View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by City Research Online

Hatzopoulos, P. & Haberman, S. (2011). A dynamic parameterization modeling for age-period-cohort mortality. Insurance: Mathematics and Economics, 49(2), pp. 155-174. doi: 10.1016/j.insmatheco.2011.02.007

CITY UNIVERSITY LONDON

EST 1894

City Research Online

Original citation: Hatzopoulos, P. & Haberman, S. (2011). A dynamic parameterization modeling for age-period-cohort mortality. Insurance: Mathematics and Economics, 49(2), pp. 155-174. doi: 10.1016/j.insmatheco.2011.02.007

Permanent City Research Online URL: http://openaccess.city.ac.uk/4070/

Copyright & reuse

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

Versions of research

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

Enquiries

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at publications@city.ac.uk.

A dynamic parameterization modeling for the age-period-cohort mortality

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

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

Abstract

An extended version of Hatzopoulos and Haberman (2009) dynamic parametric model is proposed for analyzing mortality structures, incorporating the cohort effect. A one-factor parameterized exponential polynomial in age effects within the generalized linear models (GLM) framework is used. Sparse principal component analysis (SPCA) is then applied to time dependent GLM parameter estimates and provides (marginal) estimates for a two-factor principal component (PC) approach structure. Modeling the two-factor residuals in the same way, in age-cohort effects, provides estimates for the (conditional) three-factor age-period-cohort model. The age-time and cohort related components are extrapolated using dynamic linear regression (DLR) models. An application is presented for England & Wales males (1841-2006).

Keywords: Cohort; Mortality forecasting; Generalized Linear Models; Sparse Principal Component analysis; Factor analysis; Dynamic Linear Regression; Bootstrap confidence intervals.

1. Introduction

Many mortality models have been proposed since the Gompertz law of mortality in 1825. Recent developments in mortality modeling have tended to be extrapolative in nature, with the principal components (PC) approach being prominent. Thus, Bell and Monsell (1991) extended the Ledermann and Breas (1959) approach, where a PC approach it is used in forecasting age-specific mortality rates. In a seminal paper, Lee & Carter (1992) explored a modified version of this approach for forecasting mortality rates. The main statistical tool used was least-squares estimation via singular value decomposition (SVD) of the matrix of the log age specific observed forces of mortality.

As has been mentioned by many authors, several weaknesses arise in connection with Lee & Carter (LC) method (see Pitacco, 2004). Improvements to the LC model occur when the model is adjusted by fitting a Poisson regression model to the number of deaths at each age (Brillinger 1986; Brouhns et al 2002). Renshaw and Haberman (2003a) incorporate age differential effects, introducing a double bilinear predictor structure into the LC forecasting methodology, and optimize the Poisson likelihood, as opposed to optimizing the Gaussian likelihood, as under the LC approach, and then compare the results. Also, Hyndman and Ullah (2005) use several PCs in order to capture the differential movements in age-specific mortality rates. They smooth first the observed log-mortality rates with constrained and weighted penalized regression splines and they decompose the fitted curves using functional PCA. Recently, many authors have proposed new approaches to mortality forecasts, utilizing (nonparametric) smoothing. Thus, Currie et al (2004) use bivariate penalized B-splines to smooth the mortality surface in both the time and age dimensions within a penalized GLM framework. Hyndman and Ullah (2005) smooth the observed log-mortality rates with constrained and weighted penalized regression splines. De Jong and Tickle (2006) introduce a state space framework using B-spline smoothing. Gao and Hu (2009) introduce a Generalized Dynamic Factor method and multivariate BEKK GARCH model to describe mortality dynamics under conditional heteroskedasticity. Lazar and Denuit (2009) utilize dynamic factor analysis and the methodology of Johansen cointegration to project mortality through a linear state space representation which assumes that common factors can be modelled as a multivariate random walk with drift.

Further, in many developed countries (including UK, USA, Japan and Germany), there is evidence of a cohort effect – thus, in the UK, generations born between 1925 and 1945 approximately seem to have experienced more rapid mortality decreases than earlier or later generations. Renshaw and Haberman (2006) incorporate this effect by developing an age-period-cohort version of the LC model which provides an improved fit to the data

compared to the basic LC model.

Under our proposed approach, the cohort dependent PCs are modelled as random walk time series since, conditional on age-time effects (according to Hatzopoulos and Haberman (2009) method), the cohort PCs are mean reverting stochastic processes.

The remainder of the paper is organised as follows. In section 2, we analyse the methodology proposed for the GLM and SPCA approach, for the extraction of the time and cohort related PCs, and we define the particular class of DLR models which we utilize for forecast purposes. In section 3, we illustrate an application based on the recent England &Wales males' mortality experience. In section 4, we discuss the results and provide some concluding remarks. Finally, in Appendix A, we implement a closing-out procedure for the highest ages based on a dynamic variant of the Gompertz law.

2. Methodology

2.1 Using Sparse Principal Component Analysis

In Hatzopoulos and Haberman (2009), a new parametric method for modeling the age-time mortality effects was introduced. According to this GLM approach, the optimum degree k for the orthonormal polynomials (see equation 2.1, the *age-period full model,* Hatzopoulos and Haberman (2009)) is now determined by maximizing the Bayesian (or Schwarz) Information Criterion (BIC) : $BIC = -2 \cdot log(l) + (n \cdot k) \cdot log(n \cdot a)$, where, l is the (maximum) value of the total likelihood for all the n-calendar years and all the a-ages, with $n \cdot a$ observations and $n \cdot k$ parameters. The above method produces a random (asymptotically normal) matrix of estimated parameters, $\mathbf{B}$, of order n by k. In an analogous way with the SVD approach, we apply SVD (or equivalently PC analysis) to the matrix $\mathbf{B}$ to extract the mortality dynamics in age-time effects (see Hatzopoulos and Haberman (2009)).

PC analysis is a commonly dimension reduction technique to detect possible structures in the relationship between variables, particularly by reducing the dimensionality of the data and computing a few dominant eigenvectors of the data's covariance matrix. It seeks linear combinations of the data variables (often called factors or principal components) that capture a maximum amount of variance. 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 addition, if the input data are observed over time in a cross-sectional manner, each column of the data matrix is a time series and the temporal dependence in the data is summarized in the diagonal (variances) and off-diagonal (cross-covariances) elements of the variance-covariance matrix. 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). PC analysis usually combines together into the same component variables with a similar pattern, with similar loadings indicating the equal importance of the variables. Hence, if the variables have a similar variance pattern, this will be taken as similarity in the importance of the variables, and this often leads to the first few components "averaging" the variables, and hence a failure to achieve dimension-reduction. If the input data consist of non-stationary time series, a single linear combination of all the time series can explain the variability existing in the input data and component loadings for all input variables will be similar if not all equal (Lansangan and Barrios, 2009). In such cases, having a few non-zero coefficients in the principal components would greatly improve the relevance and interpretability of the factors. In sparse PC analysis, we seek a trade-off between the two goals of expressive 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).

Moreover, in studies with various mortality experiences, using a common PCA (or SVD applied to the matrix B), results in factor loadings structures $g_i(x)$ (see Model 1, Hatzopoulos and Haberman (2009)) with

significant high values for both positive and negative loadings. This undesirable feature gives high (negative) interdependence structure for different range of ages, explaining different mortality dynamics with the same factor, resulting in spurious interpretations and forecasting (see section 3). SPCA improves this problematic dependent structure by giving a better clustering of significant high factor loading values.

Given a covariance matrix $A \in S_k$, the problem of finding a sparse factors which explains the maximum amount of variance in the data can be written as follows:

$$\text{Maximize } Tr(A \cdot X) \text{ subject to } Tr(X) = 1,\ 1' \cdot |X| \cdot 1 \le k,\ X \ge 0,\ X \in S_k$$

using semidefinite relaxation techniques to compute approximate solutions (Luss and Aspremont, 2006). The covariance matrix X is the solution of the problem, in which the eigenvectors P denote now matrix with the eigenvectors resulting from the singular value decomposition of X (with associated vector of eigenvalues Λ). Luss and Aspremont (2006) solve a penalized formulation of the problem:

$$\text{Maximize } Tr(A \cdot X) - s \cdot 1' \cdot |X| \cdot 1 \text{ subject to } Tr(X) = 1,\ X \ge 0,\ X \in S_k$$

where s is a scalar which defines the sparsity (higher values of s gives more sparsity).

Thus, sparseness can be attained in constructing principal components of non-stationary time series by imposing constraints on the estimation of the component loadings and dimension-reduction and the search for common patterns among non-stationary time series can be achieved simultaneously. Using simulation techniques, Lansangan and Barrios (2009), show that SPCA (sparse principal component analysis) can achieve sparseness while consistently recognizing the variance patterns among non stationary time series.

In the presence of cohort effects, the number of optimum factors retained for representing the time effect is very important for the identification of the cohort effects. If we choose fewer factors than the optimum number, then these disregarded time effects will be carried over as cohort effects, possible in a non-stationary manner, or, if we choose more factors, then the cohort structure will break down. Thus, keeping an optimum subset *p(<k)* of the SPCs (see below), which explains the "majority" of the variance, leads to (see Model 1, Hatzopoulos and Haberman (2009)) the *age-period model*:

$$\log(\hat{m}_x(t)) = A(x) + \sum_{i=1}^{p} g_i(x) \cdot Y_i(t) + \varepsilon_x(t) \tag{1}$$

Under a common factor analysis, $g_i(x)$ are the factor loadings and denote the covariance between the *i*-factor for age *x* and $Y_i(t)$ denotes the value of the *i*-common factor for calendar year *t*.

In order to implement the above approach, we give a method to identify the optimum number of SPCs factors retained (*p*-value on the age-period model) as well a method to identify the scalar which defines the sparsity in the SPCA (*s*-value). Different choices of *s*-values can define different age-time dynamics (and different "optimum" *p*-values).

The log-graduated central mortality rates, in the age-period model, can be alternatively decomposed (see Model 2, Hatzopoulos and Haberman (2009)) as an *age-period association model*:

$$\log(\hat{m}_x(t)) = \mu + \alpha(x) + b(t) + \sum_{i=1}^{p} f_i(x) \cdot Y_i(t) + \varepsilon_x(t) \tag{2}$$

where $b(t) = \sum_{i=1}^{p} m_i \cdot Y_i(t)$ for $m_i = L_0 \cdot e_{1,i}$. We note that the factor loadings $g_i(x)$ can be decomposed as $g_i(x) = m_i + \sum_{j=2}^{k} L_{j-1}(x) \cdot e_{j,i} = m_i + f_i(x)$, and the m_i values then can be viewed as *mean* indices of the factor loadings (or associations) for all the ages with the *i*-factor. The m_i values are also measures of the importance (weights) for the *i*-factor in the construction of the $b(t)$ term.

Confidence intervals for these m_i values can be constructed by bootstrapping techniques. Starting from the estimated GLM *k*-dimensional random vector $\underline{\beta}_t$, at time *t,* which is (asymptotically) normally distributed with variance-covariance matrix $Var(\underline{\beta}_t)$, as has been defined in section 2.1, we simulate *N* bootstrap samples $\{\underline{\beta}_t^{(i)}\}$, *i=1,2,…,N*, where $\underline{\beta}_t^{(i)}$ are realizations from the multivariate *k*-dimensional Normal distribution with parameters $(\underline{\beta}_t, Var(\underline{\beta}_t))$, for each calendar year *t*. For each bootstrap sample, the associated SPCA gives bootstrap confidence intervals for the m_i values. If 0 belongs to these confidence intervals, then we can assume that the associated components are not sufficiently important to be included in the mortality structure. An alternative method could also be used to identify the number of factors retained, if we just take the maximum of the positive or negative m_i values and express them as percentage on the *N* bootstrap sample, defining a confidence coefficient or a confidence level (CL) for each PC. If the m_i values are near all positive (or negative) (i.e. if $\mathrm{CL} \simeq 100\%$) in the bootstrap procedure, it is an indication of robustness and the significance of the particular factor. The CL gets values in the interval [50%, 100%].

We can choose the *'optimum' value s* (the scalar which defines the sparsity in the SPCA), to be the value that *maximizes the number of significant SPC, as defined by the CL criterion* (at a given significance level). Among these significant SPC (according to the CL-criterion), we can specify the key age groups by taking, for each age, the maximum association value (MAV), defined by: $g^{\max}(x) = \max_{\{i\}}\{g_i(x)\}$. It is desirable that the above method should produce distinct key age groups of $g^{\max}(x)$ neighbouring values, which belong to the same components. This desirable feature would lead to a clear picture for the mortality dynamics.

Utilizing the above bootstrapping technique we can finally obtain a *bootstrap CI* for the $g_i(x)$ values. Similar arguments could also be applied to the cohort effects for the derivation of the number of optimum cohort related factors, as will be explained below. The difference between the method used to derive the period retained factors and the cohort retained factors is that, in the case of cohort-age modelling the factor dynamics are assumed to be stationary and also we are not interested in a fine clustering of $g_i(x)$ values. Since cohort effects normally concern a wide range of relevant ages, we incorporate PCA techniques to capture the mortality dynamics in cohort manner, taking as an indicator of the number of retained PC, the associated variance explained.

2.2 Incorporation of Cohort Effects

In the presence of cohort effects, the GLM estimates $\underline{\beta}_t$ are marginal estimates of age and period effects. If we consider the residual cohort effect conditional on the already estimated age and period effects, then we have:

$$\log(E(D_{x,t})) = \log(R_{x,t} \cdot m_x(t)) = \log(R_{x,t}) + A(x) + \sum_{i=1}^{p} g_i(x) \cdot Y_i(t) + \sum_{j=1}^{k_2} b_{j-1}(c) \cdot L_{j-1}(x)$$

in which the $\log(R_{x,t}) + A(x) + \sum_{i=1}^{p} g_i(x) \cdot Y_i(t)$ term is treated as an offset, in the rectangle DEBF (Figure 1). The GLM estimates $\{\beta_{j-1}(c)\}$ of $\{b_{j-1}(c)\}$, for $c = c_1, \ldots, c_{nc}$ and *j=1,…,k*$_2$, can be estimated in a Poisson model with the fitted values from the age – period effects model treated as an offset. These estimates are conditional estimates of the cohort effect. In taking this approach to modelling the cohort effects, we follow the general guidelines adopted by Cairns et al (2008, 2009).

In the same way with the age period effects, we apply the eigenvalue decomposition to the associated covariance matrix of these conditional estimates (using ordinary PC analysis since the cohort PCs are

considered to follow a mean reverting process), leading to graduated central mortality rates in age – period – cohort effects, the *age-period-cohort full model*:

$$\log(\hat{m}_x(t)) = A'(x) + \sum_{i=1}^{p} g_i(x) \cdot Y_i(t) + \sum_{i=1}^{k_2} D_i(x) \cdot H_i(t-x)$$

and keeping a subset $q(< k_2)$ of the PCs, leads to an *age-period-cohort model*:

$$\log(\hat{m}_x(t)) = A'(x) + \sum_{i=1}^{p} g_i(x) \cdot Y_i(t) + \sum_{i=1}^{q} D_i(x) \cdot H_i(t-x) + \varepsilon_x(t-x) \tag{3}$$

defined in the parallelogram DEBF (Figure 1). Then $A'(x) = A(x) + A^c(x)$ if $\underline{A}^c = \mathbf{L}_c \cdot \underline{\bar{\beta}}_c$ is the vector of $A^c(x)$ age scores which represents the residual adjustment for the main age profile $A(x)$ after including the cohort effects. $\mathbf{L}_c = \{L_{j-1}(x)\}$ is the design matrix in cohort effects of order a by k_2, and $\underline{\bar{\beta}}_c$ denotes a k_2-dimentional vector of mean values for the columns in the matrix of the estimated parameters $\{\beta_{j-1}(c)\}$. $D_i(x) = \underline{L}_x^c \cdot \underline{e}_i^c$ if $\underline{L}_x^c$ denotes the x-row of the cohort design matrix $\mathbf{L}_c$ and $\underline{e}_i^c$ denotes the i-eigenvector from the cohort PCA. $H_i(c) = \underline{\beta}_c^r \cdot \underline{e}_i^c$ if $\underline{\beta}_c^r$ is the rescaled vector: $\underline{\beta}_c^r = \underline{\beta}_c - \underline{\bar{\beta}}_c$ and $\underline{\beta}_c$ is a k_2-dimensional random vector which denotes the PC scores from the cohort PCA. λ_i^c is the related i-eigenvalue, and $\varepsilon_x(t-x) = \sum_{i=q+1}^{k_2} D_i(x) \cdot H_i(t-x)$ is the residual unexplained variance from the cohort PCA.

If $\underline{\varepsilon}_c = \left[\varepsilon_{c,x_1}, \ldots, \varepsilon_{c,x_a}\right]'$ denotes the associated vector of errors for $c = c_1, \ldots, c_{nc}$ then the random vectors $\underline{\varepsilon}_c$ are asymptotically normally distributed with zero mean and variance-covariance matrix $\mathbf{V}_c$, estimated by $\hat{\mathbf{V}}_c = {}^{tr}\mathbf{H} \cdot {}^{tr}\mathbf{P}_c{}' \cdot Var(\underline{\beta}_c) \cdot {}^{tr}\mathbf{P}_c \cdot {}^{tr}\mathbf{H}'$, where ${}^{tr}\mathbf{P}_c = [\underline{e}_{q+1}^c, \cdots, \underline{e}_{k_2}^c]$ denotes the matrix of the last k_2-q eigenvectors and ${}^{tr}\mathbf{H} = [\underline{H}_{q+1}, \cdots, \underline{H}_{k_2}]$ denotes the matrix of the last k_2-q cohort PCs.

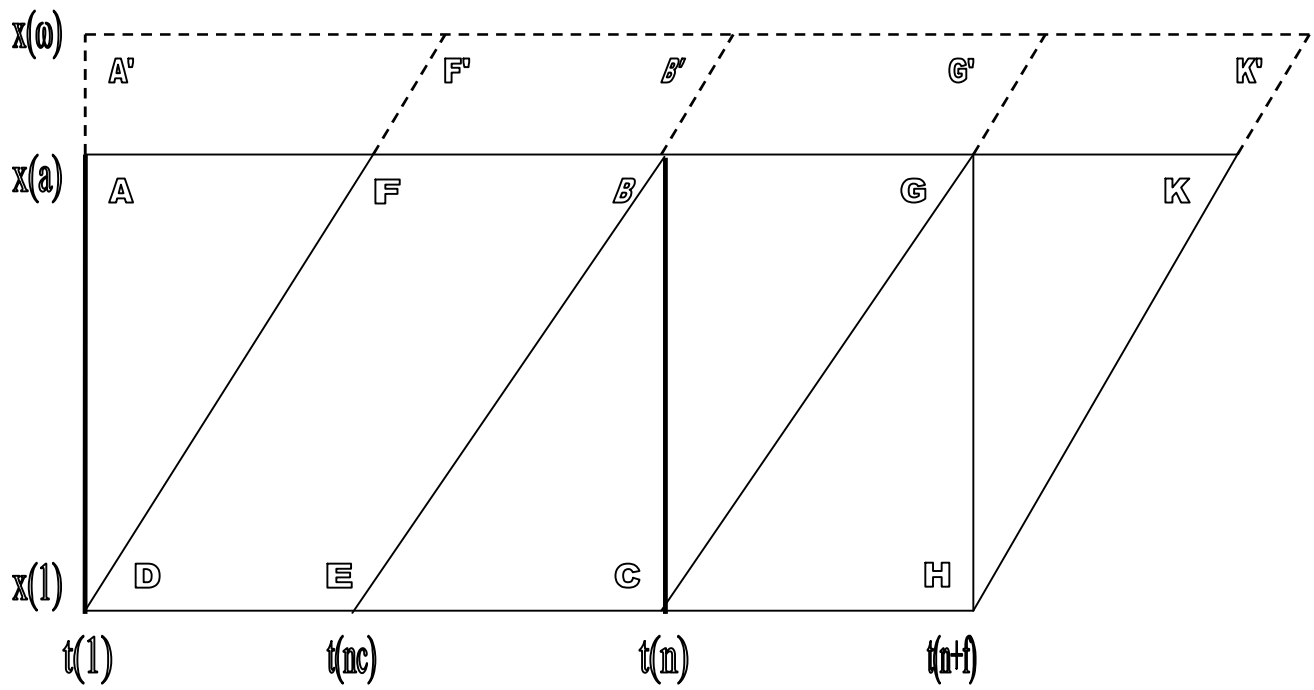


Figure 1: Diagram of observed and extrapolated data.

In the method described above for the estimation of the conditional cohort effects, we have not utilized the observed mortality data in the triangle CEB in terms of the cohort effects (Figure 1). The GLM graduation in cohort effects needs to have observations for the full age range (from ages x_1 to x_a). Any forecast method applied to the $Y_i\ (t)$ or $H_i\ (c)$ stochastic vectors will not use these observed mortality rates. In order to take into account these observed mortality rates and to obtain $\underline{\beta}_c$ estimates for all of the *n*-cohorts, we need to move from considering the data in the parallelogram DEBF to data in the parallelogram DCGF, and so we propose the introduction of the following method:

1) obtain forecast $\hat{Y}_i\ (t_{n+k})$ values in age-time effects (from rectangle ABCD), for the next *f=a-1* calendar years (rectangular BCHG), and obtain period forecast values for the central mortality rates, for *k=1,…,f* and *i=1,…,p*

2) obtain forecast $\hat{H}_i\ (c_{nc+k})$ values in age-cohort effects (from parallelogram DEBF) for the next *f* cohorts, for *k=1,…,f* and *i=1,…,q* (from parallelogram ECGB) and in combination with the forecast values from step 1, obtain forecast central mortality rates, say $\hat{m}_x(t_{n+k})$ for *k=1,…,f*, in the triangle BCG, taking account both of time and cohort effects

3) use the observed death counts, d_{x,t_n}, from the last calendar year *n*, to obtain the 'observed' values for the central exposures in the triangle BCG: $\hat{R}_{x,t_{n+k}} = \dfrac{d_{x,t_n}}{\hat{\hat{m}}_x(t_{n+k})}$ for *k=1,…,f*

4) fit the death counts $\hat{d}_{x,t}$ under the age-period-cohort full model in cohort effects, in the parallelogram DCGF, where the $\log(\hat{\hat{R}}_{x,t}) + A(x) + \sum_{i=1}^{p} g_i(x)\cdot\hat{\hat{Y}}_i\ (t)$ term is treated as an offset, for $t=t_1,\ldots,t_{n+f}$, where $\hat{\hat{R}}_{x,t} = R_{x,t}$ are the observed exposures, $\hat{\hat{Y}}_i\ (t) = Y_i\ (t)$ the GLM estimated PCs, and $\hat{d}_{x,t} = d_{x,t}$ the observed death counts, if $t=t_1,\ldots,t_n$, (area DCBF), and $\hat{\hat{R}}_{x,t_{n+k}} = \hat{R}_{x,t_{n+k}}$ $\hat{\hat{Y}}_i\ (t_{n+k}) = \hat{Y}_i\ (t_{n+k})$ and $\hat{d}_{x,t_{n+k}} = d_{x,t_n}$, if *k=1,…,f* (triangle BCG), in order to obtain GLM cohort estimates $\{\beta_{j-1}(c)\}$ of $\{b_{j-1}(c)\}$, for $c=t-x=c_1,\ldots,c_n$ and *j=1,…,k*$_3$, and also cohort PCs $\hat{H}_i(c)$ estimates, for $c=t-x=c_1,\ldots,c_n$ and *i=1,…,q* to define the *age-period-cohort expanded full model*:

$$\log(\hat{m}_x(t)) = A'(x) + \sum_{i=1}^{p} g_i(x)\cdot Y_i\ (t) + \sum_{i=1}^{k_3} D_i(x)\cdot\hat{H}_i(t-x)$$

or the equivalent *age-period-cohort expanded model*:

$$\log(\hat{m}_x(t)) = A'(x) + \sum_{i=1}^{p} g_i(x)\cdot Y_i\ (t) + \sum_{i=1}^{q} D_i(x)\cdot\hat{H}_i(t-x) + \varepsilon_x(t-x) \tag{3a}$$

for time effects $t=t_1,\ldots,t_n$, age effects $x=x_1,\ldots,x_a$, and cohort effects $t-x=c=c_1,\ldots,c_n$ (parallelogram DCGF).

The residuals are given by $\varepsilon_x(t-x) = \sum_{i=q+1}^{k_3} D_i(x)\cdot\hat{H}_i(t-x)$, and we assume that the random variables $\varepsilon_x(t-x)$ are normal with zero mean and variance $V_{x,t-x}$, which is estimated by $\hat{V}_{x,t-x} = \sum_{i=q+1}^{k_3} D_i^2(x)\cdot Var\left(\hat{H}_i(t-x)\right)$.

Under this approach, we obtain forecast estimates in the triangle BCG that utilize all of the observed mortality data. This approach can be viewed as being related to credibility theory, if we note that the cohort ($\underline{\beta}_c$) GLM conditional estimates, for the last *f* cohorts, are estimated by combining the observed mortality rates (triangle ECB) with the 'prior' ($\hat{m}_x(t_{n+k})$) forecast estimates (triangle BCG), giving different weights (credibility factor) to each cohort. Also, by using the death counts, $d_{x,n}$, from the last calendar year *n*, to obtain values for the central exposures in step 3, we utilize the distribution structure of deaths in age effects from the last calendar year. This approach can be justified by the well reported "rectangularization" feature shown by many recent mortality experiences (i.e. the increasing concentration of the probability density of the random lifetime function around the mode) and the importance of the information content of the latest data point.

Using similar arguments with the age-period association model (model structure 2), the log-graduated central mortality rates, can be alternatively decomposed as an *age-period-cohort association (expanded) model*:

$$\log(\hat{m}_x(t))=\mu'+\alpha'(x)+b(t)+c(t-x)+\sum_{i=1}^{p} f_i(x)\cdot \mathrm{Y}_i(t)+\sum_{i=1}^{q} d_i^c(x)\cdot \hat{H}_i(t-x)+\varepsilon_x(t-x) \qquad (4)$$

where $c(t-x)$ denotes the independent main cohort effect: $c(t-x)=\sum_{i=1}^{q} m_i^c\cdot \hat{H}_i(t-x)$ for $m_i^c = L_0^c\cdot e_{1,i}^c$,

$\mu'=\mu+\mu^c$ with $\mu^c=\bar{\beta}_0^c\cdot L_0^c$, $\alpha'(x)=\alpha(x)+\alpha^c(x)$ with $\alpha(x)=\sum_{j=2}^{k}\bar{\beta}_{j-1}^c\cdot L_{j-1}(x)$, and

$d_i^c(x)=\sum_{j=2}^{k} e_{j,i}^c\cdot L_{j-1}^c(x)$.

2.3 Forecasting

2.3.1 Dynamic Linear Regression models

Based on the *age-period-cohort expanded model* (model structure 3a), forecast estimates $\hat{Y}_i\ (t)$ and $\hat{\hat{H}}_i(t-x)$ are needed for the time and cohort dynamics: $Y_i\ (t)$ and $\hat{H}_i(t-x)$ respectively. For forecast purposes, we note the importance of linear and log-linear extrapolation structures (see, for example Sithole et al 2000; Renshaw and Haberman 2003b; Pitacco et al 2009). With this as background, we extend the conventional linear modeling approaches of regression analysis and low order ARIMA models that have been widely used and advocate, for each time related PC, a *specific class* of *dynamic linear regression* (DLR) models:

$$Y_i\ (t)=\alpha_i+b_{i,t}\cdot t+e_{i,t}$$

for each calendar year *t* (the so-called regressor) and for each time related PC *i=1,…,p*, with the slope being a stochastic time variable parameter that follows a random walk process: $b_{i,t}=b_{i,t-1}+\zeta_{i,t-1}$. The innovations $e_{i,t}$ and $\zeta_{i,t}$ are assumed to be white noises random variables.

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).

Experiments with various mortality experiences have shown that the time related non-stationary PCs can be represented adequately under this particular DLR model structure. The crucial feature of the model is the signal-noise ratio, or the so-called the noise-variance ratio hyper-parameters (NVR): $\sigma_{i,t}^2 = \frac{\sigma_{\zeta_{i,t}}^2}{\sigma_{e_{i,t}}^2}$. If necessary, a pre-determined NVR value can be used to control the required level of smoothness: thus, small values enhance the smoothness. The computations have been implemented in Matlab using the Captain Toolbox (Taylor, 2007).

In contrast, the cohort related PCs are modelled as independent RW plus noise time series model since, conditional on the age-time effects, we assume that the cohort related PCs are mean reverting (stationary) stochastic processes:

$$\hat{H}_i(c) = b_{i,c} + e_{i,c} \qquad \text{and} \qquad \hat{\hat{H}}_i(c) = b'_{i,c} + e'_{i,c}$$

where $b_{i,c} = b_{i,c-1} + \zeta_{i,c-1}$ and $b'_{i,c} = b'_{i,c-1} + \zeta'_{i,c-1}$, for each cohort $c=t-x$ and for each cohort related PC $i=1,\ldots,q$. As above, we employ state space models and the Kalman filter technique for providing a computationally efficient framework for carrying out the estimation.

We can define the rate of cohort-log-mortality improvement, for positive age related $D_i(x)$ values, due to the i-PC: $\Delta\hat{\hat{H}}_i(c) = 1 - \frac{\hat{\hat{H}}_i(c)}{\hat{\hat{H}}_i(c-1)}$ if $\hat{\hat{H}}_i(c)>0$, or $-\Delta\hat{\hat{H}}_i(c)$ otherwise. For negative age related $D_i(x)$ values we can define the opposite of the cohort-log-mortality improvement $\Delta\hat{H}_i(c)$. We can also define the overall rate of cohort-log-mortality improvement: $\Delta\hat{c}(c) = 1 - \frac{\hat{c}(c)}{\hat{c}(c-1)}$ if $\hat{c}(c)>0$, or $-\Delta\hat{c}(c)$ otherwise, where $c=t-x$ is the cohort index. These rates of improvement can be used to describe the dynamics in age-cohort effects, in connection with the sign of the associated PC values (or the sign of values from the main cohort trend) and the sign of the age related values.

The above modelling gives the *age-period-cohort forecast model*:

$$\log(\hat{\hat{m}}_x(t)) = A'(x) + \sum_{i=1}^{p} g_i(x)\cdot\hat{Y}_i\ (t) + \sum_{i=1}^{q} D_i(x)\cdot\hat{\hat{H}}_i(t-x) \tag{5}$$

(where $A'(x) = \mu' + \alpha'(x) = A(x) + A^c(x)$) for time effects $t=t_1,\ldots,t_{n+f}$, age effects $x=x_1,\ldots,x_a$, and cohort effects $c=t-x=c_1,\ldots,c_{n+f}$ (parallelogram DHKF).

2.3.2 Confidence Intervals

The combination of the age-period-cohort forecast model (model structure 5) with the age-period-cohort complete model (see Appendix, model structure A1) give the *age-period-cohort forecast-complete model:*

$$\log(\hat{\hat{m}}_x(t)) = A'(x) + \sum_{i=1}^{p} \hat{g}_i(x)\cdot\hat{Y}_i\ (t) + \sum_{i=1}^{q} \hat{D}_i(x)\cdot\hat{\hat{H}}_i(t-x) \tag{6}$$

There are several sources of randomness in the above modelling. The *true value* of $\log(m_x(t))$, assuming the model specification is correct, is given by

$$\log(m_x(t)) = \left(\hat{A}'(x) + e\left(A'(x)\right)\right) + \sum_{i=1}^{p}\left(\hat{g}_i(x) + e\left(g_i(x)\right)\right)\cdot\left(\hat{Y}_i\ (t) + e\left(Y_i\ (t)\right)\right) +$$
$$\sum_{i=1}^{q}\left(\hat{D}_i(x) + e\left(D_i(x)\right)\right)\cdot\left(\hat{\hat{H}}_i(t-x) + e\left(\hat{H}_i(t-x)\right)\right) + \varepsilon_x(t-x) + \mathrm{e}_x(t-x)$$

where $e(\ )$ denotes the errors in estimating the relevant quantity, $\varepsilon_x(t-x)$ is the residual error in fitting the model for age *x* and cohort *t-x* using a finite set of PC functions, and $\mathrm{e}_x(t-x)$ is the observation error which comes from the random variation of deaths in the Poisson distribution.

The *total error*, $E_{x,t-x}$, is the difference between the true value of $\log(m_x(t))$ and expression (6):

$$E_{x,t-x} = e\left(A'(x)\right) + \sum_{i=1}^{p}\left(\hat{g}_i(x) + e\left(g_i(x)\right)\right)\cdot e\left(Y_i\ (t)\right) + \sum_{i=1}^{p} e\left(g_i(x)\right)\cdot\hat{Y}_i\ (t) +$$
$$\sum_{i=1}^{q}\left(\hat{D}_i(x) + e\left(D_i(x)\right)\right)\cdot e\left(\hat{H}_i(t-x)\right) + \sum_{i=1}^{q} e\left(D_i(x)\right)\cdot\hat{\hat{H}}_i(t-x) + \varepsilon_x(t-x) + \mathrm{e}_x(t-x)$$

Assuming independence among all of those different sources of error, the *total variance* of $E_{x,t-x}$ is given by

$$\sigma^2_{E_{x,t-x}} = \sigma^2_{e(A'(x))} + \sum_{i=1}^{p}\left(\hat{g}_i^2(x) + \sigma^2_{e(g_i(x))}\right)\cdot\sigma^2_{e\left(Y_i\ (t)\right)} + \sum_{i=1}^{p}\sigma^2_{e(g_i(x))}\cdot\hat{Y}_i^2(t) +$$
$$\sum_{i=1}^{q}\left(\hat{D}_i^2(x) + \sigma^2_{e(D_i(x))}\right)\cdot\sigma^2_{e\left(\hat{H}_i(t-x)\right)} + \sum_{i=1}^{q}\sigma^2_{e(D_i(x))}\cdot\hat{\hat{H}}_i^2(t-x) + \sigma^2_{\varepsilon_x(t-x)} + \sigma^2_{\mathrm{e}_x(t-x)}$$

where $\sigma^2_{e(A'(x))}$ correspond to the variance of the error in estimating $A'(x)$ and can be estimated using the DLR model structure (see Appendix); $\sigma^2_{e(g_i(x))}$ and $\sigma^2_{e(D_i(x))}$ correspond to the variance of the error in estimating $g_i(x)$ and $D_i(x)$ respectively and can be estimated using the DLR model structure (see Appendix); $\sigma^2_{e\left(Y_i\ (t)\right)}$ denotes the variance of the error in estimating $Y_i\ (t)$ and can be estimated using the DLR model structure (see section 2.3); $\sigma^2_{e\left(\hat{H}_i(t-x)\right)}$ denotes the variance of the error in estimating $\hat{H}_i(t-x)$ and can be estimated using the DLR model structure (see section 2.3); $\sigma^2_{\varepsilon_x(t-x)}$ denotes the variance of the residual error in fitting the model using a finite set of PC functions and can be estimated by the 'residual sum' $\sum_{i=q+1}^{k_3}\hat{D}_i^2(x)\cdot Var\left(\hat{H}_i(t-x)\right)$; and $\sigma^2_{\mathrm{e}_x(t-x)}$ denotes the variance of the observation error in the Poisson distribution and can be estimated with the terms $\frac{\hat{\varphi}_{t-x}}{d_{x,t-x}}$, where φ_{t-x} is the over-dispersed parameter and can be estimated by the ratio of the deviance divided by the associated degrees of freedom for each cohort from model structure (3), and $d_{x,t-x}$ denote the observed number of deaths for each cohort *c=t-x=1,…,nc* and each age $x = x_1,...,x_a$. Outside of the cohort range, the $d_{x,t-x}$ can be substituted by the observed number of deaths from the last calendar year: $\hat{d}_{x,t-x} = d_{x,t_n-x}$ for cohorts *c=t-x=* $c_{nc+1},\ldots,c_{n+f}$, and outside the age range, the $d_{x,t-x}$ can be substituted by the observed number of deaths from the last observed age: $\hat{d}_{x,t-x} = d_{x_a,t-x}$ for ages $x = x_{a+1},...,x_\omega$.

In order to generate an interval forecast, assuming a normal distribution for the variable $E_{x,t-x}$, we can obtain the following *95% confidence interval* estimate for $m_x(t)$:

$$CI(m_x(t)) = \left(\hat{\hat{m}}_x(t)\cdot e^{-1.96\cdot\sigma_{E_{x,t-x}}}, \hat{\hat{m}}_x(t)\cdot e^{1.96\cdot\sigma_{E_{x,t-x}}}\right)$$

where $\hat{\hat{m}}_x(\mathrm{t})=\exp\left\{\hat{a}+\hat{b}_x\cdot x'+f_x+t\cdot f_{t,x}+f_{c,x}\right\}$ if $f_x=\sum_{i=1}^{p}\hat{b}_{i,x}^{p}\cdot\hat{a}_i$, $f_{t,x}=\sum_{i=1}^{p}\hat{b}_{i,x}^{p}\cdot\hat{b}_{i,t}$ and $f_{c,x}=\sum_{i=1}^{q}\hat{b}_{i,c}\cdot\hat{b}_{i,x}^{c}$.

2.3.3 Reduction Factors

From the age-period-cohort forecast-complete model (model structure 6), we obtain that $\log(\hat{\hat{m}}_x(t))=\hat{A}'(x)+\sum_{i=1}^{p}\hat{g}_i(x)\cdot\left(\hat{\alpha}_i+\hat{b}_{i,t}\cdot t\right)+\sum_{i=1}^{q}\hat{D}_i(x)\cdot\hat{b}'_{i,t-x}$. If we set the first observed year as the base year, i.e. year where *t=0*, and if we forecast backwards the RW time series $\hat{b}'_{i,0-x}$, which can be estimated by its first smoothed value: $\hat{b}'_{i,0-x}=\hat{b}'_{i,0}$, it follows that

$$\log(\hat{\hat{m}}_x(0))=\hat{A}'(x)+\sum_{i=1}^{p}\hat{g}_i(x)\cdot\hat{\alpha}_i+\sum_{i=1}^{q}\hat{D}_i(x)\cdot\hat{b}'_{i,0}$$

In this way, we can define a *reduction factor model in time effects*, as discussed by several authors including Renshaw and Haberman (2003c, 2006):

$$\hat{\hat{m}}_x(t)=\hat{\hat{m}}_x(0)\cdot RF_{x,t} \quad \text{if} \quad RF_{x,t}=\exp\left\{\left(\sum_{i=1}^{p}\hat{g}_i(x)\cdot\hat{b}_{i,t}\right)\cdot t+\sum_{i=1}^{q}\hat{D}_i(x)\cdot\left(\hat{b}'_{i,t-x}-\hat{b}'_{i,0}\right)\right\}$$

For projected cohorts, where $c=t\text{-}x>c_n$, the estimates $\hat{b}_{i,t}=\hat{b}_{i,t_n}$ and $\hat{b}'_{i,c}=\hat{b}'_{i,c_n}$ become constant, and the reduction factors are: $RF_{x,t}=\exp\left\{\left(\sum_{i=1}^{p}\hat{b}_{i,x}^{p}\cdot\hat{b}_{i,t_n}\right)\cdot t+\sum_{i=1}^{q}\left(\hat{b}'_{i,c_n}-\hat{b}'_{i,0}\right)\cdot\hat{b}_{i,x}^{c}\right\}$, i.e. on the logarithmic scale, a linear function in time effects, for each age. In addition, for projected ages, where $x>x_a$, the log-reduction factors become identical.

2.3.4 Calculation of Life Expectancy

Based on the *age-period-cohort forecast-complete model* (model structure 6) and in a similar way with Hatzopoulos and Haberman (2009), we can obtain estimates of the cohort-based expected remaining lifetime, $e_x(c)$, by constructing a life table for each cohort $c=t\text{-}x=c_1,\ldots,c_{n+f}$ and complete age range $x=x_1,\ldots,x_\omega$ (parallelogram $DHK'F'$, Figure 1), under the assumption of the constant force of mortality (CFM), for any age *x* and cohort *c*: $\mu_{x+z}(c)=\mu_x(c)$ for $0\le z<1$, and $\mu_x(c)=m_x(c)$.

Associated confidence intervals for the expected remaining lifetime can then be obtained using parametric bootstrapping. Starting from the estimated GLM *k*-dimensional random vector $\underline{\beta}_t$, for each calendar year *t,* with variance-covariance matrix $Var(\underline{\beta}_t)$, we simulate *N* bootstrap samples { $\underline{\beta}_t^{(i)}$ }, *i=1,2,…,N*, where $\underline{\beta}_t^{(i)}$ are realizations from the multivariate *k*-dimensional Normal distribution with parameters ($\underline{\beta}_t$,$Var(\underline{\beta}_t)$), for each calendar year *t*. For each bootstrap sample, we obtain estimates for model structure (6). This yields *N* realizations for the expected remaining lifetimes and then the *95% CIs for the expected remaining lifetimes* are determined by the percentiles, that is $CI_{95}=[\,p_{0.025},p_{0.975}\,]$, for each forecast cohort.

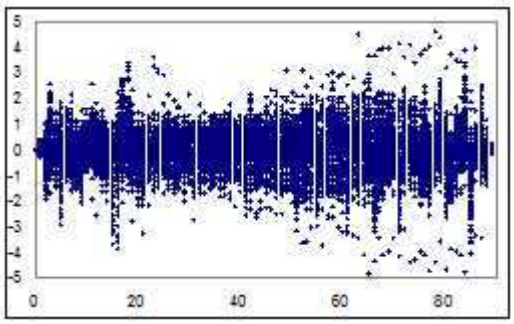
3. Applications

3.1 The data

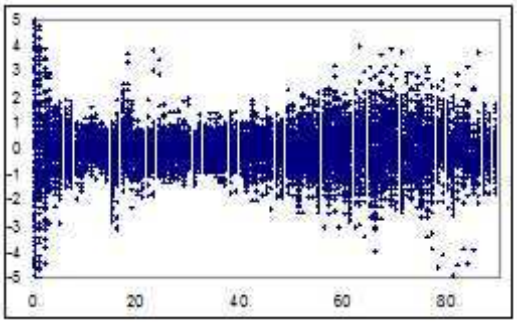
In order to illustrate the methodology, we present a case study based on England & Wales male total population mortality experience, for calendar years 1841-2006 and individual ages 0,1,…,89. The data are freely provided by the "Human Mortality Database" (www.mortality.org). We use ages up to age 89 because whilst data on numbers of deaths are generally available by single year of age above age 89, official population estimates are not available.

3.2 Residuals

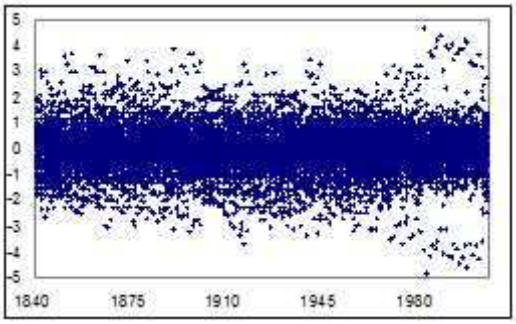
Figure 2, shows the standardized deviance residuals (SDR) plotted against age, time and cohort effects, for England & Wales for the period 1841-2006 and ages 0-89, under the *age-period model* (model structure 1). The left graphs (graphs 2.a1 2.b1 & 2.c1) correspond to the *age-period full model* (where we use all the SPCs, i.e. when $p=k$, with $k=25$, see Hatzopoulos and Haberman (2009): model structure 2.1) and the right graphs (graphs 2.a2 2.b2 & 2.c2) correspond to the age-period model (with $p=7$ SPCs). For the age-period full model, the overall patterns of the total SDR against age, time and cohort effects (graphs 2.a1 2.b1 & 2.c1) indicate an appropriate fit. We note two distinct outliers for the cohorts with years of birth 1919-1920 (graph 2.c1), leading to mostly positive SDRs for cohort 1920 and mostly negative SDRs for cohort 1919. This feature has been transferred into the age and time SDR (graph 2.a1 & 2.b1 respectively), which becomes apparent after age 60 and after the 1970s. For the age-period model, i.e. after keeping only the first 7 PCs, the SDR show clear patterns for the cohort effects, and this prompts us to investigate further the cohort effects.



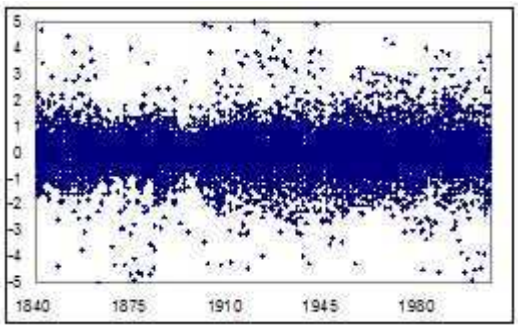
2.a1



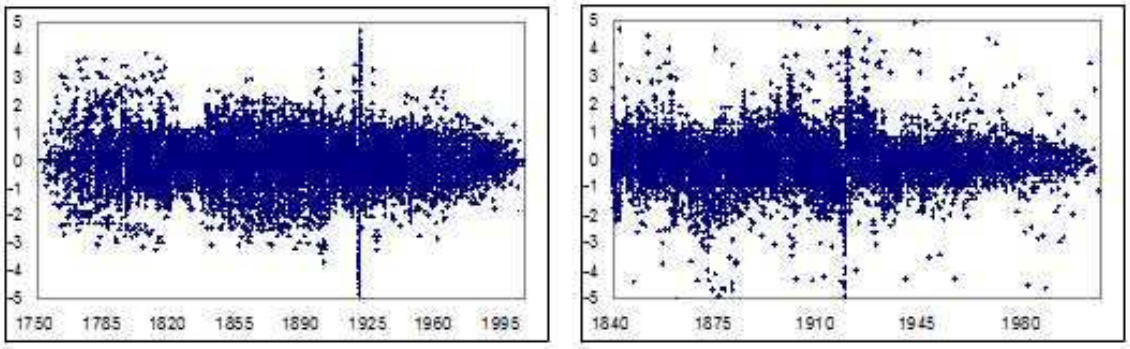
2.a2



2.b1



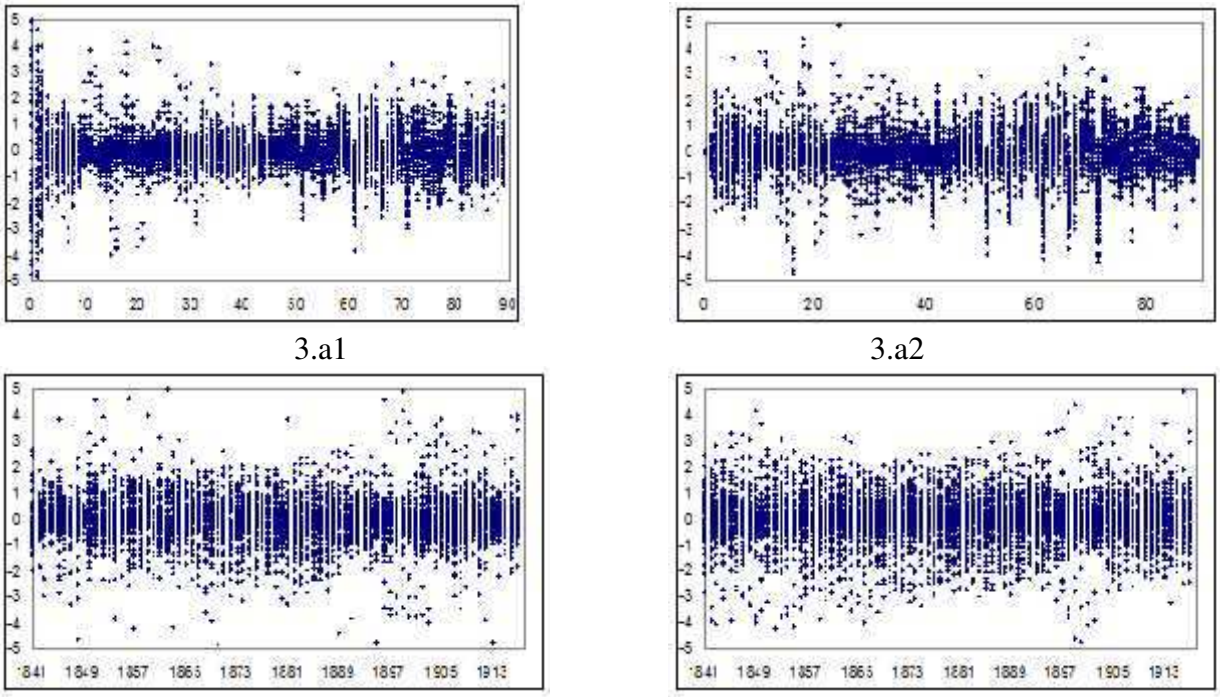
2.b2



2.c1 2.c2

Figure 2: SDR vs. age, time and cohort effects for the *age-period full model* (left) and the *age-period model* (right)

Figure 3, shows the standardized deviance residuals (SDR) plotted against age and cohort effects, under the *age-period-cohort model* (model structure 3), defined in the parallelogram DEBF (Figure 1). The left graphs (graphs 3.a1 & 3.b1) correspond to the *age-period-cohort full model* (where we use all the cohort PCs, i.e. when $q= k_2$, with k_2=20) and the right graphs (graphs 3.a2 & 3.b2) correspond to the age-period-cohort model (with q=2 cohort PCs). For the age-period-cohort full model, the overall patterns of the total SDR against age and cohort effects (graphs 3.a1 3.b1) indicate an appropriate fit. For the age-period-cohort model, the SDR do not show any age or cohort patterns (graphs 3.a2 & 3.b2).



3.a1 3.a2

3.b1 3.b2

Figure 3: SDR vs. age and cohort effects for the *age-period-cohort full model* (left) and the *age-period-cohort model* (right)

Figure 4, shows the standardized deviance residuals (SDR) plotted against age and cohort effects, under the age-period-cohort expanded model (model structure 3a), defined in the parallelogram DCGF (Figure 1). The left graphs (graphs 4.a1 & 4.b1) correspond to the *age-period-cohort expanded full model* (where we use all the cohort PCs, i.e. when $q= k_3$, with k_3=16) and the right graphs (graphs 4.a2 & 4.b2) correspond to the *age-period-cohort expanded model* (with q=3 cohort PCs). For the age-period-cohort expanded full model, the overall patterns of the total SDR against age and cohort effects (graphs 4.a1 4.b1) indicate an appropriate fit. For the age-period-cohort expanded model, the SDR do not show any age or cohort patterns (graphs 4.a2 & 4.b2).

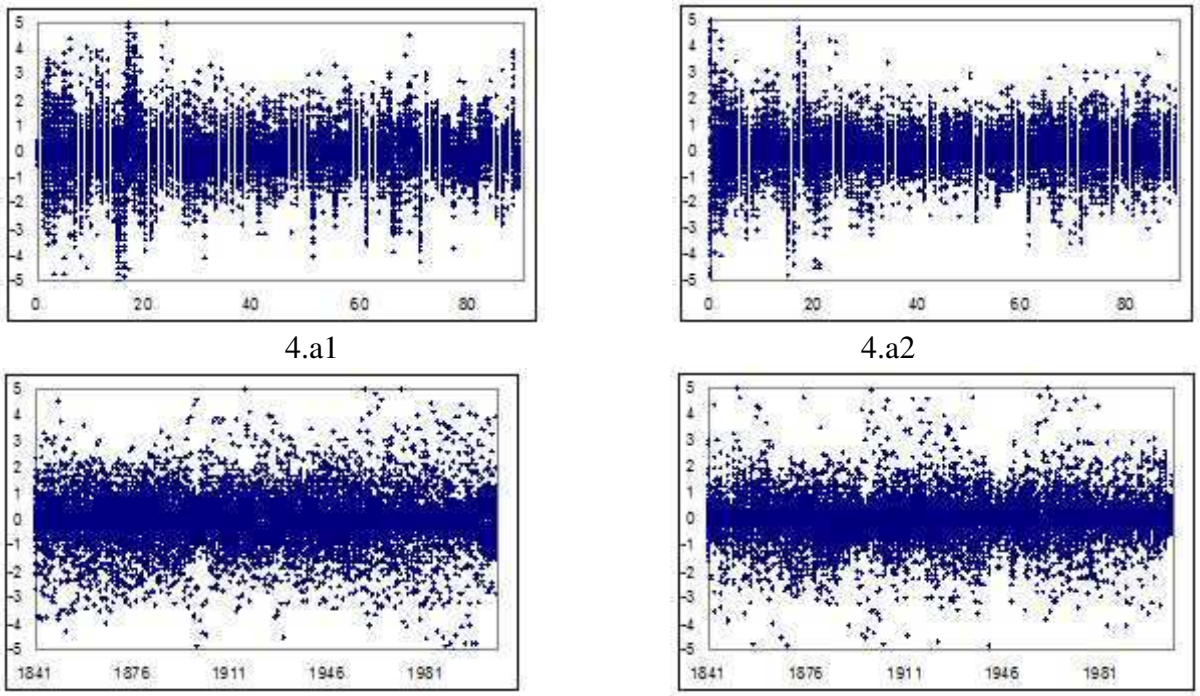


Figure 4: SDR vs. age and cohort effects for the *age-period-cohort expanded full model* (left) and the *age-period-cohort expanded model* (right)

3.4 Model Components and Forecasts

As in the *age-period model* (model structure 1), and the discussion of the construction of complete mortality tables in Appendix (model structure A1), the age main effects, $A(x)$, are modelled with a DLR model with age varying slope parameter: $A(x) = a + b_x \cdot x' + e_x$ where b_x follows an IRW model, and e_x is white noise. Figure 5, shows the evolution of the IRW parameter b_x and the main age effect. The inflection point, for this range of calendar years (1841-2006) is near the age x_{ip}=84. As expected, beyond age 89, the confidence intervals for b_x and $A(x)$ are very wide.

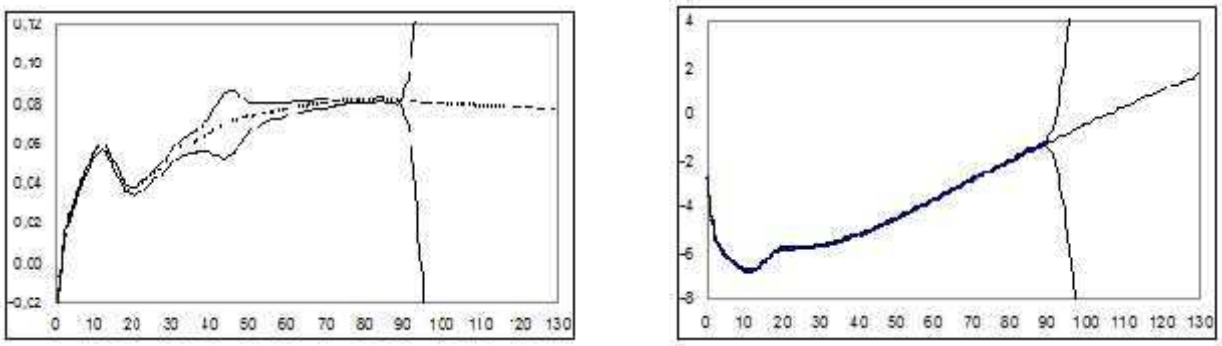


Figure 5: The IRW age varying slope parameter $\hat{b}_x$ (left) and the main age profile $\hat{A}(x)$ values (right), based on model structures (1) and (A1), with associated CIs.

Table 1, gives the eigenvalues (λ_i-values), in significant order, based on the *age-period model* (model structure 1), with the associated percentage variance explained (%var), *mean* association values m_i and associated (bootstrap) confidence level (CL) values (*N=10.000* bootstrap samples). According to the CL criterion there are seven significant SPCs (with minimum CL value 98,5%), which together explain the 100% of the total variance (with E-10 precision), with an 'optimum' sparse value *s=36.*

PC index	λ_i	% var	m_i	%CL
1	6,32E+01	8,71E-01	7,50E-02	99,9
2	8,25E+00	1,14E-01	7,33E-02	99,9
3	1,05E+00	1,45E-02	-1,06E-02	100,0
4	5,48E-02	7,55E-04	1,01E-03	100,0
5	6,01E-04	8,28E-06	-1,68E-04	99,9
6	2,77E-04	3,82E-06	-3,95E-04	98,5
7	7,39E-08	1,02E-09	-8,24E-06	99,9
25	5,49E-10	7,57E-12	-2,82E-06	56,0
8	1,36E-12	1,87E-14	1,59E-07	71,0
9	2,46E-14	3,39E-16	6,19E-09	67,8
10	5,23E-15	7,21E-17	-4,00E-10	54,0
11	3,04E-15	4,19E-17	-1,89E-10	51,3
12	4,95E-16	6,82E-18	5,51E-11	51,9
24	4,41E-16	6,07E-18	-3,38E-11	51,8
23	2,99E-16	4,13E-18	-2,17E-13	50,7
13	2,39E-16	3,29E-18	-5,12E-11	50,4
22	1,73E-16	2,38E-18	6,16E-12	50,5
21	1,08E-16	1,49E-18	-1,43E-11	53,9
14	6,04E-17	8,33E-19	-4,60E-12	51,2
15	3,81E-17	5,24E-19	-9,15E-12	50,2
20	3,16E-17	4,36E-19	1,01E-11	51,4
16	1,78E-17	2,45E-19	4,59E-12	51,3
19	7,18E-18	9,89E-20	6,97E-12	50,2
18	1,77E-18	2,43E-20	1,71E-12	51,0
17	9,07E-19	1,25E-20	-5,08E-12	51,8

Table 1: Eigenvalues based on the *age-period model* (model structure 1), with the associated percentage variance explained, m_i values and related bootstrap confidence levels (CL).

Table 2, gives the eigenvalues (λ_i^c-values), in significant order, based on the *age-period-cohort model* (model structure 3), with the associated percentage variance explained. According to the confidence level criterion (CL), there are two PCs, with minimum CL value 92,3%, which explain the 46% of the total residual variance.

PC index	λ_i^c	% var	m_i	%CL
1	3,18E-02	2,48E+01	5,84E-03	92,3
2	2,78E-02	2,16E+01	6,33E-02	99,1
3	1,39E-02	1,08E+01	-6,75E-03	50,3
4	1,15E-02	8,90E+00	-3,88E-02	53,2
5	9,16E-03	7,10E+00	4,18E-02	52,8
6	7,28E-03	5,70E+00	-2,49E-02	53,1
7	5,61E-03	4,40E+00	3,33E-02	53,5
8	4,09E-03	3,20E+00	-2,90E-02	53,2
9	3,49E-03	2,70E+00	1,27E-02	53,2
10	3,16E-03	2,50E+00	-3,00E-03	52,8
11	2,61E-03	2,00E+00	2,66E-03	51,8
12	1,71E-03	1,30E+00	-1,59E-03	52,2
13	1,44E-03	1,10E+00	1,31E-02	50,5
14	1,35E-03	1,00E+00	-4,18E-03	52,0
15	1,12E-03	9,00E-01	4,33E-03	50,3
16	7,44E-04	6,00E-01	-2,06E-02	51,8
17	6,18E-04	5,00E-01	1,13E-02	51,3
18	5,57E-04	4,00E-01	-5,28E-03	50,1
19	3,52E-04	3,00E-01	-9,63E-03	52,9
20	2,69E-04	2,00E-01	-1,21E-02	50,7

Table 2: Eigenvalues based on the *age-period-cohort model* (model structure 3), with the associated percentage variance explained, m_i^c values and related bootstrap confidence levels (CL).

Table 3, gives the eigenvalues (λ_i^c-values), in significant order, based on the *age-period-cohort expanded model* (model structure 3a), defined in the parallelogram DCGF (Figure 1), with the associated percentage variance explained. According to the confidence level criterion (CL), there are 3 significant PCs with minimum CL value 80%, which explain 68% of the total (conditional in time) residual variance.

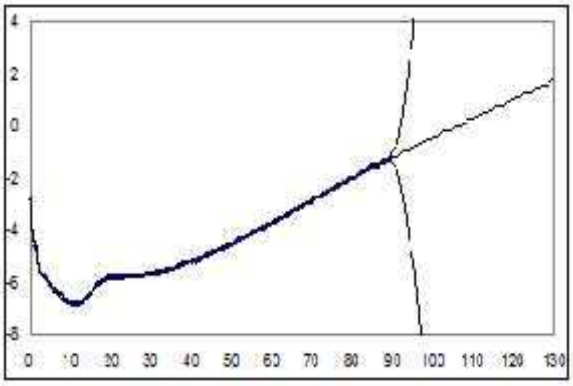
PC index	λ_i^c	% var	m_i	%CL
1	2,90E-02	2,60E+01	6,32E-02	80
2	1,99E-02	1,79E+01	-2,98E-02	92
3	1,59E-02	1,43E+01	4,93E-02	92
4	1,16E-02	1,04E+01	-5,20E-02	76
5	9,76E-03	8,80E+00	2,18E-03	77
6	6,08E-03	5,50E+00	-9,72E-03	76
7	5,04E-03	4,50E+00	2,00E-02	76
8	3,94E-03	3,50E+00	-1,28E-02	75
9	2,53E-03	2,30E+00	1,39E-02	73
10	1,95E-03	1,80E+00	-1,05E-02	71
11	1,80E-03	1,60E+00	3,59E-03	66
12	1,18E-03	1,10E+00	-5,12E-03	61
13	1,06E-03	1,00E+00	2,57E-03	54
14	7,84E-04	7,00E-01	4,99E-03	53
15	5,72E-04	5,00E-01	-5,50E-03	52
16	2,22E-04	2,00E-01	-5,27E-03	51

Table 3: Eigenvalues based on the *age-period-cohort expanded model* (model structure 3a), with the associated percentage variance explained, m_i^c values and related bootstrap confidence levels (CL).

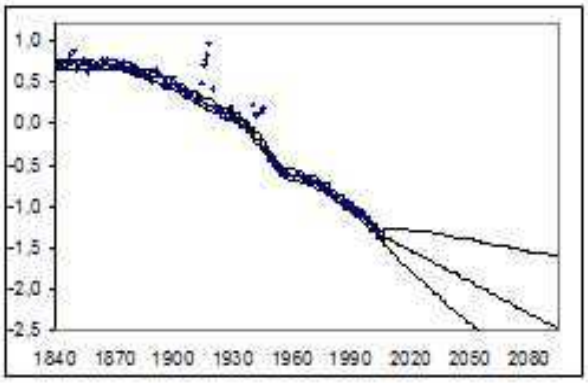
Figure 6, displays the most important components for the *age-period model* (model structure 1), according to the methodology in section 2.1, with associated forecasts and (bootstrap) CIs. For the $Y_i(t)$ values, the calendar years 1914-1919 and 1940-1945 are treated as missing values and are estimated by smooth values according to the DLR model structures (section 2.3.1).

The first two interaction terms, which account for the 100% of the total variance (with E-02 precision), describe different the dynamics of age groups. Among these first two components, we can specify the key age groups by taking, for each age, the maximum association value (MAV). According to the MAV, the first interaction term, which explains the 87,1% of the total variation (Table 1), refers to ages 0-44 and its time component shows a steady linear trend after the 1970s. The second interaction term, which explains 11,4% of the total variation, refers to ages 45+ and its time component shows a steady downward trend especially after the 1980s. In this way, we split the whole age range into two age groups which experience different dynamics over time.

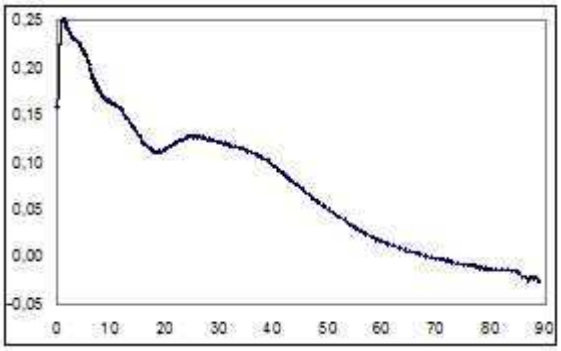
The other five interactions explain relative deviations mainly from the first interaction term. We can specify the key age groups, from the last five components, by taking, for each age, the maximum association value (MAV) among all the 7 components. Specifically, the third interaction term, according to the MAV, refers to ages 15-19, and its time component shows a rapid relative deterioration after the 1950s with a relative improvement after the 1980s. The fourth interaction term, according to the MAV, refers to ages 20-30, and its time component shows a rapid relative deterioration after the 1970s with a relative improvement after the year 2000 which could be explained by the well-known "accident hump" effect. The fifth interaction term, according to the MAV, refers to ages 10-14 and age 1 and its time component shows a relative improvement. The sixth interaction term, according to the MAV, refers to ages 31-41 and ages 5-7, and its time component shows a significant relative deterioration after the 1980s which could be influenced by the AIDS effect. The seventh interaction term, according to the MAV, refers to ages 42-48 and age 0, and its time component shows a relative deterioration after the 1960s. In Figure 6, it is clear that ages 0 and 1 have $g_i(x)$ values which differ from those of other ages. This feature indicates low dependence with the other ages, and explains the MAV for seventh and fifth interaction term respectively. Ages 5-7, referring to the sixth interaction term, experience the lowest death counts among all the ages and explain the MAV inherent to the sixth interaction term. Also, notable is the large improvement for the combined effects, as can be illustrated by the additive $b(t)$ term, for calendar years 1946-1955.



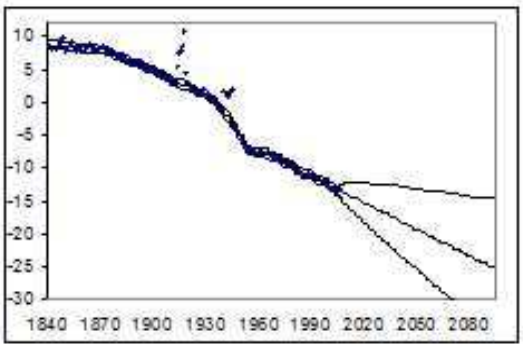
$A(x)$



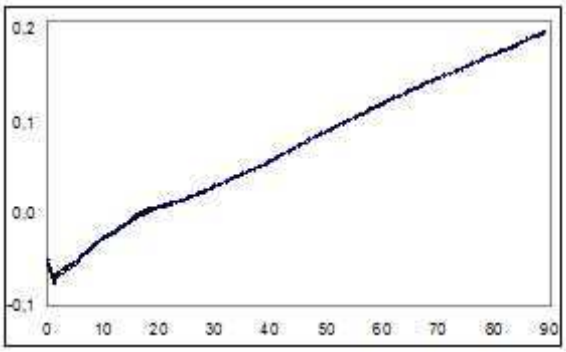
$b(t)$



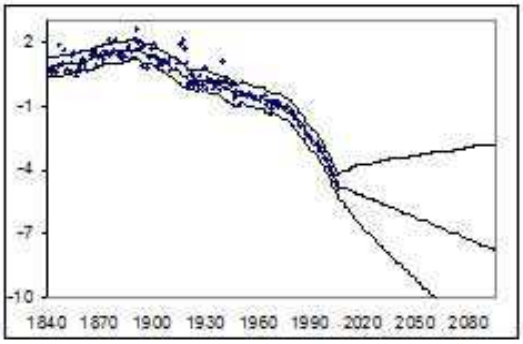
$g_1(x)$



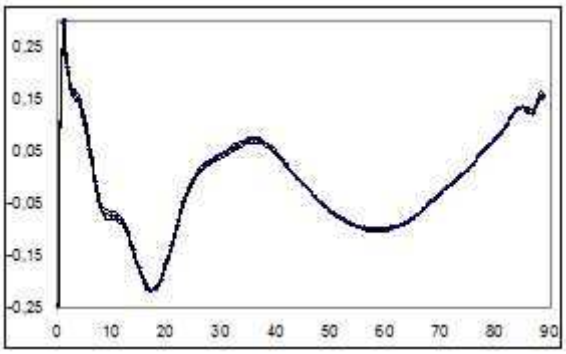
$Y_1(t)$



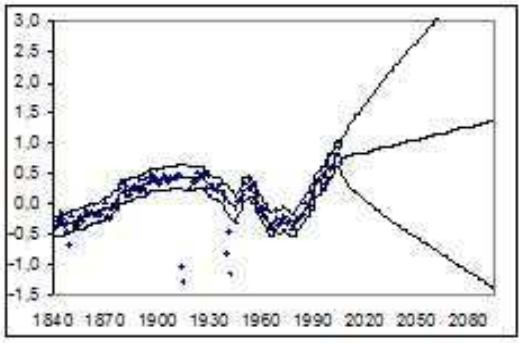
$g_2(x)$



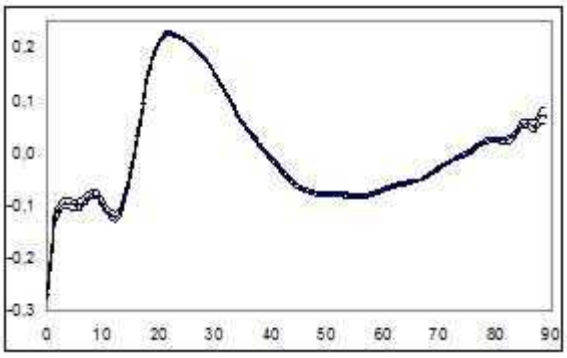
$Y_2(t)$



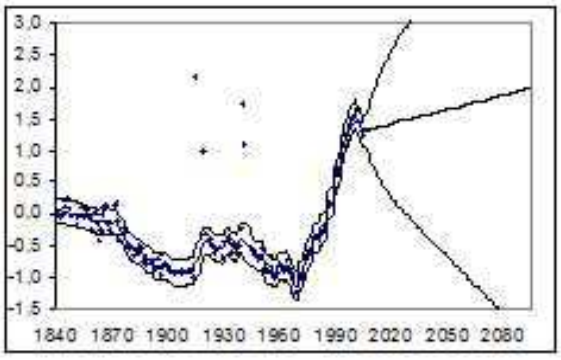
$g_3(x)$



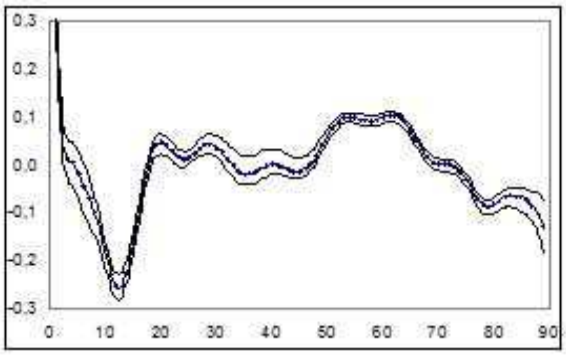
$Y_3(t)$



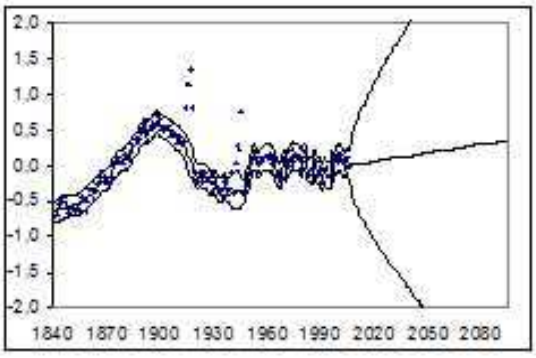
$g_4(x)$



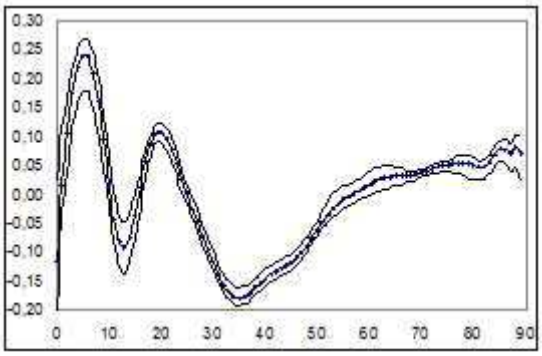
$Y_4(t)$



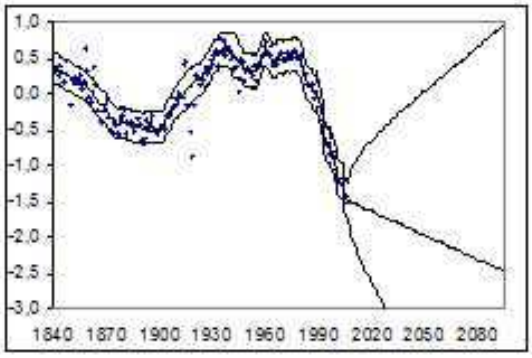
$g_5(x)$



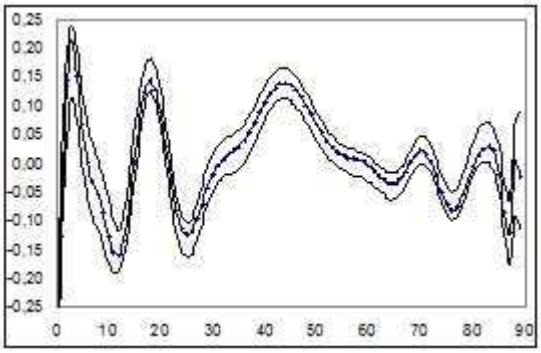
$Y_5(t)$



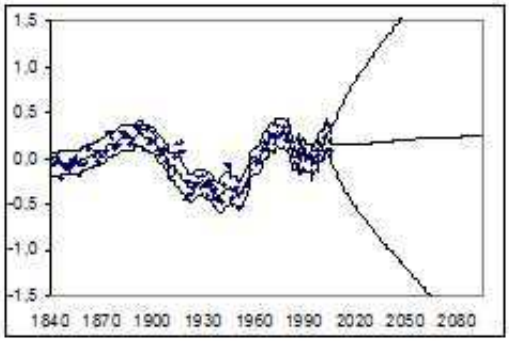
$g_6(x)$



$Y_6(t)$



$g_7(x)$

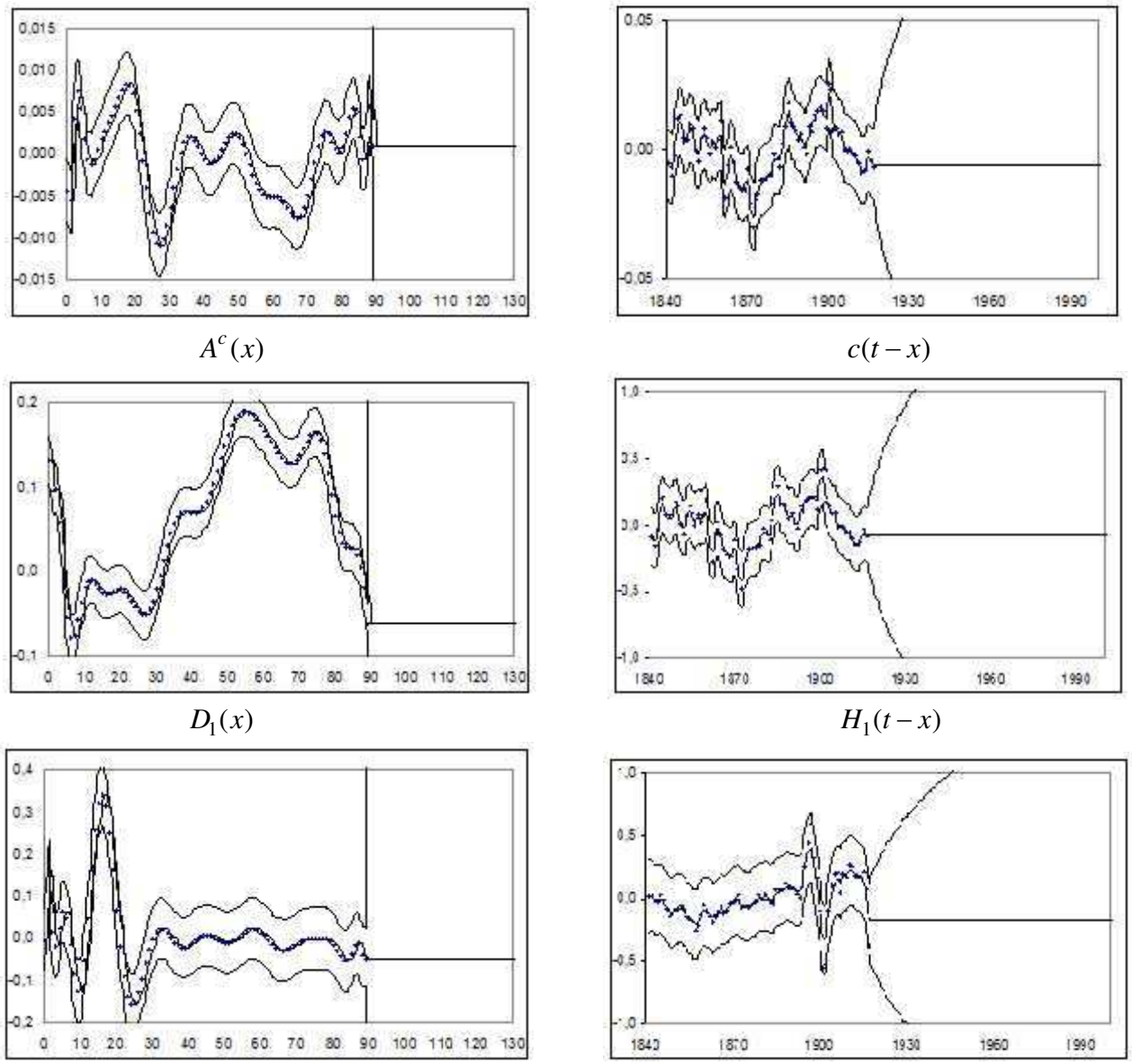


$Y_7(t)$

Figure 6: Dynamics in age-time effects based on the *age-period model* (model structure 1) in the rectangular ADCB (Figure 1). $A(x)$ (with associated CI) and $b(t)$ (with associated CI under the DLR structure) are the main additive effects, $g_i(x)$ & $Y_i(t)$ are the interaction components in significant order, $g_i(x)$ are the age related components and bootstrap associated CI (left) and estimated, smoothed and predicted $\hat{Y}_i(t)$ time related components and associated CI (right), under the DLR structure.

Figure 7, displays the most important age-cohort components, with associated forecasts and CIs under the DLR structure, in order of significance, based on the *age-period-cohort model* (model structure 3).

These graphs show the limited information we derive for the study of the cohort dynamics, without being able to utilize all of the observed mortality data.



$A^c(x)$ $c(t-x)$

$D_1(x)$ $H_1(t-x)$

$D_2(x)$ $H_2(t-x)$

Figure 7: Dynamics in age-cohort effects based on the *age-period-cohort model* (model structure 3) defined in the rectangular DEBF (Figure 1). $A^c(x)$ and $c(t-x)$ are the main additive effects, $D_i(x)$ & $H_i(t-x)$ are the interaction components in significant order, estimated, smoothed and projected $\hat{D}_i(x)$ age related components and associated CI (left) and estimated, smoothed and predicted $\hat{H}_i(t-x)$ cohort related components and associated CI (right)

Figure 8, displays the most important age-cohort components, with associated forecasts and CIs under the DLR structure, in order of significance, based on the *age-period-cohort expanded model* (model structure 3a), and defined in the parallelogram DCGF (Figure 1). The third interaction term, which explains 14,3% of the residual variance refers mostly to ages 30+, with positive values for the age related component $D_1(x)$. The cohort associated PC, $\hat{H}_1(t-x)$, has negative values for the cohorts born in the period 1904-1937, indicating a positive cohort effect for these cohorts. Particularly, for cohorts born in the period 1904-1913, the cohort rate of improvement is positive, which implies that this positive cohort effect will be more noticeable for cohorts born in that period. The second interaction term, which explains 17,9% of the residual variance, refers mostly to ages 24+, with negative values for the component $D_2(x)$. The cohort dependent PC has significant negative values for cohorts born in the period 1920-1930, indicating a negative cohort effect for these cohorts, and has significant positive values for cohorts born in the period 1904-1919 and 1931-1950, indicating a positive cohort effect for these cohorts. The first interaction term, which explains 26% of the residual variance, refers mostly to ages 10-20, with positive values for the component $D_3(x)$. The cohort dependent PC has significant positive values for the most cohorts born in the period 1903-1932 (except for cohorts born in the period 1917-1919), indicating a negative cohort effect for these cohorts. This effect seems to describe unexplained residual variance due to the World Wars I & II.

Many authors have observed and analyzed these cohort effects. Renshaw and Haberman (2006), Booth and Tickle (2008) and Cairns et. al. (2008, 2009) identify that individuals born around the period 1925 to 1935 have experienced rather larger improvements in mortality compared with people born before or after this period. Cairns et. al. (2008) note that cohorts born around 1930 have experienced strong rates of mortality improvement between ages 40 and 70 relative to, say, cohorts born 10 years earlier or 10 years later and, for its part, the cohort born around 1950 seems to have experienced worse mortality than the immediately preceding cohorts. Possible explanations for this 'golden' cohort include a healthy diet in the 1940s and early 1950s (or, to be more precise, an absence of unhealthy food), and the introduction of the National Health Service in 1948. Also, Richards et al (2005) report that the 'cohort effect' in this context is the observed phenomenon that people born in the U.K. between 1925 and 1945 (centered on the generation born in 1931) have experienced more rapid improvement in mortality than generations born on either side of this period. This study shows that there are two 'sub-cohorts' of the 1925 to 1945 cohort: an earlier group where the improvements may be largely due to smoking; and a later one where other factors, such as diet in early life, may have played a greater role. It is also notable in this study that the second 'sub-cohort' of high mortality improvement, applying to people born in the early 1940s, is found in both national population and insurance experiences. In this context, Willets (2004) comments that the cohort effect in the U.K. is perhaps less about a 'healthy generation benefiting from wartime rationing and the Welfare State', and rather more the result of preceding generations being particularly unhealthy and, indeed, 'damaged' through, for example, exposure to smoking. Thus, the second cohort component could relate to this 'smoking' effect, which describes death due to diseases heavily related to smoking, such as chronic obstructive pulmonary disease, ischemic heart disease, stroke and cancer, mainly of the lung, head and neck, oesophagus, stomach and the urinary bladder. Also, Booth & Tickle (2008), report negative cohort effects (i.e. lower mortality) that have been identified at ages 65+ for cohorts born after 1900 and Richards et al. (2007) create two cohorts for this period: years of birth 1903–1909, 1910–1923.

The overall combined effects of the above cohort dynamics can be expressed by the main cohort effect $c(t-x)$. Figure 8 reveals that the main cohort effect, in the 20th century, has significant negative values, for cohorts born in the periods 1912-1918 with positive cohort rate of improvement (implying a relative improvement for successive cohorts), and also significant negative values for the cohorts born in the period 1929-1946, indicating a positive cohort effect for both those cohorts ("golden cohort" effect). In addition, the main cohort effect has significant positive values, for cohorts born in the periods 1921-1928, indicating a negative cohort effect for these cohorts ("smoking cohort" effect). A very interesting feature is also revealed for cohorts 1918-1919 and 1946, where their values are distinct from their neighbourhoods, especially for cohort 1919. In association with their negative values these cohorts have experienced relative higher improvements in mortality compared with cohorts before or after these cohorts. The years 1918-1919 refer to time just after the 1st World War and the Spanish flu and the year 1946 refers to the time just after the 2nd World War.

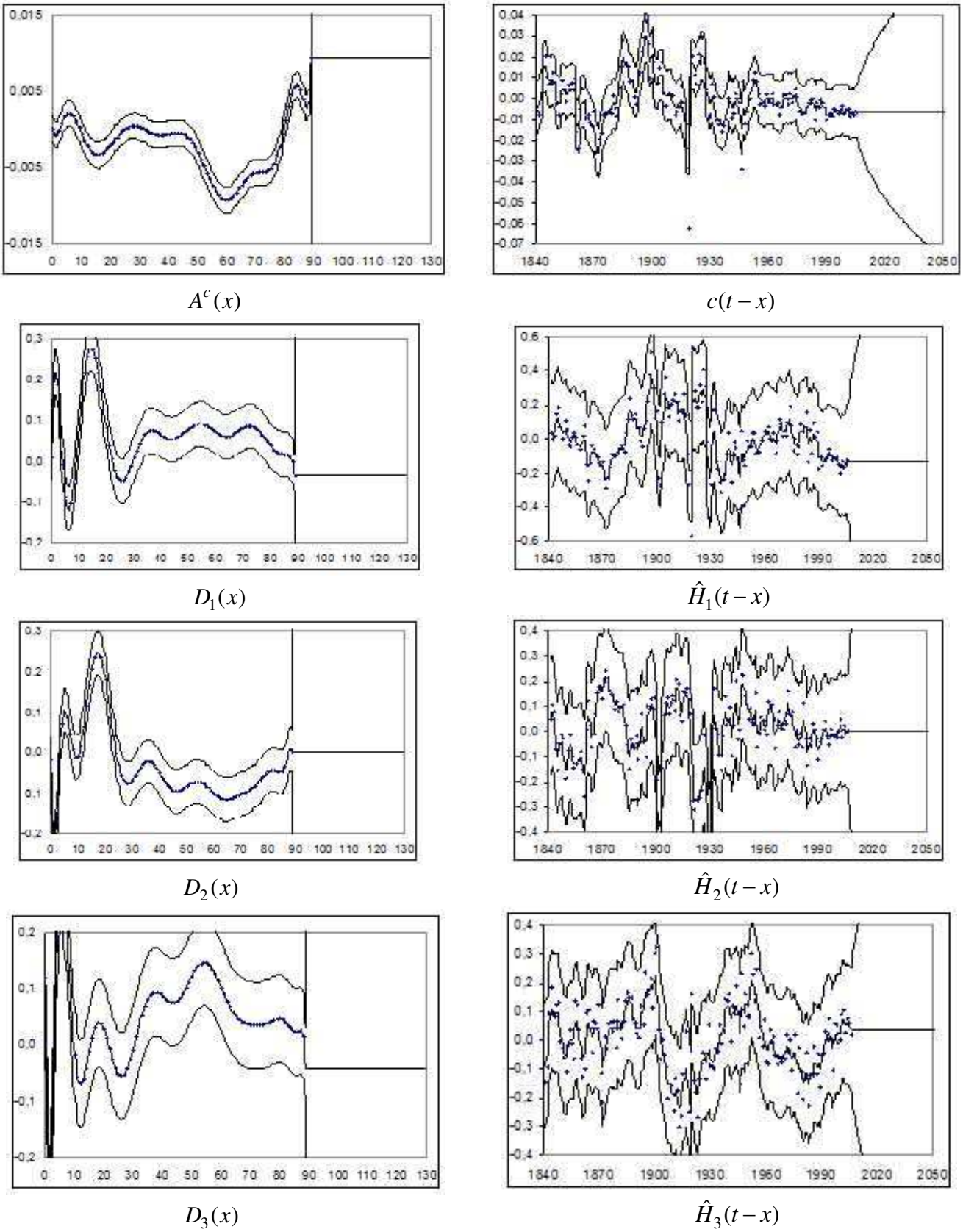
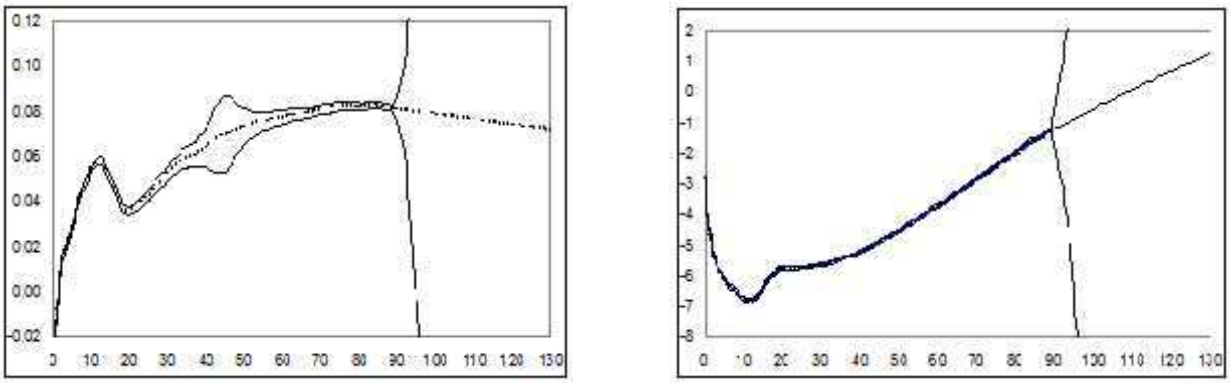


Figure 8: Dynamics in age-cohort effects based on the *age-period-cohort expanded model* (model structure 3a), in the rectangular DCGF (Figure 1). $A^c(x)$ and $c(t-x)$ are the main additive effects, $D_i(x)$ & $\hat{H}_i(t-x)$ are the interaction components in significant order, estimated, smoothed and projected $\hat{D}_i(x)$ age related components and associated CI (left) and estimated, smoothed and predicted $\hat{\hat{H}}_i(t-x)$ cohort related components and associated CI (right)

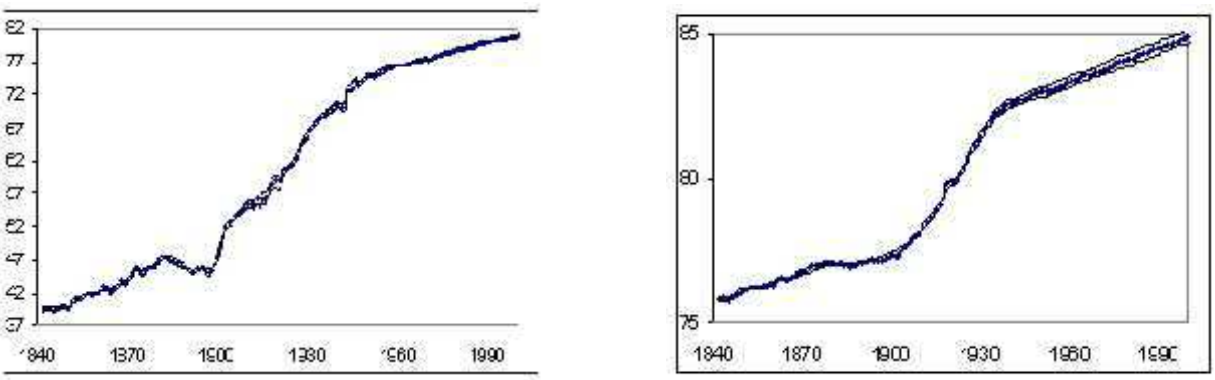
In the *age-period-cohort expanded model* (model structure 3a), and the discussion of the construction of complete mortality tables in Appendix, the age main effects, $A'(x)$, are modelled with a DLR model with age varying slope parameter: $A'(x) = a + b_x \cdot x' + e_x$ where b_x follows an IRW model, and e_x is white noise. Figure 9 shows the evolution of the IRW parameter b_x. The inflection point, for this range of calendar years (1841-2006) is again near the age x_{ip} =84.



9.1 9.2

Figure 9: The IRW age varying slope parameter $\hat{b}_x$, based on the *age-period-cohort expanded model* (model structure 3a), with associated CIs (graph 9.1), and the corresponding $A'(x)$ values with smoothed and projected estimations, with associated CIs (graph 9.2).

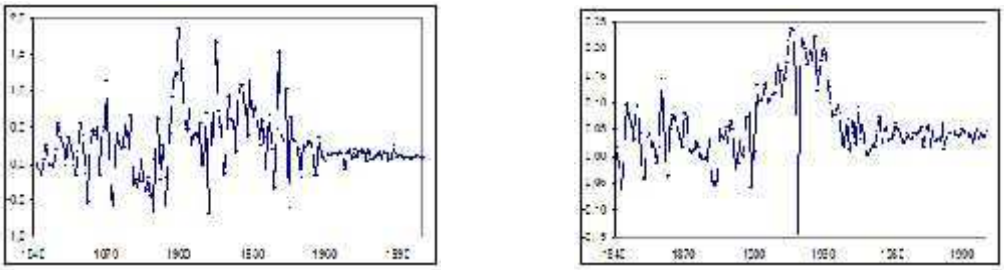
Figure 10, shows the bootstrap 95% percentile bootstrap CIs for cohort life expectancies at birth ($e_0(c)$, graph 10.1), and for cohort expected age at death given survival at age 65 ($65 + e_{65}(c)$,graph 10.2), for cohorts *c=t-x=1841,…,2006* and the complete age range $x = x_1,...,x_{130}$ (based on the *age-period-cohort forecast-complete model* (model structure 6), parallelogram $DHK'F'$, Figure 1). We note that the difference between the two curves diminishes with time, and this convergence is in accordance with the reported "rectangularization" feature for latest mortality experiences. The difference between the upper limit and the lower limit of the 95% CIs for $e_0(c)$ life expectancies have values from 0.5 to 1, and for $e_{65}(c)$ life expectancies have values from 0.1 to 0.5. Particularly, in Figure 10.1 the difference between the upper limit and the lower limit of the CIs does not become wider at the last cohorts and this feature is due to the procedure employed in section 2.2, treating the extrapolated crude rates as a real data, in the incorporation of cohort effects.



10.1 10.2

Figure 10: Bootstrap 95% percentile CIs for cohort life expectancies at birth (graph 10.1) and cohort expected age at death given survival at age 65 (graph 10.2), for E&W males, based on the *age-period-cohort forecast-complete model* (model structure 6).

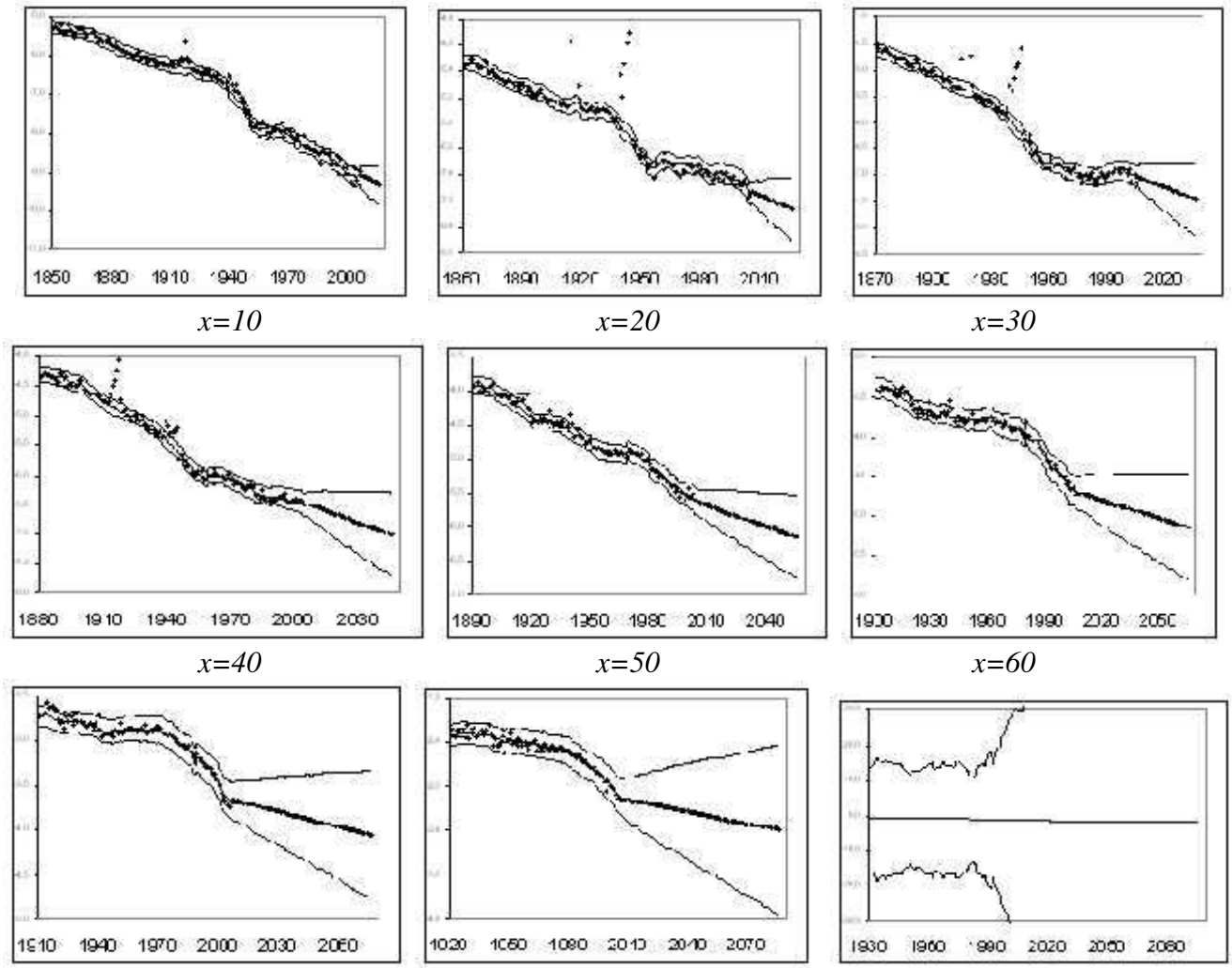
Figure 11, shows the difference between successive observed-projected cohort life expectancies at birth ($\Delta e_0(c) = e_0(c+1) - e_0(c)$) and observed-projected cohort expected age at death given survival at age 65 ($\Delta e_{65}(c) = e_{65}(c+1) - e_{65}(c)$), for cohorts *c=t-x=1841,...,2006* and the complete age range $x = x_1, ..., x_{130}$ (based on the *age-period-cohort forecast-complete model*, model structure 6), parallelogram $DHK'F'$, Figure 1). We note that for graph 11.2, the differences in cohort life expectancies have the biggest values for cohorts born in the period 1901-1932, with an outlier at cohort 1919.



11.1 11.2

Figure 11: Difference between successive observed-projected cohort life expectancies at birth (graph 11.1) and at age 65 (graph 11.2), for cohorts *c=t-x=1841,...,2006* and complete age range $x = x_1, ..., x_{130}$, for E&W males.

Figure 12, shows observed log central mortality rates, smoothed and projected log-central mortality rates and associated 95% CIs (based on the *age-period-cohort forecast-complete model*, model structure 6), (only values inside the parallelogram $DCG'F'$, Figure 1).



x=10 *x=20* *x=30*

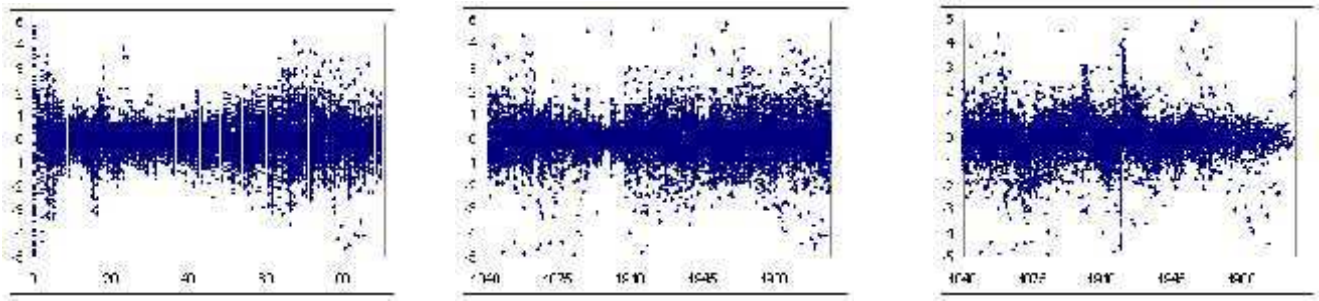
x=40 *x=50* *x=60*

x=70 *x=80* *x=90*

Figure 12**:** Observed log central mortality rates, smoothed and predicted log central mortality rates and associated CIs, for ages *x=10,20,30,40,50,60,70,80* and *90*, for E&W males.

3.5 Using Principal Component Analysis

In order to compare the SPC approach with the conventional PC approach (see Hatzopoulos and Haberman (2009)), we present the results for the same data set (E & W males), using the conventional PC analysis approach. Figure 13, shows the standardized deviance residuals (SDR) plotted against age, time and cohort effects, for England & Wales for the period 1841-2006 and ages 0-89, under the *age-period model* (model structure 1, with p=6 PCs). For the age-period model, i.e. after keeping only the first 6 PCs, the overall patterns of the total SDR against age and time (graphs 13.a 13.b) indicate an appropriate fit and show clear patterns for the cohort effects (graph 13.c), very similar results with Figure 2.



13.a 13.b 13.c

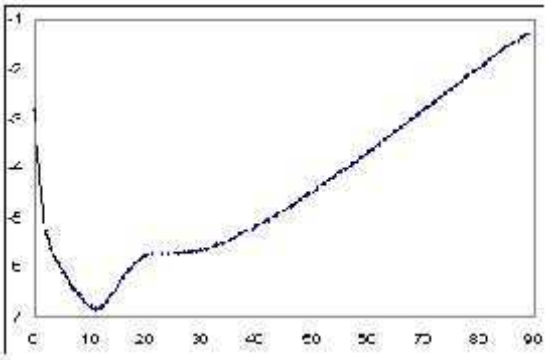
Figure 13: SDR vs. age, time and cohort effects for the *age-period model*, using PC analysis

Table 4, gives the eigenvalues (λ_i-values), in significant order, based on the age-period model (model structure 1 and PC approach), with the associated percentage variance explained (%var), *mean* association values m_i and associated (bootstrap) confidence level (CL) values. Although the CL criterion gives three only significant PCs, in order to compare the two approaches we demonstrate the first six PCs (which together explain the 99,78% of the total variance).

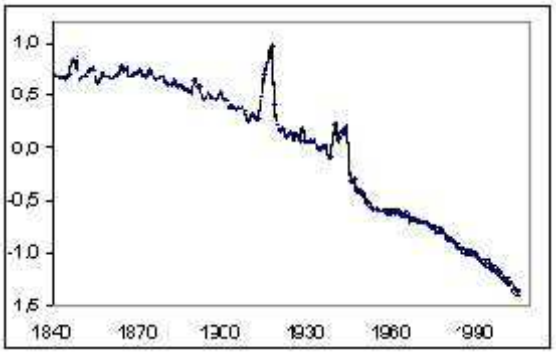
PC index	λ_i	% var	m_i	%CL
1	6,09E+01	9,49E-01	-8,77E-02	100,0
2	2,15E+00	3,36E-02	-7,42E-03	100,0
3	5,34E-01	8,33E-03	-4,90E-02	100,0
4	3,00E-01	4,67E-03	-1,91E-02	72,8
5	7,00E-02	1,09E-03	1,15E-03	95,1
6	5,12E-02	7,98E-04	1,54E-02	53,1
7	2,43E-02	3,79E-04	-1,36E-02	78,8
8	1,99E-02	3,10E-04	6,55E-03	56,3
9	1,67E-02	2,60E-04	-3,98E-03	59,9
10	1,55E-02	2,42E-04	-7,62E-03	57,5
11	1,34E-02	2,09E-04	-2,47E-03	56,4
12	1,13E-02	1,76E-04	3,28E-03	56,2
13	1,04E-02	1,63E-04	-4,65E-03	56,0
14	7,54E-03	1,18E-04	1,13E-03	55,6
15	6,06E-03	9,45E-05	-9,47E-04	55,1
16	4,48E-03	6,99E-05	2,94E-03	54,7
17	3,28E-03	5,11E-05	-2,16E-03	54,3
18	2,93E-03	4,57E-05	7,08E-04	54,0
19	2,21E-03	3,44E-05	-5,48E-04	53,4
20	1,69E-03	2,64E-05	9,71E-04	52,9
21	1,11E-03	1,72E-05	-9,91E-04	52,7
22	9,18E-04	1,43E-05	2,02E-04	50,1
23	8,19E-04	1,28E-05	-1,58E-04	50,1
24	5,77E-04	8,99E-06	-7,48E-05	50,5
25	1,17E-04	1,83E-06	-4,17E-04	50,2

Table 4: Eigenvalues based on the *age-period model* (model structure 1 with PC analysis), with the associated percentage variance explained, m_i values and related bootstrap confidence levels (CL).

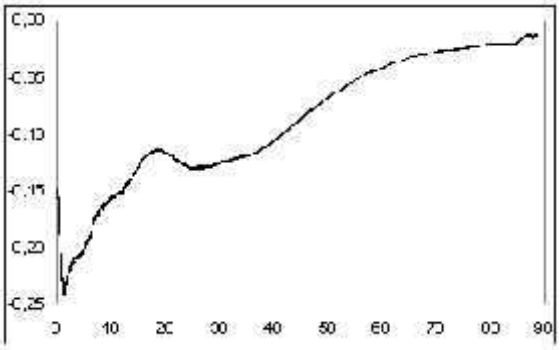
Figure 14, displays the first six components for the age-period model (model structure 1, with PC analysis). As was expected, the first interaction term dominates the mortality dynamics, which explains the 94,9% of the total variation (Table 4), and refers to the whole age range (0-89). The other five interactions explain relative deviations from the first interaction term. The second interaction term, which explains 3,36% of the total variation, refers to ages 19-30, according to the maximum association value (MAV). We note that the age component $g_2(x)$, is very similar with the $g_4(x)$ age component under the SPC approach (Figure 6), which explained as the "accident hump" effect. The third interaction term, according to the MAV, refers to ages 53-82, and its time component shows a rapid relative improvement after the 1980s. The fourth interaction term, according to the MAV, refers to ages 7-15 and 31-43. It is very complicated to explain this interaction term because it is a combined effect from those two age groups. We note, that the age groups 10-14 and 31-41 explain the fifth and sixth interaction term respectively under the SPC approach (Figure 6). The same difficulties arise for the other two interaction terms as well. The fifth interaction term, refers to ages 44-52 and 83-89. We note, that the age group 42-48 explains the seventh interaction term under the SPC approach (Figure 6). The sixth interaction term, refers to ages 1-2 and 16-18. We note, that the age group 16-18 explains the third interaction term under the SPC approach (Figure 6). Also, we note the main additive terms, $A(x)$ and $b(t)$, are almost identical under both approaches (Figure 6).



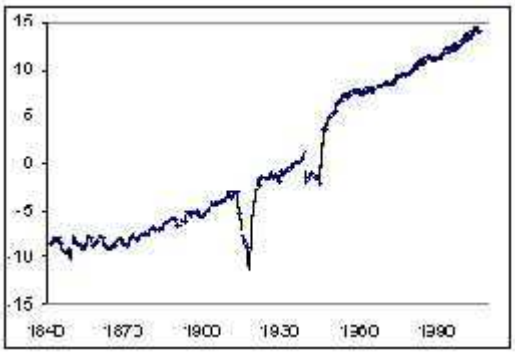
$A(x)$



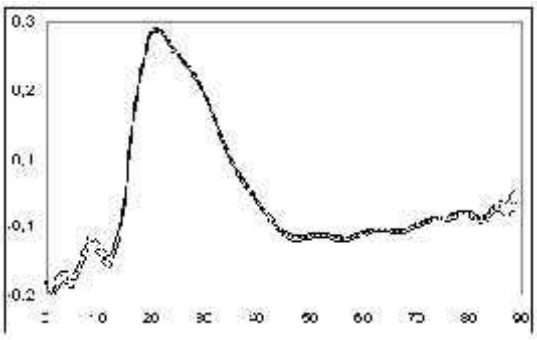
$b(t)$



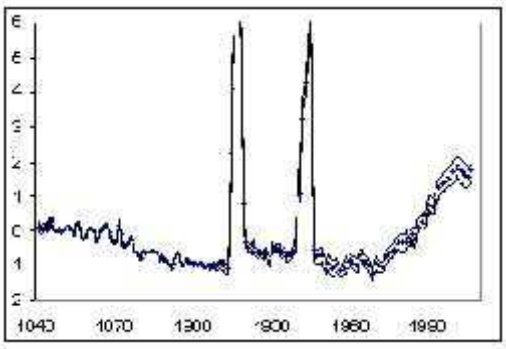
$g_1(x)$



$Y_1(t)$



$g_2(x)$



$Y_2(t)$

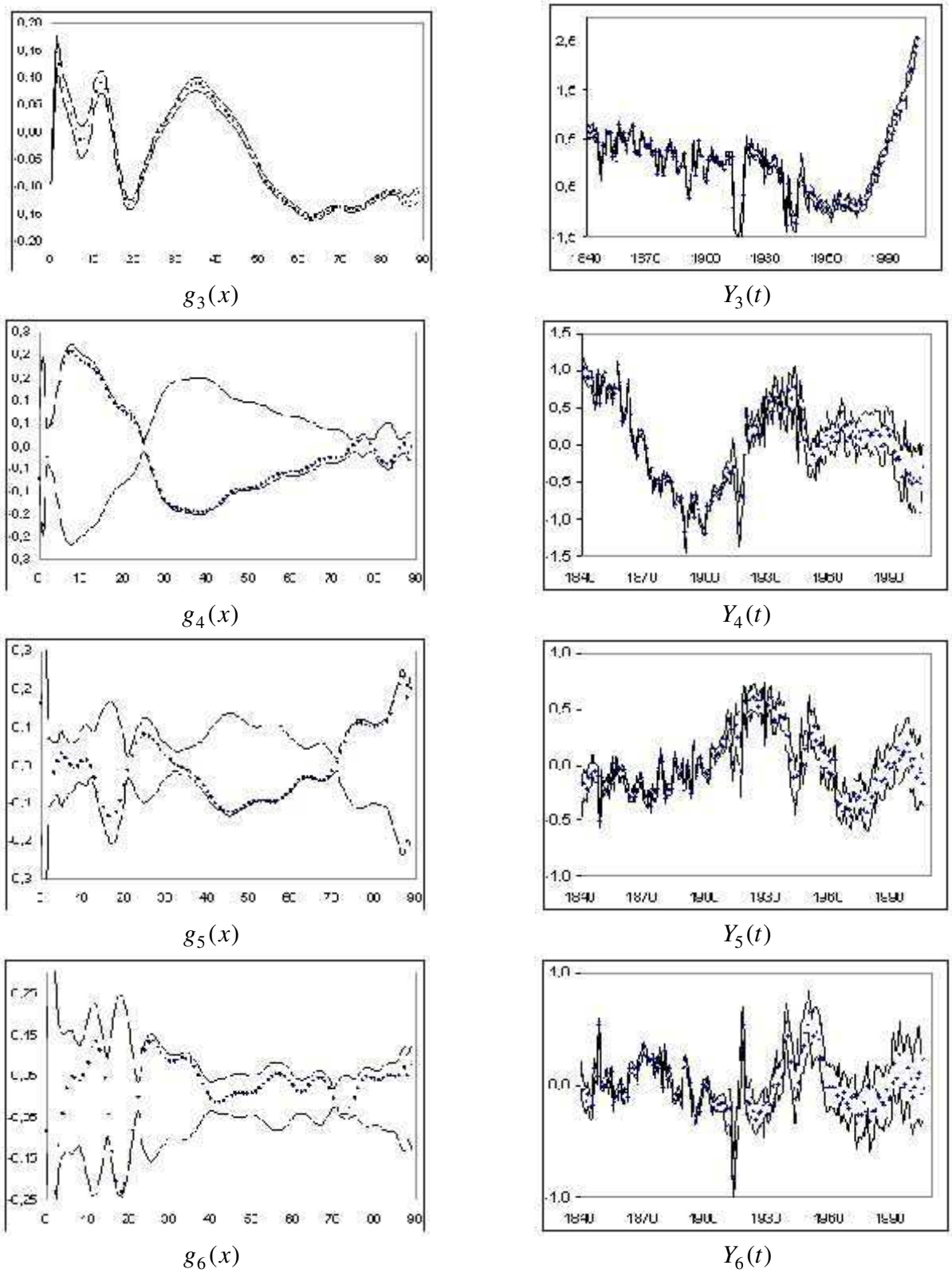


Figure 14: Dynamics in age-time effects based on the *age-period model* (model structure 1, under PC approach) in the rectangular ADCB (Figure 1). $A(x)$ (with associated CI) and $b(t)$ (with associated CI) are the main additive effects, $g_i(x)$ & $Y_i(t)$ are the interaction components in significant order, $g_i(x)$ are the age related components and bootstrap associated CI (left) and estimated, $Y_i(t)$ time related components and associated CI (right).

4. Discussion

We have investigated a new approach to the modeling and projected of mortality rates for the age-time-cohort effects. An extended version of Hatzopoulos and Haberman (2009) dynamic parametric model is presented incorporating the cohort effect. A one-factor parameterized exponential orthonormal polynomials in age effects within the GLM framework treating calendar year as a factor. Then, we apply SPCA to the time dependent GLM non-stationary time series parameter estimates in order to provide (marginal) estimates for a two-factor PC approach structure (the age-period model). Then, we model the age-period residuals by one-factor parameterized exponential orthonormal polynomials in age effects treating now cohort as a factor. Finally, we apply PCA to cohort dependent GLM parameter estimates in order to provide (conditional) estimates for a two-factor structure in age and cohort effects. Overall, in this way a three-way structure is derived in age-period-cohort effects (the age-period-cohort model).

Utilizing SPCA, the proposed method differentiates from the LC and the Hatzopoulos and Haberman (2009) method by providing estimates which are based on the eigenvalue decomposition of the sparse covariance matrix of the GLM non-stationary time series parameter estimates. SPCA with E&W male mortality experience implies two main age groups (see section 3), in contrast with the Hatzopoulos and Haberman (2009) approach where only the first PC dominates the age-time dynamics. SPCA approach avoids the high interdependence structure for different range of ages, with possible spurious interpretations and forecasting (see Figure 14), and improves this problematic dependent structure by giving a better clustering of significant high factor loading values (see Figure 6). In empirical studies with various mortality experiences we get similar results, where the first age group corresponds to childhood-middle ages and the second age group corresponds to senescent mortality. This result is in accordance with many of the well-known parameterized mathematical laws, which have been proposed for the construction of mortality tables (Pitacco et al, 2009). Thus, Heligman and Pollard (1980) introduce an eight-parameter model with three terms capturing the age pattern of mortality in childhood, young adulthood and senescence; Rogers and Planck (1983) apply a four-term nine-parameter formula with exponentially declining child mortality, a double exponential accident hump and a Gompertz term representing senescent mortality; Siler (1983) propose an eight-parameter model with three independent terms describing mortality during 'immaturity', adulthood and senescent. Also Mode and Busby (1982), and Rogers and Little (1993), inter alia, each propose a mathematical law, where they use different mathematical formulae for different age groups, usually involving 3 or 4 different age groups, in order to describe the whole mortality in age effects. Although the above parameterized mathematical laws can adequately describe the age variation in mortality experience, for each successive time period, the deficiency of this approach comes from the high interdependencies between the time series for the different parameters which cause serious difficulties for forecasting and for estimating the forecast prediction intervals. In contrast, with our proposed method, the columns of the GLM parameter time series estimates $\mathbf{B} = \{\beta_{j-1}(t)\}$ in the age-period(-cohort) model encompass a high correlation structure, and this correlation structure is exploited and modelled by the proposed method. The only difference with the other parameterized approaches is that all of the GLM parameter estimates are on the same scale and this facilitates the application of PCA to the variance-covariance matrix.

The age-period-cohort association model (model structure 4) provides estimates for a statistical association model for a three-way cross-classification table with age, time and cohort being the three main random effects. Under this approach, the components $b(t)$ and $c(t-x)$ are important terms, since they describe the overall mortality trend in time and cohort effects, providing additional insights into the overall mortality dynamics, especially in cases where many dynamic factors are present. Also, the main age random effect component $A(x)$ is very informative about the correlation structure among the ages.

We utilize DLR models for modelling and forecasting the age-time and cohort related components. The statistical treatment of the DLR models is based on the state space framework and the Kalman filter. The period dependent PCs are modelled as linear functions in time, under DLR modelling, with the slope following a RW time series process. The cohort dependent PCs are modelled as RW plus noise time series, since conditional on age-time effects, the cohort PCs are assumed to be mean reverting stochastic processes. For the construction of a complete life table, we introduce a dynamic variant of the Gompertz law, applied to the age related components, assuming that the slope term of Gompertz law is an IRW time series process, under DLR

modelling. These DLR estimates in age-period-cohort effects ensure that we have continuity and smoothness across all three variants for the mortality rates (the degree of smoothness, if desired, can be controlled by the appropriate NVR hyper-parameters).

In the literature, many authors point out that, in order to assess whether any stochastic mortality model is a good model or not, it is important to consider certain criteria against which the model can be tested. Following Cairns et. al. (2008), we consider the following key criteria:

1. *Mortality rates should be positive*; positive mortality central values are assured in our model structure since the log link function is being used with a Poisson GLM structure.
2. *The model should be consistent with the historical data*; the proposed model incorporates a great degree of flexibility in order to capture historical mortality trends and dynamic mortality changes. With dynamic linear regression modeling, we are able to utilize all of the available historic mortality patterns without the restricted assumption that the logarithms of the mortality rates are approximately a linear function of time, an assumption which imposes restrictions in the dominant time component. Further, under the age-period full model structure (see Model 1, Hatzopoulos and Haberman (2009)), the polynomial expansion approach for modelling age patterns of mortality renders great flexibility in the graduation process since most continuous functions can be approximated by a polynomial to any degree of accuracy in the form of a Taylor power series.
3. *Long-term dynamics under the model should be biologically reasonable*; under the age-period-cohort expanded model (model structure 3a), the PC scores, $Y_i\ (t)$ and $\hat{H}_i(t-x)$ estimates, describe the long-term dynamics, and in association with the age profiles, $g_i(x)$ and $D_i(x)$ respectively, they are dynamic factors and they can be labelled in association with SPCA. SPCA enables the construction of a few stochastic PCs, which describe the dynamics of different age group. The England and Wales mortality experience leads to 7 factors, where each one describes certain dynamics in the mortality evolution. For example the 4th factor can be labelled as the "accident hump" effect or the 2nd factor, in age-period-cohort effects, can be labelled as the "golden cohort effect".
4. *Parameter estimates and model forecasts should be robust relative to the period of data and range of ages employed*; Empirical studies with various mortality experiences verify this criterion.
5. *Forecast levels of uncertainly and central trajectories should be plausible and consistent with historical trends and variability in mortality data*; Figure 12, which displays observed and projected log central mortality rates for some ages, exhibits smoothed estimates which are plausible and consistent with historical and future trends.
6. *The model should be straightforward to implement using analytical methods or fast numerical algorithms*; all the computations have been implemented in Matlab.
7. *The model should be relatively parsimonious*; in the presence of cohort effects, the model is expressed by the simple model structure (3).
8. *It should be possible to use the model to generate sample paths and calculate prediction intervals*; the proposed method utilizes parametric bootstrapping and gives bootstrap confidence intervals for the mean association values (CL-criterion) and confidence intervals for the expected remaining lifetime.
9. *The structure of the model should make it possible to incorporate parameter uncertainty in simulations*; the proposed method utilizes parametric bootstrapping and for each bootstrap sample, the GLM parameters $\mathbf{B}^{(i)}$ are estimated.
10. *At least for some countries, the model should incorporate a stochastic cohort effect*; the age-period-cohort model (model structure 3) incorporates a stochastic cohort effect.
11. *The model should have a non-trivial correlation structure*; under the age-period-cohort model (model structure 3), there are *p*-time PCs and *q*-cohort PCs, leading to a non-trivial correlation structure.
12. *The model is applicable for the full age range*; the model applied to England and Wales male mortality experience over the whole range of ages (0-89).

We would like to thank Doctor Oluf Dimitri Røe (MD, PhD, Oncologist, St. Olavs University Hospital, Trondheim, Norway) for advice on the medical terminology that helps with explaining the cohort effects.

References

Bell, W.R. and Monsell, B.C. 1991. "Using Principal Components in time Series modelling and Forecasting of Age-Specific Mortality Rates." Proceedings of the American Statistical Association, Social Statistics Section pp. 154–159.

Booth, H. and Tickle, L. 2008. "Mortality modelling and forecasting: A review of methods". Annals of Actuarial Science, 3 (1&2), pp 3-43.

Brillinger D. R. 1986. "The natural variability of vital rates and associated statistics". Biometrics 42(4),pp 693-734.

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

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.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A. and Balevich, I. 2009. "A quantitative comparison of stochastic mortality models using data from England and Wales and the United States". North American Actuarial Journal 13 (1), pp 1-35.

Coale, A.J. and E.E. Kisker 1990. "Defects in data on old-age mortality in the United States: New procedures for calculating schedules and life tables at the highest ages". Asian and Pacific Population Forum, 4 (1), pp 1–31.

Currie, I.D., Durban, M. and Eilers, P.H.C. 2004. "Smoothing and forecasting mortality rates". Statistical. Modelling. 4 (4), pp 279–298.

De Jong, P. and Tickle, L. 2006. "Extending Lee–Carter mortality forecasting". Mathematical Population Studies 13 (1), pp 1–18.

Efron, B. and Tibshirani, R. 1998. "An Introduction to the Bootstrap". CRC Press, Boca Raton, FL.

Gao, Q. and Hu, C. 2009. "Dynamic factor model with conditional heteroskedasticity". Insurance: Mathematics & Economics 45 (2), pp 410–423.

Goodman, L.A. 1991. "Measures, models, and graphical displays in the analysis of cross-classified data". Journal of the American Statistical Association, vol. 86, No. 416, pp. 1085-1111.

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.

Heligman, L. and Pollard, J.H. 1980. "The age pattern of mortality", *J. Inst. Actuaries* **107** (1980), pp. 49–80.

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).

Hyndman, R. J. and Ullah, M.S. 2005. "Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach". Working Paper, Department of Economics and Business Statistics, Monash University. http://www.robhyndman.info/papers/funcfor.htm.

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

Lazar, D. and Denuit, M. 2009. "A multivariate time series approach to projected life tables". Applied stochastic models in Business & Industry 25 (6), pp 806-823.

Ledermann, S. and Breas, J. 1959. "Les Dimensions de la Mortalite´." Population 14, pp 637–682.

Lee, R.D. and Carter, L.R. 1992. "Modelling and forecasting U.S. mortality". Journal of the American Statistical Association 87, pp 659–671.

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

Mode, C. & Busby, R. C. 1982. "An eight-parameter model of human mortality – the single decrement case". *Bulletin of Mathematical Biology,* 44(5), pp 647-659.

Pitacco, E. 2004. "Survival models in dynamic context: A survey". Insurance: Mathematics & Economics 35 (2), pp 279–298.
Pitacco, E., Denuit, M., Haberman, S. and Olivieri, A. 2009. "Modelling Longevity Dynamics for Pensions and Annuity Business". Oxford University Press.

Renshaw, A.E. and Haberman, S. 2003a. "Lee–Carter mortality forecasting with age-specific enhancement". Insurance: Mathematics and Economics 33 (2), pp 255–272.

Renshaw, A.E. and Haberman, S. 2003b. "Lee–Carter mortality forecasting, a parallel generalized linear modelling approach for England & Wales mortality projections". Applied Statistics 52 (1), pp 119–137.

Renshaw, A.E. and Haberman, S. 2003c. "On the forecasting of mortality reduction factors". Insurance: Mathematics and Economics 32 (3), pp 379–401.

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.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.

Rogers, A. and F. Planck 1984. Parameterized multistage population projections. Working Paper for presentation at the Annual Meeting of the Population Association of America, Minnesota, May 3–5.

Rogers, A. & Little, J. S. 1994. "Parameterizing age patterns of demographic rates with the multiexponential model schedule". *Mathematical Population Studies,* 4(3), pp 175-195.

Siler, W. (1983). "Parameters of mortality in human populations with widely varying life spans". Statistics in Medicine, 2, pp 373-380.

Sithole, T.Z., Haberman, S. and Verrall, R.J. 2000. "An investigation into parametric models for mortality projections, with applications to immediate annuitants and life office pensioners' data". Insurance: Mathematics and Economics, 27 (3), pp 285-312.

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

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

Appendix

From various mortality investigations in developed countries, it has been observed that life expectancy at birth has increased steadily during the 20th century without showing any sign of approaching a fixed limit. In addition, it has been observed that the number of people surviving up to the older ages (for example 80 years and above) has increased considerably. Although crude mortality rates (i.e. population estimates and reported deaths) at these ages are available, they are biased by poor age reporting regarding both those alive and those who die and the pattern at the very old ages is heavily affected by random fluctuations because of the scarcity of the data. Thus, these rates lack the required quality needed for the construction of complete life tables.

Various methodologies have been proposed for estimating mortality rates at the oldest ages, projecting the mortality indices along the age axis up to the ultimate age using extrapolation techniques. Empirical observation shows that the curve of mortality rates, on a logarithmic scale, presents a concave shape for the oldest ages in contrast with the simple Gompertz law which assumes that the exponential rate of mortality increase is constant (the Gompertz law, on a logarithmic scale, is $\ln(m_x) = a + b \cdot x$). In order to capture this empirical evidence, many models have been proposed by various authors.

One of the most well known approaches is the Coale-Kisker method (Coale and Kisker, 1990), which assumes that the exponential rate of mortality increase at the oldest ages is not constant but declines linearly, a pattern empirically confirmed by a number of studies (e.g., Horiuchi and Wilmoth, 1998).

The rate of mortality increase at age x is defined as $k_x = \ln\left(\frac{m_x}{m_{x-1}}\right)$, up to age (usually =110 or 120), and, in the Coale-Kisker method, this is assumed to be a linear function of the ages. We note that in the Gompertz law, $k_x = b$.

A possible extension to the Gompertz law is obtained if we assume that the slope term , b, follows an integrated random walk (IRW) time series model in order to capture the empirical patterns of mortality at oldest ages, i.e. we assume that:

$$\ln(m_x) = a + b_x \cdot x' + e_x \qquad \text{if} \qquad \Delta b_x = \Delta b_{x-1} + \zeta_{x-1}$$

where ζ_x is white noise, $\Delta b_x = b_x - b_{x-1}$ denotes the difference operation, and the ages have been rescaled: $x' = x - x_m$ where x_m denotes the median age ($x_m = (x_1 + x_a)/2$).

This kind of modelling belongs to the class of DLR models with the slope now modelled as a stochastic age variable parameter that follows an IRW process, and such can be estimated by state space models and Kalman filter techniques. Empirical studies with mortality data supports this modelling.

The smoothed estimated rate of mortality increase now becomes $\hat{k}_x = \hat{b}_{x-1} + \Delta\hat{b}_x \cdot x'$ for $x = x_1, ..., x_a$ and the projected rate of mortality increase is estimated by $\hat{k}_{x_a+h} = \hat{b}_{x_a-1} + \Delta\hat{b}_{x_a} \cdot \left(x_a + 2 \cdot h - x_m\right)$ for $h = 1, ..., x_\omega - x_a$, i.e. as a linear function of the oldest ages.

Empirical studies with the England & Wales males population on a cross-sectional basis, with x_a=89, show that the "inflection point", say x_{ip}, where the curvature changes sign and the log-mortality curve changes from being convex to concave, i.e. the age in which the slope ($\Delta\hat{b}_x$) becomes negative, varies as follows: from 1841-1945 $x_{ip} \cong 80$, for 1945-1990 $x_{ip} \cong 40 - 50$, for 1990-2000 $x_{ip} \cong 60$, and for 2001-2006 $x_{ip} > 89$.

Due to the instability of this approach, instead of applying this model to the crude mortality rates in each calendar year, we apply the above model structure to the age profile $A(x)$ (or to the age profile $A'(x)$ in the presence of cohort effects), which captures the age mortality patterns for all calendar years combined:

$$A(x) = a + b_x \cdot x' + e_x$$

if b_x follows an IRW model, and e_x is white noise. This approach gives more robust estimates, since the $A(x)$ estimates are more reliable than the crude mortality rates, especially at the oldest ages.

Following this approach we also need smooth and extrapolated $g_i(x)$ and $D_i(x)$ values. We apply a simple DLR model based on an independent RW plus noise time series model, and employ state space models and the Kalman filter technique for estimation purposes:

$$g_i(x) = b_{i,x}^p + e_{i,x}^p \quad \text{and} \quad D_i(x) = b_{i,x}^c + e_{i,x}^c$$

if $b_{i,x}^p$ and $b_{i,x}^c$ follow an RW model and $e_{i,x}^p$ and $e_{i,x}^c$ are white noise. Applying this type of structure, we are assured of smoothness and continuity to the mortality rates in age effects.

Thus, the final smoothed and projected model structure, in log scale, can be summarized as an *age-period-cohort complete model*:

$$\log(\hat{m}_x(t)) = \hat{A}'(x) + \sum_{i=1}^{p} \hat{g}_i(x) \cdot Y_i\ (t) + \sum_{i=1}^{q} \hat{D}_i(x) \cdot \hat{H}_i(t-x) \tag{A1}$$

with $\hat{A}'(x) = \hat{a} + \hat{b}_x \cdot x'$ if b_x follows an IRW model, $\hat{g}_i(x) = \hat{b}_{i,x}^p$ and $\hat{D}_i(x) = \hat{b}_{i,x}^c$ if $b_{i,x}^p$ and $b_{i,x}^c$ follow an RW model, for $x = x_1, ..., x_a, ..., x_\omega$, and c=t-x= $c_1, \ldots, c_{n+f}$.