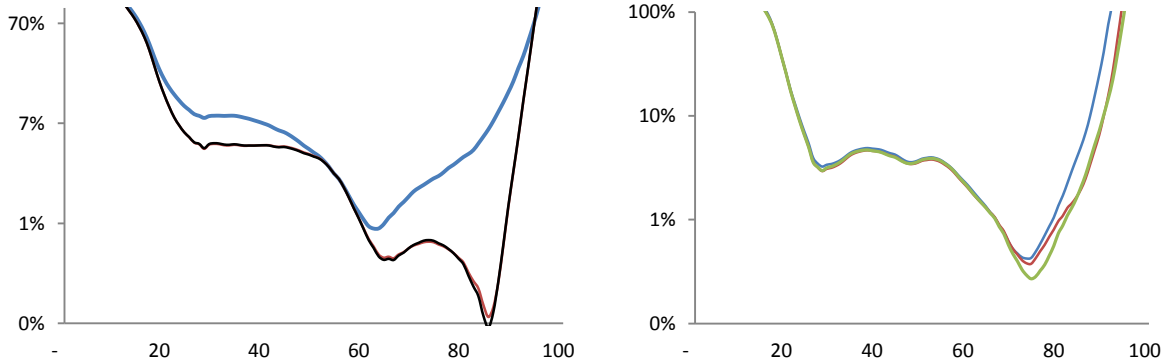
The remaining $UV_x(j)$ values have some common characteristics. The second UV in order of significance, $UV_x(2)$ values, refer to senescent mortality (mainly for ages above 65, i.e. mainly for pensioners). These $UV_x(2)$ values have more impact at ages 85-90, since at these ages we have the maximum difference from the $UV_x(1)$ values in log-scale, i.e. the maximum values of $\log(UV_x(2)) - \log(UV_x(1))$. In general, for each added component, the improvement in the UV can be quantified by the area covered between successive lines. This area can be approximated by the total differences, of successive log-UVs: $\sum_x \log(UV_x(j+1)) - \log(UV_x(j))$ for *j=1,2,…,p-1*. From Figure 7, we can conclude that this common feature is more significant for females than for males' experience (except for Sweden).

For Sweden and Denmark females experience, the third UV in order of significance, $UV_x(3)$ values, refer to middle-old ages (Sweden mostly for ages between 50-75, with maximum difference from the $UV_x(2)$ values in log-scale at age 60, and Denmark for ages between 50-75, with maximum difference at age 70). We note that this feature, for Sweden and Denmark females, is comparable with the second in significant order UV.

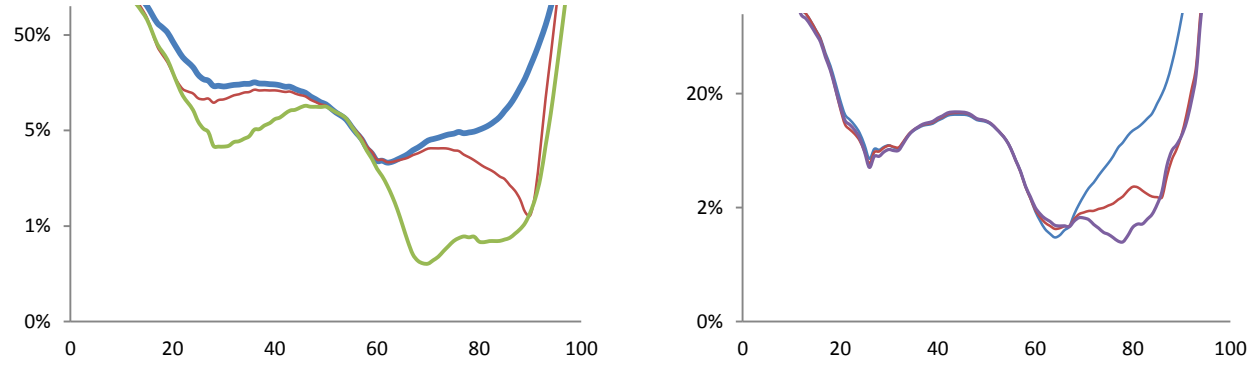
For the remaining experiences, the $UV_x(3)$ values refer to old ages (mainly for ages between 65 and 85) with the maximum difference from the $UV_x(2)$ values in log-scale at ages around 80. This common feature is much more significant for males than for females' experience and a distinctly more important for Norway and Denmark among the males experience.



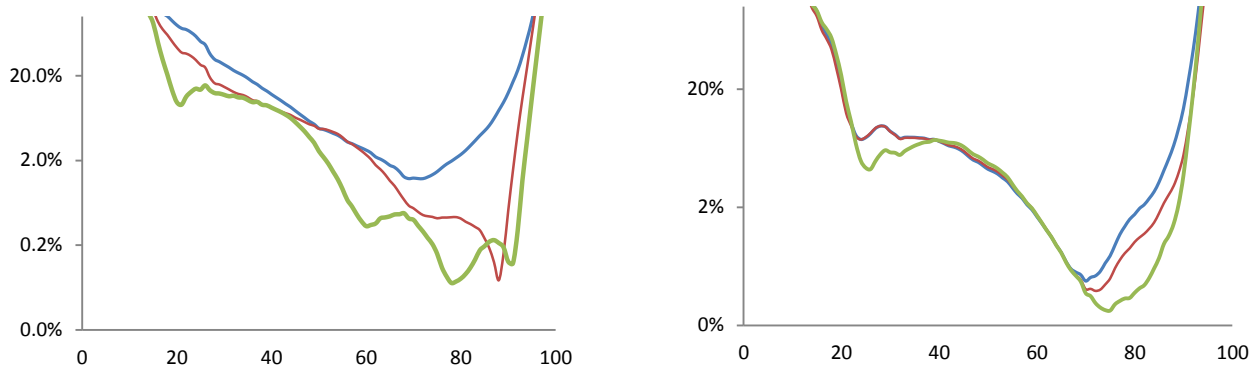
7.1 Norway $UV_x(1)$-$UV_x(2)$-$UV_x(3)$ curves (Females left – Males right)



7.2 E&W $UV_x(1)$-$UV_x(2)$-$UV_x(3)$ curves (Females left – Males right)



7.3 Sweden $UV_x(1)$-$UV_x(2)$-$UV_x(3)$ curves (Females left – Males right)



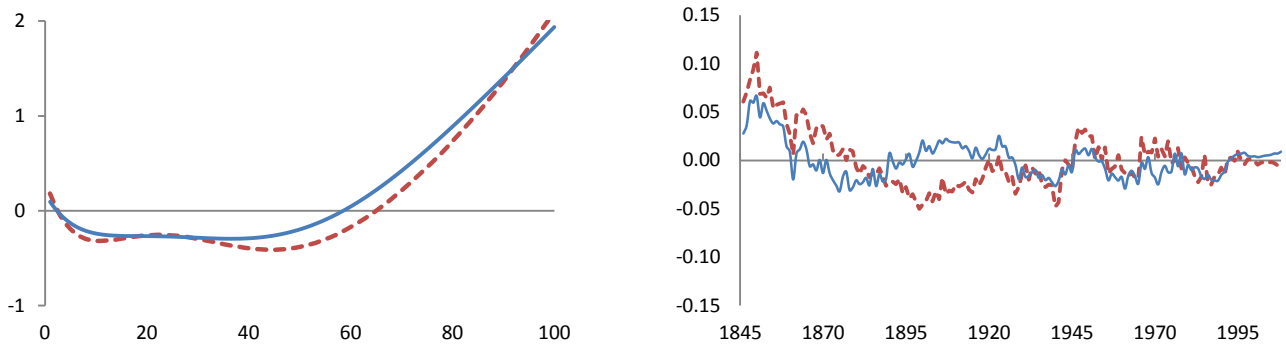
7.4 Denmark $UV_x(1)$-$UV_x(2)$-$UV_x(3)$ curves (Females left – Males right)

Figure 7: Most important Unexplained Variances $UV_x(1)$-$UV_x(2)$-$UV_x(3)$, in age effects, for each added component in significant order – lower curves correspond to lower significant components, based on model (3). The left graphs refer to females' experience and the right graphs refer to males' experience.

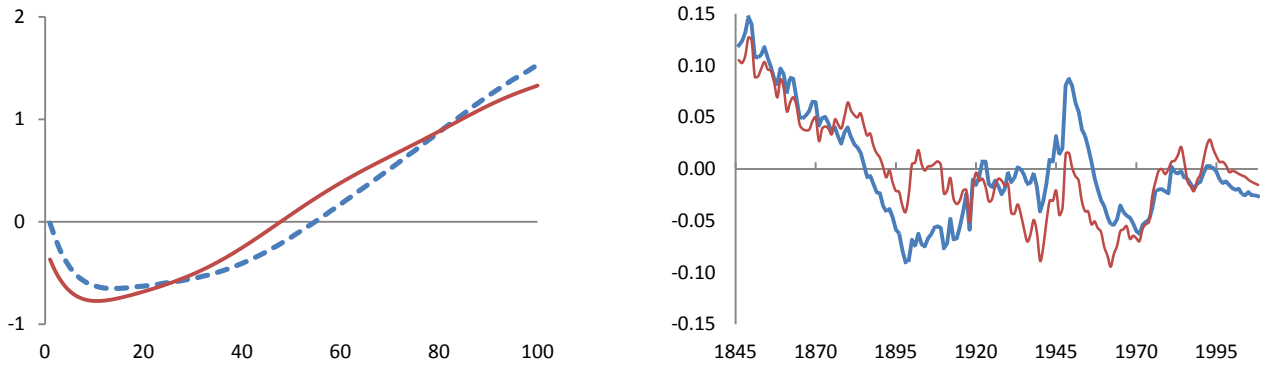
3.5 Remaining significant components

The left panel graphs in Figures 8 & 9 ($g_t(x)$ values, for $i>1$), explain the relative importance of the response of the log-log survival probability, at age x, to variations and trends in the cohort index c (as described by the $Y_i(c)$ trends), according to model (2). In combination with the results from Figure 7, we can identify and explain the deviations from the first main component. For positive $g_t(x)$ values, positive (negative) values of $Y_i(c)$ represent deterioration (improvement) in mortality relative to the first main interaction term. The dashed lines, in Figures 8 & 9, refer to females' experience.

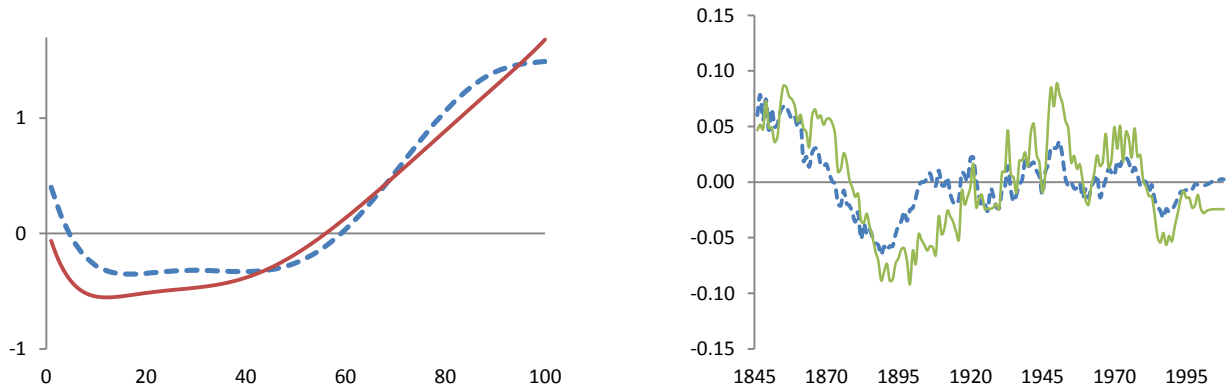
In Figure 8, all the graphs describe deviations from the main trend for the senescent mortality (pensioners' ages above 65), with maximum impact for ages 85-90, according to the second in significant order unexplained variance, $UV_x(2)$ values. This second more significant component has some common trends, as displayed by the $Y_2(c)$ dynamics (all right graphs in Figure 8). Mostly for females' experience, we can observe a select cohort effect around the year 1900, which is consistent for Britain with the work of Wong & Haberman (2004), who discuss the select cohort effect and cohort born after 1900 once they reach age 65. Particularly, for E&W, and males' experience, we observe that for years 1925-1945 there is a positive effect, which emphasizes the select cohort effect for cohorts born between 1925 and 1945. Andreev & Vaupel (2006) detects a cohort effect for E&W males and cohorts born around 1930, stretching from age 50 in the 1980 to age 70 in the late 1990s with rates of improvement on average 3.5% per annum. For females they found a similar cohort effect but it is shifted in time centering on the cohorts born in 1935, but less expressed and more blurred over the cohorts.



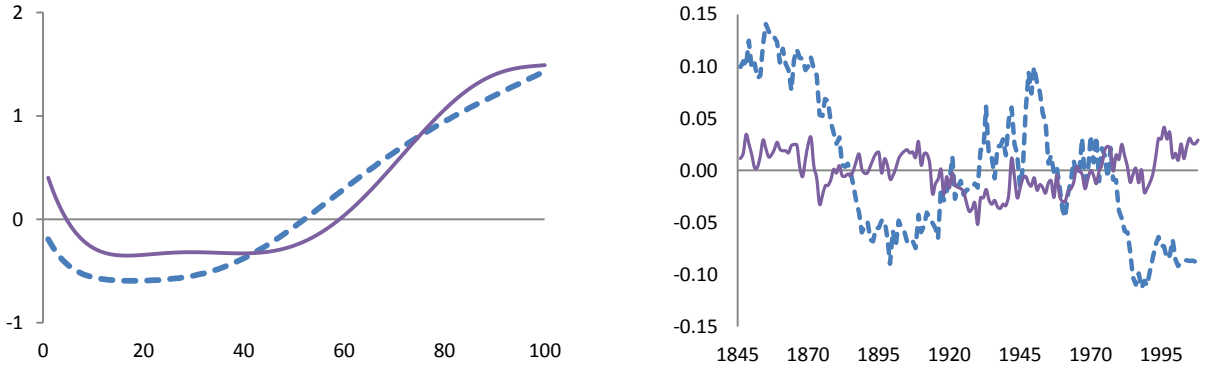
8.1 Norway $g_2(x)$ - $Y_2(c)$



8.2 E&W $g_2(x)$ - $Y_2(c)$



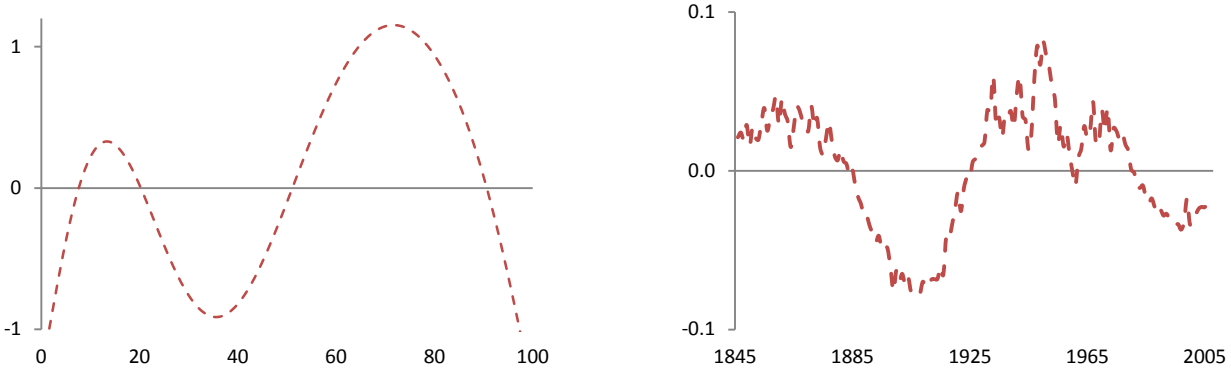
8.3 Sweden $g_2(x)$ - $Y_2(c)$



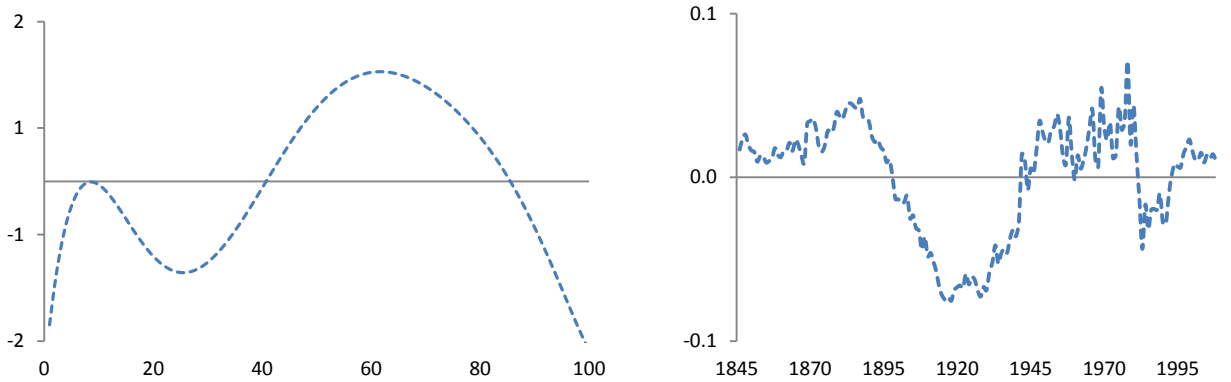
8.4 Denmark $g_2(x)$ - $Y_2(c)$

Figure 8: Dynamics in age-cohort effects, for $g_2(x)$ and $Y_2(c)$ scores, based on the *age-cohort model* (model structure 2) in the parallelogram EBDH (Figure 3). The dashed lines refer to females' experience.

Figure 9, displays a different common characteristic for Sweden and Denmark females experience, which in connection with Figure 7, refers to the middle-old ages (ages between 50 and 75). The $Y_3(c)$ trends indicate a positive effect for years 1900-1940 in the case of Sweden and for 1890-1920 in the case of Denmark, and a select cohort effect for these groups of ages.



9.1 Sweden females $g_3(x)$ - $Y_3(c)$

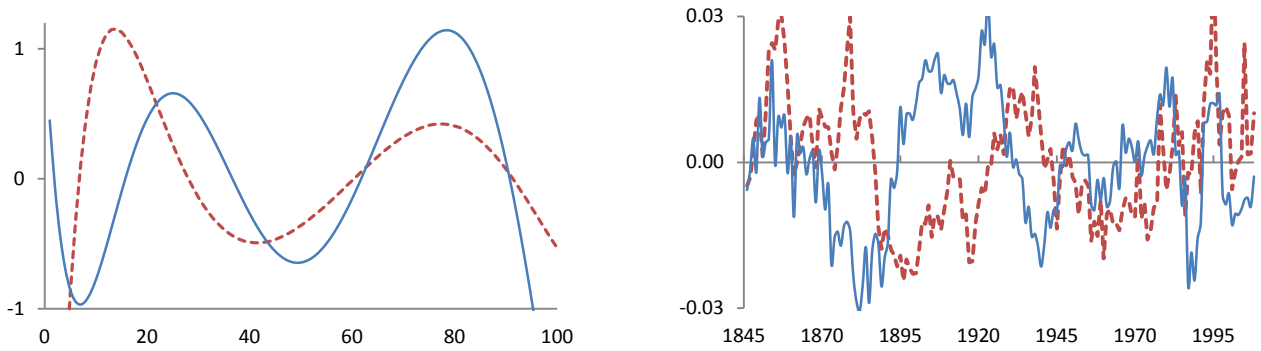


9.2 Denmark females $g_3(x)$ - $Y_3(c)$

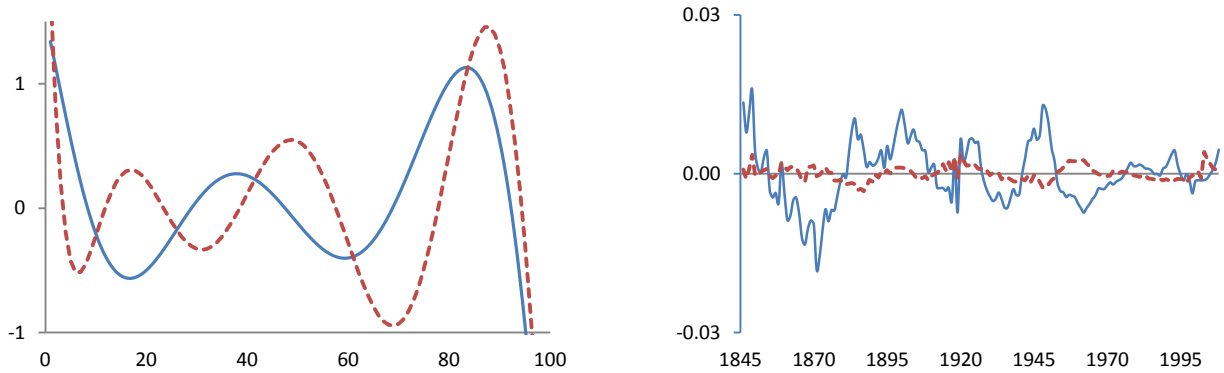
Figure 9: Dynamics in age-cohort effects, for $g_3(x)$ and $Y_3(c)$ scores, based on the *age-cohort model* (model structure 2) in the parallelogram EBDH (Figure 3), for Sweden and Denmark female experience.

Figure 10, shows another common characteristic produced, the third in significant order unexplained variance, which in connection with the results from Figure 7, refer to old ages (mainly for ages between 65 and 85) with maximum values at ages around 80. As discussed earlier, this common feature is much more significant for the males' than for the females' experience and is distinctly more important for Norway and Denmark among the males experience.

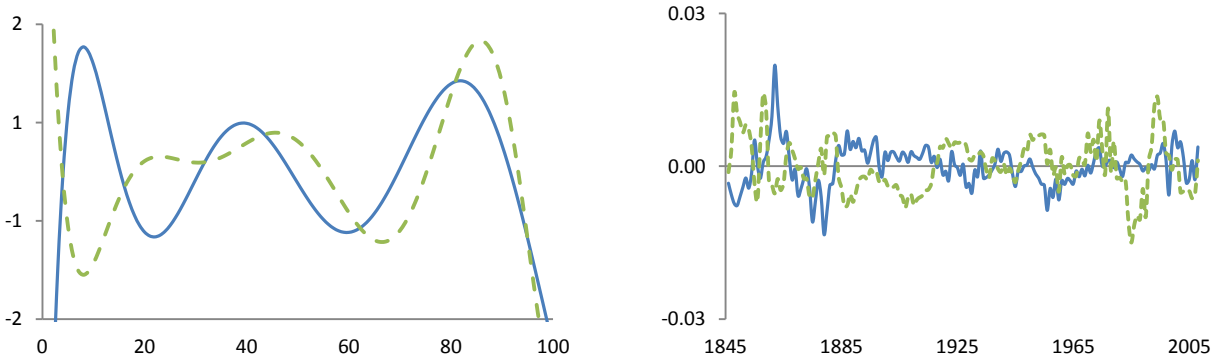
For Norway, we have a negative effect for years 1895-1930 for males and 1920-1940 for females. These results are in accordance with those of Janssen & Kunst (2005), who found that, for men, the decline stagnated among Danish, Dutch, and Norwegian birth cohorts born between 1890 and 1915 and who noted the stagnation of all-cause mortality decline for women born from 1920 onwards. For Norwegian men, the proportion of ever smokers increased for the birth cohorts 1890–1914, after which it stayed at the same level up to cohort 1930–34 (83%). There is a similar feature for the Danish population, where we have a negative effect for years 1885-1925 for males and 1915-1940 for females. For E&W, we have a negative effect for years 1880-1910 for males and 1910-1925 for females. These findings are comparable with those of Willets (2004), where the highest lung cancer mortality rates by age group and year of birth (Figures 7 & 8) are in agreement with graph 10.2. For the Swedish population, this characteristic has a negligible effect particularly for males. As noted by Richards et. al. (2005), Sweden differs from many other developed European countries in that cigarette smoking has not been the major form of tobacco consumption during the 20th century. For the remaining components, which have not been displayed and discussed, the majority of the unexplained variances refer to the neighborhood of age 20 ('accident hump' in cohort effects).



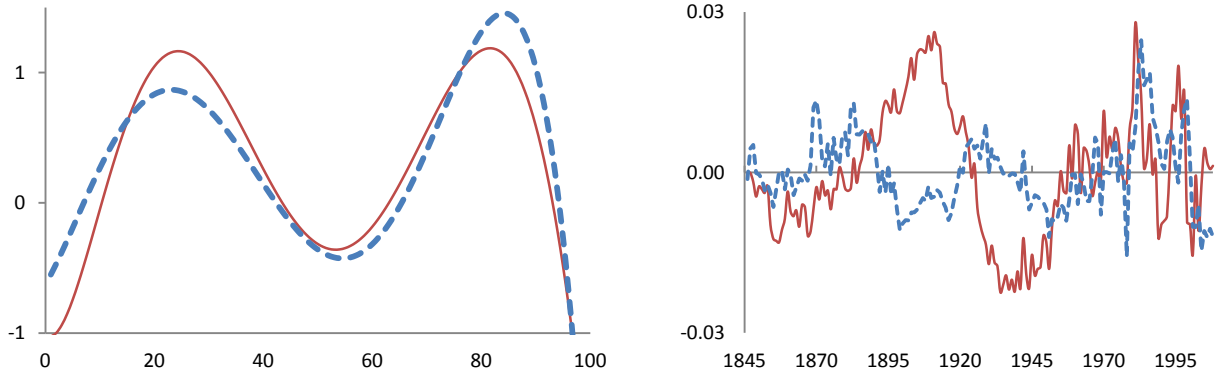
10.1 Norway $g_3(x)$ - $Y_3(c)$



10.2 E&W $g_3(x)$ - $Y_3(c)$



10.3 Sweden males $g_3(x)$ - $Y_3(c)$, Sweden females $g_4(x)$ - $Y_4(x)$



10.4 Denmark males $g_3(x)$ - $Y_3(c)$, Denmark females $g_4(x)$ - $Y_4(x)$

Figure 10: Dynamics in age-cohort effects, based on the *age-cohort model* (model structure 2) in the parallelogram EBDH (Figure 3). The dashed lines refer to females' experience.

4. Incorporation of period effects

According to the model structure (3), we have obtained fitted values for the cohort survival rates, ${}_x p_0(c)$, for the whole age range ($x=1,2,...,\omega$) in the rectangle GBCH (Figure 3), after keeping the first p-significant (most important) age-cohort interaction terms (Table 1). In the presence of time effects, the GLM estimates $\underline{\beta}_c$ in model (1) are marginal estimates of age and cohort effects. If we consider the residual time effect conditional on the already estimated age and cohort effects, in the rectangle GBCH, then the survival functions

$$ {}_x p_0(t) = {}_x p_0(c+x) = \exp\left(-\int_0^x \mu_{u,c+u} \cdot du\right) = \exp\left(-\int_0^x \mu_{u,t-x+u} \cdot du\right) $$

for a particular year t are independent for each age x, as they refer to a different population for each age x (age x last birthday). Given a year t, for each age x, the ${}_x p_0(t)$ values correspond to a different cohort $c=t-x$.

Corresponding to the age-cohort model structure, if we consider the vector of cumulative independent totals, $Z_x = \sum_{i=0}^{x} \Delta_i(t-x)$, where the random variable $\Delta_i(t-x)$ denotes the number of deaths aged i for cohort $t-x$, the cumulative responses have mean $m_x = {}_x q_0(t) = 1 - {}_x p_0(t)$ and covariance matrix $V(\mathbf{m}) = diag\{V_1(m_1),...,V_\omega(m_\omega)\}$, where $V_x(m_x) = {}_x q_0(t) \cdot (1 - {}_x q_0(t))$ for $x=1,2,...,\omega$. According to McCullagh & Nelder (1989, page 325), the integral

$$ Q(m_x; Z_x) = \int_{z_x}^{m_x} \frac{z_x - t}{\sigma^2 \cdot V_x(m_x)} \cdot dt $$

behaves like a log-likelihood and since the components of $\mathbf{Z}$ are independent, the quasi-likelihood for the complete data is the sum of the individual contributions: $Q(\mathbf{m};\mathbf{z}) = \sum_{x=1}^{\omega} Q(m_x; z_x)$, and matching quasi-deviance $D(\mathbf{m};\mathbf{z}) = -2 \cdot \sigma^2 \cdot Q(\mathbf{m};\mathbf{z})$. In our case, the individual log-likelihood for each age x equals

$$ Q(m_x; z_x) = \frac{1}{\sigma^2} \cdot \left\{ z_x \cdot \log\frac{z_x}{m_x} + (1 - z_x) \cdot \log\frac{1 - z_x}{1 - m_x} \right\} $$

The conventional estimate of σ^2 is a moment estimator:

$$ \tilde{\sigma}^2 = \frac{1}{\omega - k_2} \sum_{x=1}^{\omega} \frac{(z_x - m_x)^2}{V_x(m_x)} $$

If we treat the term $A(x) + \sum_{i=1}^{p} g_i(x) \times Y_i(c)$ from model (3) as an offset, in the rectangle GBCH (Figure 3), we can model the survival function ${}_x p_0(t)$ in a similar way (in complementary log-log scale), leading to the model structure

$$ \log(-\log({}_x \hat{p}_0(t))) = A(x) + \sum_{i=1}^{p} g_i(x) \times Y_i(c) + \mathbf{L} \times \underline{\hat{b}}_t \qquad (4) $$

where the design matrix $\mathbf{L}$ is an orthonormal polynomial of degree k_2 (Table 1), in age affects, and $\hat{\underline{b}}_t$ is the vector of the estimated parameters of dimension k_2. For the estimation of the unknown parameter vectors $\underline{\beta}_t$ for each successive year, we minimize the quasi-deviance function

$$D(\mathbf{m};\mathbf{z}) = -2 \cdot \sum_{x=1}^{\omega} \left\{ z_x \cdot \log \frac{z_x}{m_x} + (1 - z_x) \cdot \log \frac{1 - z_x}{1 - m_x} \right\}$$

The model structure (4) subsequently can be applied to each successive year in the rectangle GBCH, to derive the matrix $\hat{\mathbf{B}}_t$ of the estimated parameters. The optimum degree k_2 for the model structure (4), can be determined by investigating the likelihood ratio test for all the *n*-years combined, defined by

Likelihood ratio = - 2 · (total log-likelihood with k-1 parameters) + 2 · (total log-likelihood with k parameters)

Under this test, possible maximum values (or local maximum values) of the likelihood ratios' will give the optimum model (Table 1, k_2 values). In the same way as for the age-cohort model (3), we apply SPCA to the matrix $\mathbf{B}_t$ of the estimated parameters, in the rectangle GBCH, according to model (4), to extract the mortality dynamics in age-time effects, leading to the age-cohort-period model structure

$$\log(-\log({}_x\hat{p}_0(t))) = A'(x) + \sum_{i=1}^{p} g_i(x) \cdot Y_i(c) + \sum_{i=1}^{q} g_i^t(x) \cdot Y_i^t(t) + \varepsilon_x(t) \qquad (5)$$

where $A'(x) = A(x) + A^t(x)$, if $\underline{A}^t = \mathbf{L} \cdot \overline{\beta}_t$ is the vector of age scores which represents the residual adjustment for the main age profile $A(x)$ after including the period effects, and $\varepsilon_x(t)$ is the residual error in fitting the model (4) for age *x* at year *t* using a finite set of SPCs.

In order to derive the *q*-significant time effects, in model structure (5), we define, for each added component, the Unexplained Variance (UV) of each age. For each age, we find the variance of the time series of survival rates and the variance of the error between the models predicted survival rates and the actual survival rates, after having subtracted the offset term from model (3), in the rectangle GBCH (Figure 3). The ratio of these two variances then defines the UV:

$$UV_x(i) = \frac{\mathrm{var}\left(error^t(x, j) \right)}{\mathrm{var}\left({}_x\overset{o}{p}'_0(t) \right)}$$

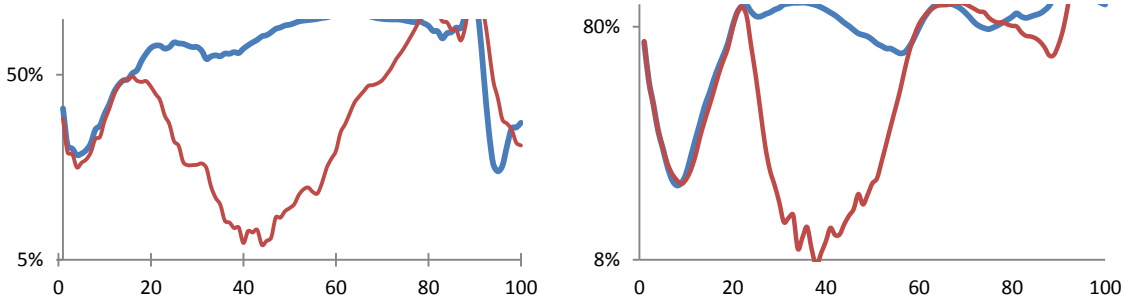
where ${}_x\overset{o}{p}'_0(t)$ denote the observed survival rates after having subtracted the offset term $A(x) + \overset{p}{\underset{i=1}{\mathring{a}}} g_i(x) \times Y_i(c)$ from model (3), in the rectangle GBCH (Figure 3), i.e.

$${}_x\overset{o}{p}'_0(t) = \exp\left(-\exp\left(\log(-\log({}_x\overset{o}{p}_0)) - A(x) + \sum_{i=1}^{p} g_i(x) \cdot Y_i(c) \right) \right)$$

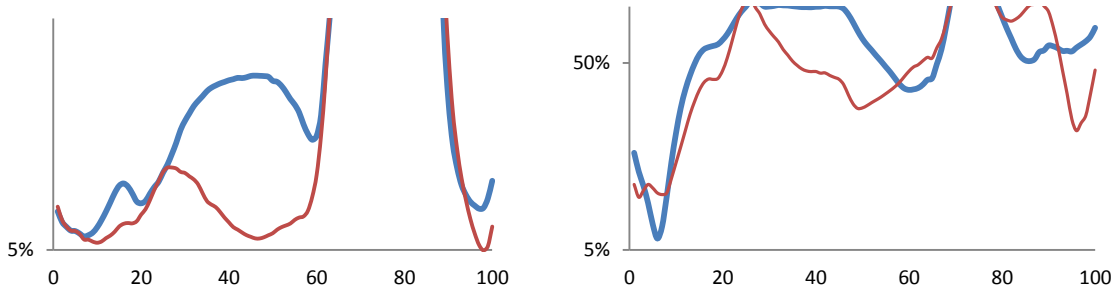
and $error^t(x, j)$ denote the error under model (5) with *j*-components, after having subtracted the offset term $A(x) + \sum_{i=1}^{p} g_i(x) \times Y_i(c)$, in the rectangle GBCH, for *j=1,2,…,q*, i.e.

$$error^t(x, j) = \exp\left(-\exp\left(\log(-\log({}_x\overset{o}{p}_0)) - A'(x) + \sum_{i=1}^{p} g_i(x) \cdot Y_i(c) - \sum_{i=1}^{j} g_i^t(x) \cdot Y_i^t(t) \right) \right)$$

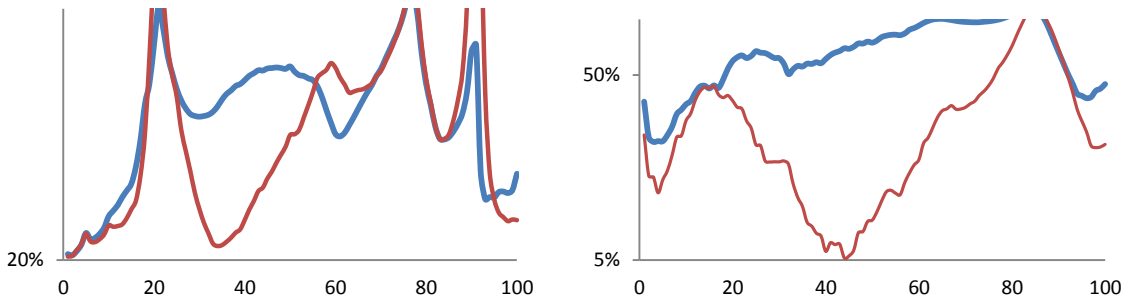
This measure of fit will reveal, for each $Y_i^t(t)$ time trend, the contributions of each group of ages. Figure 11 shows the UVs, in log-scale, for each added component based on model (5), in time effects. In combination with the left graphs in Figure 12 ($g_i^t(x)$ values), which explain the relative importance of each age *x*, to variations and trends in the time index *t* ($Y_i^t(t)$ trends), we can identify deviations from the age-cohort model structure.



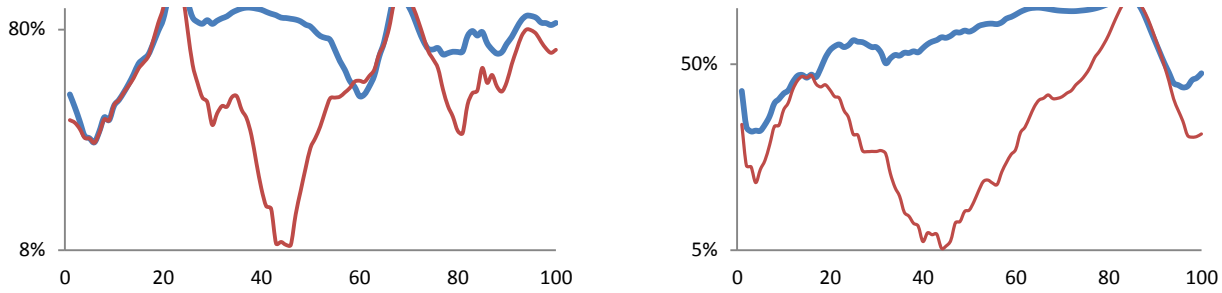
11.1 Norway (Females – Males)



11.2 E&W (Females – Males)



11.3 Sweden (Females – Males)



11.4 Denmark (Females – Males)

Figure 11: Unexplained Variance, in age effects, for each added age-period component, based on model (5).

From Figures 11 & 12, we note that the mortality experiences under investigation share common characteristics.

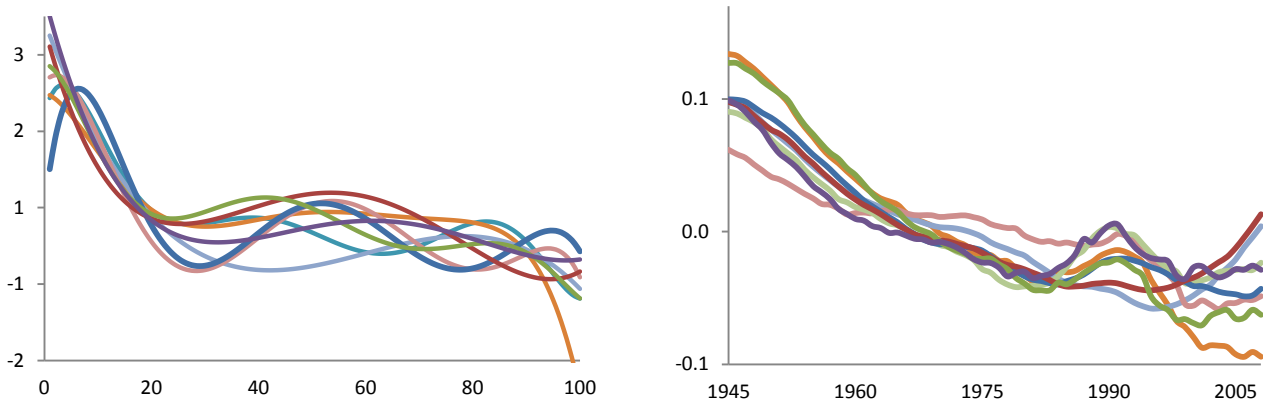
Figure 11, and left graph in Figure 12.1. In Figure 11, the upper lines in all of the graphs show lower values of the UVs at young ages. Besides, in Figure 12.1 the $g_i^t(x)$ values, which explain the relative importance of each age *x*, to variations and trends in the time index *t* ($Y_i^t(t)$ trends), have maximum values at the young ages. From the $Y_1^t(t)$ trends, which show the time trend for the young ages, it is clear that the rates of mortality improvements, since 1947, for these ages are gradually diminishing, and during the 1970s have stagnated and, in some cases, have later deteriorated.

Many other authors describe equivalent mortality dynamics in time effects. Horiuchi and Wilmoth (1998), discuss mortality dynamics, noting that "just when the age-specific death rates of the young became so low (that their further decline could contribute little to increasing life expectancy), the rates of decline at the older ages began to accelerate". Glei and Horiuchi (2007), state that, for industrialized countries during recent decades, the gain in life expectancy is mainly due to changes in adult mortality, and moreover, among those populations, the rate of mortality decline does not vary markedly across adult ages. Lee (2003) comments that "the mortality age schedule has changed shape between the first half of the 20th century, when the mortality decline was much more rapid for the young than for the old, and the second half, when there is little difference among the rates of decline above age 20 or so". Also, White (2002), discusses that "the predominant types of mortality reduction shifted from curing infectious diseases that heavily affect the young to degenerative diseases that largely affect the elderly". Also Hatzopoulos and Haberman (2013) illustrate the linear trend for the young ages (particularly after the 1970s) and the accelerated rate of improvement for the remaining ages (especially for the males after the 1980s), based on pooled mortality experience from 19 countries. According to Andreev & Vaupel (2006) "starting with the 1970s mortality decline took different paths in different countries which can be broadly classified as predominantly period and age shifting patterns of mortality decline. The age shifting pattern is characterized by progressive shifting of high rates of improvement into the higher ages"

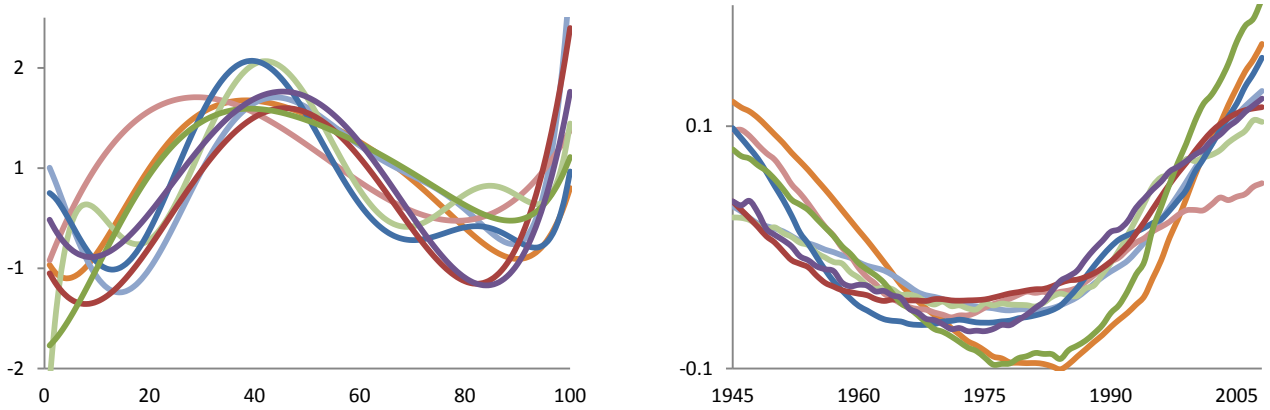
The second most common characteristic refers to ages 30's-40's, as shown by the lower lines in all the graphs in Figure 11, and in the left panel graph in Figure 12.2. The $Y_2^t(t)$ trend, shows the time trends for this range of ages, where it is clear that the relative improvements for these ages have stagnated until 1980's and after that have deteriorated.

Murphy (2009) discusses possible reasons for this characteristic in Britain and states that "considerable attention has been drawn to the relatively poor mortality performance of young adults of cohorts born in the 1950s and 1960s following a substantial deterioration in the rate of mortality improvement for young adults in the 1980s, which has been used to highlight the apparently privileged position of the 'golden generations'. A major reason for this was increases in mortality from HIV/AIDS, external causes and substance abuse".

In addition, we identify age-time residual effects that refer to ages around the age of 20 (accident hump time effect) for most of the remaining SPCs in age-time effects.



12.1 Norway-E&W-Sweden-Denmark $g_1^t(x)$ - $Y_1^t(t)$

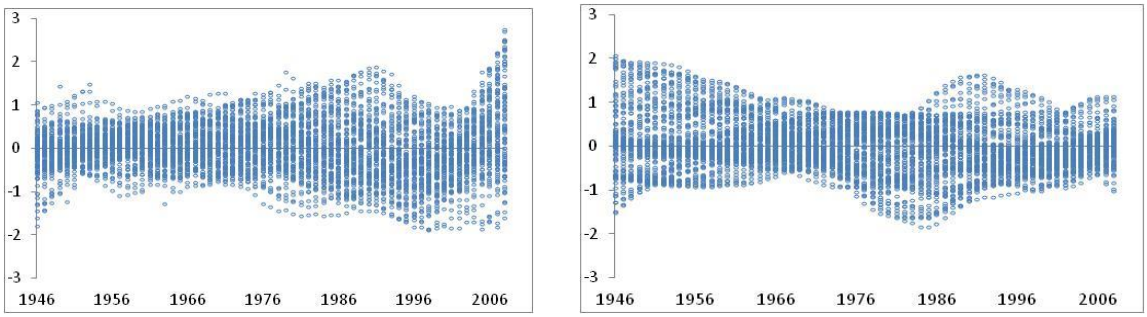


12.2 Norway-E&W-Sweden-Denmark $g_2^t(x)$ - $Y_2^t(t)$

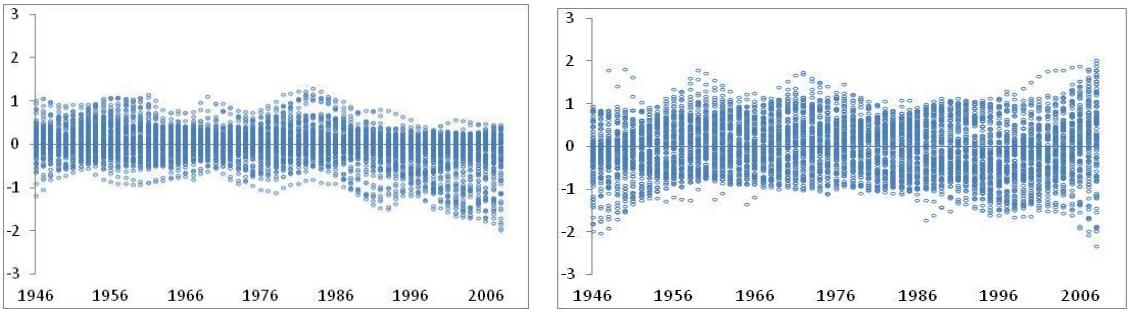
Figure 12: Dynamics in age-period effects (Norway-E&W-Sweden-Denmark both sexes) based on the *age-cohort-period model* (5) in the rectangle GBCH (Figure 3). $g_i^t(x)$ & $Y_i^t(t)$ are the residual interaction (period) components, $g_i^t(x)$ are the age related components and $Y_i^t(t)$ are the period related components.

Table 1 displays the Variance Explained (VE) (orthogonal-triangular decomposition of the $\mathbf{Y}^t$ matrix), for the common two characteristics, under model structure (5) for calendar years 1945-2008. Table 1 provides information on the VE: $VE(Y_1^t(t))$ values for the first SPC and $VE(Y_2^t(t))$ values for the second SPC.

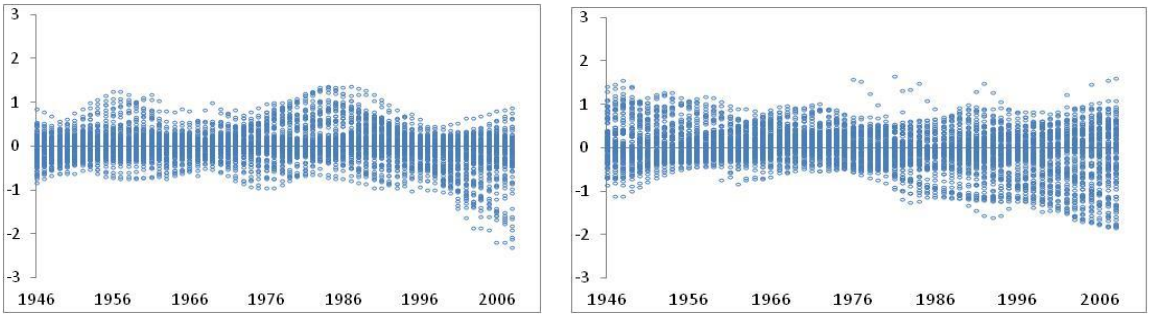
Figure 13 shows the Standardized Deviance Residuals (SDR) plotted against time effects, for calendar years 1846-2008, according to the *age-cohort-period model* (5). The left hand graphs refer to female experience. For all of the graphs, the overall patterns of the SDR against time indicate an appropriate fit.



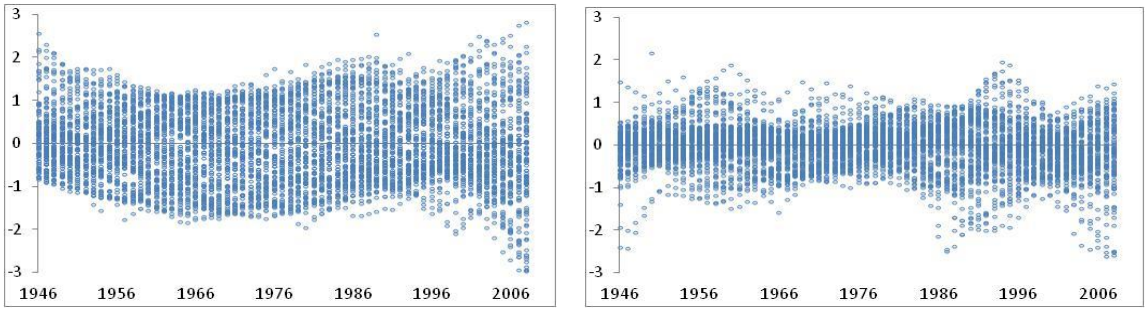
13.1 Norway (Females – Males)



13.2 E&W (Females – Males)



13.3 Sweden (Females – Males)



13.4 Denmark (Females – Males)

Figure 13: SDR vs. time effects for the age-cohort-period model (5), left graphs for females and right graphs for males.

If we extrapolate the time dependent components $Y_i^t(t)$'s, in the rectangle HCDI we can derive the graduated-forecast survival values in age-cohort-time effect, defined in the parallelogram GJDH:

$$\log(-\log({}_x\hat{p}_0(c))) = A'(x) + \sum_{i=1}^{p} g_i(x)\cdot Y_i(c) + \sum_{i=1}^{q} g_i^t(x)\cdot \hat{Y}_i^t(t) + \varepsilon_x(c+x) \qquad (6)$$

5. Discussion

The study based on Norway, England & Wales, Sweden and Denmark, for both sexes, total population mortality experience, for calendar years 1846-2008 and individual ages x=0,…,99. The data are freely provided by the "Human Mortality Database" (www.mortality.org). All the computations have been implemented in Matlab (2013). The proposed model fitting proceeds as follows:

- Calculate, in the trapezium EBCH, the crude cohort probability, p_x, of survival between ages x and $x+1$, given that is alive age x, and the corresponding crude cohort survival probability function ${}_x p_0 = \prod_{i=0}^{x-1} p_i$ (for cohorts 1846-2008 and each available age x=1,…,100).
- For each successive cohort inside the parallelogram EBCF (i.e. for cohorts 1846-1908 with full range of ages), graduate the survival probability function, using the complementary log-log link function, subject to particular constraints, to get a random matrix of GLM estimated parameters (model 1).
- Use the iterative procedure described in section 3.1, to extrapolate the data inside the parallelogram FCDH and derive a random matrix of GLM parameters (for cohorts 1909-2008).
- Apply SPCA to the matrix of both GLM parameters, to derive the age-cohort model structure 3 inside the parallelogram EBDH (cohorts 1846-2008).
- Consider the residual time effect conditional on the already estimated age and cohort effects, in the rectangle GBCH (i.e. for calendar years 1947-2008 with full range of ages), and graduate the (conditional) survival probability function to derive a random matrix of time-depended GLM estimated parameters (model structure 4).
- Apply SPCA to the time-depended matrix, in the rectangle GBCH, to extract the mortality dynamics in age-time effects and extrapolate the time dependent components in the rectangle HCDI.
- Derive the age-cohort model structure 6, in the parallelogram GJDH (for post-war cohorts 1947-2008 and full range of ages).

We have investigated a new approach to the modeling and projected of survival probabilities for the age-cohort-time effects. We introduce a one-factor parameterized simple orthonormal polynomials in age effects, within the GLM framework, treating year of birth as a factor. Then, we apply SPCA to the cohort dependent non-stationary time series parameter estimates in order to provide (marginal) estimates for a two-factor SPC approach structure (the age-cohort sparse model). Then, we model the age-cohort residuals by a one-factor parameterized orthonormal polynomials in age effects treating now calendar year as a factor. Finally, we apply SPCA to time dependent GLM parameter estimates in order to provide (conditional) estimates for a two-factor structure in age and time effects. Overall, in this way a three-way structure is derived in age-cohort-period effects (the age-cohort-period model). Many authors make use of three-way structure, like Russolillo et. al. (2011).

All of the mortality experiences investigated, according to the quasi-deviance profile have the same optimum number of parameters (k_1=9). Also, experiences with other countries give the same result. Under the 4 constraints described in section 2, the minimum number of polynomial parameters allowed in model (1) is k_1 =5. Although, the model with k_1=5 parameters can describe the core age-cohort trends, it does not embrace the differential mortality effects (remaining interaction terms) in comparison with the model which utilizes the k=9 parameter structure.

In contrast with the age-time-cohort structure (Hatzopoulos and Haberman, 2011 & 2013) where the calendar year of death was the predominant effect, the proposed method, where the year of birth is the leading effect, produces a single smooth SPC which captures the vast majority of the variance explained. This involves one major non-stationary principal component to explain the whole range of age dynamics, in cohort effects. The remaining minor SPCs are mean reverting (stationary) stochastic processes. In age-time-cohort structure (Hatzopoulos and Haberman, 2011), the SPCA with E&W male mortality experience implies two main age groups. In contrast, Hatzopoulos and Haberman (2013) consider the post-war time period (1947-2006) and utilize pooled data from 19 countries worldwide: they derive two common significant interaction terms which summarize the basic mortality dynamics, from all the pooled countries, in age and time effects. These two common interaction terms split the age range in two groups, young and adult.

Although PC analysis is a classic tool for analyzing multivariate data one of the key shortcomings of PC analysis is that these factors are linear combinations of all variables, that is, all factor coefficients (or loadings) are non-zero (Luss and Aspremont, 2006). In the case of a non-stationary time series, simultaneous drifting of the series may register as correlations between the columns, thus potentially influencing the components

(Lansangan and Barrios, 2009). In such cases, having only a few non-zero coefficients in the principal components would greatly improve the relevance and interpretability of the factors. SPCs describe the overall mortality trend in cohort and time effects, providing additional insights into the different mortality dynamics, especially in cases where many dynamic factors are present.

The scores of the log-log transformation for the survival function, in age effects and cohort manner, inherit some important characteristics. They are monotonically increasing in all the age range, concave for the first 10 ages, convex after the ages of 50, and finally tend to be linear for senescent ages (specifically for the ages above 65 for females and above 70 for males). This last characteristic is equivalent to the Gompertz Law, which provides additional insight and simplicity for the construction of complete life tables. All of the above characteristics are preserved in the forecast method advocated (section 3), leading to more credible results.

The proposed model produces common mortality effects (Figure 7 & 10) with similar trends, in age-cohort (Figure 8, 9 & 10) and (conditionally) in age-period effects (Figure 11 & 12). Figure 6 describes the rates of mortality improvements in a smooth way. A common feature is the stagnation, or even deterioration, of the mortality improvements at the decades of 1900's-1920', and a relative faster rates of improvements basically for decades of 30s-50s. The transformation took place mostly in the 1960s when the death rates for many countries were stagnant or increasing.

Figures 7 & 10 show distinctive common age groups who are dominant for the related mortality dynamics. The first more important relative deviation from the main component, refers to senescent (pensioners) mortality dynamics (Figure 8), describing a select cohort effect mostly for women population around the year 1900 and at late 1940s – early 1950s almost all graphs show a pure cohort effect. Also, an additional select cohort effect is found for Swedish & Danish female population (Figure 9), for middle ages (ages between 50 and 75) and cohort years around 1900s-1930s for Sweden and 1890s-1910s for Denmark.

A different pure cohort effect is detected by another common characteristic referring to old ages (Figure 10), mainly for ages between 65 and 85, with more significant scores for males (distinctively for Norway and years 1895-1930, and for Denmark and years 1920-1940). For E&W, we have a negative effect for years 1880-1910 for males and 1910-1925 for females. This effect is likely to be associated mostly with cigarette consumption leading to lung cancer mortality rates. For Swedish population, this characteristic has a negligible effect, particularly for males, due to the fact that many Swedes use snuff which is placed in the mouth rather than being smoked.

For the (conditionally) age-period effects, Figure 11 exhibits the time trend for the young ages. We can see that these post-war rates of mortality improvements have gradually diminished, and during the 1970s have stagnated and in some cases later have deteriorated relative to the main trend. These results are supported by the previous work of Andreev & Vaupel (2006) and also by White (2002), who comments that "the predominant types of mortality reduction shifted from curing infectious diseases that heavily affect the young to degenerative diseases that largely affect the elderly". The second common (conditionally) age-period characteristic, presented in Figure 12, refers to the ages 30-49, where it is clear that the relative improvements in mortality have stagnated until the 1980s and after that have deteriorated. A possible reason for this is the increase in mortality from HIV/AIDS, suicide, accidents, external causes and substance abuse.

Additionally, for the remaining components, the majority of the unexplained variance refers to the ages around 20 ('accident hump' effect). This effect is either integrated in the age-cohort model (3) or in the (conditional) age-period model structure (5). For Norway males, Sweden and Denmark for both sexes, the proposed method incorporates the 'accident hump' effect in the age-cohort model, and for E&W and Norway females the model includes the 'accident hump' effect in the (conditional) age-period model structure (5).

Di Cesare et. al. (2009) have explored which of the four models of mortality dynamics (Lee-Carter, Booth-Macdonald-Smith variant of Lee-Carter, Age-Period-Cohort, or Bayesian) have best explained trends in three different causes of death in E&W: lung cancer; influenza, pneumonia and bronchitis; and motor vehicle accidents (neither). Their conclusions are respectively: lung cancer (cohort); influenza, pneumonia and bronchitis (period); and motor vehicle accidents (neither). They conclude that "no single model is most accurate

for all three causes and the best model depended on the underlying mechanisms driving the trends". The method proposed in this paper extracts and forecasts the "smoking effect" component under an age-cohort structure, as well as select and poor cohort effects, related to the pensioners' mortality dynamics. The (conditional) age-period model structure captures mostly mortality differentials from influenza, HIV/AIDS, suicide, accidents, external causes and substance abuse. The 'accident hump' effect can be included in both structures.

It is clear that the proposed model structure which has been described in this paper has not dealt with confidence intervals and related risk management issues. Our main purpose has been to extract the key characteristics that describe the log-log transform of the cohort survival probabilities in age-cohort-period effects. An open question for subsequent work is whether this model structure can effectively integrate risk management techniques for longevity risk measures, especially when using many interaction terms.

References

Alai, D.H. and Sherris, M. 2014. "Rethinking Age-Period-Cohort Mortality Trend Models". Scandinavian Actuarial Journal, 2014 (3), pp 208-227.

Andreev, K. F. and Vaupel, J. W. 2006. "Forecasts of cohort mortality after age 50". MPIDR Working Paper WP-2006-012. (2006).

Barker, D. J. P., Osmond, C. and Law, C. M. 1989. "The intrauterine and early postnatal origins of cardiovascular disease and chronic bronchitis". Journal of Epidemiology and Community Health 43: pp 237–240.

Bell, W.R. 1997. "Comparing and Assessing Time Series Methods for Forecasting Age-Specific Fertility and Mortality Rates." Journal of Official Statistics 13(3), pp 279–303.

Beltra´n-Sa´nchez, H., Crimmins, E. M. and Finch, C. E. 2012. "Early cohort mortality predicts the rate of aging in the cohort: a historical analysis". Journal of Developmental Origins of Health and Disease, 3 (5), pp 380–386.

Benjamin, B., and Pollard, J.H. 1993. "The Analysis of Mortality and other Actuarial Statistics". London, England: Faculty of Actuaries and Institute of Actuaries.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M., 2008. "Mortality density forecasts: An analysis of six stochastic mortality models". Pensions Institute Discussion Paper PI-0801, Pensions Institute, Cass Business School.

Di Cesare, M. and Murphy, M. 2010. "Forecasting mortality, different approaches for different causes of deaths? The cases of lung cancer; influenza, pneumonia, and bronchitis; and motor vehicle accidents". British Actuarial Journal, 15 (Supplement): pp 185-211

Charlton, J., and Murphy, M. (eds.). 1997. "The Health of Adult Britain 1841–1994. Office of Population Censuses and Surveys, Decennial Supplement Series DS, No. 12 and 13. London: HMSO.

Davey Smith, G. and Lynch, J. 2004. "Commentary: Social capital, social epidemiology, and disease aetiology". International Journal of Epidemiology 33: pp 691–700.

Derrick, V. P. A. 1927. "Observation on (1) error on age on the population statistics of England and Wales and (2) the changes of mortality indicated by the national records (with discussion)". Journal of the Institute of Actuaries 58: pp 117–159.

Finch, C. E., and Crimmins, E. M. 2004. Inammatory Exposure and Historical Changes in Human Life-Spans, Science, 17 September 2004, Volume 305.

Fletcher, R. and Powell, M.J.D. 1963. "A Rapidly Convergent Descent Method for Minimization". Computer Journal, Vol. 6, pp. 163-168.

Gavrilova, N.S., and Gavrilov, L. A. 2011a. "Season of Birth and Exceptional Longevity: Comparative Study of American Centenarians, Their Siblings, and Spouses". Journal of Aging Research, Volume 2011, Article ID 104616, 11 pages, http://dx.doi.org/10.4061/2011/104616

Gavrilova, N.S., and Gavrilov, L. A. 2011b. "Ageing and Longevity: Mortality Laws and Mortality Forecasts for Ageing Populations" [In Czech: "Stárnutí a dlouhov kost: Zákony a prognózy úmrtnosti pro stárnoucí populace"]. Demografie, 53(2): pp 109-128.

Glei, D. and Horiuchi, S. 2007. "The narrowing sex differential in life expectancy in high-income populations: Effects of differences in the age pattern of mortality". Population Studies, vol. 61 (2), pp. 141-159.

Harvey, A. 1991. "Forecasting, Structural Time Series Models and the Kalman Filter". Cambridge University Press.

Hatzopoulos, P. and Haberman, S. 2009. "A parameterized approach to modelling and forecasting mortality". Insurance: Mathematics and Economics, 44 (1), pp 103-123.

Hatzopoulos, P. and Haberman, S. 2011. "A dynamic parameterization modelling for the age-period-cohort mortality". Insurance: Mathematics and Economics, 49 (2), pp 155-174.

Hatzopoulos, P. and Haberman, S. 2013. "Common mortality modeling and coherent forecasts. An empirical analysis of worldwide mortality data". Insurance: Mathematics and Economics, 52 (2), pp 320-337.

Horiuchi, S. and Wilmoth, J.R. 1998. "Deceleration in the age pattern of mortality at older ages". Demography, vol 35 (4), pp 391-412.

Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at: http://www.mortality.org. (data downloaded at November, 2009).

Janssen, F. and Kunst, A. E. 2005. "Cohort patterns in mortality trends among the elderly in seven European countries". International Journal of Epidemiology 34: 1149–1159: pp 1950–99.

Kermack, W. O., McKendrick, A. G. and McKinlay P. L. 1934. "Death-rates in Great Britain and Sweden: Some general regularities and their significance". The Lancet 223: pp 698–703. (Reprinted in International Journal of Epidemiology 30: pp 678–683).

Lansangan, J.R. and Barrios, E.B. 2009. "Principal Components Analysis of Nonstationary Time Series Data". Statistics and Computing 19 (2), pp 173-187.

Lee, R. D. 2003. "Mortality forecasts and linear life expectancy trends", in T. Bengtsson, ed., "Perspectives on Mortality Forecasting: Vol. III. The Linear Rise in Life Expectancy: History and Prospects". Social Insurance Studies No.3, Swedish National Social Insurance Board, Stockholm, pp 19-39.

Luss, R. and Aspremont, A. 2006. "DSPCA: a Toolbox for Sparse Principal Component Analysis". Mathematical Subject Classification: 90C90, 62H25, 65K05.

Mardia, K. V., Kent, J. T. and Bibby, J. M. 1997. "Multivariate Analysis". Academic Press.

MATLAB Release 2013b, The Math Works, Inc., Natick, Massachusetts, United States.

McCullagh, P. and Nelder, J.A. 1989. “Generalized linear models”. (2nd ed.) London: Chapman Hall.

Meza, R., Pourbohloul, B. and Brunham, R. C. 2010. “Birth cohort patterns suggest that infant survival predicts adult mortality rates”. Journal of Development Origins of Health and Disease, 1 (3), pp 174-183.

Murphy, M. 2009. “The ‘golden generations’: Fact or fiction?”. British Actuarial Journal 15, Supplement, pp 151-184.

Renshaw, A. E. and Haberman, S. 2006. “A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors”. Insurance: Mathematics and Economics, 38 (3), pp 556-570.

Richards, S. J., Kirkby, G. and Currie, I.D. 2005. “The importance of year of birth in two-dimensional mortality data”. British Actuarial Journal, 12 (I), pp 5–61.

Richards, S. J. 2008. “Detecting year-of-birth mortality patterns with limited data”. J. R. Statist. Soc. A, 171, pp 279–298.

Russolillo, M., Giordano, G. and Haberman, S. 2011. "Extending the Lee-Carter model: a three-way decomposition". Scandinavian Actuarial Journal, 6, pp 96-117.

Taylor, C., J. 2007. Engineering Department, Lancaster University, Lancaster, LA1 4YR, United Kingdom. Web http://www.es.lancs.ac.uk/cres/captain/.

White, K.M. 2002. “Longevity advances in high-income countries, 1955-96”. Population and Development Review 28 (1), pp 59-76.

Willets, R.C. 2004. “The cohort effect: insights and explanations”. British Actuarial Journal, 10 (4), pp 833-877.

Wilmoth, J.R., K.F. Andreev, D. Jdanov, and D.A. Glei. 2007. "Methods protocol for the Human Mortality Database. Version 5. Available at http://www.mortality.org/Public/Docs/MethodsProtocol.pdf."

Wong-Fupuy, C. and Haberman, S. 2004. “Projecting mortality trends: Recent developments in the UK and the US”. *North American Actuarial Journal*, 8: pp 56-83.

Yue, J. C., Yang, S.S., and Huang, H-C. 2008. “A study of the Lee-Carter model with a jump”. Living to 100 Symposium January 7-9, 2008. Featured Papers.

Zou, H., Hastie, T., and Tibshirani, R. 2006. “Sparse Principal Component Analysis”. Journal of Computational and Graphical Statistics, 15 (2), pp 265-286.