FORECASTING MORTALITY IN THE EVENT OF A STRUCTURAL CHANGE[*]

Edviges Coelho

Universidade Nova de Lisboa and Statistics Portugal

Instituto Nacional de Estatística
Av. António José de Almeida
1000-043 Lisboa, Portugal
Telephone: +351 218426100
Email: edviges.coelho@ine.pt

Luis C. Nunes[**]

Universidade Nova de Lisboa

Faculdade de Economia
Campus de Campolide
1099-032 Lisboa, Portugal
Telephone: +351 213801600
Fax: +351 213886073
Email: lcnunes@fe.unl.pt

November 26, 2009

[*] The opinions expressed in this paper are solely those of the authors and do not necessarily represent those of the institutions with which they are affiliated.
[**] Corresponding author.

Abstract

In recent decades, life expectancy in developed countries has risen to historically unprecedented levels driven by an unforeseen decline in mortality rates. The prospects of further reductions are of fundamental importance in various areas. In this context, this paper proposes a new approach to forecast future mortality and life expectancy in the event of a structural change. We show how recent advances in statistical testing for structural changes can be used to arrive at a properly specified forecasting model. Specifically, the results of tests for a change in the trend of the general level of mortality and for the presence of a unit-root are used to identify the appropriate model to be estimated and used to carry out the projections. The proposed methodology is applied to post-1950 time series mortality data for 18 developed countries. Structural changes in the rate of decline in overall mortality are found for almost every country considered and especially in the male populations. These are associated with a more accentuated decline in mortality in recent years. We also illustrate how accounting for such a change can lead to a major impact in mortality and life expectancy forecasts over the next decades.

Keywords: Life expectancy; Mortality forecasting; Lee-Carter model; Structural change; Unit root.

1. INTRODUCTION

Over the past decades, life expectancy in developed countries has risen to historically unprecedented levels. The prospects of further future reductions in mortality rates are of fundamental importance in various areas such as demography, actuarial studies, public health, social insurance planning, and economic policy. Over the last years, significant progress has been made in mortality forecasting (for a recent review see Booth, 2006). The most popular approaches to long-term forecasting are based on the Lee and Carter (1992) model. It describes the time-series movement of age-specific mortality as a function of a latent level of mortality, also known as the overall mortality index, which can be forecasted using simple time-series methods. The method was initially used to forecast mortality in the U.S.,[1] but since then it has been applied to many other countries (amongst others see Fígoli, 1998; Tuljapurkar and Boe, 1998; Lee, 2000; Carter and Prskawetz, 2001; Lee and Miller, 2001; Booth *et al.*, 2002; Brouhns and Denuit, 2002; Renshaw and Haberman, 2003; Lundstrom and Qvist, 2004; and Koissi *et al.*, 2006).

The original Lee-Carter model has also received a number of criticisms (see the discussion in Lee and Miller, 2001) and several extensions have been proposed in the literature (see Booth *et al.*, 2006). One major issue concerns the stability of the model over time. Since the method is usually applied to long time-series there is a risk that important structural changes may have occurred in the past. And any neglected structural change in the estimation period may result in forecasts that have a tendency to deviate from the future realizations of the mortality index, leading to potentially large long-term forecast errors. In fact, historically, mortality in the U.S. has not always declined in a linear way as depicted in Lee and Carter (1992) for the period 1900-1989. These authors also re-estimated their random walk with drift model for the mortality index for several shorter and more recent periods and concluded that there was some instability. As noted by Lee (2000), if one had used their method to extrapolate the mortality index backward in time, one would arrive at

[1] The U.S. Census Bureau has also adopted this method to compute life expectancy projections (see Hollmann *et al*., 2000).

too high mortality rates for the beginning of the 19th century. Tuljapurkar and Boe (1998) provide evidence that the rate of mortality decline has varied substantially from one decade to the next in Canada, Mexico, and the U.S. Other studies also document that there has been a systematic overestimation of the projected mortality rates in many countries (Stoto, 1983, Koissi *et al.*, 2006). In a multi-country comparison of several versions of the Lee-Carter method, Booth *et al*. (2006) find significant differences in the forecasting performance when alternative fitting periods are used, providing evidence of different trends in the mortality rate.

In this paper we propose a new consistent approach to forecast mortality rates and life expectancies in the event of a structural change in the trend of the overall mortality index. The proposed method is based on recent advances in testing and estimating structural change models, and allows one to arrive at an adequate time-series forecasting model that can be estimated using the available sample. In particular, by following the proposed method, one is able to detect if a change is present or not, to estimate the date where the change occurs, and to determine if the appropriate forecasting model should be based on the levels or the first-differences of the mortality index series. The last distinction is important as the two alternatives lead to substantial differences in the amplitudes of the forecast confidence intervals.

The proposed methodology is applied to post-1950 time-series mortality data for 18 developed countries. Structural changes in the rate of decline in the overall mortality rate are found for almost every country considered and especially in the male populations. We consider as an example the case of Portuguese male mortality and show that accounting for a structural change leads to a major impact in mortality and life expectancy forecasts over the next decades.

The paper is organized as follows. Section 2 presents a brief review of the Lee-Carter method and its extensions. The proposed approach to forecast mortality in the presence of structural change is described in Section 3. The empirical application is presented in Section 4. Finally, Section 5 concludes.

2. THE LEE-CARTER METHOD

2.1. The demographic model

The Lee-Carter method combines a demographic model of age-specific mortality rates with statistical time-series forecasting methods. Let $\mu_x(t)$ denote the mortality force at age *x* during calendar year *t*, and $D_{x,t}$ denote the number of deaths recorded at age *x* during year *t* from an exposure-to-risk $E_{x,t}$. We assume that age-specific mortality rates are constant within bands of age and time, but allowed to vary from one band to the next. Specifically, given any integer age *x* and a calendar year *t*, it is assumed that

$$\mu_{x+\tau}(t) = \mu_x(t) \text{ for } 0 \leq \tau < 1 . \tag{1}$$

Under (1), estimates of $\mu_x(t)$, denoted as $\hat{\mu}_x(t)$, are given by the following ratio:

$$\hat{\mu}_x(t) = \frac{D_{x,t}}{E_{x,t}} . \tag{2}$$

The Lee-Carter (1992) model assumes that the observed force of mortality for any age *x* is driven by a common time-varying component, denoted by k_t, which is also referred to as the overall mortality index. More precisely, the following relation is assumed:

$$\ln \hat{\mu}_x(t) = \alpha_x + \beta_x k_t + \varepsilon_x(t) \tag{3}$$

where $\hat{\mu}_x(t)$ is given by (2), α_x are the age-specific parameters that affect the overall level of $\ln \hat{\mu}_x(t)$ over time, β_x are the age-specific parameters that characterize the sensitivity of $\ln \hat{\mu}_x(t)$ to changes in the mortality index k_t, and $\varepsilon_x(t)$ represent error

terms capturing particular age-specific historical influences not explained by the model. These are assumed to have zero mean and constant variance σ_ε^2.

Lee and Carter (1992) propose a two-stage procedure to estimate the model given by equation (3). First, a least-squares solution to estimate α_x, β_x, and k_t is found. Since the model is clearly underdetermined, some normalization constraints are imposed to obtain a unique solution: the sum of the β_x over all ages equals one and the sum of the k_t over time equals zero. As a consequence, the α_x will be equal to the average values of $\ln \hat{\mu}_x(t)$ over time. The solution is found using the singular value decomposition. Since the model is written in terms of log mortality, the observed total number of deaths in each year will not equal the sum of the fitted deaths by age. To ensure this equality, the k_t are estimated a second time, taking the α_x and β_x estimates from the first step, such that for each year *t*, given the actual age distribution for the population, the implied number of deaths equals the observed number of deaths.

The homoskedasticity assumption on the error term $\varepsilon_x(t)$, implied by the singular value decomposition estimation of the parameters α_x, β_x, and k_t, has been considered fairly unrealistic (Wilmoth, 1993; Alho, 2000) since the logarithm of the observed force of mortality is much more volatile at old ages, where the number of deaths and of those exposed to risk are relatively small. To circumvent this problem, Brouhns *et al.* (2002) propose an alternative model that keeps the Lee-Carter log-bilinear form for the force of mortality, but replaces the least-squares approach with a Poisson regression for the number of deaths. Specifically, they consider the following model describing the distribution of $D_{x,t}$:

$$D_{x,t} \sim Poisson\left(E_{x,t}\mu_x(t)\right) \quad \text{with} \quad \mu_x(t) = \exp\left(\alpha_x + \beta_x k_t\right), \tag{4}$$

where the parameters α_x, β_x, and k_t retain the meaning originally attributed by the Lee-Carter model. These parameters can be estimated by maximum likelihood still subject to the model identification constraints that the sum of the β_x over all ages

equals unity and the sum of the k_t over time equals zero. The second stage in the estimation of the k_t, required in the classical Lee-Carter model, is no longer necessary with this approach.

Given the estimated α_x and β_x in equation (3) or (4), the problem of forecasting age-specific mortality rates and, consequently, life expectancies, is reduced to forecasting the mortality index k_t.

2.2. Modeling the index of mortality

Lee and Carter (1992) propose using the Box-Jenkins methodology to arrive at an appropriate ARIMA model to forecast the index of mortality. In that methodology, the first step always consists of determining if some transformation of the series is necessary to induce stationarity before identifying and estimating the forecasting model. Many studies have arrived at an ARIMA(p, 1, q) model, that is, a stationary ARMA(p, q) model fitted to the first-differences of the mortality index. Series whose first-differences are stationary are also called difference-stationary, and are described by a unit-root in their autoregressive representation. The usual approach to check this is by analysing the behaviour of the empirical autocorrelation functions. However, it is also possible to use formal statistical tests. The most popular are variants of the original Dickey and Fuller (1979) tests for the presence of a unit-root.

In the Lee and Carter (1992) study of U.S. mortality, the authors arrive at an ARIMA (0,1,0) model, that is, a random walk with drift:

$$k_t - k_{t-1} = d + u_t, \tag{5}$$

where the drift d gives the mean annual change in k_t and u_t are i.i.d. errors. The drift d is estimated as the time-average of the series of first-differences $\Delta k_t = k_t - k_{t-1}$. Given any initial value for k_t, this model implies that:

$$k_{t+s} = k_t + ds + u_{t+1} + \cdots + u_{t+s}, \tag{6}$$

that is, future values of the mortality index equal the sum of a deterministic linear trend component with slope d, plus an error component given by the sum of several shocks u. Since part of the uncertainty when forecasting k_{t+s} comes from the cumulative sum of the shocks, $u_{t+1} + \cdots + u_{t+s}$, it follows that the forecast uncertainty grows with the forecast horizon s.

If for a particular mortality index series the unit-root/difference-stationary hypothesis is rejected, the alternative hypothesis that should be considered would be that of stationarity around a linear trend capturing the linear decline in mortality. In such a case, the mortality index would be described by the following trend-stationary model:

$$k_t = d_0 + d_1 t + u_t, \tag{7}$$

with u_t being a stationary ARMA process. In this equation, the d_1 parameter represents the slope of the linear trend. Since this model is specified and estimated in the levels of k_t, not the first-differences, it will in general produce forecasts that are different from the previous unit-root model. Also, given that in this model the deviations from the linear trend, given by u_t, are stationary, the forecast uncertainty will no longer have a tendency to increase with the forecast horizon.

Unfortunately, the usual procedures based on the analysis of the autocorrelation function or on unit-root tests to decide about the appropriate class of forecasting model to be used, are only valid if the linear trend, and in particular the trend slope d in model (6) or d_1 in model (7), has remained constant over time. In the next section we discuss in more depth the implications of the presence of a structural change in the trend of the mortality index and propose a suitable approach to cope with this possibility when forecasting the mortality index.

3. ALLOWING FOR A STRUCTURAL CHANGE

A large literature on structural change models has emerged in the last years (see the surveys by Perron 2006, 2008). An important result is that, when producing

long-term forecasts, any neglected or wrongly placed structural change occurring in the estimation phase may result in forecasts with a tendency to deviate from the future realizations of the series, resulting in potentially large forecast errors. It is also known that the correct approach to test for and estimate a model with a structural change is highly dependent on the stationarity properties of the time-series process. In particular, if the mortality index series k_t is known to be trend-stationary then i) the tests for the presence of a structural change in the trend and ii) the estimation of the forecasting model should both be carried out using regressions based on the levels of the series. On the other hand, if it is known that the k_t series is non-stationary with a unit-root, first-differences should be used instead. In general, the results obtained by following these two alternatives can be quite different in terms of a presence or not of a structural change and, in case one is detected, in terms of its dating.

As mentioned above, the traditional approaches to detect the presence of a unit-root and the need to first-difference the data are based on examining the empirical autocorrelation functions or on performing standard unit-root tests. However, as shown by Perron (1989), these approaches are not valid if a structural change is present, as they have a tendency to wrongly suggest a unit-root when the series is, in fact, stationary around a broken trend. As a solution, Perron (1997) proposes a modified unit-root test that is valid in the presence of a structural change at a known date. Zivot and Andrews (1992), Perron and Rodríguez (2003), and other authors have extended the test to the case where the change date is not known and must be estimated. However, when no change is actually present, several problems arise. As shown in Harris *et al.* (2009) (henceforth HHLT) there will be severe efficiency losses. Moreover, the estimated change date suggested by these tests may be spurious since, as shown by Nunes *et al.* (1995), the presence of a unit-root may lead one to erroneously find a structural change when there is none. In fact, when using the Zivot and Andrews (1992) unit-root test, the estimated change date is not even consistent for the true date if a change does occur.[2] Another issue is that the correct critical values to implement these tests become dependent on whether a structural change is present or not.

[2] An example where such an inappropriate approach is followed is Chan *et al.* (2008).

We propose a solution to this dilemma in the context of mortality forecasting by following a simple and consistent sequential approach based on recent statistical results in the literature of structural change. In a first step, since it is not possible to know beforehand if there is a unit-root or not, we apply a test for a structural change in the trend proposed by Harvey *et al.* (2009) (henceforth HLT) that is valid regardless of whether the series is difference-stationary or trend-stationary. Secondly, as in HHLT, the result of this structural change test is used to decide on the appropriate unit-root test to use: with or without allowing for a structural change. Thirdly, if according to the HLT test a change is present, the break date is estimated taking into account the result of the HHLT test, that is, a) using the first-differences of the series if a unit-root is found or b) using the levels of the series otherwise. Finally, based on the conclusions of the previous steps, the appropriate ARIMA model is fitted to the k_t series and used to produce the forecasts. In the following subsections we describe these structural change and unit-root tests in more detail.

3.1. Testing for a change in the trend

As mentioned above, we use the HLT test for the existence of a structural change in the trend of the mortality index k_t when it is not known *a priori* if the series has a unit-root or not. The date of the break, if present, is estimated from the available sample. The test is relatively simple to implement since it only requires estimating the following two linear regression models by least squares:

$$k_t = \alpha + \beta t + \gamma DT_t(\tau) + u_t, \quad t = 1, \ldots, T, \tag{8}$$

and

$$\Delta k_t = \beta + \gamma DU_t(\tau) + u_t, \quad t = 2, \ldots, T, \tag{9}$$

where the change dummy variables are defined as $DT_t(\tau) = t\text{-}T_B$ if $t > T_B$ and $DT_t(\tau) = 0$ if $t \leq T_B$, $DU_t(\tau) = 1$ if $t > T_B$ and $DU_t(\tau) = 0$ if $t \leq T_B$, with $T_B = [\tau T]$ denoting the possible trend change date, and τ the associated change date fraction with $\tau \in (0, 1)$.

Equation (8) corresponds to the trend-stationary case with stationary shocks u_t and k_t fluctuating around a linear trend with slope β subject to a change of magnitude γ occurring at a given date T_B. Equation (9) corresponds to the case where the shocks to k_t are non-stationary with a unit-root, so that the first-differences, $\Delta k_t = k_t - k_{t-1}$, fluctuate around the mean, given by the drift parameter β which is also subject to a change of magnitude γ occurring at date T_B. In this last equation, the shocks u_t to Δk_t are also stationary. In both equations, the null hypothesis of no change in the trend slope corresponds to H_0: $\gamma = 0$.[3]

If k_t is known to be trend-stationary, a valid test for the presence of a structural change in the trend slope can be based on the t-statistic for testing H_0 in equation (8), which we denote as $t_0(\tau)$. To allow for serial correlation in the error term, an auto-correlation robust t-statistic must be used. Since the change date τ is not known *a priori*, but may be inferred from the data itself, the test statistic is computed by the maximum of the sequence of $t_0(\tau)$ statistics for all possible change fractions τ as:

$$t_0^* = \sup_{\tau \in \Lambda} | t_0(\tau) | \qquad (10)$$

where the supremum is taken over a set $\Lambda = [\tau_L , \tau_U]$, with $0 < \tau_L < \tau_U < 1$. HLT suggest setting the *trimming* parameters τ_L and τ_U equal to 0.1 and 0.9 respectively.

On the other hand, if it is known that k_t is difference-stationary, one should use the corresponding t-statistic from equation (9), which we denote as $t_1(\tau)$. As in the previous case, given that the change date is not known *a priori*, the test statistic is given by

$$t_1^* = \sup_{\tau \in \Lambda} | t_1(\tau) |. \qquad (11)$$

The change fractions at which the test statistics t_0^* and t_1^* attain their maximum will be denoted as τ_0^* and τ_1^*, respectively.

[3] The model described here corresponds to Model A in HLT. It is straightforward to allow for a simultaneous change in the slope and the level of the trend (Model B in HLT) by including an extra dummy variable in each regression. However, if the mortality index does not show any abrupt changes in its level, it is not necessary to consider such a model.

Finally, the HLT test, which is valid for both unit-root and trend-stationary processes, is computed as a weighted average of the t_0^* and t_1^* tests:

$$t_\lambda^* = \lambda\, t_0^* + m\,(1-\lambda)\, t_1^* \tag{12}$$

with the weight λ given by

$$\lambda = \exp[\, -\,(\, 500\, S_0^*\, S_1^*\,)^2\,] \tag{13}$$

and S_0^* and S_1^* denoting the KPSS (see Kwiatkowski *et al.*, 1992) stationary test statistics calculated from the OLS residuals of equations (8) and (9) when evaluated at the τ_0^* and τ_1^* change dates respectively.[4] As shown by HLT, the 5% asymptotic critical value of t_λ^* equals 2.563 provided the constant m is set equal to 0.853.

3.2. Unit-root test

As mentioned above, standard unit-root tests have several shortcomings when it is not known *a priori* if a structural change has occurred or not. However, as shown in HHLT, the HLT t_λ^* structural change test in (12) can be used as a pre-test. If it rejects the null hypothesis of no change, an optimal test for a unit-root in the presence of a structural change should be used. If it does not reject, then an optimal unit-root test without allowing for a structural change should be used. We now describe these unit-root tests in more detail.

For the case where a structural change is detected with the HLT test, the estimated change fraction using the first-differenced model, τ_1^*, is used to estimate model (8) by GLS as in Perron and Rodriguez (2003). For $\tau = \tau_1^*$, equation (8) can be rewritten as

$$k_t = X_t(\tau_1^*)\ \theta_0 + u_t\,, \quad t = 1, \ldots, T, \tag{14}$$

[4] As shown in HLT, λ converges asymptotically to 1 if the series is trend-stationary or to 0 if it is difference-stationary at sufficiently fast rates so that the correct test statistic, t_0^* or t_1^* respectively, is selected.

where X_t (τ_1^*) = (1, t, DT_t (τ_1^*)) and $\theta_0 = (\alpha\ ,\ \beta\ ,\ \gamma\ )'$. The GLS estimation is implemented by estimating with OLS the following quasi-difference transformation of equation (14):

$$k_{c,t} = X_{c,t}(\tau_1^*)\ \theta_0 + u_{c,t}\ ,\quad t = 1, \ldots, T, \tag{15}$$

where $k_{c,t} = k_t - \rho\, k_{t-1}$ and $X_{c,t}$ (τ_1^*) = X_t (τ_1^*) - $\rho\, X_{t-1}$ (τ_1^*) for $t = 2, \ldots, T$, $k_{c,t} = k_1$ and $X_{c,t}$ (τ_1^*) = X_1 (τ_1^*) for $t = 1$, and $\rho = 1 - c/T$ where c is the quasi-difference parameter. The value of c can be chosen according to a local power criterion as explained in HHLT. Let $\hat{\theta}_c$ denote the OLS estimator of θ_0 in equation (15) and $\tilde{u}_t = k_t - X_t(\tau_1^*)\,\hat{\theta}_c$ the residuals. The unit-root test is finally obtained by estimating the following ADF regression:

$$\Delta\tilde{u}_t = \phi\tilde{u}_{t-1} + \sum_{j=1}^{p}\delta_j\Delta\tilde{u}_{t-j} + e_{p,t},\ t = p+2,\ldots,T\ . \tag{16}$$

The number of lags p can be chosen by a modified Akaike criterion (see HHLT for details). Values of c and critical values for several significance levels and estimated change fractions needed for the implementation of this test can be found in HHLT and Carrion-i-Silvestre *et al.* (2008).

When no structural change is found by the HLT test, a unit-root test without allowing for a change should be used such as the ADF-GLS optimal test proposed by Elliott *et al.* (1996). This test is computed as above but with the following two modifications: a) the structural change dummy variable is excluded from the $X_t(\tau_1^*)$ vector and b) the optimal value of c in this case equals 13.5.

4. EMPIRICAL APPLICATION

In this section, the methodology proposed above is illustrated by an application to the following 18 developed countries: Austria, Belgium, Canada,

Denmark, England and Wales, Finland, France, Ireland, Italy, Japan, Netherlands, Norway, Portugal, Spain, Switzerland, Sweden, United States of America, and West Germany. We use time-series data for the number of deaths, $D_{x,t}$, and exposure-to-risk, $E_{x,t}$, by single year of age, sex, and calendar year from 1950 until the most recent available year.[5] Data were obtained from the Human Mortality Database (HMD) (University of California and Max Planck Institute).[6]

We begin by applying the Brouhns *et al*. (2002) extension of the Lee-Carter method described in Section 2 to obtain estimates of the model parameters and the mortality index series for the male and female populations in each country. Next, we apply the procedure proposed in Section 3 to test for the presence of structural changes and unit-roots in the estimated mortality trends. Finally, we illustrate how these results can be used to arrive at appropriate forecasting models by considering the case of male mortality in Portugal. We compare the resulting mortality and life expectancy forecasts with those obtained using alternative models.

4.1. Mortality index estimation

The estimated mortality indices k_t for the male and female populations in each country are plotted in Figures 1.a-c and confirm the downward trend in mortality over time observed in all countries. A more detailed visual inspection of the graphs also seems to suggest that after the mid-1970s, the rate of decline, especially for male mortality, might have become more accentuated in a number of countries. Examples of perceptible structural changes can be found in the mid-1970s for Belgium, Canada, and West Germany, mid-1980s for Italy, Norway, and Sweden, mid-1990s for Denmark, and around 2000 for Ireland and Netherlands.

However, as explained in the previous section, in the eventual possible of a unit-root, there is a risk that such apparent structural changes may be nothing more

[5] The latest available years are 2005 for Austria, Canada, and U.S.A., 2006 for Belgium, England & Wales, France, Ireland, Italy, Netherlands, Spain, and West Germany, and 2007 for Denmark, Finland, Japan, Norway, Portugal, Switzerland, and Sweden. For West Germany data are available only since 1956.

[6] The $E_{x,t}$ are based on annual population estimates with a small correction that reflects the timing of deaths during the interval (see HMD 2007).

than illusions in the data.[7] Next, we present formal and valid tests for the genuine presence of structural changes in the mortality indices.

4.2. Structural change and unit-root tests

The traditional Box-Jenkins approach to forecasting would begin by analysing the stationarity of each series. For all the estimated k_t, the corresponding empirical autocorrelation functions approach zero very slowly while, on the contrary, the empirical partial autocorrelation functions cut off abruptly at lag one.[8] These are typical behaviours of non-stationary unit-root series and, according to the usual Box-Jenkins approach, a clear indication that the k_t series should be first-differenced, leading therefore to ARIMA(p,1,q) forecasting models. However, as explained above, this analysis is not valid in the presence of structural changes.

We now follow the sequential approach described in Section 3 to identify the correct forecasting model in the potential presence of a structural change. The results of the different structural change and unit-root tests for males and females are summarized in Table 1.

We begin by looking at the results of the t_0^* and t_1^* structural change tests.[9] For males, according to both tests, there is evidence of a structural change in the trend slope in all countries, except Finland and Spain. In these two countries, evidence is favourable to a structural change only in the case of the t_0^* test. The results of the robust t_λ* test confirm the presence of a structural change in all countries except Finland and Spain.

For females, the situation is quite different. Evidence is in general favourable to a structural change in the trend slope when using the t_0^* test, except for Finland, Spain, and Switzerland. However, for the t_1^* test, which is valid only if a unit-root is present, evidence of a structural change is favourable only for Austria, Ireland, and

[7] The problem of spurious breaks was first raised, from a graphical perspective, by Hendry and Neale (1991). A statistical explanation of the phenomena was given by Nunes *et al.* (1995, 1996).
[8] The same behaviour is obtained after removing a linear trend from the k_t series.
[9] Since there is no reason or evidence pointing to simultaneous changes in the level and slope of the downward trend in mortality, the structural change tests are performed using model A in HLT, which considers only a change in the trend slope.

Japan. The results of the t_λ* test confirm the presence of a structural change only in Austria, Ireland, Italy, Japan, and Sweden.

The next step consists of testing for a unit-root. In every case that the result of the t_λ* test points to a structural change, the appropriate unit-root test to use is the ADF-GLS test allowing for a change and denoted as ADF-GLS-break in Table 1. For males, except for Finland and Spain, this is the appropriate unit-root test to use, while for females it should be used only for Austria, Ireland, Italy, Japan, and Sweden. The ADF-GLS test not allowing for a break should be used in all the other cases. In almost all cases, the evidence supports the unit-root hypothesis. The few exceptions are the cases of males in Denmark and females in Japan and Sweden.

Finally, for the cases where in the first step a structural change was detected, the date when that change occurred is estimated in accordance with the conclusion of the appropriate unit-root test. Wherever a unit-root is present, the selected estimator of the structural change date corresponds to the date where the t_1^* statistic in (11) attains its maximum. In the few cases where a unit-root is rejected, the selected estimator of the structural change date corresponds to the date where the t_0^* statistic in (10) attains its maximum. These dates are presented in the last column of Table 1. Although not presented, for almost every case where a structural change is identified, the corresponding estimated magnitude of the change in the trend slope, given by γ in (8) or (9), is negative. The only exceptions are the cases of males and females in Japan and females in Sweden.

In summary, there is significant evidence supporting the presence of structural changes in the evolution of male mortality associated with a more accentuated decline in the overall mortality rate in recent years for almost every country considered. In contrast, evidence of structural changes in female mortality is found only for a few countries. It is interesting to note that in the case of the European countries these structural changes have all taken place in the second half of the sample, that is, after the mid-1970s. The evidence of a unit-root found in almost every case considered implies that the uncertainty regarding the future evolution of mortality will increase with the forecast horizon.

4.3. Forecasting Portuguese male mortality

To illustrate the impact of the results of the unit-root and structural change tests obtained in the previous sub-section in terms of mortality and life expectancy forecasts, we consider as an example the case of the Portuguese male population. According to the results of those tests (see Table 1), the adequate forecasting model to consider is an ARIMA(p,1,q) allowing for a change in the drift in 1996, that is, an ARMA(p,q) model for the first-differences of k_t including as deterministic regressors a constant term and a step dummy variable as in equation (9). The orders p and q of this model are identified by analysing the residual autocorrelation functions, and using the Ljung-Box Q-tests and the Akaike and Schwarz information criteria. The final estimated model is the following ARIMA(0,1,1) model allowing for the structural change in 1996:

$$k_t - k_{t-1} = -1.37 - 1.72 DU_t^{1996} + u_t, \text{ with } u_t = \varepsilon_t - 0.53\varepsilon_{t-1}, \qquad (17)$$

where $DU_t^{1996} = 1$ if $t > 1996$ and $DU_t^{1996} = 0$ if $t \leq 1996$, and ε_t is a white noise. More detailed estimation results for this model appear in the first column of Table 2. Using this model we obtain forecasts of the mortality index k_t for the period 2008-2050. These are presented in Figure 2 together with the corresponding 95% confidence bands.[10] As expected, the model predicts a decline in the mortality index at a pace that is consistent with the more pronounced decline in mortality in the final years of the 20th century and beginning of the 21st century. The fact that the confidence bands widen is a direct consequence of the unit-root non-stationarity of this model.

To better understand the impact of allowing for a structural change in the mortality index by following the approach proposed in this paper, we also considered models estimated with different specifications of the structural change and the unit-

[10] The computed confidence intervals only account for the variability coming from the fitted ARIMA model: parameter estimation errors and future shocks. Other sources of variability could be allowed for by using bootstrap methods as in Brouhns *et al.* (2005), Koissi *et al*. (2006), and Renshaw and Haberman (2008).

root. Estimation results for these models also appear in Table 2. In Figure 5 we plot the resulting point forecasts for the period 2008-2050 for some of these models. The optimal model forecasts the greater decline in the mortality index.

We begin by analysing the random walk with drift model, which was found to be the optimal ARIMA model for k_t in the original work of Lee and Carter for the US population and in many other applications of their method. This model appears in the second column of Table 2. However, according to the residual autocorrelation functions and the Ljung-Box Q-statistics, this model does not fit well the Portuguese mortality index. The unit-root model without allowing for a structural change that best fits the mortality index series is an ARIMA(1,1,0) model, which appears in the third column of Table 2.[11] The corresponding forecasts from this model appear in Figure 3. The first thing to note is that, in this model, the forecasted rate of decline in mortality is, by construction, basically given by the average rate of decline during the whole estimation period. Therefore, the more accentuated decline in mortality since the end of the 20th century until the end of the estimation sample is not translated into the future projections. Consequently, the projected decline in mortality is much less when compared with the optimal model, which allowed for a structural change in 1996. Another important difference is that the confidence bands for the projected mortality index are wider for this model. This is a consequence of the poorer fit of this model relative to the optimal structural change model.

We also considered trend-stationary models. When no break in trend was allowed for, the analysis of the residuals suggested an ARMA(1,1) model. However, this model resulted in an auto-regressive root very close to 1, which basically confirms the previously obtained unit-root tests results. Because the model also includes a linear trend, the estimated near unit-root would, in fact, lead to a quadratic trend. We conclude that this model is not able to adequately describe the mortality index.

Finally, we have considered a trend-stationary model allowing for a change in the trend slope as in equation (8). According to the t_0^* test statistic, the estimated

[11] An ARIMA(0,1,1) model was also reasonable but did not fit the model as well. The point forecasts of these models were very similar but with the simple random walk with drift model generating wider confidence bands.

break date is in 1973. The best model in this case is an ARMA(1,0) including as deterministic regressors a constant term, a linear trend, and a change dummy defined as $DT_t^{1973} = t - 1973$ if $t > 1973$ and $DT_t^{1973} = 0$ if $t \leq 1973$. Estimation results are presented in the fourth column of Table 2. The corresponding mortality index forecasts are graphed in Figure 4. The forecasted decline in mortality is also not as great as in the optimal forecasting model. Regarding the confidence bands, as expected, these do not grow as the forecasting horizon grows, since this model assumes stationarity around a broken trend. Thus, for longer horizons, this model will produce forecast intervals with the least amplitude, but obviously underestimating future uncertainty, given the results of the sequential testing procedure that led to the optimal forecasting model.

The mortality index forecasts are also used to produce forecasts of the age-specific mortality rates $\mu_x(t)$ using the previously estimated α_x and β_x in equation (4). From these, we obtain forecasts for life expectancy at birth and at age 65 for each model. These appear in Table 2 (see also Figure 6). In accordance with the forecasts obtained for the mortality index, the optimal model with a unit-root and a change in 1996 gives the highest projected gains in life expectancy during the next four decades. Life expectancy at birth is expected to increase by almost 9 years and life expectancy at age 65 by 6 years. The predicted value for life expectancy at birth in 2050 is 84.8 years with a 95% confidence interval of 84.0 to 85.6 years. Life expectancy at age 65 is predicted to grow from 16.7 years in 2007 to 22.7 years in 2050 with a ± 0.6 years 95% confidence interval. Also, in agreement with the results obtained regarding the confidence bands for the mortality forecasts, the amplitude of the confidence bands for the 2050 life expectancy forecasts is substantially smaller for the optimal model than for the unit-root models without allowing for a structural change, but larger when compared with the trend-stationary model with a break.

5. CONCLUSIONS

In this paper we propose a new methodology to forecast future mortality and life expectancy in the possible presence of a structural change in the context of the

Lee-Carter model. Specifically, the results of tests for a change in the trend of the general level of mortality and of the presence of a unit-root are used to identify the appropriate ARIMA model to be estimated and used to carry out the projections.

We apply the proposed procedure to post-1950s time-series mortality data for 18 developed countries. We find significant evidence of structural changes in the rate of decline in the overall male mortality rate for almost every country considered. These are associated with a more accentuated decline in mortality in recent years. In contrast, female mortality decline has remained stable for the majority of the countries. We also show, by studying the case of Portuguese male mortality, that accounting for such structural changes in a forecasting model can lead to major impacts in mortality and life expectancy forecasts over the next decades.

The sequential testing approach to detect unit-roots and structural changes was applied in the context of the typical Lee-Carter model. However, it should be possible to extend it to other variants of this model that allow, for instance, more than one latent mortality index, so as to capture dissimilar evolutions of mortality for different ages. Another extension of the proposed approach would be to consider the possibility of more than one structural change. These are left for future research.

ACKNOWLEDGMENTS

The authors thank Jorge Bravo and David Harvey for allowing us to use their codes for Poisson estimation and structural change testing, respectively.

REFERENCES

Alho, J. (2000). Discussion. *North American Actuarial Journal*, 4(1), 91-93.

Booth, H. (2006). Demographic Forecasting: 1980 to 2005 in Review. *International Journal of Forecasting*, 22, 547– 581.

Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee-Carter under Conditions of Variable Mortality Decline. *Population Studies*, 56, 325– 336.

Booth, H, Hyndman, R.J., Tickle, L., and Jong, P. (2006). Lee-Carter Mortality Forecasting: a Multi-country Comparison of Variants and Extensions. *Demographic Research*, 15, 289 – 310.

Brouhns, N. and Denuit, M. (2002). Risque de Longévité et Rentes Viagères II. Tables de Mortalité Prospectives pour la Population Belge. *Belgium Actuarial Bulletin*, 2(1), 2002.

Brouhns, N., Denuit, M. and Vermunt, J.K. (2002). A Poisson Log-linear Regression Approach to the Construction of Projected Lifetables. *Insurance: Mathematics and Economics*, 31, 373-393.

Brouhns, N., Denuit, M. and Van Keilegom, I. (2005). Bootstrapping the Poisson Log-bilinear Model for Mortality Forecasting. *Scandinavian Actuarial Journal*, 3, 212 -224.

Carrion-i-Silvestre, J. L., Kim, D. and Perron, P. (2008). GLS-based Unit Root Tests with Multiple Structural Breaks both Under the Null and the Alternative Hypotheses, *Econometric Theory,* 25, 1754-1792.

Carter, L.R. and Prskawetz, A. (2001). Examining Structural Shifts in Mortality Using the Lee-Carter Method. Max Planck Institute for Demographic Research, WP 2001-007.

Chan, W.-S., S.-H. Li, and S.-H Cheung (2008). Testing Deterministic Versus Stochastic Trends in the Lee-Carter Mortality Indexes and Its Implications for Projecting Mortality Improvements at Advanced Ages. Paper presented at the Living to 100 and Beyond Symposium, Orlando, FL, January 7–9, 2008.

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the Estimators for Autoregressive Time-series with a Unit-root. *Journal of the American Statistical Association*, 74, 427-431.

Elliot, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient Tests for an Autoregressive Unit-root, *Econometrica*, 64, 813-836.

Fígoli, M.G.B. (1998). Modelando e Projectando a Mortalidade no Brasil. *Revista Brasileira de Estudos de População*, 15(1), 95-113.

Harvey, D.I, Leybourne, S.J. and Taylor, A.M.R. (2009). Simple, Robust and Powerful Tests of the Changing Trend Hypothesis, *Econometric Theory*, 25, 995-1029.

Harris, D., Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2009). Testing for a Unit-root in the Presence of a Possible Break in Trend, *Econometric Theory*, 25, 1545-1588.

Hendry, D.F., and Neale, A.J. (1991). A Monte Carlo Study of the Effects of Structural Changes on Tests for Unit-roots, in "Economic Structural Change, Analysis, and Forecasting", P. Hackl and A. H. Westlund (eds.), Springer-Verlag, 95-119.

Hollmann, F.W., Mulder, T. and Kallan, J. (2000). Methodology and Assumptions for the Population Projections of the United States: 1999 to 2100. Population Division Working Paper no. 38, U.S. Census Bureau, Washington.

Human Mortality Database, HMD. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), Available at www.mortality.org or www.humanmortality.de (data downloaded on 17 Oct. 2008).

Human Mortality Database, HMD (2007). Methods Protocol for the Human Mortality Database. J.R. Wilmoth, K. Andreev, D. Jdanov, and D.A. Glei with the assistance of C. Boe, M. Bubenheim, D. Philipov, V. Shkolnikov, P. Vachon. Available at: www.mortality.org/Public/Docs/MethodsProtocol.pdf.

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

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P.J. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3), 159–178.

Lee, R. (2000). The Lee-Carter Method for Forecasting Mortality, with various Extensions and Applications. *North American Actuarial Journal*, 4(1), 80-93.

Lee, R. and Carter, L. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87(419), 659-675.

Lee, R. and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38(4), 537-549.

Lundstrom, H. and Qvist, J. (2004). Mortality Forecasting and Trend Shifts: an Application of the Lee-Carter Model to Swedish Mortality Data. *International Statistical Review*, 72(1), 37-50.

Nunes, L.C., Kuan, C.-M. and Newbold, P. (1995). Spurious Change. *Econometric Theory*, 11(4), 736-749.

Nunes, L.C., Newbold, P. and Kuan, C.-M. (1996). Spurious Number of Changes. *Economics Letters*, 50(2), 175-178.

Perron, P. (1989). The Great Crash, The Oil Price Shock, and the Unit-root Hypothesis. *Econometrica*, vol. 57, no. 6, 1361-1401.

Perron, P. (1997). Further Evidence on Changing Trend Functions in Macroeconomic Variables. *Journal of Econometrics*, 80, 355-385.

Perron, P. (2006). Dealing with Structural Breaks, in "Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory", K. Patterson and T.C. Mills (eds.), Palgrave Macmillan, 278-352.

Perron, P. (2008). Structural Change, in "The New Palgrave Dictionary of Economics", 2nd ed, S. Durlauf and L. Blume (eds.), Palgrave Macmillan.

Perron, P and Rodríguez, G. (2003). GLS Detrending, Efficient Unit-root Tests and Structural Change. *Journal of Econometrics*, 115, 1-27.

Renshaw, A.E. and Haberman, S. (2003). Lee-Carter mortality forecasting: a parallel generalized linear modeling approach for England and Wales mortality projections. *Applied Statistics, Journal of the Royal Statistical Society*, Series C, 52(1), 119-137.

Renshaw, A.E. and Haberman, S. (2008). On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee-Carter modelling, *Insurance: Mathematics and Economics*, 42(2), 797-816.

Stoto, M. A. (1983). The Accuracy of Population Projections. *Journal of the American Statistical Association*, 78(381), 13- 20.

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

Zivot, E. and Andrews, D.W.K. (1992). Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *Journal of Business & Economic Statistics*, 10(3), 251-270.

Wilmoth, J.R. (1993). Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality change. Technical Report, University of California, Berkeley. U.S.A.

Table 1. Structural change and unit-root tests

Males	Structural Change Tests			Unit-Root Tests		Conclusion	
Country	t_0*	t_1*	t_λ*	ADF-GLS	ADF-GLS-break	I(0)/I(1)	Break Date
Austria	17.21**	4.85**	4.49**	-0.93	-1.83	I(1)	1983
Belgium	15.91**	3.70**	3.16**	-0.64	-1.72	I(1)	1976
Canada	14.97**	5.59**	4.77**	-0.26	-2.01	I(1)	1975
Denmark	26.10**	7.40**	10.83**	-0.46	-3.24***	I(0)	1994
England & Wales	17.40**	5.01**	4.27**	-0.64	-1.90	I(1)	1979
Finland	8.98**	2.23	1.90	-1.41	-1.70	I(1)	-
France	12.41**	3.13**	3.93**	-1.15	-2.39	I(1)	1985
Ireland	10.04**	8.02**	6.81**	-1.02	-1.40	I(1)	1999
Italy	18.10**	5.80**	4.95**	-0.95	-2.51	I(1)	1983
Japan	8.92**	3.44**	2.94**	-0.80	-1.41	I(1)	1955
Netherlands	9.78**	5.37**	4.58**	-0.68	-1.18	I(1)	2000
Norway	19.54**	5.64**	4.81**	-1.27	-2.00	I(1)	1988
Portugal	8.32**	3.77**	3.18**	-0.10	-1.83	I(1)	1996
Spain	3.49**	2.77	2.32	-2.24	-3.20***	I(1)	-
Switzerland	10.03**	4.00**	3.41**	-0.92	-1.79	I(1)	1990
Sweden	22.86**	5.35**	4.83**	-1.21	-1.66	I(1)	1988
United States	14.58**	4.55**	4.04**	-0.78	-2.79	I(1)	1968
West Germany	13.72**	4.88**	4.16**	-0.83	-1.73	I(1)	1975

Females	Structural Change Tests			Unit-Root Tests		Conclusion	
Country	t_0*	t_1*	t_λ*	ADF-GLS	ADF-GLS-break	I(0)/I(1)	Break Date
Austria	12.82**	3.10**	2.66**	-0.94	-1.42	I(1)	1983
Belgium	7.69**	2.00	1.71	-1.43	-2.00	I(1)	-
Canada	5.23**	1.97	1.68	-1.77	-2.80	I(1)	-
Denmark	6.46**	2.55	2.17	-1.25	-2.31	I(1)	-
England & Wales	5.90**	2.32	2.19	-1.41	-2.05	I(1)	-
Finland	2.44	1.91	1.63	-1.53	-1.83	I(1)	-
France	3.97**	1.40	1.42	-1.78	-2.61	I(1)	-
Ireland	7.12**	5.22**	4.46**	-1.74	-1.38	I(1)	1999
Italy	12.05**	2.61	5.41**	-0.78	-1.77	I(1)	1983
Japan	5.75**	4.26**	4.19**	-1.46	-4.32***	I(0)	1955
Netherlands	6.52**	1.81	1.54	-1.16	-1.80	I(1)	-
Norway	4.29**	1.57	1.34	-2.24	-2.57	I(1)	-
Portugal	9.86**	2.88	2.46	-0.43	-2.20	I(1)	-
Spain	1.59	1.45	1.24	-3.02	-2.96	I(1)	-
Switzerland	1.94	1.60	1.36	-1.92	-2.28	I(1)	-
Sweden	5.19**	1.44	3.35**	-1.00	-3.51***	I(0)	1984
United States	3.76**	1.96	1.67	-1.50	-2.05	I(1)	-
West Germany	7.26**	2.34	1.99	-1.56	-2.05	I(1)	-

Notes: ** denotes rejection of the null hypothesis of no structural change at the 5% level. *** denotes rejection of the null hypothesis of a unit-root at the 5% level. I(0) denotes trend-stationarity and I(1) denotes a unit-root process. The following critical values were used: 2.56 for t_0^* and t_λ^*, and 3.00 for t_1^* (Table 1 in HLT); -3.19 for ADF-GLS (Table I in Elliot *et al.*, 1996); and a set of critical values that depend on the break fraction (between -3.45 and -3.09) for ADF-GLS-break (Table 1 in Carrion-i-Silvestre *et al.*, 2008).

Table 2. Estimated forecasting models for the Portuguese male mortality index

	(1) ARIMA(0,1,1) break in 1996	(2) ARIMA(0,1,0) no break	(3) ARIMA(1,1,0) no break	(4) ARMA(1,0) break in 1973
Constant	-1.37** (0.16)	-1.70** (0.35)	-1.69** (0.23)	38.57** (2.90)
DU_t^{1996}	-1.72** (0.38)			
t				-0.92** (0.16)
DT_t^{1973}				-1.12** (0.21)
$AR(1)$			-0.40** (0.13)	0.59** (0.13)
$MA(1)$	-0.53** (0.12)			
Log-Likelihood	-121.65	-132.65	-127.86	-125.74
AIC	4.45	4.77	4.64	4.63
SBC	4.56	4.81	4.71	4.78
Q(4) p-value	0.509	0.011	0.840	0.113
No. Obs.	56	56	56	56
Forecasts of life expectancy at birth				
2030	80.9 [80.2 , 81.7]	78.7 [76.8 , 80.5]	78.7 [77.4 , 79.9]	78.8 [78.2 , 79.2]
2050	84.8 [84.0 , 85.6]	81.1 [78.7 , 83.3]	81.1 [79.5 , 82.6]	81.6 [81.1 , 82.1]
Forecasts of life expectancy at age 65				
2030	20.0 [19.5 , 20.5]	18.5 [17.3 , 19.7]	18.5 [17.7 , 19.3]	18.5 [18.2 , 18.9]
2050	22.7 [22.1 , 23.3]	20.1 [18.5 , 21.6]	20.1 [19.1 , 21.1]	20.5 [20.1 , 20.8]

Notes: ** denotes significance at the 1% significance level. Standard errors appear in parentheses. 95% forecast confidence intervals appear in square brackets.

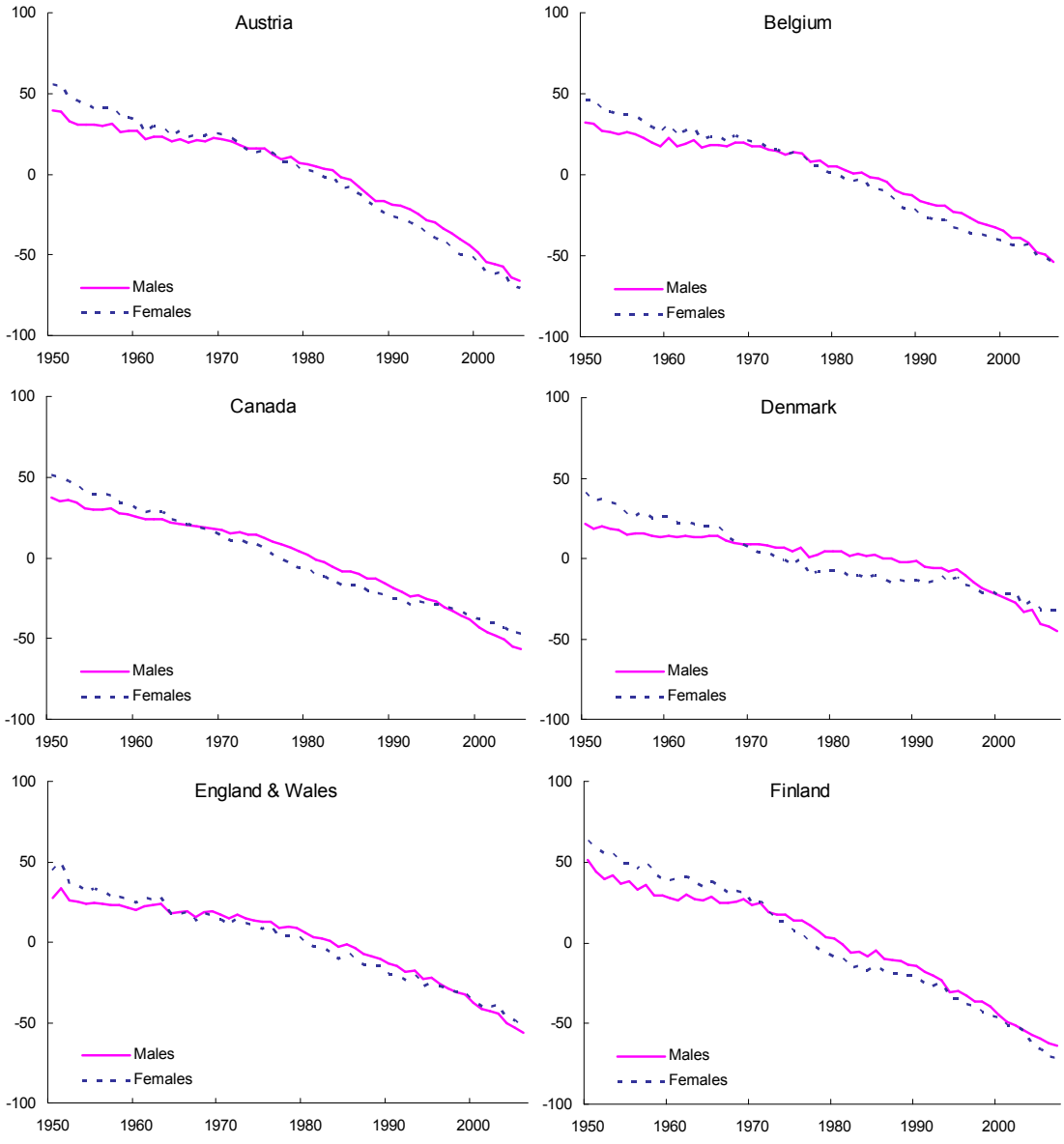

Figure 1.a. Estimated overall mortality indices

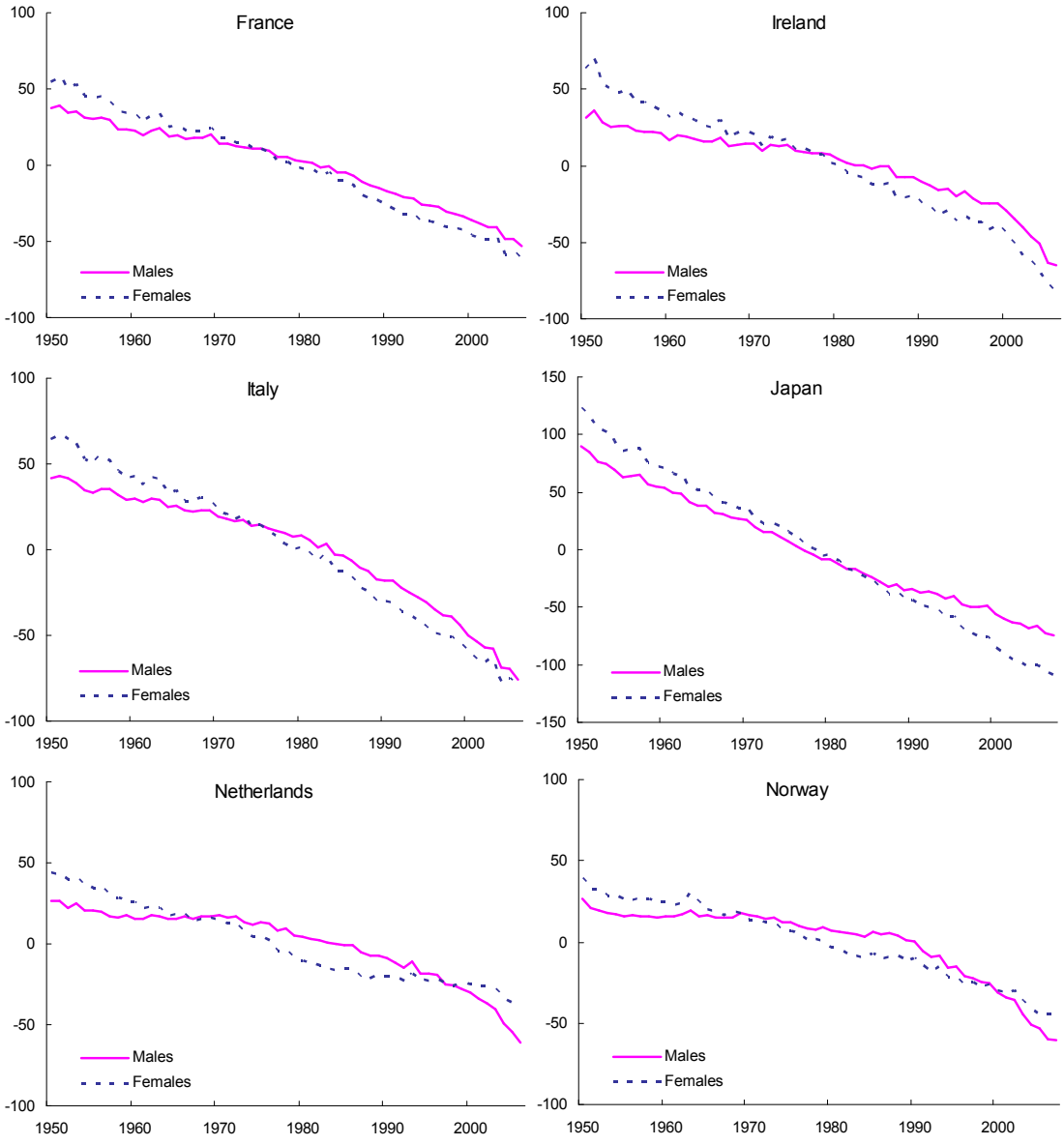

Figure 1.b. Estimated overall mortality indices (continued)

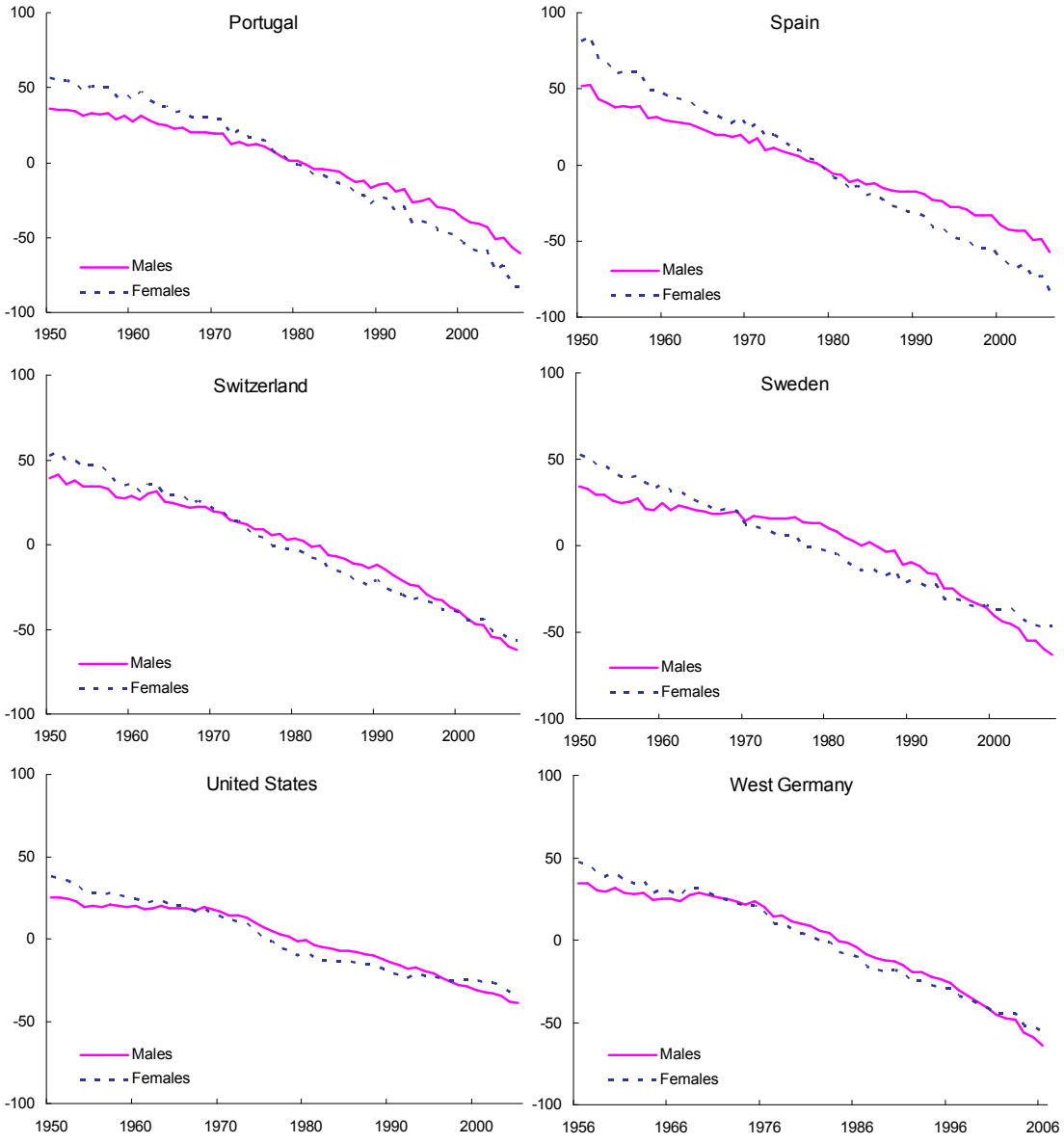

Figure 1.c. Estimated overall mortality indices (continued)

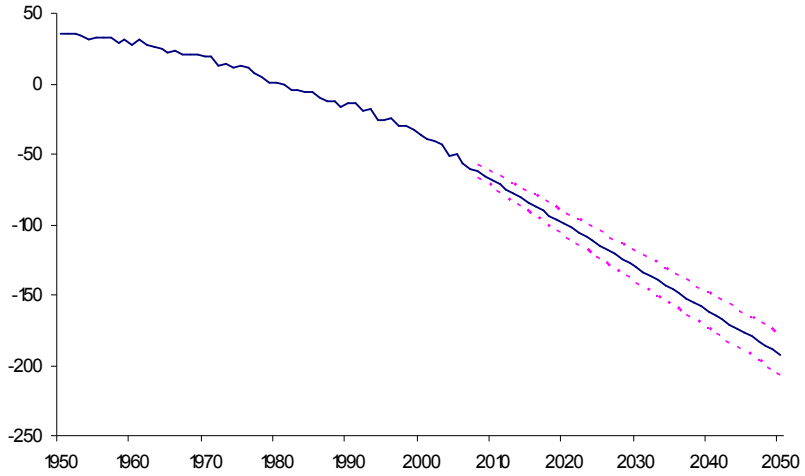

Figure 2. Portuguese males mortality index (1950-2007) and forecasts from an ARIMA(0,1,1) with a break in drift in 1996 (2008-2050)

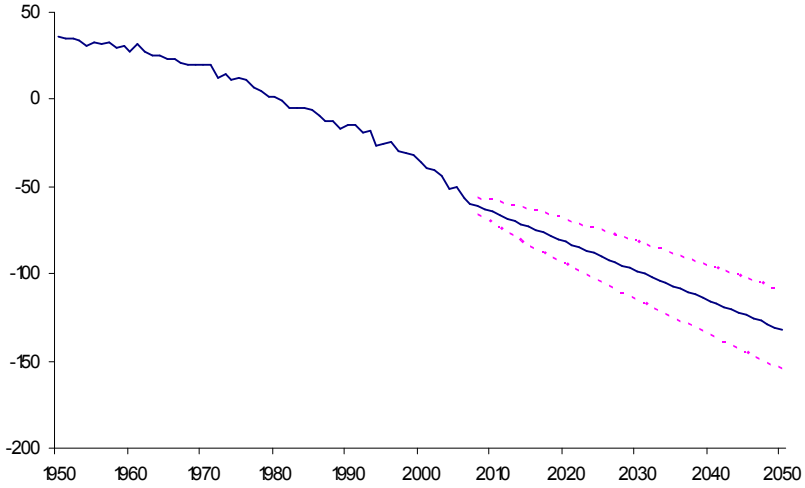

Figure 3. Portuguese males mortality index (1950-2007) and forecasts from an ARIMA(1,1,0) with no break (2008-2050)

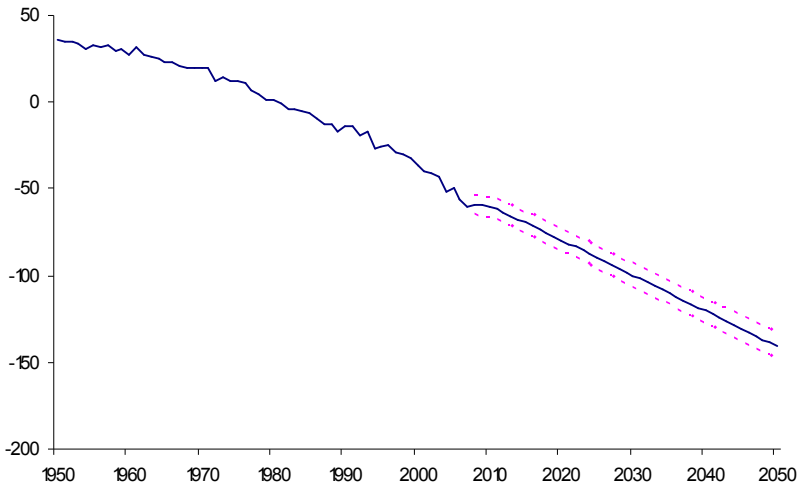

Figure 4. Portuguese males mortality index (1950-2007) and forecasts from an ARMA(1,0) with a break in trend slope in 1973 (2008-2050)

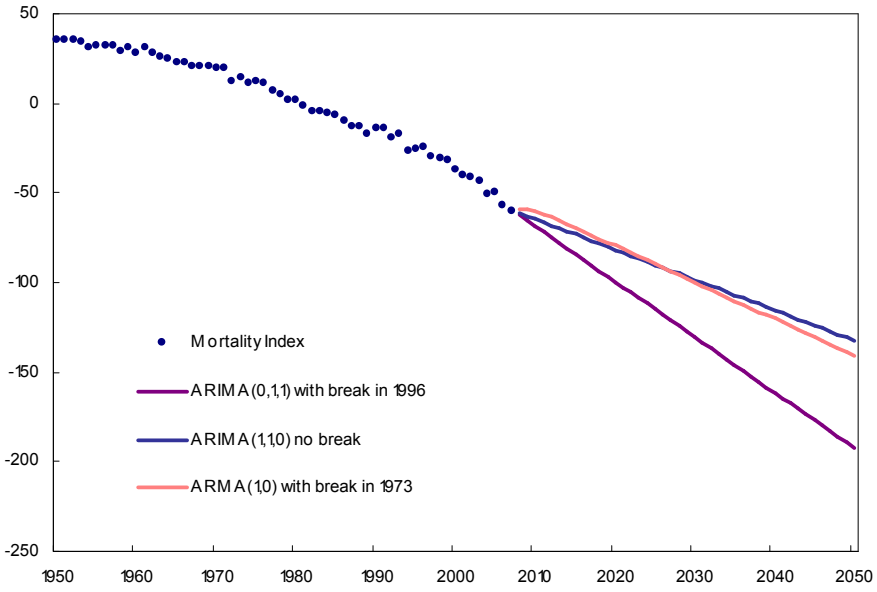

Figure 5. Portuguese males mortality index (1940-2007) and forecasts for several models (2008-2050)

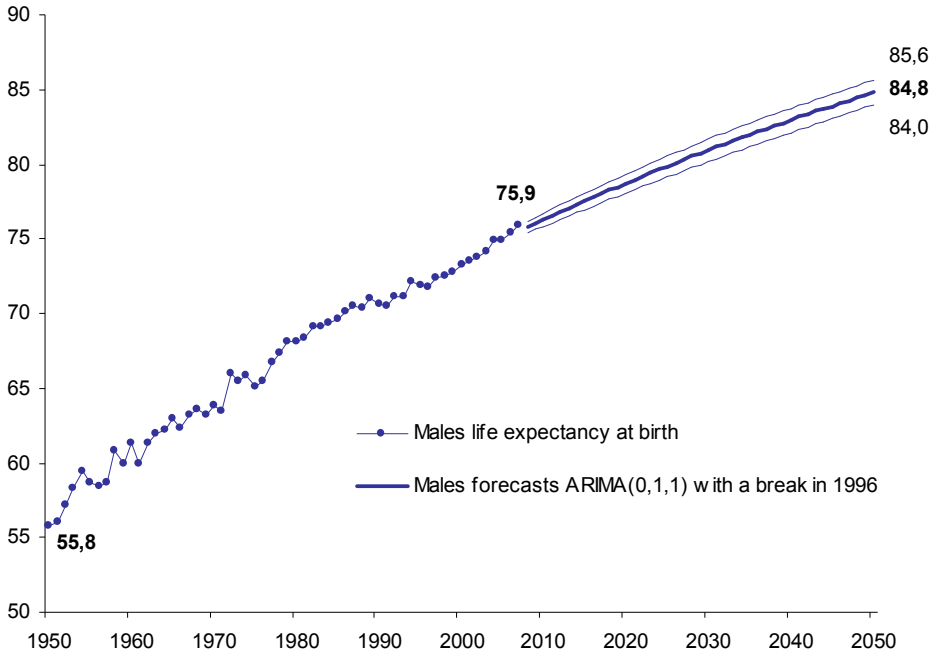

Figure 6. Portuguese males life expectancy at birth (1950-2007) and forecasts from an ARIMA(0,1,1) with a break in drift in 1996 (2008-2050).