A parameterized approach to modeling and forecasting 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

A new method is proposed of constructing mortality forecasts. This parameterized approach utilizes Generalized Linear Models (GLMs), based on heteroscedastic Poisson (non-additive) error structures, and using an orthonormal polynomial design matrix. Principal Component (PC) analysis is then applied to the cross-sectional fitted parameters. The produced model can be viewed either as a one-factor parameterized model where the time series are the fitted parameters, or as a principal component model, namely a log-bilinear hierarchical statistical association model of Goodman (1991) or equivalently as a generalized Lee-Carter model with *p* interaction terms. Mortality forecasts are obtained by applying dynamic linear regression models to the PCs. Two applications are presented: Sweden (1751-2006) and Greece (1957-2006).

Keywords: Mortality forecasting; Generalized Linear Models; Principal Component analysis; Dynamic linear regression; Bootstrap confidence intervals.

1. Introduction

The changes in mortality rates over time need to be accurately measured and projected, in order to inform the choice of mortality bases for a variety of disciplines and applications (e.g. in the life insurance and pensions industries, government policy and planning). Downward mortality trends have been gradually decline since the start of the twentieth century, but there is evidence that such trends may not be uniform across the age groups. As mentioned by Wilmoth (2000), these gains in longevity are the result of a complex array of changes (standards of living, public health, personal hygiene and medical care), with different factors playing major or minor roles in different time periods.

Mortality rates exhibit strong age patterns and various researchers have developed methods to capture this structure and use it in forecasting. Two important, general approaches have been developed for modeling and forecasting: the curve fitting approach and the principal component (PC) approach. The model fitting approach involves fitting parametric curves to the age-specific mortality rates and the PC approach involves computing PCs to obtain a linear transformation of the data in lower dimension and simplified structure.

The fitting of curves to mortality rates has a long history in demography and actuarial science, going back to the efforts of DeMoivre (1725) and Gompertz (1825). The objectives have usually been to estimate mortality curves with limited data, or to graduate (smooth) irregular curves of directly estimated rates. Thus, Cramer and Wold (1935) fit Makeham curves to cross-sectional Swedish mortality data for ages 30 and above, and then extrapolate the fitted parameters using logistic functions of time. Similarly, McNown and Rogers (1989) fit an eight-parameter Heligman-Pollard (1980) curve in age effects to cross-sectional US mortality for 86 years. After differencing to achieve stationarity, the eight univariate time series are estimated by ARIMA models and extrapolation is then used for forecasts. McNown and Rogers (1992) forecast total mortality and five cause-specific mortality rates, using a non-linear model, similar to the Heligman-Pollard (1980) curve with nine parameters, by fixing six parameters and modelling only the time dependence of level parameters by univariate ARIMA models.

In an analogous manner, multivariate time series models are applied by Hagnell (1991). As mentioned by many authors (see, e.g. Brouhns et. al. (2002), Pitacco (2004)), the estimated parameters through time are often strongly dependent so that univariate extrapolations may be misleading. Building a multivariate time series for

the parameters is theoretically possible but seriously complicates the approach and can lead to computational intractability. Alternatively, Renshaw et. al (1996) propose a two-explanatory variable model with two multiplicative terms: a graduation term and an age-specific trend adjustment term.

As discussed by Bell (1997), application of PCs analysis to mortality rates goes back at least to Ledermann and Breas (1959), who perform their analysis not on time series data, but with life table data from many countries, both developed and developing. Sinamurthy (1987) pursues a similar analysis with fertility data from different countries. Bozik and Bell (1987) develop a PC approach for time series forecasting of age-specific fertility rates. This approach has been extended by Bell and Monsell (1991), where it is used in forecasting age-specific mortality rates. Lee & Carter (1992) explore a modified version of this approach for forecasting mortality rates, which Lee (1992, 1993) has also applied to fertility rates. Lee and Carter propose a simple model for describing the secular change in mortality as a function of a single time index:

$$z_{tx} = \log(m_{t,x}) - \sim_x = \ _x \cdot (kappa)_t + £_{t,x}$$

where $_x \cdot (kappa)_t = s \cdot u(x) \cdot v(t)$ and $s, u(x), v(x)$ denote the singular value and respective left and right singular vectors of z_{tx}. The time component $(kappa)_t$ is an index of the level of mortality, the 'dominant temporal signal' and captures the overall time trend in log mortality at all ages, $\sim_x$ is a set of age-specific constants describing the general pattern of mortality by age, $_x$ is a set of age-specific constants describing the relative speed of change at each age (and related to the proportion of the change in overall log mortality attributable to age *x*), and $£_{t,x}$ is the residual at age *x* in year *t* and denotes the deviation of the model from the observed log-central death rates and is assumed to be a white noise process with 0 mean and relatively small variance (Lee, 2000).

The main statistical tool is least-squares estimation via singular value decomposition (SVD) of the matrix of the log age specific observed forces of mortality, applying the constraints $\sum_t (kappa)_t = 0$, $\sum_x \ _x = 1$ in order to ensure identifiability of the model, and estimating the parameter $\sim_x$ by the average log-central death rates at each age over time, together with Box-Jenkins ARIMA modeling of the time series component $(kappa)_t$. We note that in the literature, prominence is given to the random walk with drift time series models, as initially advocated by Lee and Carter (1992). Once $_x$ and $(kappa)_t$ are estimated, under the above constraints, $(kappa)_t$ is further adjusted to ensure that the actual total deaths are identical to the total expected deaths for each calendar year within the data set, as a basis for comparing actual and expected deaths.

The Lee and Carter (1992) and Bozik and Bell (1987) approaches can be viewed as being similar. They differ in the way that Lee-Carter approach first subtracts out age-specific means ($\sim_x$) and uses one PC. As discussed by Bell (1997), given that a one-PC approximation is to be used, subtracting out the age-specific means is definitely recommended. Since the use of two PCs provides, by definition, the best two-dimensional linear approximations, the Lee and Carter approximation has an accuracy that lies between the accuracy provided by approaches using one and two PCs obtained without removing means. Subtracting the means is useful with low-dimensional PC approximations, but it becomes less important when more PCs are used, and is unnecessary if all of the PCs are used as in Bell and Monsell (1991).

As has been mentioned by many authors, several weaknesses arise in connection with LC method. The LC method implicitly assumes that the errors are homoskedastic. This assumption, which follows from the ordinary least squares estimation method used as the main statistical tool, seems to be unrealistic, as the logarithm of the observed mortality rate is much more variable at older ages than the younger ages, because of the much smaller number of deaths observed at old and very old ages (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). Tabeau et al (2001) noted that the LC model can be viewed as a statistical association model for a two-way cross-classification table, and as such belongs to the class of generalized linear modeling and can be

estimated by maximum likelihood estimation for Poisson-distributed errors. Brouhns et. al. (2002) and Renshaw and Haberman (2003a) have each implemented similar alternative approaches to mortality forecasting, based on heteroscedastic Poisson (non-additive) error structures.

Violations also exist in the LC model from the assumption that the age component $_x$ is invariant over time, and that the mortality improvement at all ages will follow a fixed pattern (Lee and Miller 2001). In the LC method, the SVD estimates reduce to, in essence, simple means and standard deviations and these estimates are insensitive to the more subtle aspects of mortality trends and it has been argued that they form a crude basis for mortality projections (Jong & Tickle, 2006). In fact, the mortality experience of the industrialized world over the course of the twentieth century would suggest substantial age-time interactions: the two dominant trends affecting different age groups at different times (Booth et. al. 2002). Several modifications have been proposed to cope with these limitations of the LC model. Renshaw and Haberman (2003a) incorporate the age differential effects, introducing a double bilinear predictor structure, (age-specific enhancement) into the LC forecasting methodology by optimizing the Poisson likelihood, as opposed to optimizing the Gaussian likelihood, as under the LC approach, and compare the results. The LC approach fails to capture and project the most recent upturn in crude mortality rates, roughly in the age range 20-39 years, as established in the raw data over the last quarter of the century and widely attributed to increases in the number of suicides and deaths related to HIV infection and AIDS. This is particularly noticeable in the case of UK male assured lives (with policies of duration 2+ years) where the 1st singular value accounts for 74% of the total variance. However, while the fitting of such augmented systematic model structures is straightforward, the generation of the subsequent univariate time series forecasts, based on such structures, is potentially problematic (Renshaw and Haberman 2003a). In many developed countries (including UK, US, 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 than the basic LC model. Similarly, Yue et al (2008) apply the Bell (1997) approach to several countries, and propose a jump model to the mortality rates, including two second-order interaction terms (age-period and age-cohort). They suggest that the reduction shift of ages for different time periods can be treated as a "cohort" effect, introducing an age-period-cohort (APC) model. Also, Hyndman and Ullah (2005) use several PCs 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 principal component analysis.

Further, the LC model assumes that the logarithms of the mortality rates are approximately a linear function of time (and the mortality rates of all ages eventually go to 0). Several authors indicate a marked departure from linearity in the dominant time component, casting doubt on the 'universal pattern' of decline and hence the general applicability of the linear forecasting model. Thus, some authors prefer to have a restricted fitting period. Thus, Booth et. al. (2002) propose a method for determining the optimal fitting period in order to address non-linearity in the time component.

Recently, many authors have proposed new approaches to mortality forecasts, utilizing (nonparametric) smoothing. 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.

Under our proposed approach, we build a generalized linear model, modelling the actual number of deaths as independent realizations of Poisson random variables, for each calendar year (or cohort) independently, treating age as an explanatory variate, and using a log link function and (orthonormal) polynomial predictor structure in age effects. The polynomial approach in modelling age patterns of mortality can be justified by the fact that most functions can be approximated by a polynomial to any degree of accuracy in the form of a Taylor power series. Thus, we utilize a parametric smoothing approach and we graduate the crude mortality rates with orthonormal polynomials in age effects, treating calendar year (or cohort) as a factor. Under this approach, we condense the information contained in the past experience into a set of critical parameters (parameter estimates),

which contain as much information as possible and explain the inherent variation of the mortality rates. The modelling of the estimated parameters trends in time is accomplished by applying PC analysis to the table of the parameters estimates, where each row denotes the number of years, and each column denotes the number of fitted parameters needed for each year. In this way, we reduce the dimensionality of the problem by focusing on the linear combinations of the estimated parameters leading to the PC terms. These are then used for modelling and forecasting purposes. We utilize dynamic linear regression (dlr) models, for modelling and forecasting the PCs. The statistical treatment of the dlr models is based on the state space framework and the Kalman filter.

The remainder of the paper is organised as follows. In section 2, we analyse the methodology proposed for the parameterized approach of modelling central mortality rates and we define the particular class of dlr models which we utilize for forecast purposes. In section 3, we illustrate two applications based on Sweden and Greece mortality experiences and incorporate an ex-post study to define the optimum fitting period for the long-historical mortality data from Sweden. In section 4, we offer a discussion and some concluding remarks.

2. Methodology

The data for analysis, which are denoted by (d_{xt}, R_{xt}), comprising the observed number of deaths, d_{xt}, with matching central exposures to the risk of death, R_{xt}, defined over rectangular data grids (*t*,*x*), with *t* ranging over the individual calendar years range $[t_1, t_n]$ and *x* ranging over the (grouped) age range $[x_1, x_n]$. If we assume that for any (grouped) age *x* and calendar year *t*, the force of mortality is constant, i.e. $\mu_{x+c}(t) = \mu_x(t)$ for $0 \le c < g$, where *g* is the width of the age grouping (for integer ages *g*=1), then the force of mortality is identical with the central rate of mortality: $\mu_x(t) = {}_g m_x(t)$. We model, the response variates, the actual number of deaths as independent realizations of Poisson random variables, D_{xt}, conditional on R_{xt}, i.e. $D_{xt} \sim Poisson(R_{xt} \cdot \mu_x(t))$ (Brillinger, 1986), for each calendar year independently. In this way, the calendar time enters the model as a factor.

In many actuarial mortality investigations, the data available do not consist of the actual deaths and the exposures based on individual lives. In studies of insured lives or annuitants for example, each policyholder may have more than one policy and any claim may subsequently give rise to more than one "death". The actual data available, for this kind of investigation, are the number of policies, ceasing through death and the corresponding exposed to risk based on policies. Similarly, overdispersion can occur in modelling data when sampling a Poisson process (Cox, 1967) over an interval whose length is not fixed, but is itself random (the alternative Gamma distribution for the central exposures is discussed in Renshaw et. al., 1997). Therefore, in these situations, a simple Poisson process does not describe the real process under which the data are generated. In the context of a Poisson GLM, this feature is described as over-*dispersed* Poisson process. Renshaw (1992), describes a methodology of joint modelling of the mean and of the dispersion, using the over-dispersed Poisson model for policies, such that

$$E(D_{xt}) = m_x(t) \cdot R_{xt} \quad \text{and} \quad Var(D_{xt}) = \varphi_t \cdot E(D_{xt})$$

where the over-dispersed parameter φ_t is independent of the response variate, and is the theoretical equivalent of the empirical variance ratio *r* discussed by Forfar et al (*1988*) and can be estimated by the ratio of the quasi-deviance divided by the associated degrees of freedom. If φ_t is treated as a constant, it does not have any effect on the values of the parameter estimates in the linear predictor, only on the standard errors and confidence intervals, which might differ seriously if this effect is not taken into account (Cox, *1983*). Renshaw (*1991*) describes the implementation of $\mu_x(t)$ graduations, using generalized linear model techniques, based on these distributional assumptions. Using log link predictor formulae, which is the canonical link for the Poisson distribution; the estimates are unique. The predictor structure is:

$$\eta_x(t) = \log(E(D_{xt})) = \log(R_{xt} \cdot m_x(t)) = \log(R_{xt}) + \log(m_x(t))$$

in which the $\log(R_{xt})$ term is treated as an offset. This implies that the rates of mortality are modelled as an exponential function of age effects. Treating age as an explanatory variate, we model the linear predictor, for

each calendar year independently, using an orthonormal polynomial structure in age effects. That is, we investigate predictor structures of the type $\sum_{j=1}^{k} b_{j-1}(t) \cdot L_{j-1}(x)$, where $L_{j-1}(x)$ denotes an orthonormal, (zero-centered for *j=2,…,k*), polynomial of degree *j-1*, for *j=1,2,…,k*, leading to *graduated* mortality rates:

$$\hat{m}_x(t) = \exp\left\{ \sum_{j=1}^{k} \quad _{j-1}(t) \cdot L_{j-1}(x) \right\} \tag{2.1}$$

for each calendar year independently. The random variables $_{j-1}(t) = \hat{b}_{j-1}(t)$ are the fitted variables at time *t*, for *j=1,2,…,k*. With $\_(t) = \underline{\hat{b}}(t)$ we denote the multivariate random variable at time *t*, where $\underline{b}(t) = (b_0(t), b_1(t), \cdots, b_{k-1}(t))' \in R^k$ the unknown vector of the parameters. The integrals, which are needed for the calculation of the norm of the polynomials, are computed using discretization. Under this modelling, the design matrix $\mathbf{L} = (L_0(x), L_1(x), ..., L_{k-1}(x))$ (of order *x* by *k*) consists of *k*-dimensional orthonormal vectors. We note that, in comparison with LC-methodology, this modelling ensures that the actual total deaths are identical to the total expected deaths for each calendar year.

The choice of employing orthonormal polynomials to form the linear predictor's vector basis is justified by the fact that the standard errors of the multivariate random variable $\_(t)$, for each calendar year *t*, are on the same scale, i.e. the standard errors of the parameters have about the same values, and this assists the use of PC analysis on the variance-covariance matrix of the $\_(t)$. We note that, the variance-covariance matrix has the form $Var(\_(t)) = (\mathbf{L}'\mathbf{V}_t^{-1}\mathbf{L})^{-1}$, where $\mathbf{V}_t = diag(u_{xt})$, $u_{xt} = Var(D_{xt}) \cdot dr_{xt} = \hat{\cdot}_t \cdot E(D_{xt}) \cdot dr_{xt} = \hat{\cdot}_t \cdot R_{xt} \cdot \hat{m}_x(t) \cdot dr_{xt}$, and dr_{xt} the associated deviance increments defined us $dr_{xt} = \hat{\cdot}_t \cdot \left[d_{xt} \cdot \log(\frac{d_{xt}}{R_{xt} \cdot \hat{m}_x(t)}) - (d_{xt} - R_{xt} \cdot \hat{m}_x(t)) \right]$.

A restriction for the standard errors of the estimated parameter vector $\_(t)$ in time should be that they must represent only random fluctuations and not be associated with any model risk or any sampling errors. Thus, assuming that the model fit is adequate, and from the fact that the standard errors depend on the size of the expected deaths, then the total central exposures, in each calendar year under investigation must have close values, or equivalently the population must show a relative stationarity.

There are several methods suggested in the literature for determining the optimum degree *k* for the orthonormal polynomials, for example the Akaike Information Criterion (AIC) : $AIC = 2 \cdot k - 2 \cdot log(L)$, the Bayesian (or Schwarz) Information Criterion (BIC) $BIC = -2 \cdot log(L) + k \cdot log(n)$, where *L* is the maximum value of the likelihood, *k* is the number of parameters and *n* is the number of observations. Alternatively, we can use the deviance profile criterion which depends on the difference between the scaled deviances of two nested models following a chi-square distribution, with degrees of freedom equal to the difference between the independent parameters estimated under the two models, and we can then examine *p*-values under the hypothesis that the nested model is the correct one (Kendall and Stuart, 1967). For all GLMs (except for the linear model with the Normal distribution), the difference in the scaled deviance is known to be distributed as a chi-square only asymptotically, and rather little is known about how good the asymptotic approximation is for small sets of data. However, under the GLM approach, we can investigate the optimum (parsimonious) degree *k* by examining the appropriateness of the scaled deviance residuals, in age and time effects, in association with various statistical tests (chi-square test, run test, sign test). The inclusion of more parameters does not alter significantly the parameter estimates of the nested models, due to the orthogonality of the design matrix. Certainly, heavily over parameterised models should be avoided, in order to remove only random fluctuations

from the crude mortality rates and to justify the cross-sectional graduation process, and also if we need a parsimonious parameterized model structure.

As has been pointed out (see Pitacco, 2004), the uncertainty in modelling and forecasting mortality rates is attributable to random fluctuations, "process risk", uncertainty in estimating the values of the parameters, "parameter risk" or to uncertainty in the choice or structure of the model, "model risk". Under the proposed parameterized approach, in order to minimize the "model risk", we can investigate the appropriateness of the 'law of mortality' by various statistical tests (chi-square test, run test, sign test) and utilize scaled deviance residuals.

Fitting an orthonormal polynomial structure in age effects, for each calendar year, we produce a matrix, of order *n* by *k*, of $\beta_{j-1}(t)$ entries: $\mathbf{B} = \{\beta_{j-1}(t)\}$, for *t*=1,2,… *n* and *j=1,2,…,k.* and a matrix, of order *n* by *x*, with entries $\hat{m}_x(t)$: $\mathbf{M} = \exp\{B \cdot L'\}$ or $\log(\mathbf{M}) = B \cdot L'$, for *t*=1,2,… *n* and $x = x_1, \ldots, x_2$.

Noting the properties of GLMs, each cross-sectional vector of the estimated parameters is a *k*-dimensional random variable which follows asymptotically a multivariate *k*-dimensional normal distribution: $\_(t) \sim N_k(\underline{b}(t), \mathbf{\Sigma}_t)$, where $\mathbf{\Sigma}_t$ is the associated covariance matrix, at each different calendar year *t*. Let $\_{j-1} := (\ _{j-1}(1),\ _{j-1}(2), \cdots,\ _{j-1}(n))'$ denote the random process, for each *j=1,2,...,k.*

In this context, the modelling of the estimated parameters trends in time is accomplished by applying PC analysis (in association with the covariance matrix) to the table of the fitted parameters $\mathbf{B} = \{\beta_{j-1}(t)\}$, where each row denotes the number of years, and each column denotes the number of fitted parameters needed for each year. We usually use the rescaled vectors, say $\_^r_{j-1} = \_{j-1} - \bar{\ }_{j-1}$, for each *j=1,2,...,k,* where $\_{j-1} = (\ _{j-1}(1),\ _{j-1}(2), \cdots,\ _{j-1}(n))'$ and $\bar{\ }_{j-1}$ the arithmetic mean value of the $\_{j-1}$ vector (treating them as constants), producing the rescaled matrix $\mathbf{B}^r = \mathbf{B} - \underline{\bar{\beta}}' = \{\beta^r_{j-1}(t)\} = \{\beta_{j-1}(t) - \bar{\beta}_{j-1}\}$, where $\underline{\bar{\beta}} = \left(\bar{\beta}_0, \bar{\beta}_1, ..., \bar{\beta}_k\right)'$ is the vector of the mean values, for each *j=1,2,...,k* and for each calendar year *t=1,2,…,n,* with associated covariance matrix **Σ**.

PC analysis involves the computation of the singular value decomposition of a data set, usually after mean centering the data for each attribute. PC analysis can be used to reduce the dimensionality of a data set by retaining those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones (Mardia et. al. 1997). Such low-order components often contain the "most important" aspects of the data and we keep a "small" subset of them, say *p(<k),* which explains the "majority" of the variances (see section 3.4). Thus, we apply eigenvalue decomposition, to the covariance matrix **Σ,** with the associated matrix of eigenvectors $\mathbf{P} = [\underline{e}_1, \underline{e}_2, \cdots, \underline{e}_k]$ and vector of eigenvalues $\underline{\Lambda} = \left[\mathsf{a}_1, \mathsf{a}_2, ..., \mathsf{a}_k\right]'$. The decomposition produces a matrix of PC scores, of order *n* by *k*: $\mathbf{Y} = \mathbf{B}^r \cdot \mathbf{P}$, leading to the equations $\underline{Y}(t) = \mathbf{P}' \cdot \_^r(t)$, where $\underline{Y}(t) = \left(Y_1(t), Y_2(t), ..., Y_k(t)\right)'$ the vector of the PC scores in year *t*, and $Y_i(t) = \_^r(t)' \cdot \underline{e}_i$ for each calendar year *t=1,2,…,n* and each PC *i=1,2,…,k*. We note that the vectors of PC scores, $\underline{Y}(t)$, are normally distributed since they are linear combinations of the *k*-dimensional normally distributed vector random variables $\_^r(t)$. Thus, the result of the principal components analysis to the fitted GLM parameters is the creation of a new multivariate random process, the principal component random vectors, $\underline{Y}(t)$, which are normally distributed, and then we examine the PCs for possible trends in time. The PC analysis ensures that $Cov(Y_i, Y_l) = 0$, for $i \neq j$, and, from the normality property, we obtain that the stochastic processes $Y_i(t)$, for *i=1,2,…,k,* are independent each other.

The vector of the graduated log-central death rate at each age in year *t* is then of the form

$$\log(\hat{\underline{m}}_t) = \mathbf{L}\cdot \_(t) = \mathbf{L}\cdot\left(\bar{\_} + \_^{r}(t)\right) = \mathbf{L}\cdot\bar{\_} + \mathbf{L}\cdot\_^{r}(t) = \mathbf{L}\cdot\bar{\_} + \mathbf{L}\cdot\mathbf{P}\cdot\underline{Y}(t) = \underline{A} + \mathbf{G}\cdot\underline{Y}(t) \qquad (2.2)$$

where $\underline{A} = \mathbf{L}\cdot\bar{\_}$ or $A(x) = \underline{L}'(x)\cdot\bar{\_}$, $\mathbf{G} = \mathbf{L}\cdot\mathbf{P} = \left(\underline{g}_1, \underline{g}_2, ..., \underline{g}_k\right)$ with columns $\underline{g}_i = \mathbf{L}\cdot\underline{e}_i$ and values $g_i(x) = \underline{L}'(x)\cdot\underline{e}_i$, and where $\underline{L}(x) = \left(L_0(x), L_2(x), ..., L_k(x)\right)'$ denotes the vector of the design matrix's *x*-row. We note that the matrix of eigenvectors $\mathbf{P}$ is a projection matrix which transforms the matrices $\mathbf{B}^r$ and $\mathbf{L}$ into PC scores and age-specific scores respectively. Thus, we derive the structure

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

Keeping a "small" subset *p(<k)* of the PCs, which explains the "majority" of the variance (see section 3.4), leads to:

$$\log(\hat{m}_x(t)) = A(x) + \sum_{i=1}^{p} g_i(x)\cdot Y_i(t) + £_x(t) \qquad \text{(model 1)}$$

The disturbance term $£_x(t) \sim N(0, v_x)$, is the error component at age *x* in year *t* and denotes the deviation of the model represented by the excluded PCs, which are normally distributed with zero mean and variance v_x, estimated by $\hat{v}_x = \sum_{i=p+1}^{k} g_i(x)^2 \cdot ª_i$. The model that we have derived can be viewed as a variant of the Lee-Carter model with *p* interaction terms. According to the LC method, the first PC, $Y_1(t)$, denotes the index of the level of mortality that captures the overall time trend in log mortality at all ages. In that case, the time components $Y_i(t)$ are stochastic Gaussian processes, which are linear combinations of the (rescaled) cross-sectional fitted parameters and the eigenvectors. $A(x)$ is a set of age-specific constants which are linear combinations of the mean fitted parameters and rows of the design matrix, describing the general pattern of mortality by age. $g_i(x)$ is a set of age-specific constants which are linear combinations of the eigenvectors and the rows of the design matrix. The first function $g_1(x)$ denotes the proportion of the change in overall log mortality attributable to age *x*. The remaining interaction terms $g_i(x)\cdot Y_i(t)$, for *i=2,…,p*, incorporate the age differential effects, including age-period interaction terms. The importance of each component can be measured by the PC variances $ª_i$, for *i=2,…,p*, which define the goodness of approximation as judged by the ratios $l_p^2 = \frac{ª_1 + ... + ª_p}{ª_1 + ... + ª_k}$, where *p* the number of components used and $ª_1 \geq ... \geq ª_p$. The PCs can be considered also as 'factors' (common characteristics as analysed in factor analysis) that describe certain features of the mortality process. These 'factors' can be labelled in agreement with the age components $g_i(x)$.

From the non-reduced model (1), after some simple algebraic manipulations we obtain:

$$\log(\hat{m}_x(t)) = «_x + \sum_{j=1}^{k}\left(\sum_{i=1}^{k} e_{j,i}\cdot Y_i(t)\cdot L_{j-1}(x)\right) = «_x + \sum_{i=1}^{k}\left(e_{1,i}\cdot L_0(x) + \sum_{j=2}^{k} e_{j,i}\cdot L_{j-1}(x)\right)\cdot Y_i(t)\text{, i.e.}$$

$$\log(\hat{m}_x(t)) = «_x + b(t) + \sum_{i=1}^{k} f_i(x)\cdot Y_i(t) \Leftrightarrow$$

$$\log(\hat{m}_x(t)) = \alpha + \beta(x) + b(t) + \sum_{i=1}^{k} \kappa_i \cdot \gamma_i(x) \cdot \mu_i(t) \qquad (2.3)$$

where $\alpha = \bar{\beta}_0 \cdot L_0$, $\beta(x) = \sum_{j=2}^{k} \bar{\beta}_{j-1} \cdot L_{j-1}(x)$, $b(t) = L_0 \cdot \sum_{i=1}^{k} e_{1,i} \cdot Y_i(t)$,

$$Y_i(t) = \sum_{j=1}^{k} e_{j,i} \cdot \kappa_{j-1}(t), \qquad f_i(x) = \sum_{j=2}^{k} e_{j,i} \cdot L_{j-1}(x),$$

$$\gamma_i(x) = \frac{f_i(x)}{\sqrt{\sum_{x=x_1}^{x_n} f_i^2(x)}}, \quad \mu_i(t) = \frac{Y_i(t)}{\sqrt{\sum_{t=t_1}^{t_n} Y_i^2(t)}} = \frac{Y_i(t)}{\sqrt{\lambda_i \cdot (n-1)}} \quad \text{and} \quad \kappa_i = \sqrt{\lambda_i \cdot (n-1)} \cdot \sqrt{\sum_{x=x_1}^{x_n} f_i^2(x)}$$

The model now can be viewed as a member of the class of log-linear models, namely a hierarchical statistical association model for a two-way cross-classification table with age and time being the two main random effects, similar to the so-called (unweighted) *association model* of Goodman (1991).

The component α is the overall mean, the components $\beta(x)$ and $b(t)$ are the age main effect and time main effect, which comprise the independent or additive model, and the $\gamma_i(x)$ and $\mu_i(t)$ are viewed as scores relating to the x^{th} age and t^{th} calendar year categories respectively, which embody the interactions between age and time. $\beta(x)$ gives the age profile and represents deviations from the overall mortality mean α, that are attributable to age *x*, and $b(t)$ is an index of the level of mortality in time effects that captures the overall time trend in log mortality at all ages (with values that represent deviations from the overall mortality mean that are attributable to calendar year *t*). The components $\gamma_i(x)$ describe relative deviations from the general age pattern of mortality (the independent component) and they indicate the sensitivity of the logarithm of the force of mortality at age *x* to variations and trends in the time index $\mu_i(t)$. They modify the main age profile and represent the age-specific patterns of mortality changes. The shape of the $\gamma_i(x)$ profile tells which rates respond rapidly and which slowly over which period of time, in response to particular trends in $\mu_i(t)$. For negative $\gamma_i(x)$ values, increasing (decreasing) values of $\mu_i(t)$ represent a faster rate of improvement (deterioration) relative to independent model, and for positive $\gamma_i(x)$ values, increasing (decreasing) values of $\mu_i(t)$ represent a faster rate of deterioration (improvement) relative to independent model with relative degree of deterioration (improvement) as indicated by the first derivative of the $\mu_i(t)$'s.

For identifiability reasons, the age and time scores are zero-centered normed scores subject to the constraint that the age and time scores are internally orthogonal:

$$\sum_{x=x_1}^{x_n} v_i(x) = 0, \quad \sum_{x=x_1}^{x_n} v_i^2(x) = 1, \quad \sum_{t=t_1}^{t_n} \mu_i(t) = 0, \quad \sum_{t=t_1}^{t_n} \mu_i^2(t) = 1,$$

$$\sum_{x=x_1}^{x_n} v_i(x) \cdot v_j(x) = 0 \quad \text{and} \quad \sum_{t=t_1}^{t_n} \mu_i(t) \cdot \mu_j(t) = 0 \quad \forall i \neq j \qquad (2.4)$$

The parameters κ_i (see formulae 2.3) are the intrinsic association parameters that measure the importance of the age-time interaction scores in explaining the deviations from the independent model. The parameters κ_i incorporate the age variations, in comparison with the square root of the variances λ_i that take account only of

the time variations. The ¤$_i$ parameters are singular values of the matrix of the residuals (from the independent model) D, defined by $D=(d_{xt})=\left(\log(\hat{m}_x(t))\text{-}\ll -\sim(x)-b(t)\right)$, with associated singular vectors ¬$_i(x)$ and µ$_i(t)$. The eigenvalues parameters ¤$_i^2$, can be used as an index of the importance of each interaction term and define the goodness of approximation as judged by the ratios $r_p^2=\frac{¤_1^2+...+¤_p^2}{¤_1^2+...+¤_k^2}$, where *p* is the number of interaction terms used : $¤_i^2\geq...\geq¤_p^2$. The ratio offers a guide as to how many interaction terms are to be included in the model, applying a similar methodology from PC analysis. A possible criterion for the optimum choice of the number of eigenvalues could be to look at the pattern of eigenvalues (a scree plot) and see if there is a natural breakpoint, or to identify clusters of eigenvalues and keep those clusters that explain the majority of the deviations from the additive model. A final choice would be motivated by the possible interpretations of the associated ¬$_i(x)$ and µ$_i(t)$ values. After retaining the most important interaction terms, we derive the model structure:

$$\log(\hat{m}_x(t))=«+\sim(x)+b(t)+\sum_{i=1}^{p}¤_i\cdot¬_i(x)\cdot\ µ_i(t)+£_x(t) \qquad \text{(model 2)}$$

Thus, forecasts for the µ$_i(t)$ time series (or equivalently, for the $\hat{Y}_i(t)$ time series) will produce forecast values for the mortality rates. Any forecast stochastic model for the PCs can lead to predicted mean values, say $\underline{\hat{Y}}(t)$, for each calendar year, with associated estimated variance-covariance matrices $Var(\underline{\hat{Y}}(t))$. From the predicted mean values of the PCs, we can obtain the 'predicted values' of the original parameter estimates from the mathematical relationship: $\hat{\mathbf{B}}^r=\hat{\mathbf{Y}}\cdot\mathbf{P'}$, under the assumptions that the eigenvector matrix $\mathbf{P}$ will not change its structure during the forecasting period. From the Cayley-Hamilton Theorem we have, for the relationship $\mathbf{\Sigma=P\Lambda P'}$, that, if a real function *f(x)* is given by a power series: $f(x)=a_0+a_1\cdot x+a_2\cdot x^2+...$ then $f(\mathbf{\Sigma})=\mathbf{P}f(\mathbf{\Lambda})\mathbf{P'}$. This property implies that, any changes in the covariance matrix $\mathbf{\Sigma}$, during the forecast period, of the form $f(x)=a_0+a_1\cdot x+a_2\cdot x^2+...$, will not affect the matrix of eigenvectors $\mathbf{P}$. It seems, under this property, that the predicted values $\hat{\mathbf{B}}^r$ (and also the predicted $\hat{\mathbf{Y}}$ values) have a degree of robustness in terms of structural changes of the covariance matrix $\mathbf{\Sigma}$ during the forecast period. The linear transformation $\hat{\mathbf{B}}^r=\hat{\mathbf{Y}}\cdot{}^r\mathbf{P'}$, where the matrix ${}^r\mathbf{P}$, of order *k* by *p*, denotes the reduced form of the matrix of eigenvectors by keeping only the first *p* eigenvectors, leads to the set of equations $\hat{\ }_{\_}^{r}(t)={}^r\mathbf{P}\cdot\underline{\hat{Y}}(t)$ for each calendar year *t*, which form a new, *k*-dimensional, normally distributed, random process. Thus, alternatively, the predicted log-mortality can be viewed as a one-factor parameterized model, where the time series are the fitted parameters:

$$\log(\hat{\hat{m}}_x(t))=\sum_{j=1}^{k}\hat{\ }_{j-1}(t)\cdot L_{j-1}(x) \qquad \text{(model 3)}$$

where $\hat{\ }_{j-1}(t)=\sum_{i=1}^{p}e_{j,i}\cdot\hat{Y}_i(t)+\bar{\ }_{j-1}$. The estimated variance-covariance matrices of the cross-sectional estimated parameter vectors, for each calendar years *t* are

$$Var\left[\hat{\ }_{\_}^{r}(t)\right]={}^r\mathbf{P}\cdot Var\left[\underline{\hat{Y}}(t)\right]\cdot{}^r\mathbf{P'}={}^r\mathbf{P}\cdot diag\left(Var\left[\hat{Y}_1(t)\right],...,Var\left[\hat{Y}_p(t)\right]\right)\cdot{}^r\mathbf{P'} \qquad (2.5)$$

and the estimated variance of the predicted log-mortality rates structure, for any grid of values (x,t) are

$$Var\left[\log(\hat{\hat{m}}_x(t))\right]=Var\left[\underline{L}'(x)\cdot\hat{\ }_{\_}(t)\right]=\underline{L}'(x)\cdot Var\left[\hat{\ }_{\_}(t)\right]\cdot\underline{L}(x) \qquad (2.6)$$

If the random vectors $\hat{\underline{Y}}(t)$ belong to the class of *linear estimators* in the observations $Y_i(t)$, then the random vectors $\hat{\_}(t)$ and $\log(\hat{\hat{m}}_x(t))$ are also asymptotically normally distributed, and confidence intervals can be derived.

Also, from the predicted mean values and the estimated variances of the PCs, we can obtain the 'predicted values' and the estimated variances of the time main effects, under the independence assumption for the predicted mean values of the PCs:

$$\hat{b}(t) = L_0 \cdot \sum_{i=1}^{p} e_{1,i} \cdot \hat{Y}_i(t) \qquad \text{and} \qquad Var\left[\hat{b}(t)\right] = L_0^2 \cdot \sum_{i=1}^{p} e_{1,i}^2 \cdot Var\left[\hat{Y}_i(t)\right] \tag{2.7}$$

For forecast purposes, for each PC, we advocate a specific class of dynamic linear regression (dlr) models:

$$\hat{Y}(t) = a(t) + b(t) \cdot t + e(t) \tag{2.8}$$

for each calendar year *t* (the so-called regressor), with stochastic time variable parameters (level and slope) that follow an autoregressive-moving average process. Newbold and Bos (1985) argue that it is difficult to distinguish among different members of the autoregressive-moving average class of models to represent stochastic parameter behaviour, and their view is that, for the great majority of problems met in practice, a regression model with stochastic parameters following a first-order autoregressive process should provide an adequate representation of the available data, though on occasion one or more additional autoregressive terms may be required. Experiments with various mortality experiences have shown that the PCs can be represented adequately under the dlr model structures, with the stochastic parameters following a first order autoregressive process, for stationary PCs, or a random walk process, for non-stationary PCs:

$$a(t) = \ddot{}_1 \cdot a(t-1) + \dagger(t) \qquad \text{and} \qquad b(t) = \ddot{}_2 \cdot b(t-1) + ¸(t) \tag{2.9}$$

All of the components are stochastic and the disturbances driving them ($\pm_e^2$, $\pm_\dagger^2$ and $\pm_¸^2$) are assumed to be mutually uncorrelated. The crucial feature of the model is the signal-noise ratio (or noise-variance ratio hyper-parameters) $q_1 = \dfrac{\pm_\dagger^2}{\pm_e^2}$ and $q_2 = \dfrac{\pm_¸^2}{\pm_e^2}$.

If the correlation coefficients $\ddot{}$ are less than 1 in absolute value, then the model is called a first-order autoregressive model, where the strength of the correlation between two values of the time series is decreasing as their distance apart in time increases. If $\ddot{}$ is equal to 1 the model is called a 'random walk' model (the "border line non-stationary" case) and if both the $\ddot{}$ coefficients are equal to one then the model reduces to the *local linear* model with reduced form - an ARIMA(0,2,2) process (Harvey, 1991). Random walk models are best thought of as describing a situation in which changes from one period to the next are white noise and are therefore unpredictable in terms of past changes. In this case, we assume that the (Gaussian) PCs incorporate local linear trends, and we remove those linear trends in time using dlr models and Kalman filter techniques. In the special case when $\pm_¸^2$=0, the model reduces to an ARIMA(0,1,1) model plus constant term (Harvey, 1991). This particular case has been found to be very popular in modelling the PCs trends. However, it is possible that $\ddot{}$ exceeds one in absolute value, in which case the model is called "explosive", but these models are not useful for representing real data in the social sciences (Newbold and Bos, 1985).

The effect of $\dagger(t)$ is to allow the level of the trend to shift up and down, while $¸(t)$ allows the slope to change. The larger the variances, the greater the stochastic movements in the trend. If the white noise of the stochastic parameter is equal to zero, then we consider the corresponding parameter constant and, if both the white noises of the stochastic parameters are equal to zero, then the model reduces to an ordinary linear regression model.

The computations have been implemented in Matlab using the Captain Toolbox (Taylor, 2007).

Dynamic linear regression models are simple models from the wider class of models known as principal structural time series models. Structural time series models are set up in terms of components, such as trends and cycles, which have a direct interpretation. The principal univariate structural 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).

Renshaw & Haberman (2003b) use a GLM approach and model time as a known covariate, having first established a well-defined origin. They introduce a break-point or hinge so that greater flexibility is achieved in capturing the more recent age-specific time trends. The pattern in the estimated k_t (the 1st PC) in the LC model essentially comprises two linear segments, hinged in the mid-1970s, for England & Wales males. Yue et al (2008) propose a jump model for the mortality rates, and include a cut-off point (or jump) for the second-order interaction term using an ordinary regression approach. They find that the modified method achieves a much lower Mean Absolute Percentage Error (MAPE) compared with the LC model (which utilizes only the 1st PC) and they also consider the possibility of three or more interaction terms. Both of these papers utilize linear spline modelling, the first approach applying linear splines to the 1st interaction term and the second method using linear splines for the 2nd interaction term. Further, Sithole et al (2000) discuss the desirable characteristics of projected mortality profiles. There are comments in favour of an extrapolation, which is linear, or approximately linear, possibly after a logarithmic transformation. They argue that the use of splines with higher degrees (eg quadratic or cubic) could lead to extrapolated trajectories with turning points, which could give distorted projections.

The proposed modelling in the current analysis can be viewed as a generalization, in a dynamic context, of the above approaches. Each PC can be modelled with dynamic linear trends. The dlr modelling utilizes all of the historical mortality data, and so longer term forecasts can be produced.

In our analysis, we obtain estimates of the period-based expected remaining lifetime, $e_x(t)$, by constructing a life table for each calendar year. In the general case, where the table is based on age grouping (called an abridged life table), under the assumption of the constant force of mortality (CFM), then for any (grouped) age *x* and calendar year *t* the force of mortality is constant: $\mu_{x+c}(t) = \mu_x(t)$ for $0 \le c < n$, where *n* is the width of the age grouping, and $\mu_x(t) = {}_n m_x(t)$. The probability that an individual from age-group *x*, in calendar year *t*, does not reach the next age-group is ${}_n q_x(t) = 1 - \exp(-n \cdot \mu_x(t))$. The hypothetical expected number of persons alive at the start of each *n*-year age interval *x*, at calendar year *t*, is given by the iterative formula $l_{x+n}(t) = l_x(t) \cdot (1 - {}_n q_x(t))$, with an arbitrary initial value $l_0(t)$, called the radix. The corresponding number of person-years ${}_n L_x(t) = n \cdot \left(l_x(t) + \left(1 - \frac{{}_n a(x)}{n} \right) \cdot (l_{x+n}(t) - l_x(t)) \right)$, where ${}_n a(x) = \frac{\left(\frac{{}_n q_x(t)}{\mu_x(t)} - n \cdot (1 - {}_n q_x(t)) \right)}{{}_n q_x(t)}$ is the average of the *n*-year period lived in the interval by those who died in that interval, under the CFM assumption. Therefore, the total future lifetime of the $l_x(t)$ persons who attain age *x*, in calendar year *t*, is $T_x(t) = \sum_{i \ge x} {}_n L_i(t)$ and the expected future remaining lifetime for individuals from age-grouping *x* is $e_x(t) = \frac{T_x(t)}{l_x(t)}$.

Associated confidence intervals (CI) for the expected remaining lifetime can be obtained using parametric bootstrapping, such that, taking into account the sampling errors in the GLM parameters and the forecast errors

in the projected PCs. Starting from the observations d_{xt} we simulate N bootstrap samples $\{d_{xt}^{(i)}\}$, $i=1,2,\ldots,N$, where $d_{xt}^{(i)}$ are realizations from the Poisson distribution with parameters (d_{xt}). For each bootstrap sample, the GLM parameters $\mathbf{B}^{(i)}$ are estimated and the associated PCs are projected on the basis of the dlr models selected from the original data. This yields N realizations for the expected remaining lifetimes and then the 95% CI are the percentile intervals $CI_{95} = [\, p_{0.025}, p_{0.975} \,]$, for each forecast year. The bootstrap confidence interval avoids the normal assumption and is more reliable than the standard normal interval (Efron and Tibshirani, 1998). The approach used corresponds to the semi-parametric bootstrap as described by Renshaw and Haberman (2008) and by Pitacco et al (2008).

3. Applications

3.1 The data

In order to illustrate the methodology, studies from two countries are conducted:

- Sweden male-female mortality experience, calendar years 1751-2006
- Greece male-female mortality experience, calendar years 1957-2006.

These countries have been selected because Sweden has the longest available mortality experience in contrast with the Greece mortality experience which is relatively short. The data are cross-classified by individual calendar year and age grouping ([0,5), [5,9),…,[95, →)) for Sweden and ([0,5), [5,9),…,[80, 85)) for Greece. Let $[x]:=[x,x+5)$, for $x=0,5,10,\ldots$. The data are freely provided, for Sweden by the "Human Mortality Database"(www.mortality.org) and for Greece by the General Secretariat of National Statistical Service of Greece (www.statistics.gr).

3.2 Optimal fitting period

We investigate model structures of the type $\log(\text{«}_x(t)) = \sum_{j=1}^{k} {}_{j-1}(t) \cdot L_{j-1}(x')$, for each calendar year separately, where $L_{j-1}(x')$ denotes *Legendre* (zero-centered for $j=2,\ldots,k$) orthonormal polynomials of degree $j-1$, for $j=1,2,\ldots,k$, and x' the transformed values for the (grouped) ages, so that $x' \in [-1,1]$. By monitoring the scaled deviance residuals in age and time effects, and examining the p-values of the associated statistical tests, the optimal *degree k* for Sweden males is found to be 15 and for females 14, and for Greece males and females the optimal degree is found to be 12. Especially for the case of Sweden, where the available data series is long, we find that these optimal degrees hold for different choices of investigation period.

Many authors (eg. Booth et al, 2006; Keyfitz, 1991) agree that forecast accuracy is highly dependent on the particular period or population. Various mortality investigations suggest there is evidence that the choice of the base period may not be negligible in forecasting. Booth et al (2002) determine the optimal fitting period, for Australian mortality data, in order to address non-linearity in the time component, under the Lee-Carter method. They note that a major problem with the Lee-Carter method is the assumption that the age component is invariant over time and for Australian data they find that the optimal fitting period is 1968-99, where the assumption of invariance holds fairly well. Many other mortality investigations suggest that the age component invariance applies to only the second half of the 20th century, i.e. since 1950 (Tuljapurkar et. al., 2000; Lee and Miller, 2001). Koissi et. al. (2006), conduct an ex-post study to measure the Lee-Carter model performance for Nordic countries. They study how well life expectancy at birth would have been predicted (monitoring the difference between actual and forecast) if the model was used in the past and hypothetical projections were made based on variable period intervals. They use 35 observation periods, all starting from year 1955: period(i)=[1955; 1963+i], i=1,…,35 and the results suggest that, if the time series of past data is chosen on a period long enough, the model would produce satisfactory predictions, in terms of the predicted life expectancy at birth for selected years (1975, 1980, 1990, 1999).

Thus, each different period interval could produce different model identification and so different forecasts. For Sweden mortality data, we find that the optimum dlr model, for all the PCs, is the model with fixed slope and a random walk process for the level (i.e. an ARIMA(0,1,1) model plus constant). This kind of model satisfies the statistical criteria when testing the effect of changing the fitting periods. Although we could utilize all of the available historic mortality data, we can define a process of statistically determining the optimal fitting period, in respect of how well the log-central mortality rates would have been predicted if the model were used for different starting calendar years. We conduct an ex-post study for different starting calendar years: period(i,y)=[i; y], for i=1751,1752,…,y-20 (20 is the number of different age groups in order to apply PC analysis) and y=1971,1976,1981,1986,1991,1996,2001. For each y, the model performance is evaluated with the mean square error (MSE), of the observed log-central mortality rates and the projected log-central mortality rates, for all the ages. Thus, for fixed y, each starting year i gives a MSE value evaluated for all of the ages and all of the calendar years [y+1; 2006], which are of length 35,30,25,20,15,10,5 respectively for y=1971,1976,1981,1986,1991,1996,2001. In order to combine the MSE values of the 7 different projected periods, we calculate the average values obtained for each starting calendar year i, averaging over for different y values. This procedure will produce an average MSE value (AMSE) for each starting calendar year i, from 1751 to 1951.

Figure 1 shows the AMSE for Sweden males and females, when the model structure consists of the first 3 most important interaction terms, and also, for comparison reasons, we demonstrate the AMSE results when the model structure consists of only the single most important interaction term. It is found that the optimum starting year, when utilizing the first 3 more important interaction terms, is 1869 for both males and females. In the case where we use only the 1st PC, for males the optimum starting year is 1857 and for females the minimum AMSE value is the last year (1951). However, in order to make long forecasts, we use as a starting year the calendar year 1809 where the AMSE exhibits a local minimum (very close to the value of 1951). Figure 1 shows these AMSE values. We note that for all of the starting calendar years, for both genders, the AMSE is lower when utilizing 3 or 4 interaction terms (PCs) rather than 1 interaction term (1 PC). That is, these ex-post investigations show that the method improves the forecasts for log-central mortality rates, when using more than 1 PC, as we would expect.

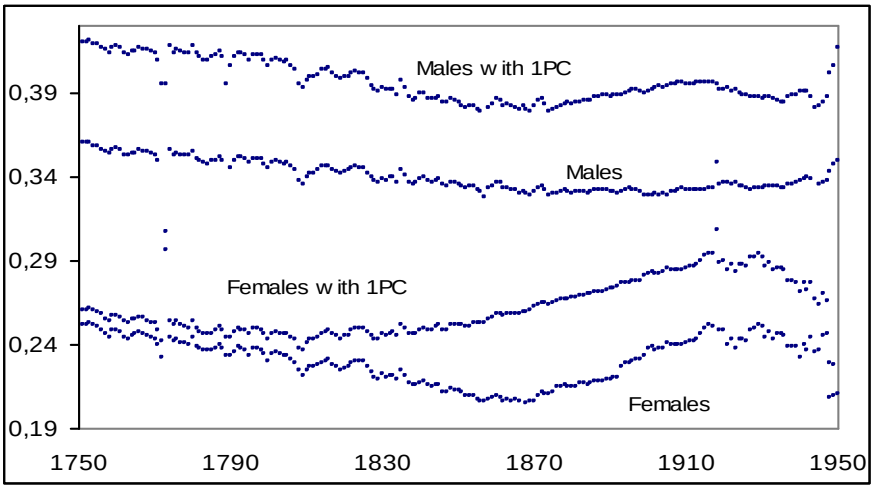


Figure 1: Average mean square error (AMSE) values vs. starting calendar year for Sweden

3.3 Residuals

Figure 2 shows the deviance residuals plotted against age and time, under the optimal fitting period 1869-2006 for Sweden and the fitting period 1957-2006 for Greece. The overall patterns of the total deviance residuals indicate identically distributed and independent residuals. The outliers in calendar years 1992-1996, in Greece, especially for males, refer mostly to age groups [45]-[55] and we believe that this is probably a cohort effect due to the World War II (1940-45) and possibly due to the Greek civil war (1946-49).

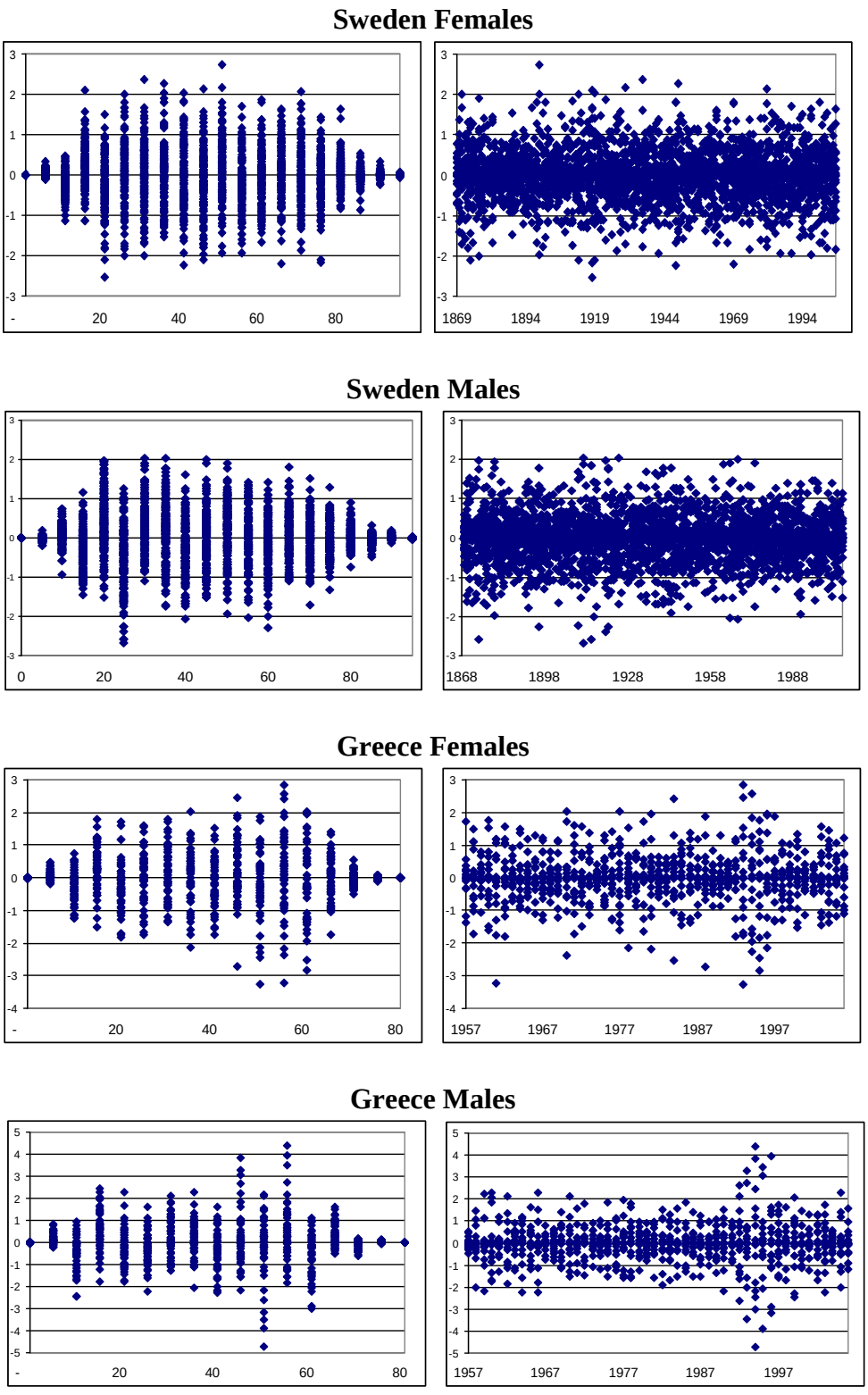


Figure 2: Deviance residuals vs. age and calendar year, for Sweden and Greece (Females-Males)

3.4 Model Components and Forecasts

Table 1, gives the eigenvalues (λ_i^2-values) based on model structure (2), with the associated percentage variance and cumulative variance (r_i^2-values) explained. After detecting the pattern of the λ_i^2 eigenvalues (scree plot) in association with the possible interpretations of the associated $\gamma_i(x)$ and $\kappa_i(t)$ values, we retain, for the case of Sweden, 3 interaction terms (explaining 97,4% of the residual variation for males and 98,5% for females) and for Greece we keep 4 interaction terms (explaining the 96,6% of the residual variation for males and the 92,9% for females).

Figure 3 displays the most important components, with associated forecasts and CIs under the dlr structure, in significant order, for Sweden and Greece, based on model structure (2). For Sweden, for both genders, the most important interaction term, $v_1(x)$ and $\kappa_1(t)$ values, shows high values (in absolute tems) for the very young and for the very old ages: this explains the deviations (from the additive model) and trends for the very young (mainly for age groups [0]-[10]) in contrast with the very old (above age group [80[). The positive slope for the $\kappa_1(t)$ trend in association with negative (positive) $v_1(x)$ values for the very young (old) ages indicates the relative improvement (deterioration) for these ages from the independent model. This term accounts for about 94% of the total residual variations. The 2nd most important interaction term explains the deviations and trends mainly for the age groups [15]-[30], for both males and females. The $\kappa_2(t)$ trend for these ages shows a relative deterioration from the independent model after the 1960s (negative slope after the 1960s in association with the negative $v_2(x)$ values). This term explains about 3,3% and 2,2% of the females and males total residual variations respectively. This feature reproduces the well known "accident hump", occurring at these ages

The 3rd most important interaction term ($\kappa_3(t)$ and $v_3(x)$ values) refers mainly to age groups [35],[40] for males, showing a relative improvement after the 1980s, and for females, it refers mainly to age groups [65]-[75] showing a relative improvement after the 1950s. It explains about 1,5% and 1,2% of the females and males residual variations respectively.

For Greece, the 1st interaction term explains the deviations and trends for the very young (mainly for age groups [0]-[10]) in contrast with the very old for females only (above age group [75[). It explains about 79,3% and 90,3% of the females and males residual variations respectively. The 2nd most important interaction term ($\kappa_3(t)$ and $v_3(x)$ values), for females, explains the relative improvement mainly for the age groups [65],[70] for the last few years, accounting for 6,6% of the total variations. For males, it explains the relative deterioration, from the independent model, mainly for the age groups [25]-[45] after the 1990s in contrast with the age groups [65]-[75], accounting for the 2,8% of the residual variations. The 3rd most important interaction term ($\kappa_2(t)$ and $v_2(x)$ values), for both genders, refers mainly to age group [20]. This probably explains the "accident hump", accounting for about 3,5% and 2% of the females and males residual variations respectively. We note that the other high $v_2(x)$ (absolute) values have been accounted for and explained in the previous more significant interaction terms. Thus, we may conclude that, if we had used the PC variances, represented by the α_i's values, instead of the eigenvalues parameters λ_i^2, to define the number of interaction terms needed to be included in the model structure, the interpretation of the results would be difficult or even misleading. This is because we would have taken the second interaction term as the second most important term and we would have tried to interpret the high (absolute) $v_2(x)$ values before the $v_3(x)$ values of the 3rd interaction term. The 4th most important interaction term ($\kappa_4(t)$ and $v_4(x)$ values), for males refers to the relative deterioration after

the 1990s mainly for the age groups [50]-[55], accounting for about 1,5% of the residual variations. For females, it refers to the relative deterioration after the 1980s mainly for the age groups [25]-[35]. This term accounts for about 3,5% of the residual variations (we note that this effect for females is "wearing off" since the level is constant and the slope is being modelled by a first order autoregressive process, implying that the stationary $\mu_4(t)$ process tends to a fixed value).

Table 2 gives the parameter estimates and associated standard errors, under dlr modelling, for PCs, based on model structure (2). For Sweden, all of the PCs are modeled with a constant slope and a random walk process for the level (i.e. ARIMA(0,1,1) models plus constant). For Greece females, for the 1st PC, both level and slope are constant and so the model reduces to a linear regression model; for the 4th PC the level is constant and the slope is modelled by a first order autoregressive (stationary) process; and for the remaining PCs the slopes are constant and the levels are modelled by a random walk processes. For Greece males, for the 1st PC, and for the 4th PC, the levels are constant and the slopes are modelled by random walk processes, and for the remaining PCs, the slopes are constant and the levels are modelled by random walk processes.

Sweden

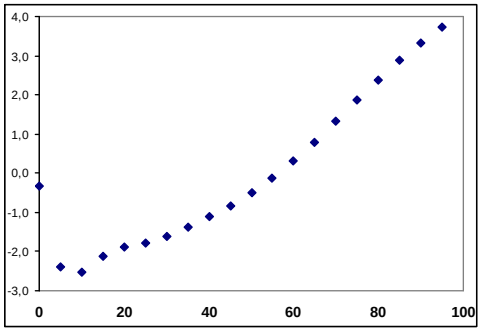
Females				**Males**			
PC index	$¤_i^2$		r_i^2	*PC index*	$¤_i^2$		r_i^2
1	592,72	93,62%	93,62%	**1**	485,14	93,91%	93,91%
2	21,15	3,34%	96,96%	**2**	11,37	2,20%	96,11%
3	9,64	1,52%	98,49%	**4**	6,47	1,25%	97,37%
4	2,44	0,38%	98,87%	**3**	4,93	0,95%	98,32%
5	1,78	0,28%	99,15%	**5**	2,39	0,46%	98,78%
10	1,19	0,19%	99,82%	**12**	1,11	0,21%	99,00%
6	1,08	0,17%	99,32%	**7**	1,03	0,20%	99,20%
7	0,83	0,13%	99,45%	**9**	0,89	0,17%	99,37%
9	0,64	0,10%	99,63%	**11**	0,86	0,17%	99,54%
8	0,47	0,07%	99,53%	**10**	0,60	0,12%	99,66%
13	0,47	0,07%	99,98%	**8**	0,60	0,12%	99,77%
11	0,38	0,06%	99,88%	**6**	0,54	0,10%	99,88%
12	0,20	0,03%	99,91%	**13**	0,42	0,08%	99,96%
14	0,10	0,02%	100,00%	**14**	0,18	0,03%	99,99%
				15	0,04	0,01%	100,00%

Greece

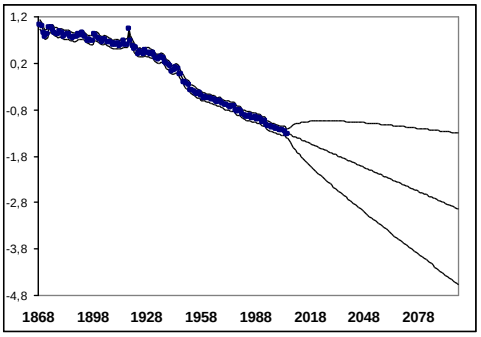
Females				**Males**			
PC index	$¤_i^2$		r_i^2	*PC index*	$¤_i^2$		r_i^2
1	25,21	79,3%	79,3%	**1**	36,01	90,3%	90,3%
3	2,09	6,6%	85,9%	**3**	1,12	2,8%	93,1%
2	1,12	3,5%	89,4%	**2**	0,79	2,0%	95,1%
4	1,12	3,5%	92,9%	**4**	0,58	1,5%	96,6%
5	0,60	1,9%	94,8%	**5**	0,46	1,2%	97,7%
7	0,51	1,6%	96,4%	**6**	0,40	1,0%	98,7%
6	0,42	1,3%	97,7%	**7**	0,18	0,4%	99,2%
8	0,27	0,8%	98,6%	**11**	0,10	0,3%	99,4%
9	0,18	0,6%	99,2%	**10**	0,07	0,2%	99,6%
11	0,15	0,5%	99,6%	**8**	0,07	0,2%	99,8%
12	0,07	0,2%	99,9%	**9**	0,06	0,1%	99,9%
10	0,04	0,1%	100,0%	**12**	0,04	0,1%	100,0%

Table 1: Eigenvalues based on model structure (2), with the associated percentage and cumulative residual variance explained.

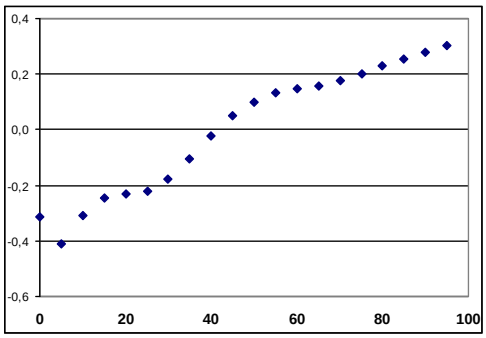
Sweden Females



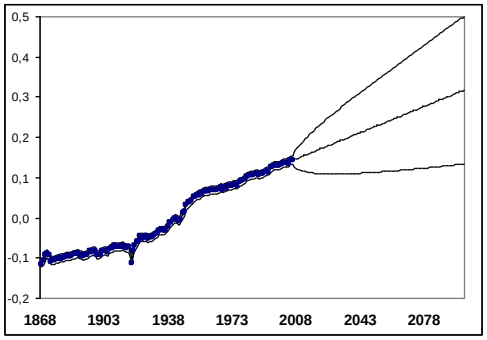
$\sim(x)$



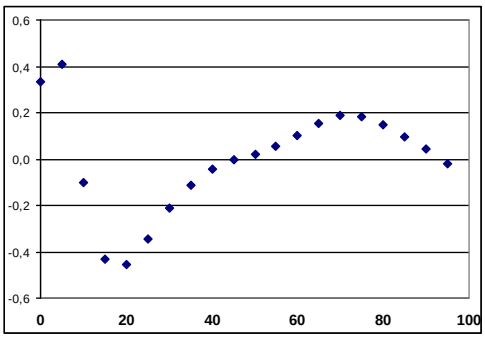
$b(t)$



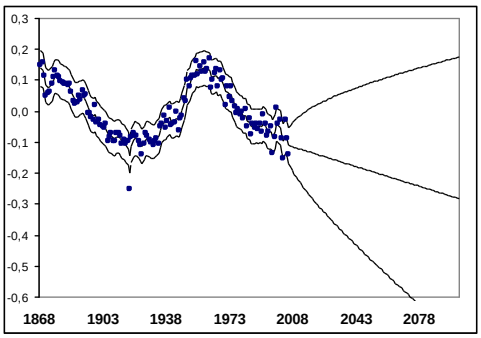
$v_1(x)$



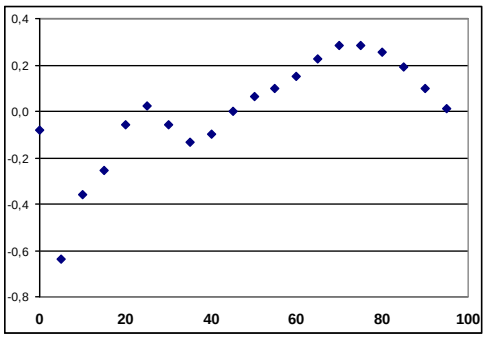
$\mu_1(t)$



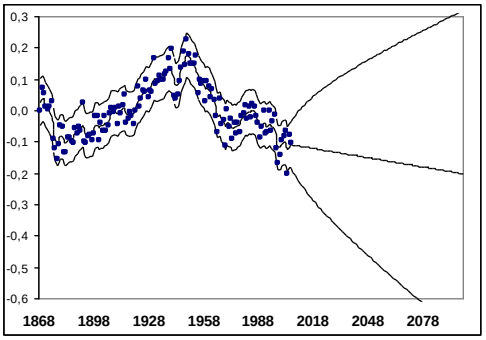
$v_2(x)$



$\mu_2(t)$

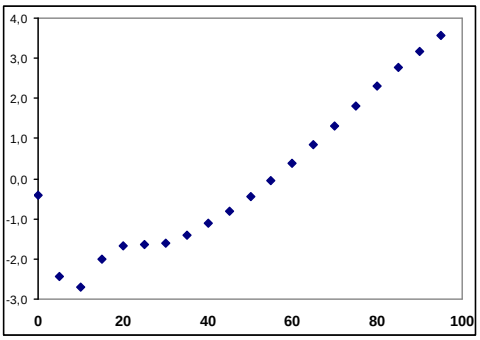


$v_3(x)$

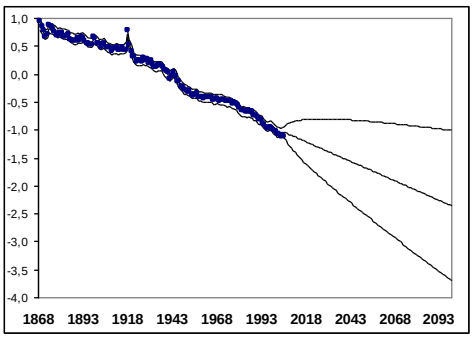


$\mu_3(t)$

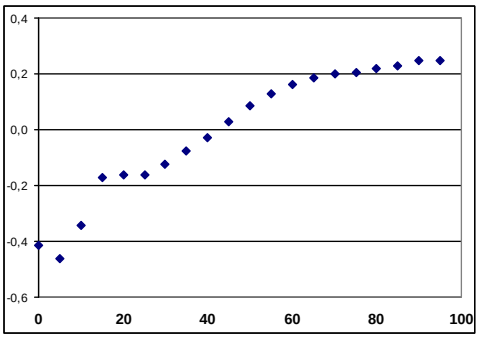
Sweden Males



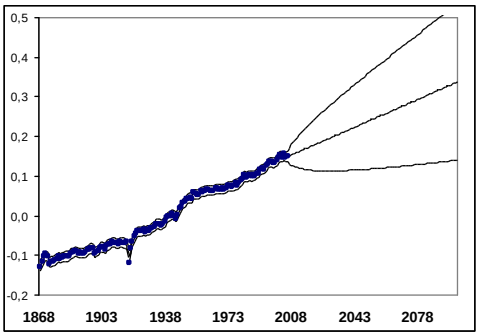
$\sim(x)$



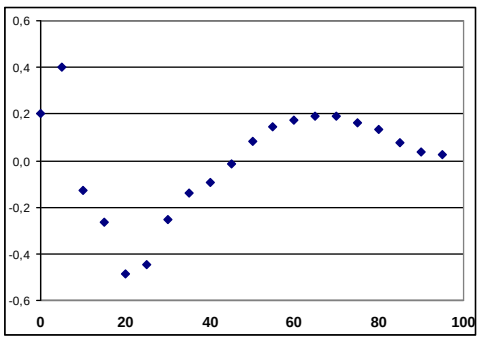
$b(t)$



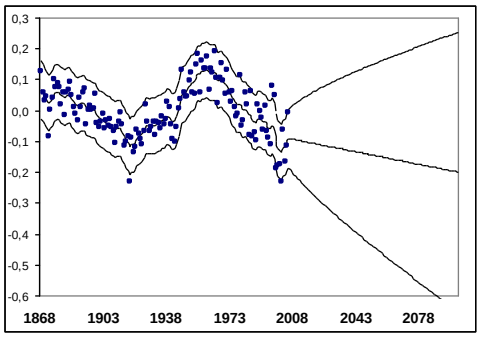
$v_1(x)$



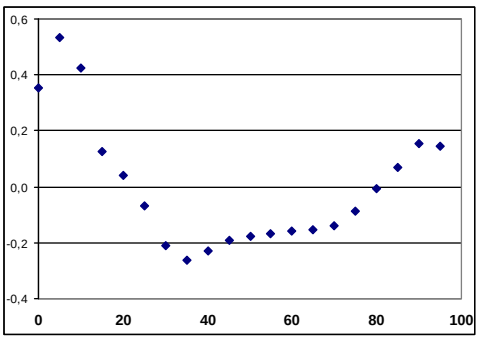
$\mu_1(t)$



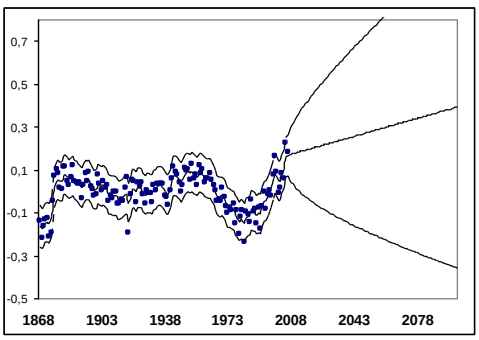
$v_2(x)$



$\mu_2(t)$

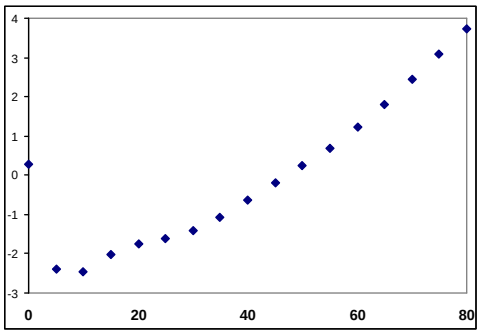


$v_4(x)$

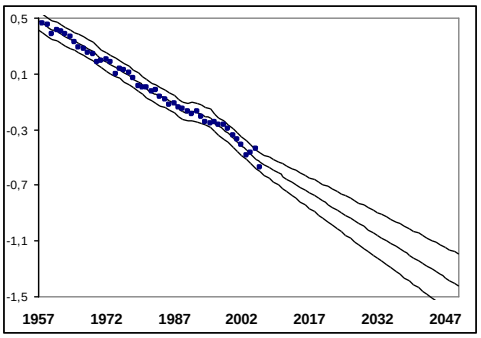


$\mu_4(t)$

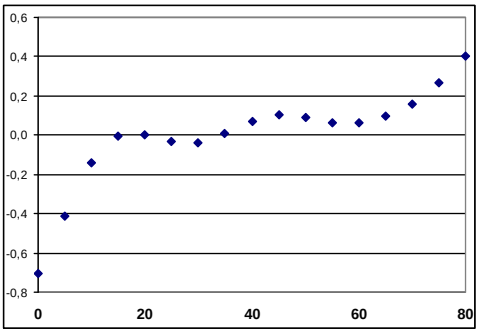
Greece Females



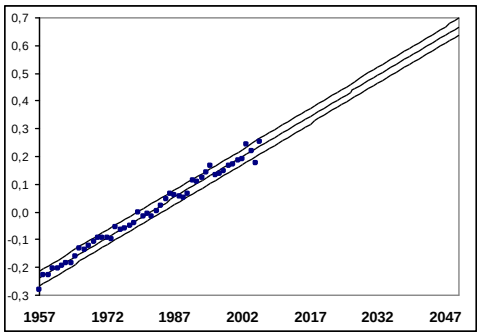
$\sim(x)$



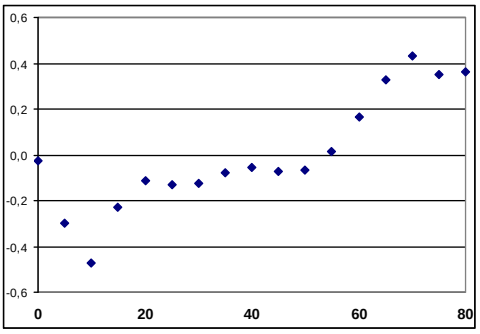
$b(t)$



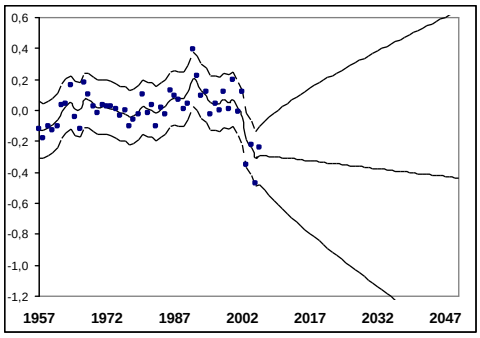
$v_1(x)$



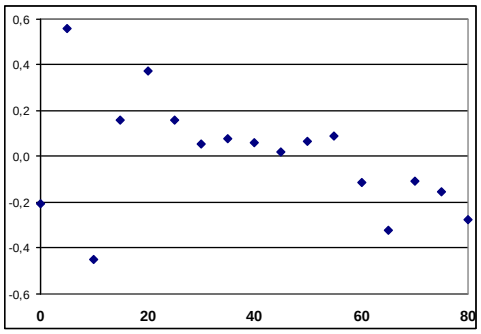
$\mu_1(t)$



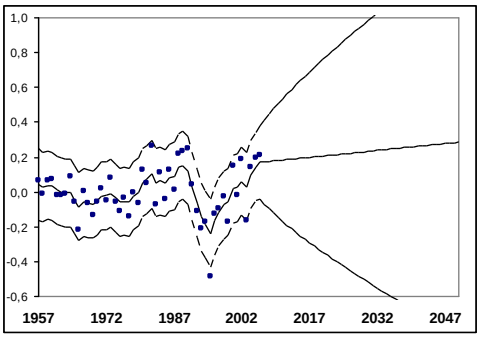
$v_3(x)$



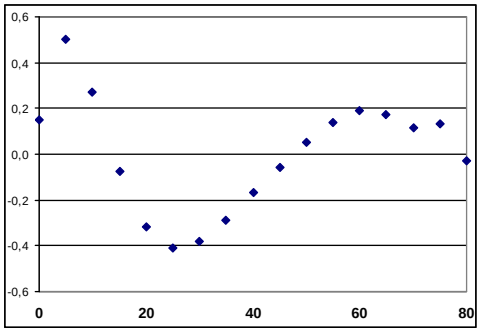
$\mu_3(t)$



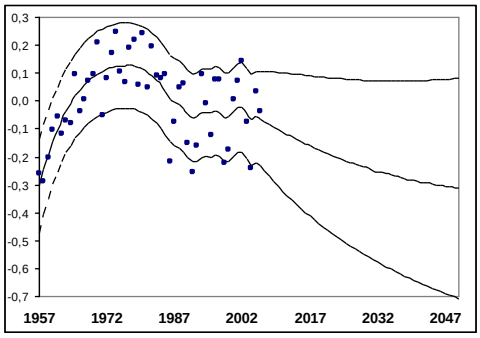
$v_2(x)$



$\mu_2(t)$

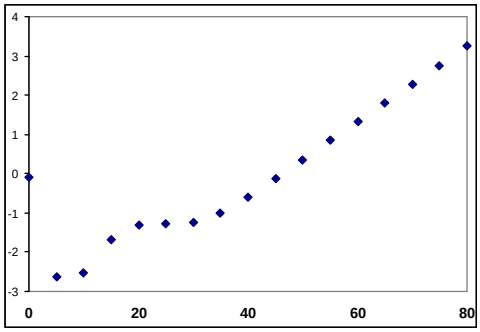


$v_4(x)$

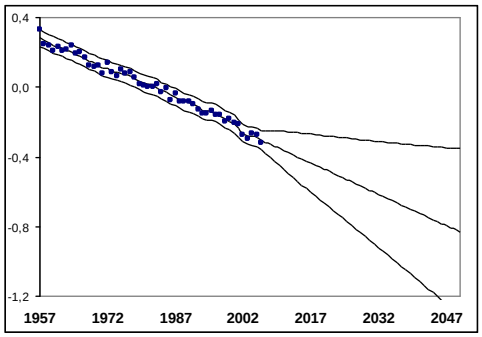


$\mu_4(t)$

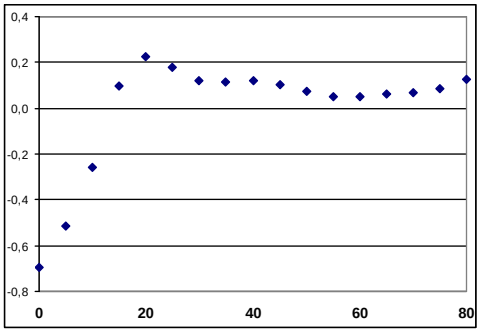
Greece Males



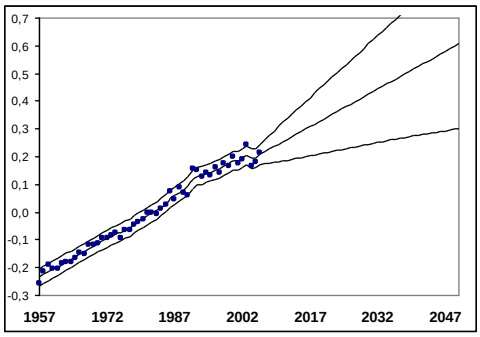
$\sim(x)$



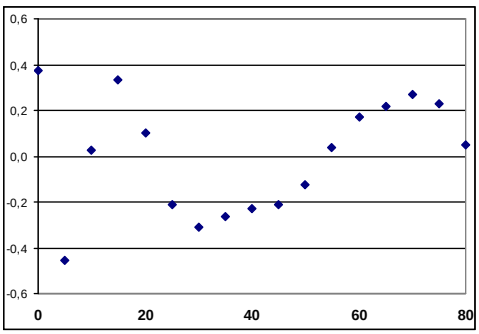
$b(t)$



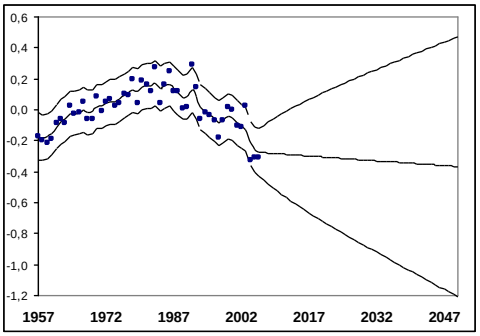
$v_1(x)$



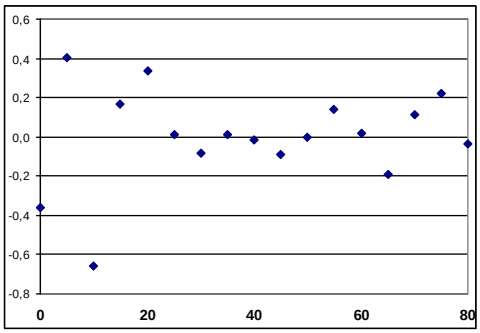
$\mu_1(t)$



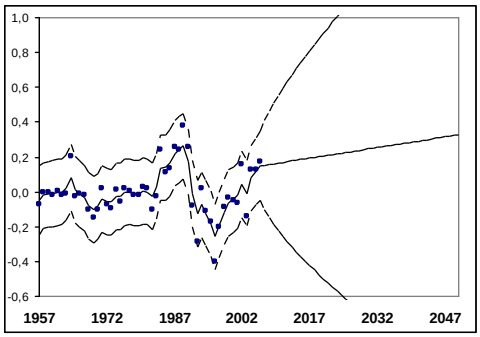
$v_3(x)$



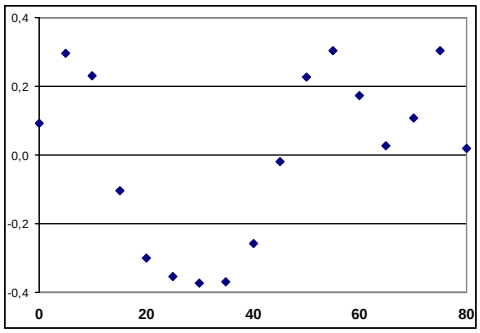
$\mu_3(t)$



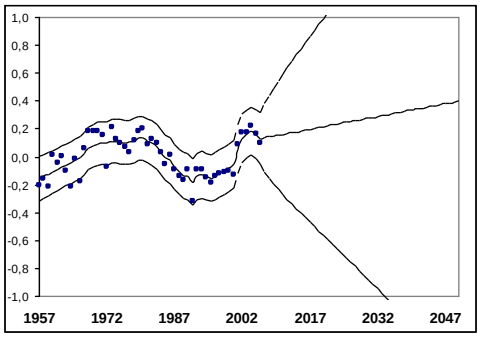
$v_2(x)$



$\mu_2(t)$



$v_4(x)$



$\mu_4(t)$

Figure 3: The first more important components, in significant order, predicted PC and associated CI, based on model structure (2), for Sweden (Females-Males) and Greece (Females-Males)

Table 2, gives the parameter estimates and standard errors, under the dlr modelling for the PCs based on model structure (2).

	Sweden		**Greece**	
	Female	**Male**	**Female**	**Male**
PC1				
Level	RW	RW	-0,2482 (0,0038)	-0,243 (0,0077)
Slope	0,0019 (0,0006)	0,002 (0,0007)	0,0097 (0,0001)	RW
q_1	3,483 (0,211)	2,658 (0,213)	0	0
q_2	0	0	0	0,000279 (0,364)
PC2				
Level	RW	RW	RW	RW
Slope	-0,0018 (0,0016)	-0,0012 (0,0015)	0,0026 (0,0088)	0,0041 (0,0116)
q_1	0,5626 (0,13)	0,18 (0,16)	0,493 (0,232)	1 (0,0041)
q_2	0	0	0	0
PC3				
Level	RW	RW	RW	RW
Slope	-0,001 (0,0018)	-0,0013 (0,0011)	-0,0034 (0,0085)	-0,0021 (0,0066)
q_1	0,4272 (0,144)	0,0553 (0,16)	0,579 (0,226)	0,5025 (0,23)
q_2	0	0	0	0
PC4				
Level			-0,368 (0,051)	-0,1752 (0,0414)
Slope			AR(1)	RW

	$\ddot{}_2$ =0,948 (0,546)	
q_1	0	0
q_2	0,000068 (0,0496)	0,000483 (0,1705)

Table 2: Parameter estimates and standard errors, under the dynamic linear modelling, for PCs, based on model (2).

Figure 4, shows the observed and bootstrap projected period life expectancies at birth and bootstrap projected life expectancies at age 65, for males (bottom lines)-females (top lines), using 1-PC (solid line) and associated 95% percentile CIs, and with *p*-PCs (dashed line) and associated 95% percentile CIs, based on model structure (2). We note that in all the cases the model with 1-PC shows lower values that the model with *p*-PCs.

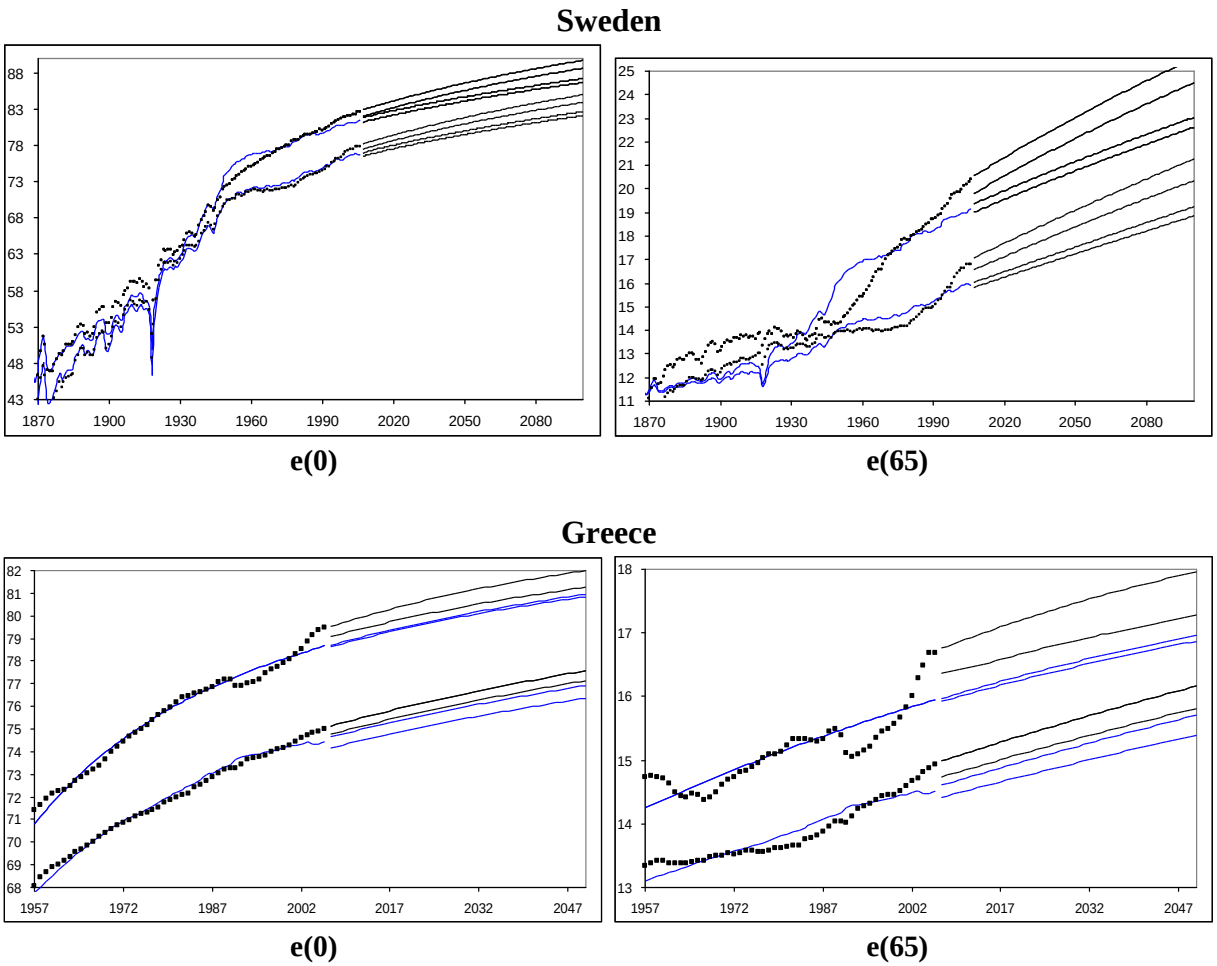


Figure 4: Observed and projected period life expectancies at birth and age 65, with 1-PC (solid lines) and *p*-PCs (dashed lines), for males (lower curves) and females (upper curves), with associated bootstrap projected life expectancies and bootstrap 95% percentile CIs

Tables 3 and 4 give observed and projected bootstrap median (period and cohort) of life expectancies at birth and age 65 respectively, for selected calendar years and associated bootstrap 95% percentile CIs with *p*-PCs, and observed and projected bootstrap median of life expectancies at birth and age 65 respectively with 1 PC and associated bootstrap 95% percentile CIs. The cohort based life expectancies correspond to a sequence of cohort-based life tables where each table refers to the cohort born in year *t*, using the diagonal arrangement for the corresponding central mortality rates: ${}_5m_0(t)$, ${}_5m_5(t+5),\ldots,{}_5m_x(t+x),\ldots$

Sweden										
Females	**1875**	**1900**	**1925**	**1950**	**1975**	**2000**	**2025**	**2050**	**2075**	**2100**
observed and projected bootstrap median of e(0)	46,8	53,5	63,7	72,4	78,0	81,9	84,2	86,2	87,9	89,3
95% bootstrap percentile CI of projected e(0)							0,9	1,0	1,1	1,1
observed and projected bootstrap median of e(0) with 1 PC	47,1	52,0	62,4	74,4	77,9	80,9	82,9	84,4	85,8	86,9
95% bootstrap percentile CI of projected e(0) with 1 PC							0,6	0,6	0,6	0,5
observed and projected bootstrap median of cohort e(0)	47,6	54,5	65,0	74,3	79,1	82,7				
95% bootstrap percentile CI of projected cohort e(0)			0,5	0,9	0,5	0,9				
Males										
observed and projected bootstrap median of e(0)	42,9	50,6	61,8	70,4	72,1	77,1	79,5	81,4	83,1	84,5
95% bootstrap percentile CI of projected e(0)							0,8	0,9	1,0	1,1
observed and projected bootstrap median of e(0) with 1 PC	42,9	50,1	61,0	70,4	72,8	76,3	78,2	79,8	81,2	82,5
95% bootstrap percentile CI of projected e(0) with 1 PC							0,5	0,5	0,5	0,5
observed and projected bootstrap median of cohort e(0)	43,5	51,3	62,5	71,2	73,3	77,8				
95% bootstrap percentile CI of projected cohort e(0)			0,4	0,4	0,3	0,4				
Greece										
Females										
observed and projected bootstrap median of e(0)					75,0	78,1	80,6	81,7		
95% bootstrap percentile CI of projected e(0)							0,6	0,7		
observed and projected bootstrap median of e(0) with 1 PC					75,1	78,2	79,8	80,9		
95% bootstrap percentile CI of projected e(0) with 1 PC							0,1	0,1		
observed and projected bootstrap median of cohort e(0)					75,9					
95% bootstrap percentile CI of projected cohort e(0)					0,3					
Males										
observed and projected bootstrap median of e(0)					71,2	74,3	76,1	77,4		
95% bootstrap percentile CI of projected e(0)							0,4	0,4		
observed and projected bootstrap median of e(0) with 1 PC					71,3	74,2	75,5	76,7		
95% bootstrap percentile CI of projected e(0) with 1 PC							0,5	0,6		
observed and projected bootstrap median of cohort e(0)					76,1					
95% bootstrap percentile CI of projected cohort e(0)					0,4					

Table 3: Observed and projected bootstrap median (period and cohort) of life expectancies at birth, for selected calendar years and associated bootstrap 95% percentile CIs with *p*-PCs, and observed and projected bootstrap median of life expectancies at birth with 1 PC and associated bootstrap 95% percentile CIs

Sweden										
Females	**1875**	**1900**	**1925**	**1950**	**1975**	**2000**	**2025**	**2050**	**2075**	**2100**
observed and projected bootstrap median of e(65)	11,8	13,1	14,0	14,3	17,5	19,8	21,3	22,7	24,0	25,1
95% bootstrap percentile CI of projected e(656)							0,8	0,9	1,0	1,0
observed and projected bootstrap median of e(65) with 1 PC	11,4	11,9	13,3	16,1	17,4	18,9	20,0	21,0	21,9	22,8
95% bootstrap percentile CI of projected e(65) with 1 PC							0,4	0,4	0,4	0,4
observed and projected bootstrap median of cohort e(65)	12,0	13,3	14,8	15,6	18,2	20,3				
95% bootstrap percentile CI of projected cohort e(65)			0,3	0,9	0,4	0,8				
Males										
observed and projected bootstrap median of e(65)	10,7	12,2	13,4	14,0	14,1	16,4	17,6	18,7	19,8	20,8
95% bootstrap percentile CI of projected e(656)							0,6	0,7	0,8	0,9
observed and projected bootstrap median of e(65) with 1 PC	11,4	11,8	12,7	14,1	14,6	15,8	16,6	17,4	18,3	19,1
95% bootstrap percentile CI of projected e(65) with 1 PC							0,2	0,3	0,3	0,4
observed and projected bootstrap median of cohort e(65)	10,8	12,2	13,7	14,3	14,5	16,6				
95% bootstrap percentile CI of projected cohort e(65)			0,2	0,2	0,2	0,3				
Greece										
Females										
observed and projected bootstrap median of e(65)					14,9	15,7	17,1	17,7		
95% bootstrap percentile CI of projected e(656)							0,6	0,7		
observed and projected bootstrap median of e(65) with 1 PC					15,0	15,8	16,4	16,9		
95% bootstrap percentile CI of projected e(65) with 1 PC							0,1	0,1		
observed and projected bootstrap median of cohort e(65)					15,2					
95% bootstrap percentile CI of projected cohort e(65)					0,3					
Males										
observed and projected bootstrap median of e(65)					13,6	14,5	15,4	16,0		
95% bootstrap percentile CI of projected e(656)							0,3	0,3		
observed and projected bootstrap median of e(65) with 1 PC					13,6	14,5	15,0	15,6		
95% bootstrap percentile CI of projected e(65) with 1 PC							0,2	0,3		
observed and projected bootstrap median of cohort e(65)					13,5					
95% bootstrap percentile CI of projected cohort e(65)					0,1					

Table 4: Observed and projected bootstrap median (period and cohort) of life expectancies at age 65, for selected calendar years and associated bootstrap 95% percentile CIs with *p*-PCs, and observed and projected bootstrap median of life expectancies at age 65 with 1 PC and associated bootstrap 95% percentile CIs

In Appendix A, we provide $_{j-1}(t)$, $\hat{}_{j-1}(t)$ values and associated 95% CI, based on model structure (3). Appendix B, shows log central mortality rates, projected log-central mortality rates and associated 95% CIs.

4. Discussion

We have investigated a new approach to the modeling and forecasting of mortality age-specific rates. The proposed method possesses a number of important characteristics. It provides estimates that are robust and involves the projection of several independent Gaussian PCs by dynamic linear regression modeling, supported by Kalman filter techniques. It utilizes the normality properties, thus giving increased precision and optimal statistical significant testing and setting of prediction intervals. The univariate time series analysis simplifies significantly the forecast procedure in contrast with the LC method, where the univariate time series forecasts, based on such structures, are potentially problematic (Renshaw and Haberman 2003a). The method incorporates a great degree of flexibility in order to capture historical mortality trends and dynamic mortality changes, without the restricted assumption that the logarithms of the mortality rates are approximately a linear function of time which imposes restrictions in the dominant time component. With dynamic linear regression modeling, we are able to utilize all of the available historic mortality patterns and are able to make forecasts for longer forecast periods. Many authors (eg. Booth et. al., 2006; Keyfitz, 1991) agree that forecast accuracy is highly dependent on the particular data set, in terms of time period or population. We describe an ex-post study for statistically determining the fitting period, based on the MSE deviations of the log-central mortality rates. The results of these ex-post studies indicate that the method improves the forecasts for the log-central mortality rates when using more than 1 interaction term.

The method also provides prediction intervals for the log-central mortality rates. Booth et. al. (2006) have shown that, even if the forecast life expectancy is accurate, compensating age-specific errors can be relatively substantial and in the long-term lead to unrealistic forecasts of the age pattern of mortality. As discussed by the authors, while accuracy in forecasting life expectancy may be important, it is not sufficient, and to gain an understanding of forecast error, the evaluation of the error in log death rates is essential.

The similarity with the LC model structure allows the current model to be viewed as a variant of the LC model with p interaction terms, but differentiates it from the LC method by providing estimates which are based on the SVD of the covariance matrix of the GLM parameter estimates, under Poisson over-dispersed modeling and a log link function. These estimates are unique and based on sufficient statistics. No adjustments and prior restrictions are required for identifiability of the parameter estimates. This method ensures identifiability and a unique solution. In contrast with the LC model, we utilize multiple interaction terms, allowing for more flexible mortality changes and improved forecasts.

The model also provides estimates for a hierarchical statistical association model for a two-way cross-classification table with age and time being the two main random effects, similar to the so-called *association model* of Goodman (1991) with p sets of scores. Under this view, the component $b(t)$ is an important term, since it describes the overall mortality trend in time effects, providing additional insight to the overall mortality trend, especially when the 1st interaction term does not explain the majority of the mortality variation. The singular value parameters of the matrix of the residuals, from the independent model, can be used as indices of the importance of each interaction term and define the goodness of approximation, incorporating both the time and age variations – this is in contrast with the LC method that indexes only the time variation. The mortality study of Greece shows the importance of these indices, since they bring out the importance of each interaction term in order of significance, giving a clearer picture of the mortality variations and trends. Also, the interpretations in comparison with the LC method are different. In this context, the interaction terms explain deviations from the independent model and the 1st interaction term is mainly governed by the very young ages since these ages have shown noticeably greater improvement (relative to the independent model) than the other ages in the recent mortality history. This feature characterizes the first interaction term and accounts for the majority of the residual variation in mortality. High absolute values for the age component scores indicate the increased importance of the PC trend. The remaining interaction terms may be explained similarly.

The proposed model structure also gives estimates and forecasts for the GLM parameters, under the parameterised model structure (3). We condense the information contained in the past experience into a set of

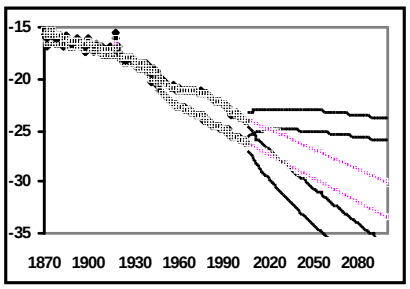
critical parameters (GLM parameter estimates), which contain as much information as possible and explain the inherent variation of the mortality rates.

The generalization of this approach to other situations, where cross-sectional time series data incorporate a parametric model for each period, is feasible. Possible extensions could also involve analyses of mortality by cause of death and of fertility rates. In terms of future work we intend to provide an extension model for the coherent modeling and forecasting of comparative mortality experiences. We are also intending to explore alternative orthonormal polynomial expansions.

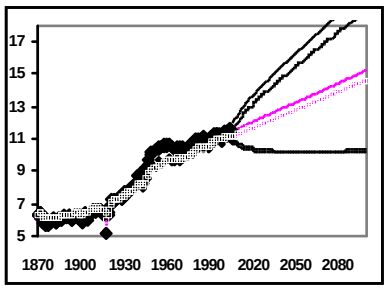
All the computations were implemented in Matlab.

Appendix A: $_{j-1}(t)$, $\hat{}_{j-1}(t)$ values and associated CI, for Sweden and Greece (Females are shown with dashed lines and Males with solid lines)

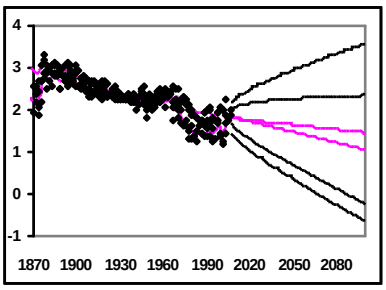
Sweden



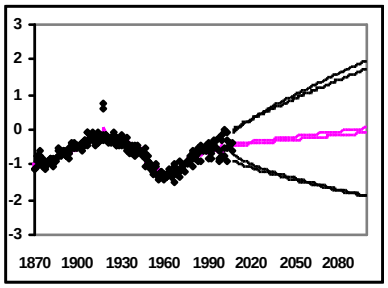
$\hat{}_0(t)$



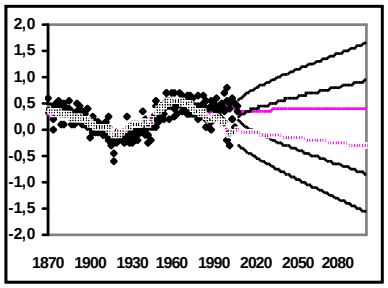
$\hat{}_1(t)$



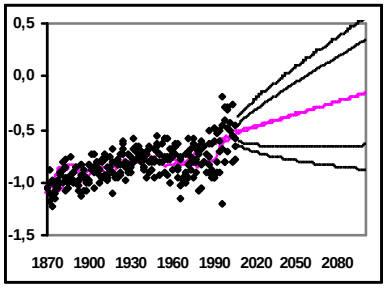
$\hat{}_2(t)$



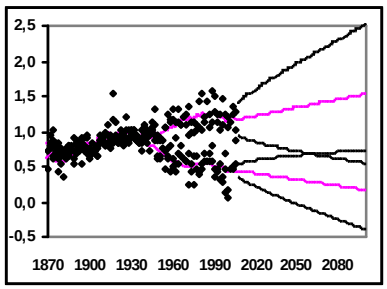
$\hat{}_3(t)$



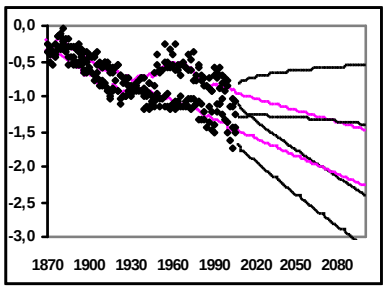
$\hat{}_4(t)$



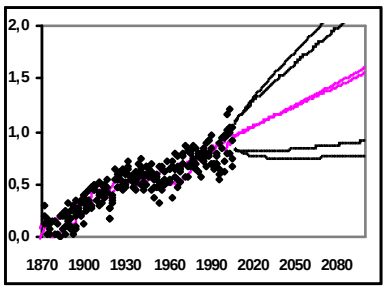
$\hat{}_5(t)$



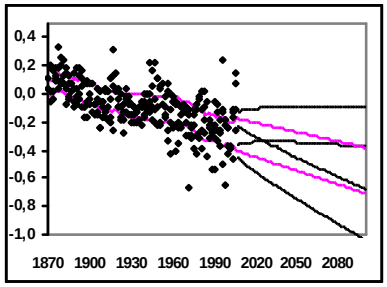
$\hat{}_6(t)$



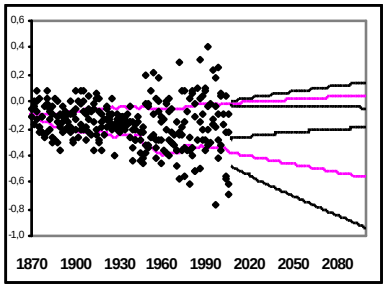
$\hat{}_7(t)$



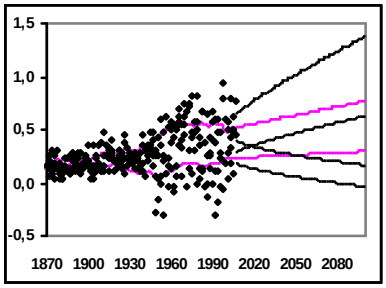
$\hat{}_8(t)$



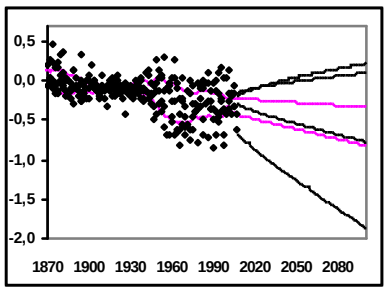
$\hat{}_9(t)$



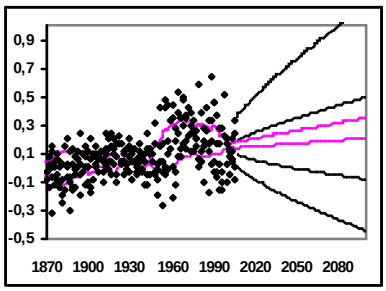
$\hat{}_{10}(t)$



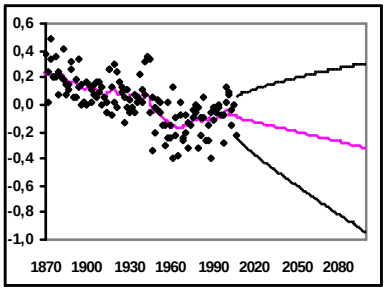
$\hat{}_{11}(t)$



$\hat{}_{12}(t)$

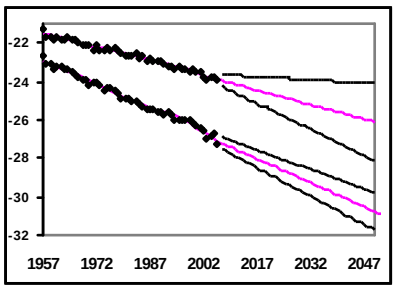


$\hat{}_{13}(t)$

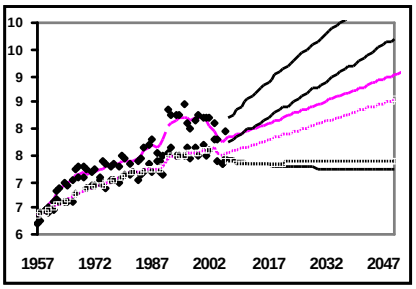


$\hat{}_{14}(t)$

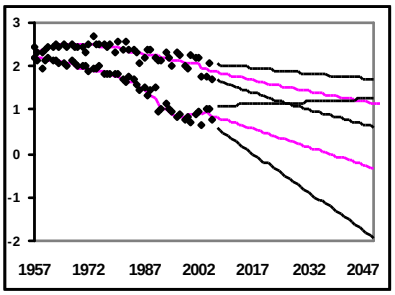
Greece



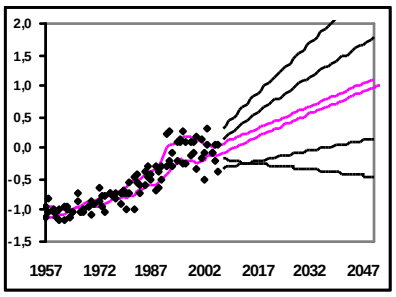
$\hat{}_{0}(t)$



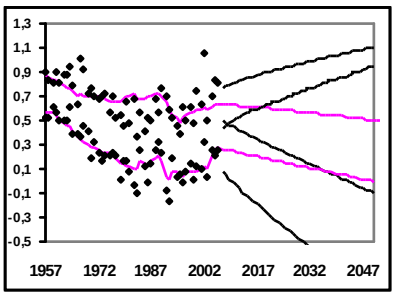
$\hat{}_{1}(t)$



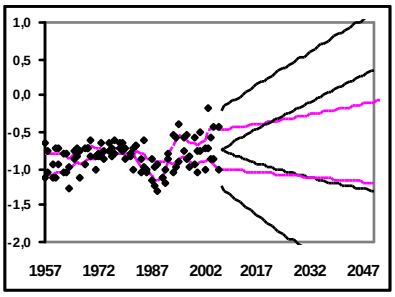
$\hat{}_{2}(t)$



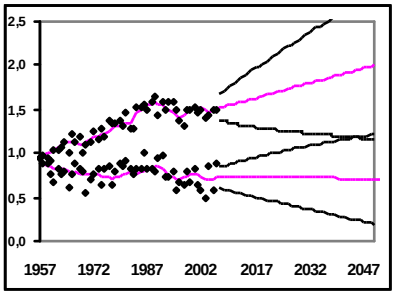
$\hat{}_{3}(t)$



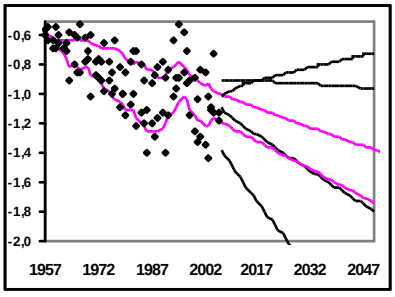
$\hat{}_{4}(t)$



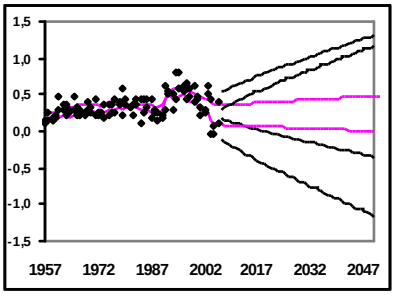
$\hat{}_{5}(t)$



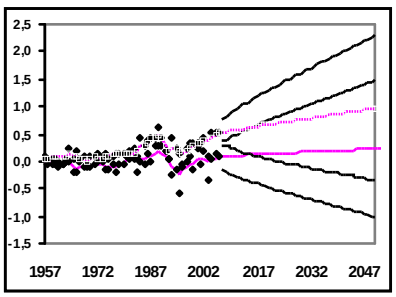
$\hat{}_{6}(t)$



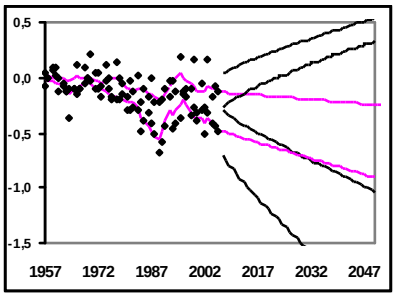
$\hat{}_{7}(t)$



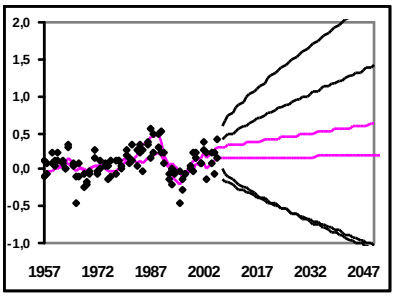
$\hat{}_{8}(t)$



$\hat{}_9(t)$



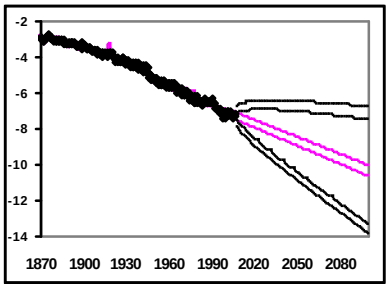
$\hat{}_{10}(t)$



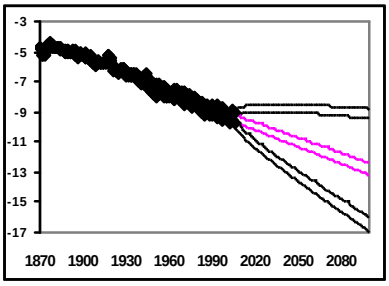
$\hat{}_{11}(t)$

Appendix B: Log central mortality rates, predicted log central mortality rates and associated CI, for Sweden and Greece (Females are shown with dashed lines and Males with solid lines)

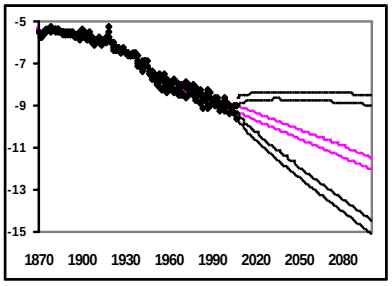
Sweden



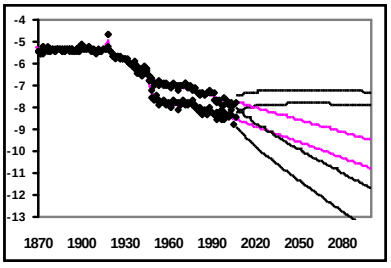
x=0-4



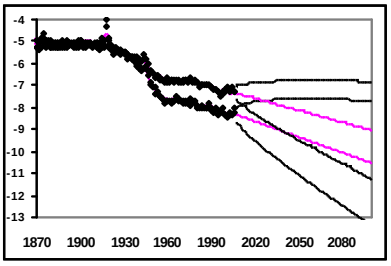
x=5-9



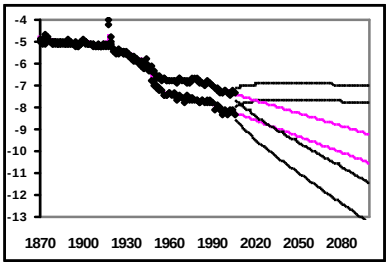
x=10-14



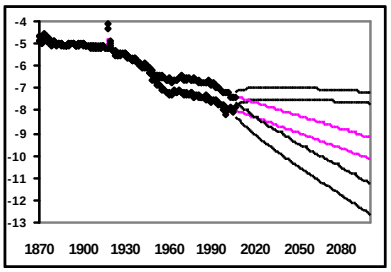
x=15-19



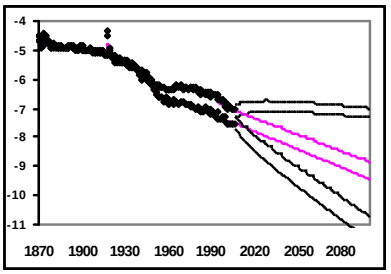
x=20-24



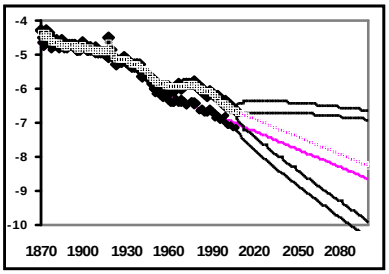
x=25-29



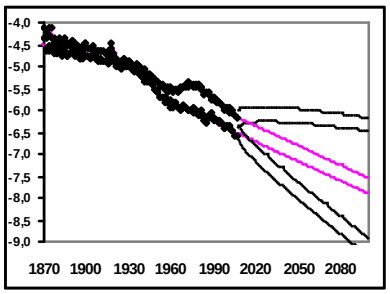
x=30-34



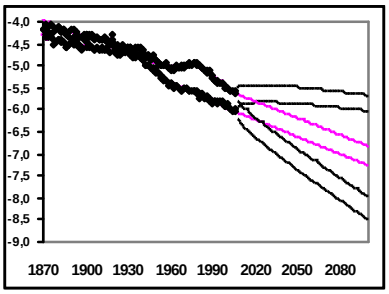
x=35-39



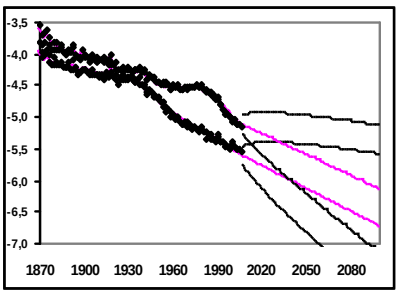
x=40-44



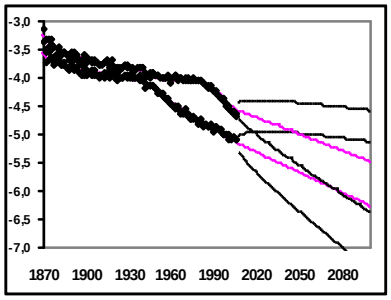
x=45-49



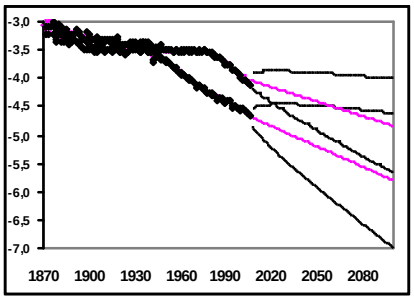
x=50-54



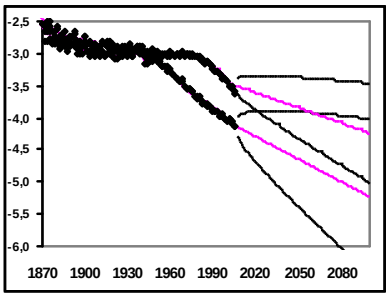
x=55-59



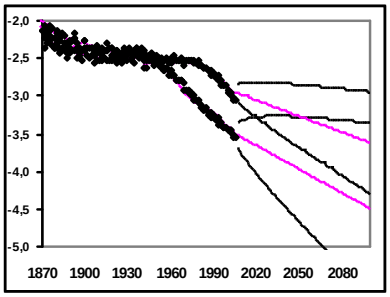
x=60-64



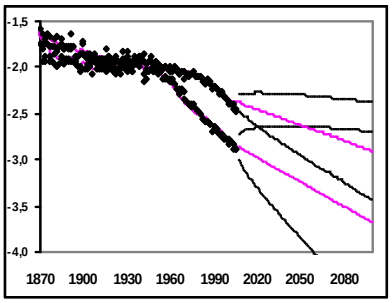
x=65-69



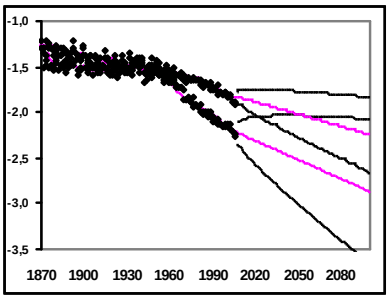
x=70-74



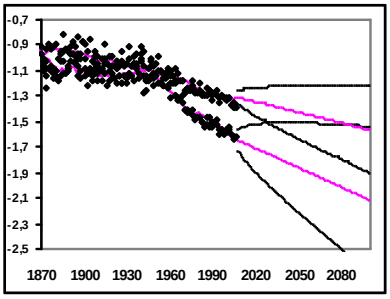
x=75-79



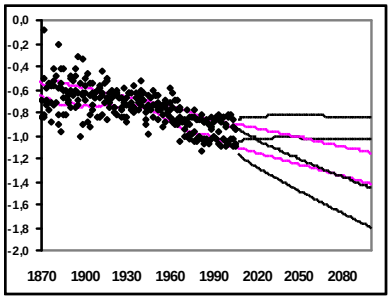
x=80-84



x=85-89

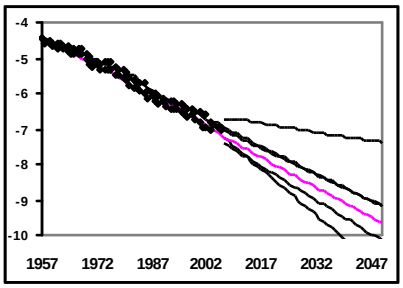


x=90-94

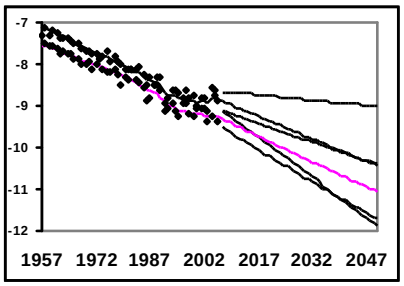


x=95+

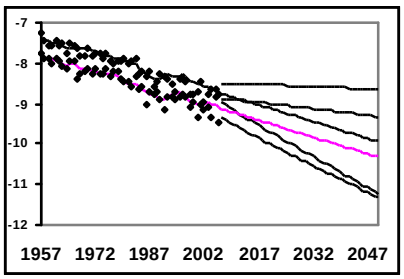
Greece



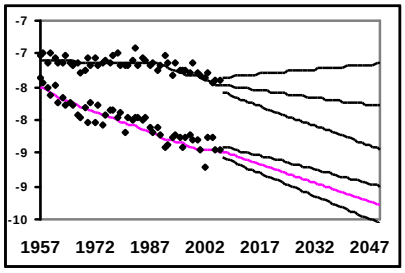
x=0-4



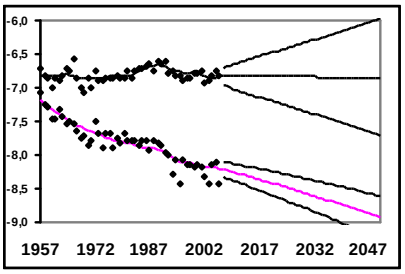
x=5-9



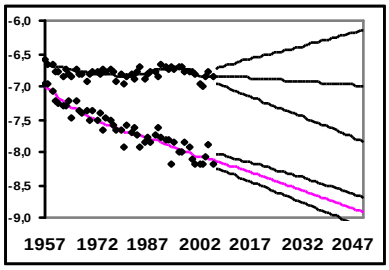
x=10-14



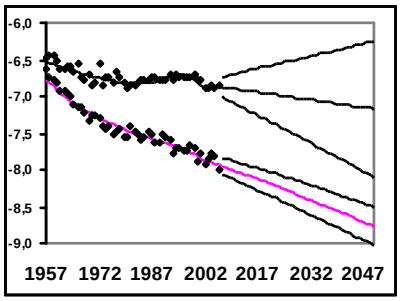
x=15-19



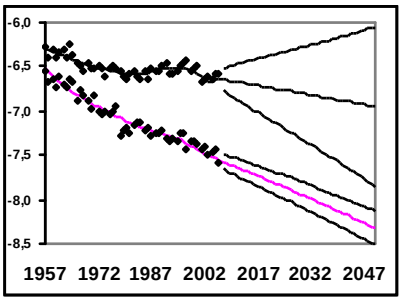
x=20-24



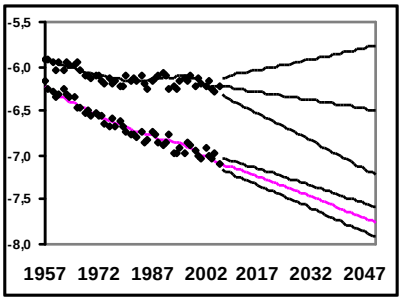
x=25-29



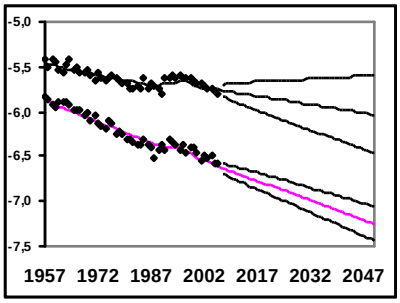
x=30-34



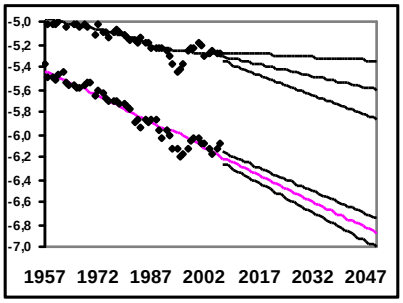
x=35-39



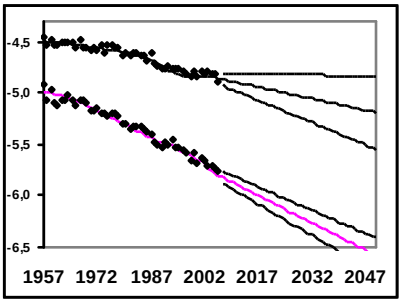
x=40-44



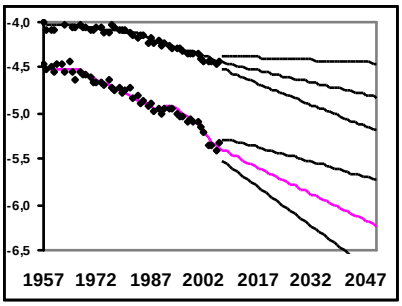
x=45-49



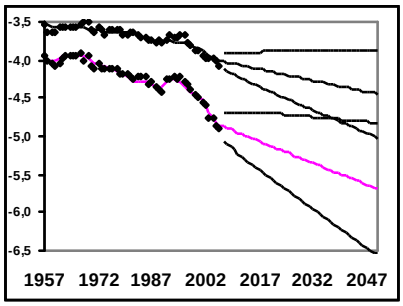
x=50-54



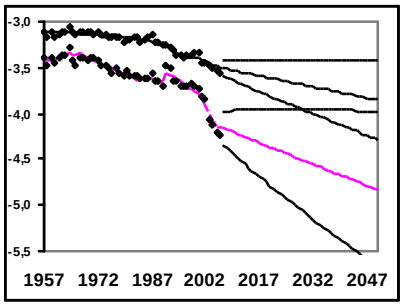
x=55-59



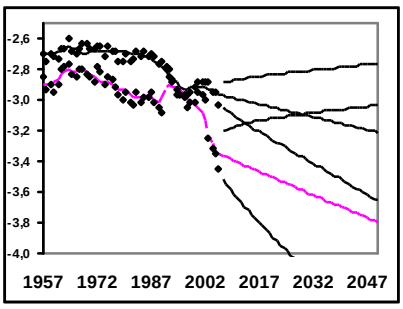
x=60-64



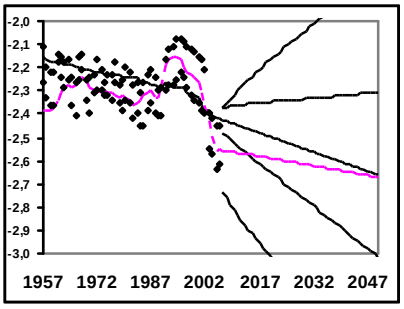
x=65-69



x=70-74



x=75-79



x=80-84

References

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

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., Hyndman, R. J., Tickle, L., et al. (2006). "Lee-Carter mortality forecasting: a multicountry comparison of variants and extensions." Demographic Research, 15(9), pp.289-310.

Booth, H., Maindonald, J. and Smith, L. 2002. "Applying Lee–Carter under conditions of variable mortality decline". Population Studies 56, pp 325–336.

Bozik, J.E. and Bell, W.R. 1987. "Forecasting Age Specific Fertility Using Principal Components." Proceedings of the American Statistical Association, Social Statistics Section pp. 396–401.

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 regression approach to the construction of projected lifetables". Insurance: Mathematics and Economics 31 (3), pp 373–393.

Cox, D.R. 1967. "*Renewal Theory*". Methuen, London .

Cox, D.R. 1983. "Some remarks on overdispersion". Biometrika 70(1), pp 269-274.

Cramer, H. and Wold, H. 1935. "Mortality variations in Sweden. A study in graduation and forecasting". *Skandinavisk Actuarie Tidskrift*, 18, pp 161-206.

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.

De Moivre, A. 1725. "Annuities on Lives: Or, the valuation of annuities upon any number of lives, as also, of reversions". London. Quoted in Smith, D. and Keyfitz, N. (1977), *Mathematical demography*: Selected Readings, Biomathematics, vol. 6, Springer-Verlag, New York, 1977.

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

Forfar, D.O., McCutcheon, J.J. and Wilkie, A.D. 1988. “On graduation by mathematical formula”. Journal of Institute of Actuaries 115 (1), pp 1–135.

General Secretariat of National Statistical Service of Greece. Available at: http://www.statistics.gr/.

Gompertz, B. 1825. “On the nature of the law of human mortality and on a new method of determining the value of life contingencies”. Philosophical Transactions of Royal Society 115, pp. 513-585.

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.

Haberman, S. and Renshaw, A.E. 2008. “Mortality, longevity and experiments with the Lee-Carter model”. Lifetime Data Analysis, 14(3):286-315.

Hagnell, M. 1991. “A multivariate time series analysis of fertility, adult mortality, nuptiality and real wages in Sweden 1751-1850: A comparison of some different approaches”. Journal of Official Statistics, 7, pp 437-455.

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

Heligman, L. and Pollard, J.H. 1980. “The age pattern of mortality”. Journal of the Institute of Actuaries 107 (1), 49–80.

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 June, 2008).

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.

Kendall, M. G., & Stuart, A. 1967. “The advanced theory of statistics” (Vol. 2). London: Griffin

Keyfitz, N. 1991. “Experiments in the projection of mortality”. Canadian Studies in Population 18(2), pp. 1-17.

Koissi, M-C, Shapiro, A. F. and Hognas, G. 2006. “Evaluating and extending the Lee–Carter model for mortality forecasting: Bootstrap confidence interval”. Insurance: Mathematics and Economics 38 (1), pp 1-20.

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.

Lee, R D. 1993. “Modelling and Forecasting the Time Series of US Fertility: Age Patterns, Range, and Ultimate Level.” International Journal of Forecasting 9, pp 187–202.

Lee, R.D. 2000. “The Lee–Carter method for forecasting mortality, with various extensions and applications”. North American Actuarial Journal 4(1), pp 80–93.

Lee, R.D. and Miller, T. 2001. “Evaluating the performance of the Lee–Carter method for forecasting mortality”. Demography 38 (4), pp 537–549.

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

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

McNown, R and Rogers, A. 1989. “Forecasting Mortality: A Parameterized Time Series Approach.” Demography 26(4) pp 645–660.

McNown, R and Rogers, A. 1992. “Forecasting Cause-Specific Mortality Using Time Series Methods.” International Journal of Forecasting 8, pp 413–432.

Newbold, P. and Bos, T. 1985. “Stochastic parameter regression models”. Series: Quantitative Applications in the Social Sciences. Sage University Papers. Series / Number 07-051.

Pitacco, E. 2004. “Survival models in dynamic context: A survey”. Insurance: Mathematics & Economics 35 (2), 279–298.

Pitacco, E. Denuit, M., Haberman, S and Olivieri A-M. 2008. “Modelling longevity dynamics for pensions and annuity business”. Oxford University Press. In press.

Renshaw, A.E. 1991. “Actuarial graduation practice and generalised linear and non-linear models”. Journal of the Institute of Actuaries 118 (2), pp 295–312.

Renshaw, A.E. 1992. “Joint modelling for actuarial graduation and duplicate policies”. Journal of the Institute of Actuaries 119 (1), pp 69 – 85.

Renshaw, A.E., Haberman, S. and Hatzopoulos, P. 1996. “The modelling of recent mortality trends in United Kingdom male assured lives”. British Actuarial Journal 2 (2), pp 449–477.

Renshaw AE, Haberman S and Hatzopoulos P. 1997. “On the duality of assumptions underpinning the construction of life tables”. Astin Bulletin 27 (1), pp 5-22.

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. “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.

Renshaw, A.E., and Haberman, S. 2008. “On Simulation-Based Approaches to Risk Measurement in Mortality with Specific to Poisson Lee-Carter Modelling.” Insurance: Mathematics and Economics, 42 (3), pp 797 -816

Sinamurthy, M. 1987. “Principal components representation of ASFR: Model of fertility estimation and projection”. CDC research monograph number, 16 (pp. 655-693). Cairo Demographic Center.

Sithole, T., 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 & Economics 27 (3), pp 285–312.

Tabeau, E., van den Berg Jeths, A. and Heathcode, C. (Eds.) 2001. “A Review of Demographic Forecasting Models for Mortality. Forecasting in Developed Countries: From description to explanation”. Kluwer Academic Publishers.

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

Tuljapurkar, S., Li, N. and Boe, C. 2000. “A Universal Pattern of Mortality Decline in the G7 Countries”. Nature 405(June), pp 789–792.

Wilmoth, J.R. 2000. “Demography of longevity: past, present, and future trends”. Experimental Gerontology 35, pp 1111-1129.

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