Cass Business School
City of London

Faculty of Actuarial Science and Statistics

Lee Carter Mortality Forecasting: Application to the Italian Population

Steven Haberman
and
Maria Russolillo

Actuarial Research Paper No. 167

November 2005

ISBN 1 901615-93-6

Cass Business School
106 Bunhill Row
London EC1Y 8TZ
T +44 (0)20 7040 8470
www.cass.city.ac.uk

“Any opinions expressed in this paper are my/our own and not necessarily those of my/our employer or anyone else I/we have discussed them with. You must not copy this paper or quote it without my/our permission”.

Lee-Carter mortality forecasting: application to the Italian population*

Steven Haberman† and Maria Russolillo‡

November 2005

Abstract

In this paper we investigate the feasibility of using the Lee-Carter methodology to construct mortality forecasts for the Italian population. We fit the model to the matrix of Italian death rates for each gender from 1950 to 2000. A time-varying index of mortality is forecasted in an ARIMA framework and is used to generate projected life tables. In particular we focus on life expectancies at birth and, for the purpose of comparison, we introduce an alternative approach for forecasting life expectancies on a period basis. The resulting forecasts generated by the two methods are then compared.

Keywords: Lee-Carter methodology; Mortality forecasting; Time series; Life expectancy

JEL Classification: C53, G22

* We wish to thank Emilia Di Lorenzo for her helpful and constructive comments.
† Faculty of Actuarial Science and Statistics, Cass Business School, City University, London, e-mail: s.haberman@city.ac.uk
‡ Department of Mathematics and Statistics, University of Napoli "Federico II", e-mail: mrussoli@unina.it

1. Introduction and motivation

1.1. Mortality on the move

During the 20th century life expectancy has increased dramatically. The Human Mortality Database shows that Italian life expectancy at birth from 1900 to 1999 rose from 41.90 to 82.26 years for females and from 41.65 to 76.12 for males. Moreover, the trends in mortality rates for many industrialised countries have also been downwards for several years. Usually we view such mortality improvements in an optimistic way: according to the statistics we live longer than our ancestors. But these changes clearly affect pricing and reserve allocation for life annuities and represent one of the major threats to a social security system that has been planned on the basis of more modest life expectancy. Even when using updated mortality tables, these trends in mortality reduction present risks for insurers. This is because these tables do not take these trends into account. Put otherwise, the risk is of underestimating the survival probability, thus determining inappropriate premiums. This risk, is known in the actuarial literature as Longevity Risk, that being the risk derived from a future mortality rate which, ex post, does not reflect the forecasted one: see Brouhns, Denuit, Vermunt (2002b). To face this risk, it is necessary to build projected tables including this trend. Thus, reasonable mortality forecasting techniques have to be used to consistently predict the trends (Brouhns, Denuit, Vermunt, 2002a).

In this paper, we investigate how the Lee Carter approach can be used to forecast mortality (Lee and Carter, 1992; Lee, 2000; Lee and Miller, 2000; Lee and Miller, 2001), by using the Italian mortality experience of the past half-century. We follow the methodology of Renshaw and Haberman (2003a), which is the inspiration for the paper.

There were two reasons for selecting the Lee-Carter model in our work. Firstly, this model represents one of the most influential recent developments in the field of mortality forecasts. Secondly, the important feature of this model is that for a precise value of the time index k, we can define a complete set of death probabilities that allow us to calculate all of the life table. Once we estimate the parameters, depending on age $\{\alpha_x, \beta_x\}$, they stay constant and invariant through time. Hence, when we know k, we can use the parameters for any year of interest. Another important feature that drove us to choose this model is that traditional projection models provide the forecaster with point estimates of future mortality rates. On the contrary, the LC method allows for uncertainty in forecasts (the so-called longevity risk).

The paper is organised as follows. Section 2 describes the Lee-Carter method for mortality projection and introduces the notation used in this paper. Model fitting on Italian mortality data is illustrated, with particular attention to the re-estimation of k_t. The standard Box and Jenkins methodology to generate an ARIMA model for the mortality index k_t is discussed in Section 3. Section 4 is devoted to forecasting the index of mortality, which is used to generate associated life table values. Next a comparison between the LC and the alternative approach to forecast life expectancies at birth is examined. Concluding comments are presented in Section 5.

2. Lee-Carter mortality forecasting methodology

2.1. The model

The Lee-Carter method is a powerful approach to mortality projections which describes the log of a time series of age-specific death rates $m_{x,t}$ as the sum of an age-specific component α_x, that is independent of time and

another component that is the product of a time-varying parameter k_t, reflecting the general level of mortality, and an age-specific component β_x, that represents how rapidly or slowly mortality at each age varies when the general level of mortality changes:

$$(1) \quad \ln(m_{x,t}) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}$$

This interesting alternative for forecasting mortality was proposed in 1992 by Lee and Carter, who published a new method extrapolating long-run forecasts of the level and age pattern of mortality, based on a combination of statistical time series methods and parametric approach.

2.2. Notation and data

In this contribution we fit the Lee-Carter model to the matrix of Italian death rates, from year 1950 to 2000. Then we use the forecasts of this single parameter to generate forecasts both of the level and of the age distribution of mortality for the next 25 years. In particular we focus on life expectancies at birth and, for the purpose of comparison, we introduce an alternative approach for forecasting life expectancies on a period basis.

The data for the Italian population, supplied by the Human Mortality Database, is divided by gender (Wilmoth et Al., 2000). Rather than using the entire dataset, we consider a subgroup of death rates for five-year age groups under 105 years old, so as to only cover five-year groups with a sample size significant enough for our analysis. The same is repeated for the corresponding exposure to risk. We denote the "Number of deaths" and the "Exposure to risk" by two 5×1 matrices, where the first number refers to the age interval, and the second number refers to the time interval (Elandt-Johnson and Johnson, 1980). For each gender and for each calendar year: $t = t_1, t_1 + 1, \ldots\ldots\ldots, t_1 + h - 1 = t_n$, where $h = t_n - t_1 + 1$, we consider all the ages $x = x_1, x_2, \ldots\ldots\ldots, x_k$, grouped in classes as

$[0,1-4,5-9,10-14,.....,95-99,100-104]$. From these data we construct an array of crude rates of deaths $m_{x,t} = \frac{d_{x,t}}{e_{x,t}}$.

2.3. Model fitting

The LC model cannot be fitted by ordinary regression methods, because there are no given regressors; thus in order to find a least squares solution to the equation (1) we use a close approximation, suggested by Lee and Carter (1992), to the singular value decomposition (SVD) method, assuming that the errors are homoschedastic. To obtain a unique solution, we impose that the sum of the β_x coefficients is equal to 1.0, and that the sum of the k_t parameters is equal to zero.

Under these assumptions, it can be seen that the α_x coefficients must be simply the average values over time of the $\ln(m_{x,t})$ values for each x.

We estimate α_x as the logarithm of the geometric mean of the crude mortality rates, averaged over all t, for each x:

$$(2) \quad \alpha_x = \tfrac{1}{h}\sum_{t=t1}^{tn} \ln m_{xt} = \ln\left[\prod_{t=t1}^{tn} m_{xt}^{\frac{1}{h}}\right]$$

Furthermore, k_t must equal the sum over age of $(\ln(m_{x,t}) - \alpha_x)$. All that remains, is to estimate the $\beta_x s$. We found each β_x by regressing $(\ln(m_{x,t}) - \alpha_x)$ on k_t, without a constant term, separately for each age group x. More precisely, we estimate β_x from $(\ln m_{xt} - \alpha_x) = \beta_x k_t^{(1)} + \varepsilon_{xt}^{'}$ (where $k_t^{(1)}$ refers to the k_t estimated above) using the least squares estimation, i.e. choosing β_x to minimize $\sum_{x,t}\left(\ln m_{xt} - \alpha_x - \beta_x k_t^{(1)}\right)^2 \Rightarrow \beta_x = \frac{\sum_{t=t1}^{tn} k_t^{(1)}(\ln m_{xt} - \alpha_x)}{\sum_{t=t1}^{tn} k_t^{(1)2}}$. The raw estimates of α_x, β_x and k_t are inserted in the Appendix A.

Here α_x describes the general age shape of the age specific death rates $m_{x,t}$, while k_t is an index that describes the variation in the level of mortality to t. The β_x coefficients describe the tendency of mortality at age x to change when the general level of mortality (k_t) changes. When β_x is large for some x, then the death rate at age x varies substantially when the general level of mortality changes (as with $x = 0$ for infant mortality, for example) and when β_x is small, then the death rates for that age vary little when the general level of mortality changes (as is often the case with mortality at older ages).

The Lee Carter model also assumes that all the age specific death rates move up or down together, although not necessarily by the same amounts, since all are driven by the same period index, k_t. Although not all occurrences of β_x need to have the same sign, in practice all the $\beta_x s$ do have the same sign, at least when the model is fit over fairly long periods. As shown in the Appendix A, the $\beta_x s$ for both females and males have the same sign, which is positive. In Fig. 1, the values of β_x, as determined with the SVD, are plotted against x, for each case separately i.e. by gender.

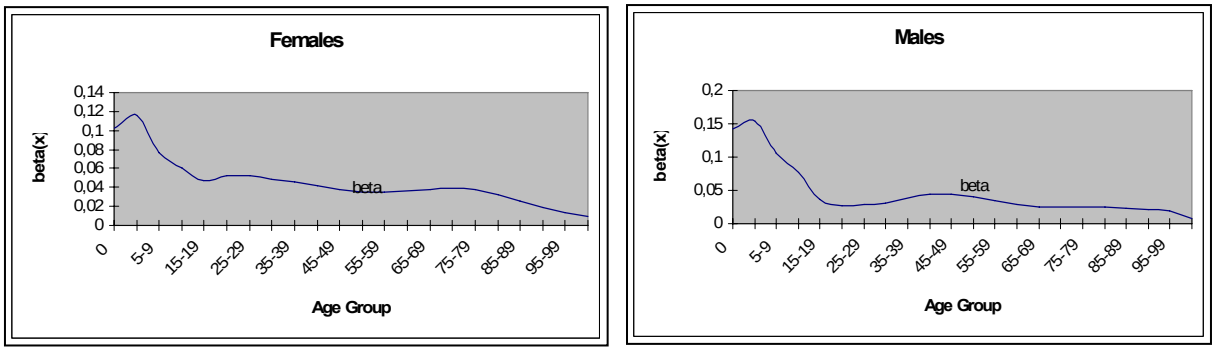

Fig.1 Beta versus age

From Fig.1 we can see that when β_x is large for some x, then the death rate at age x varies significantly when the general level of mortality changes (again, as with $x = 0$ for infant mortality) and when β_x is small,

then the death rate at that age varies little when the general level of mortality changes. This often the case with mortality at older ages.

2.4. Reestimating kt

Because the first stage estimation is based on logs of death rates rather than the death rates themselves, sizable discrepancies can occur between predicted and actual deaths. To guarantee that the fitted death rates will lead to the actual numbers of deaths, when applied to given population age distribution, we have reestimated k_t in a second step, taking the α_x and β_x estimates from the first step. To correct for this, we apply the methodology from Section 3 of Lee and Carter (1992). We thereby find a new estimate for k by an iterative search, adjusting the estimated k_t so that the actual total observed deaths $\sum_{x=x1}^{xk} d_{xt}$ equal the total expected deaths $\sum_{x=x1}^{xk} e_{xt} e^{(\alpha_x+\beta_x k_t)}$, for each year t.

The iterative method proceeds as follows:

1) We compare the total expected deaths $\sum_{x=x1}^{xk} e_{xt} e^{(\alpha_x+\beta_x k_t^{(1)})}$ to the actual total observed deaths $\sum_{x=x1}^{xk} d_{xt}$ in each period.

2) This comparison reveals one of three possible states:

   (i) If $\sum_{x=x1}^{xk} e_{xt} e^{(\alpha_x+\beta_x k_t^{(1)})} > \sum_{x=x1}^{xk} d_{xt}$, we need to decrease the expected deaths, adjusting the estimated k_t so that the new estimate of k_t, say $k_t^{(2)}$, will be: $k_t^{(2)} = k_t^{(1)}(1-d)$, if $k_t^{(1)} > 0$ (where $k_t^{(1)}$ is the first estimate of k_t) ; $k_t^{(2)} = k_t^{(1)}(1+d)$, if $k_t^{(1)} < 0$, where d is a small number.

   (ii) If $\sum_{x=x1}^{xk} e_{xt} e^{(\alpha_x+\beta_x k_t^{(1)})} = \sum_{x=x1}^{xk} d_{xt}$, we stop here the iterations.

(iii) If $\sum_{x=x1}^{xk} e_{xt} e^{(\alpha_x + \beta_x k_t^{(1)})} < \sum_{x=x1}^{xk} d_{xt}$, we need to increase the expected deaths adjusting the estimated k_t so that : $k_t^{(2)} = k_t^{(1)}(1+d)$, if $k_t^{(1)} > 0$; $k_t^{(2)} = k_t^{(1)}(1-d)$, , if $k_t^{(1)} < 0$.

3) Go back to Step 1.

As Lee and Carter (1992) point out, this approach differs from the direct SVD estimates. This is because the low death rates of youth contribute far less to the total deaths, yet when fitting the log-transformed rates they are weighted equivalently to the high death rates of the older ages. It is also worth noting that differences in population age group sizes also results in different weights in the second-stage estimation of k.

2.5. First application and comments

We have run this iterative process 1000 times using a VBA macro and Microsoft Excel to find the new estimate of k, shown in the Appendix B.

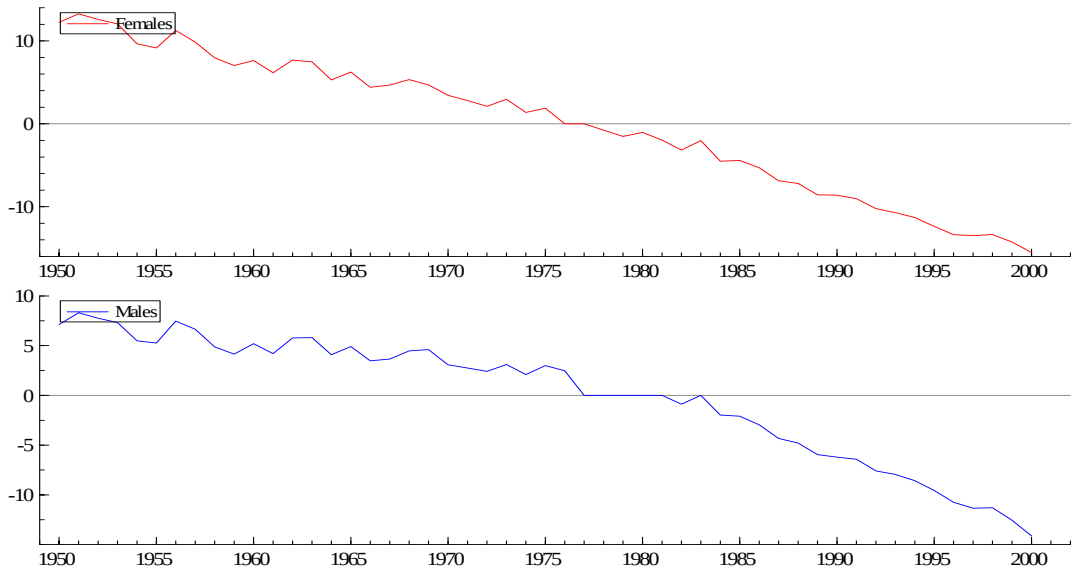

Fig.2 Re-estimates of k

Fig.2 plots estimates of k, for females and males; as shown, k declines roughly linearly from 1950 - 2000, more for females than for males. If we look at the values of k, shown in Appendix B, k declines at about the same pace during the first half of the period as it does during the second half. It also is striking that short-run fluctuations in k do not appear much greater in the first part of the period than they do in the second, with the exception of the male series in the first years. We can note that these results are consistent with the findings of Lee and Carter (1992) in their analysis of the total USA population. Both these features of k (its linear decline and its relatively constant variance) are very convenient for forecasting purposes. We can see from the re-estimated k_t that mortality improved in Italy. For the purposes of comparison with other countries, for example Britain (as presented in Renshaw and Haberman, 2003a), we can see that the Italian improvement is more pronounced. This is probably due to the fact that mortality was initially higher in Italy than in Britain, making the relative improvement greater and therefore more apparent. If we compare male to female mortality we might expect to see the same effect. Male mortality is higher than female mortality, thus possible improvements in male mortality could again be more evident than improvements in female mortality in an analogous way to the country comparison.

3. ARIMA methodology

3.1. Modelling mortality index

The estimated time-dependent parameter k_t can be modelled as a stochastic process; we thus used the standard Box and Jenkins methodology (identification-estimation-diagnosis) to generate an appropriate ARIMA (p,d,q) model for the mortality index k_t (Box and Jenkins, 1976; Hamilton, 1994).

Considering the time series given by the reestimated k_t, we need to identify a correct model, for our series, among the general class of ARIMA models. The procedure to construct the model goes through different iterative phases to arrive at a model that fits our data well (Francis X. Diebold, 2004; Makridakis, Wheelwright, Hyndman, 1998). The phases are the following:

1) Preliminary analysis of the series and possible transformation.
2) Identification of the order of the model.
3) Parameter estimation.
4) Evaluation of the model.

In the first step, we analyse the general pattern of the time series, as is illustrated in Fig. 2. A clear, almost linear, trend emerges, indicating that mortality enjoyed a steady erosion over the years.

The input series for an ARIMA needs to be stationary, that is, it should have a constant mean, variance, and autocorrelation through time. Therefore, the series usually needs to be differenced first until it is stationary. The number of times the series needs to be differenced to achieve stationarity is reflected in the d parameter. In order to determine the necessary level of differencing, one should examine the plot of the data and autocorrelogram, that displays graphically and numerically the autocorrelation function (ACF). We examine the ACF of the series and choose the value of d that gives rise to a rapid decrease of the ACF towards zero.

3.2. Identification phase

In the Identification phase, after we made the series stationary, we also need to decide how many autoregressive parameters (p) and/or moving average parameters (q) are necessary to yield an effective, but still parsimonious model of the process. We experimented with twelve models,

based on combinations of the p and q parameters varying between zero and two. The sample autocorrelations and partial autocorrelations, together with related diagnostics, provided graphical aids to model selection. This complemented our automatic identification criteria, the Akaike and Schwarz information criterion per model. To guide model selection we use these two criterions even though the SIC usually selects more parsimonious models due to its greater concern over the number of parameters to be estimated. Using a model selection strategy involving not just examination of AIC and SIC, but also examination of autocorrelations and partial autocorrelations, we are led to choose the ARIMA (0,1,0) for males and an ARIMA (0,1,1) for female. For males a model with an ar(1) term added could be marginally superior, but we preferred a random walk with drift on grounds of parsimony. We examine the general pattern of the time series for both genders in Fig.2, and we saw that a clear, decreasing trend emerges for each, indicating that the series are not stationary in mean. We are led to the same conclusions if we look at the autocorrelation function or the partial autocorrelation functions in Fig.3 (females) and 4 (males).

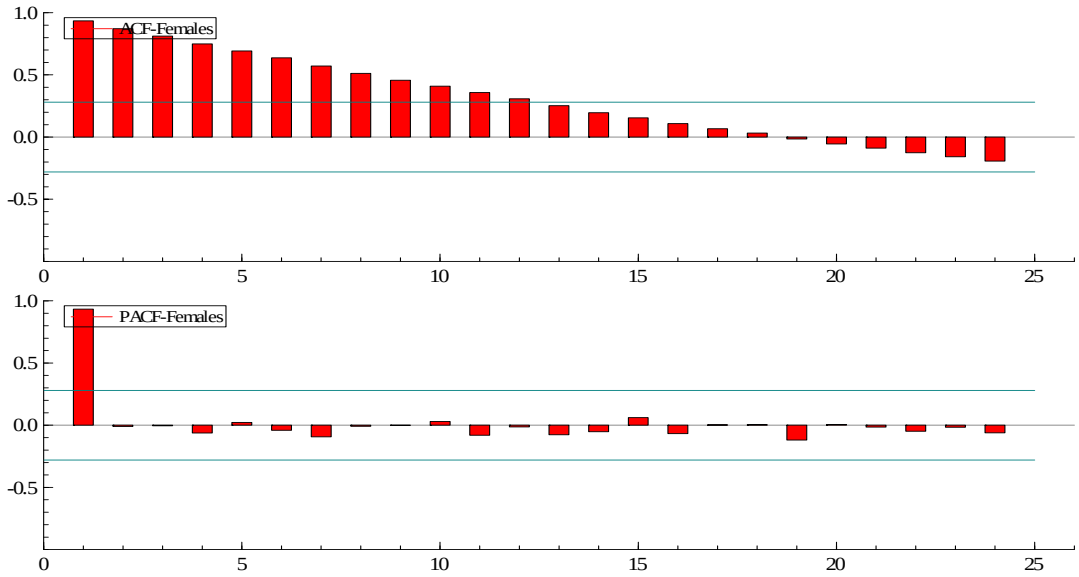

Fig.3 Female autocorrelation and partial autocorrelation function

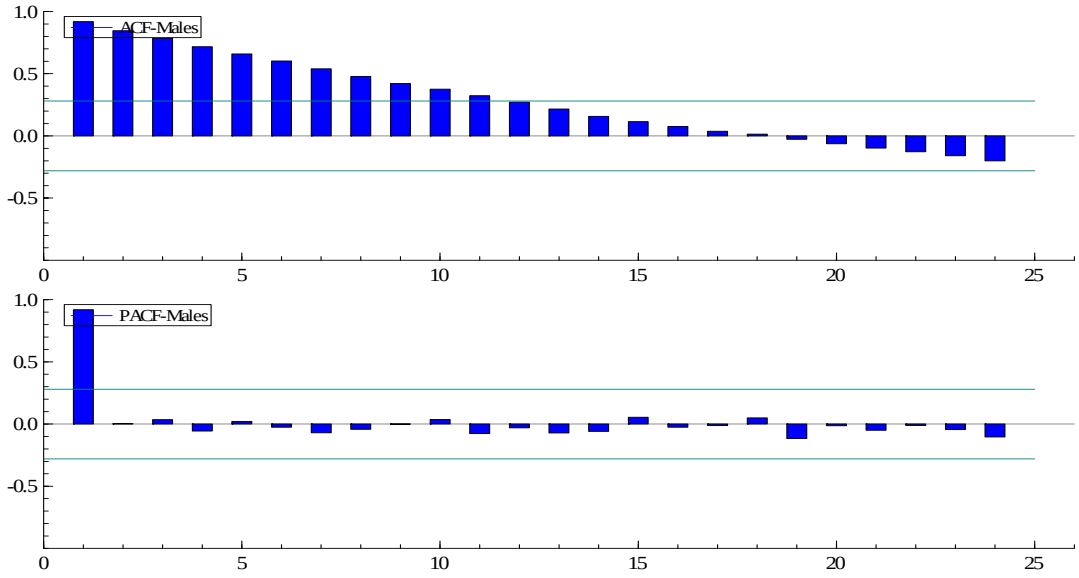

Fig.4 Male autocorrelation and partial autocorrelation function

As we can see, if we look at the graph of the autocorrelation function (ACF), this approaches zero gradually rather than abruptly. On the contrary, the partial autocorrelation function (PACF) cuts off abruptly; specifically, at displacement 1, the partial autocorrelations are significant while coefficients on all longer lags are zero. This is a clear sign of a nonstationary series.

Thus, following the Box and Jenkins methodology, we considered the differenced series, which we show in Fig.5

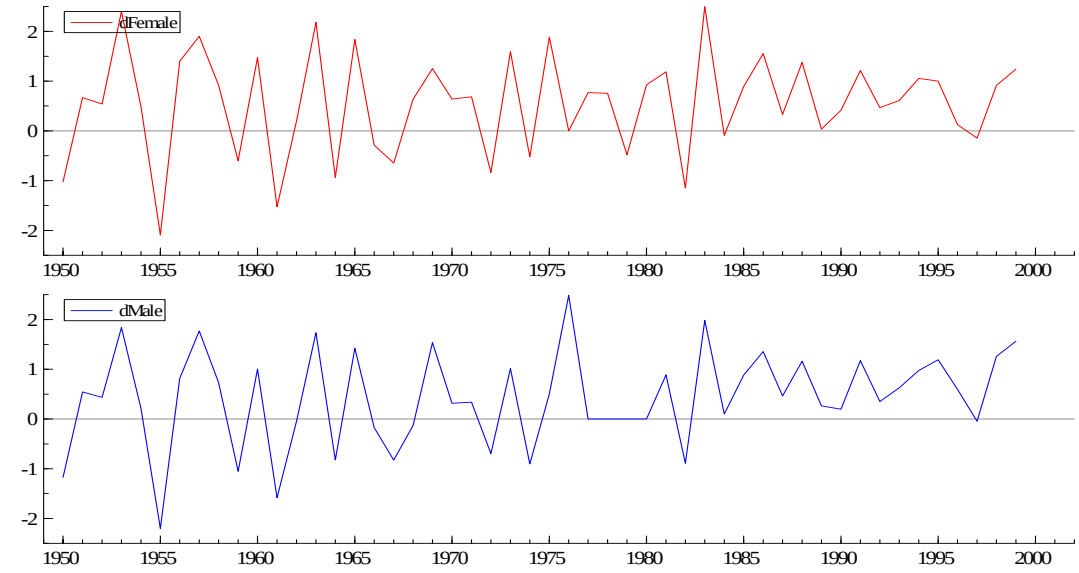

Fig.5 Differenced female and male series

After differencing the series, the nonstationarity in mean seems to be eliminated. Also the autocorrelation and partial autocorrelation functions (Fig.6), become consistent with the hypothesis of a stationary series. Because of the decreasing trend, when we estimated our model we also took a constant into consideration.

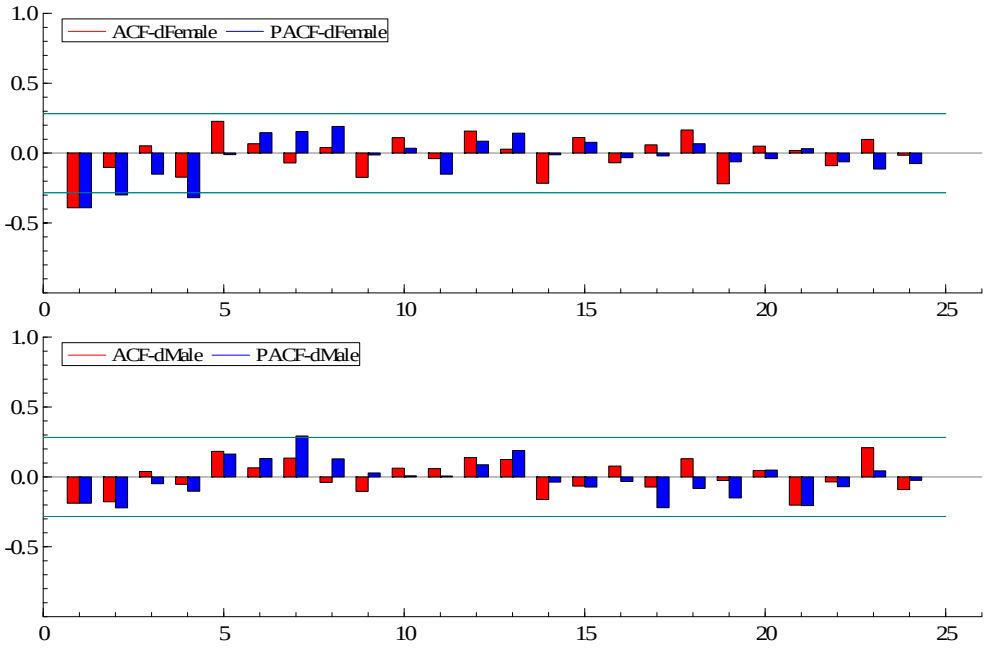

Fig.6 Autocorrelation and partial autocorrelation functions after differencing the series

3.3. Parameters estimation

Concerning the third phase, there are several different methods for estimating parameters. All of them should produce very similar estimates, but may be more or less efficient for any given model. Model parameters are estimated using statistical software, in our case time series estimation was performed by EViews using a least squares procedure. The k_t index for males was modelled as an ARIMA (0,1,0) process, i.e.:

$$K_t = K_{t-1} + \lambda + \varepsilon_t$$

and for females as an ARIMA (0,1,1) process, i.e.:

$$K_t = K_{t-1} + \lambda + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

The constant terms λ indicate the average annual change of k_t. It is this change that drives the forecasts of the long-run change in mortality. θ represents the moving average term.

The estimated parameters for both genders, and their standard errors, appear in the table below:

Male ARIMA (0,1,0)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
λ	-0.424882	0.137488	-3.090321	0.0033

Female ARIMA (0,1,1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
λ	-0.566485	0.045168	-12.54168	0.0000
θ	-0.644956	0.108801	-5.927839	0.0000

The autoregressive parameter φ is equal to zero in both cases; as we see from the t-statistics, the other parameters are significant. Furthermore, the Ljung-Box test and the residual plot guide us towards retaining the chosen model due to its good fit to the data.

For comparison, we note that Renshaw and Haberman (2003a), fitted the same ARIMA (1,1,0) process for males and females using the LC model, obtaining parameters estimates of $\varphi = -0{,}532$ and $\lambda = -0{,}3041$ for males and of $\varphi = -0{,}572$ and $\lambda = -0{,}3525$ for females. This was based on data for England and Wales over the period 1950-1998, and results in parameters which are comparable with our above estimates.

3.4. Evaluation of the model

The evaluation of the model aims at verifying that the model identified and estimated in the previous phases is adequate. If it is not, we have to suggest an alternative model. The objective of diagnostic checking is to ascertain whether the model "fits" the historical data well enough.

To verify that the model we have previously identified and estimated fits the historical data well, we perform a number of analyses. We fit different models to the matrix of Italian death rates from 1950 to 1985, thereby using a 35 years in-sample period, to generate out-of-sample forecasts for the next 15 years. After fitting a range of models in-sample, we compute the Root of Mean Square Error (RMSE) for each ARIMA model and we find that the models we have chosen (ARIMA (0,1,0) for males and ARIMA (0,1,1) for females) are the ones with the lowest RMSE. This indicates that these are the models which best approximate the historical data.

4. Projecting lifetables

4.1. Traditional method

Now we can use the ARIMA (0,1,1) and ARIMA (0,1,0) models to generate the forecasts of the index of mortality k_t for the next 25 years based on the period 1950-2000. Appendix C lists these values for both genders.

Figure 7 and 8, instead, plot the past values of k along with the forecasts based on the time series model and the associated confidence intervals, for females and males respectively. It is worth noticing that we have used the Lee-Carter method for calculating the prediction intervals that concentrates just on variability due to kappa. The other sources of variability could be allowed for by using a bootstrap method: see Brouhns, N., Denuit, M., Van Keilegom (2005).

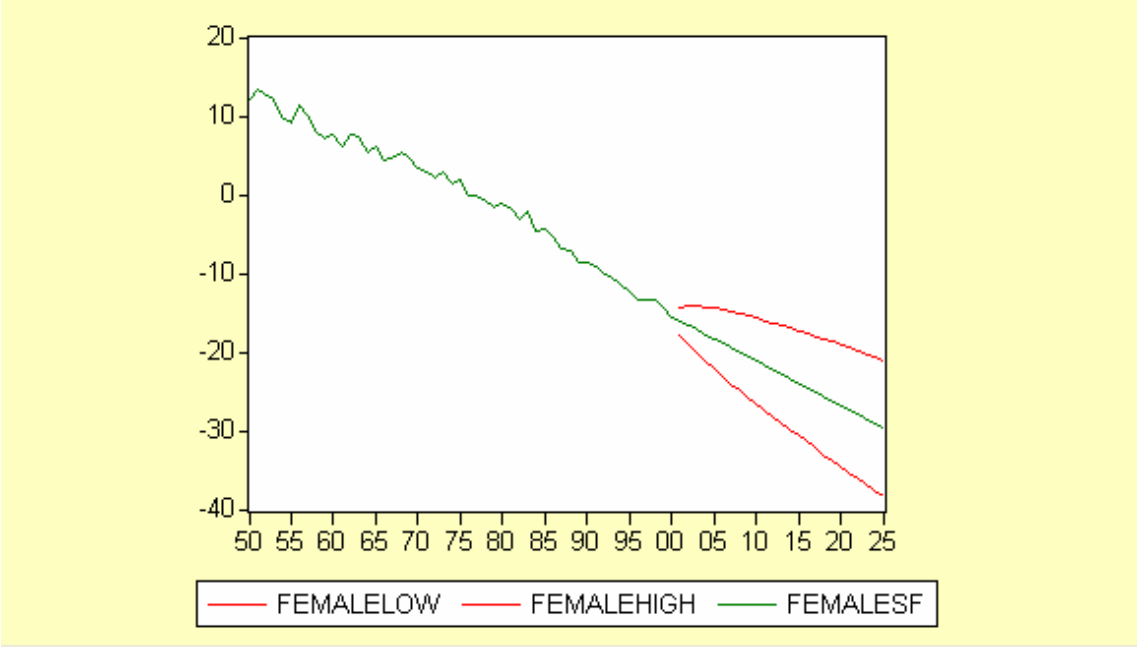

Fig. 7 Forecasts of Female Mortality Index k with confidence interval

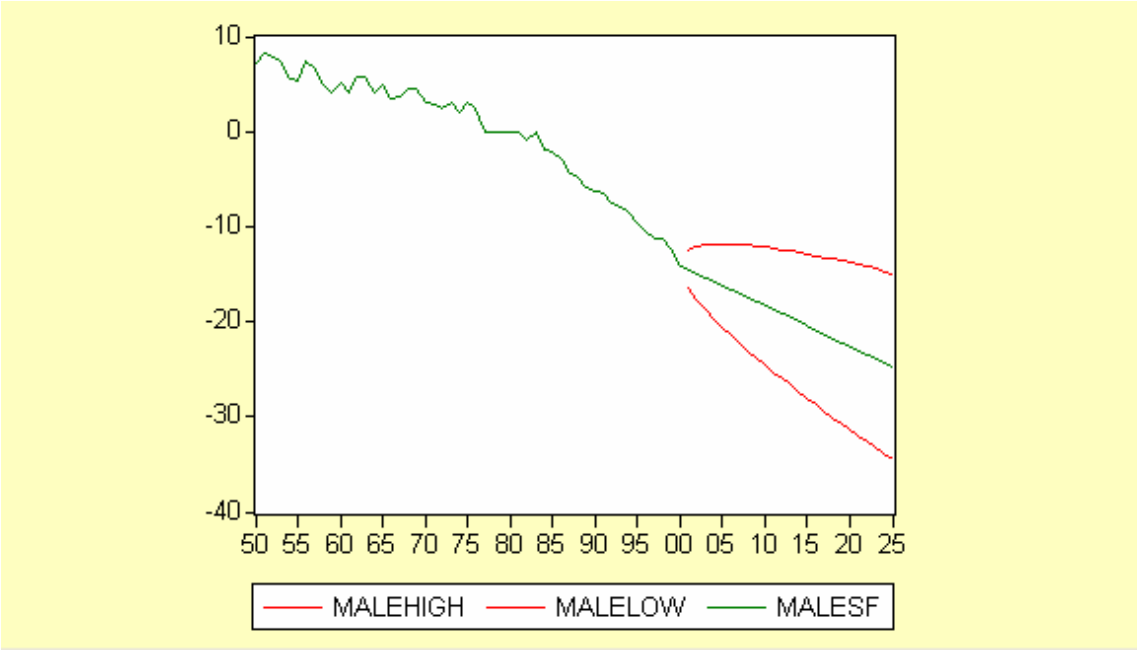

Fig. 8 Forecasts of Male Mortality Index k with confidence interval

Once we have forecasted the index of mortality, we can generate associated life table values at five-year intervals. First we insert the projected k_{2000+s}, $s = 1,2,.....,25,$ into the formulas

$$(3) \quad \overset{\circ}{m}_{x,2000+s} = \hat{m}_{x,2000} \exp\left\{ \overset{\wedge}{\beta}_x \left(\overset{\circ}{k}_{2000+s} - \hat{k}_{2000} \right) \right\}$$

to compute forecast mortality rates by alignment to the latest available empirical mortality rates $\hat{m}_{x,2000}$.

Figure 9 shows the shapes of the mortality rates that we forecast for the females generations born in years 2001 and 2025. It is worth noticing that the mortality rates for age groups 1 - 4 and 5 - 9 become virtually identical by 2001 and 2025.

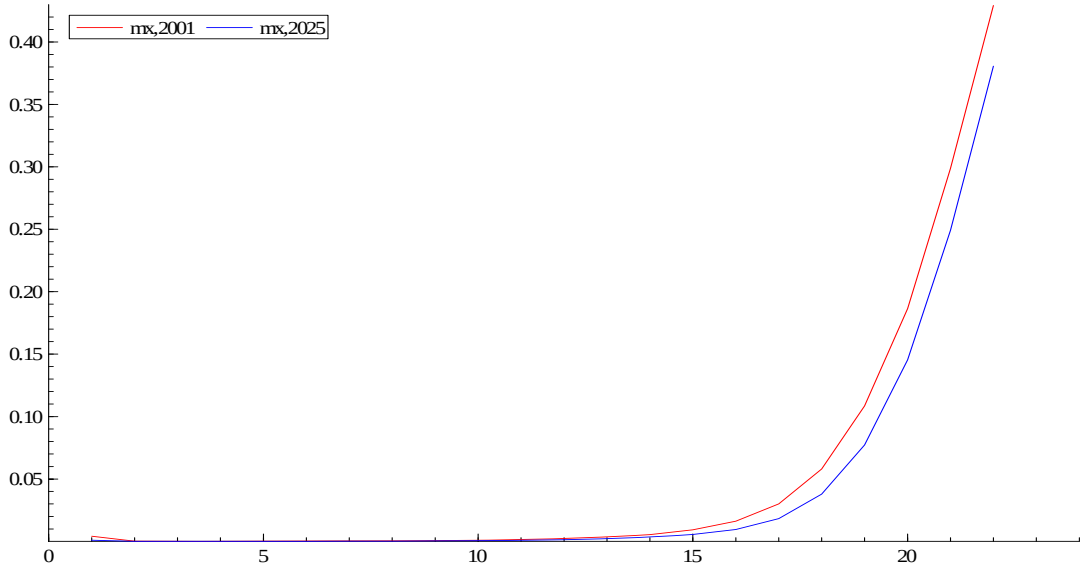

Fig. 9 Forecasted mortality rates for the female generations born in years 2001 and 2025

From these projected mortality rates, we can build projected life tables and compute life expectancy at birth: see Keyfitz N. (1977).

Thus, we convert the life table death rates, m_x, into probabilities of death, q_x. Let f_x be the average number of years lived within the age interval $[x, x+1)$ for people dying at that age. As in Renshaw and Haberman (2003a), we assume that $f_x = \frac{1}{2}$ for all age group except age 0 (for $x = 0$ we fix $f_x = 0{,}15$ for males and $f_x = 0{,}16$ for females). We then compute q_x from m_x and f_x according to the formula,

$$(4) \quad q_x \cong \frac{w_x m_x}{1 + f'_x w_x m_x}, \qquad x = x_0, x_1, ..., x_{k-2},$$

for $x = 0{,}1-4{,}5-9, ..., 100-104$, $w_{xi} = x_{i+1} - x_i$, $k = 22$ and $f'_x = 1 - f_x$.

To complete the life table calculation, let p_x be the probability of surviving from age x to $x+1$.

Therefore,

$$(5) \quad p_x = 1 - q_x,$$

for all five-year age groups up the age of 104.

From q_x calculated by (4) and an arbitrary l_0 (in our case we make it equal to 100000) the life table is constructed by working down the column of l's and d's, applying the recurrence equations

$$(6) \quad l_{x+wx} = l_x(1-q_x), \qquad x = x_0, x_1, ..., x_{k-2},$$

$$(7) \quad d_x = l_x - l_{x+w_x} = l_x q_x, \qquad x = x_0, x_1, ..., x_{k-2},$$

where l_x indicates the number of survivors and d_x is the distribution of deaths by age in the life table population.

The person-years lived by the life-table population in the age interval $[x, x+1)$ are

$$(8) \quad L_x = w_x(l_x - f'_x d_x), \qquad x = x_0, x_1, ..., x_{k-2}.$$

The person-years remaining for individuals of age x equal

$$(9) \quad T_{x_i} = \sum_{x=x_i}^{x_{k-1}} L_x$$

imply that life expectancy is given by

$$(10) \quad e_{x_i} = T_{x_i} \Big/ l_{x_i}.$$

Appendix D lists forecasts of life expectancy at birth obtained using the Lee-Carter model and also shows forecasts obtained with the alternative method which will be discussed later.

4.2. The alternative approach to forecast life expectancy

The method seen above allowed us to compute life expectancies from forecasted mortality rates. In that approach we found an appropriate ARIMA time series model for the mortality index k_t and then we used that mortality model to generate forecasts of the mortality rates. From the forecasts of mortality rates it was straightforward to calculate life tables and life expectancy at birth.

Now we introduce an alternative approach by modelling and forecasting life expectancy directly; we perform a time series analysis of the annual life expectancies at age x to generate forecasts directly. In particular, we consider annual life expectancies at birth for the Italian population, supplied by the Human Mortality Database and divided by gender, from 1950 to 2000. As before, we use the standard Box and Jenkins methodology to generate an appropriate ARIMA (p,d,q) model for our time series, represented in this case by the males and females life expectancies at birth.

In this case the life expectancies are intrinsically viewed as a stochastic process and are estimated and forecasted within an ARIMA time series model. We find that an appropriate model for males and females is ARIMA (1,1,1):

$$(11)\quad \nabla e_t = \varphi_1 \nabla e_{t-1} + \lambda + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

where ∇ is the differencing operator and $\{\varepsilon_t\}$ denotes white noise.

The fitted ARIMA (1,1,1) model generates sex-specific life expectancy forecasts directly. Appendix D shows forecasts of life expectancy at birth, comparing the results obtained using the Lee-Carter methodology and the alternative approach. Both approaches are illustrated in Figure 10, which shows life expectancy at birth from 1950 to 2000 and forecasts from 2001 to 2025. As shown the forecasts based on the LC model are dominated by the forecasts obtained under the direct time series approach (for both genders), thus bearing out the conservative nature of the life expectancy under the LC approach. We want to stress that our results are consistent with the findings of Lee and Carter (1992) and Renshaw and Haberman (2003a), in their forecasting of life expectancies in the USA and in England and Wales, respectively.

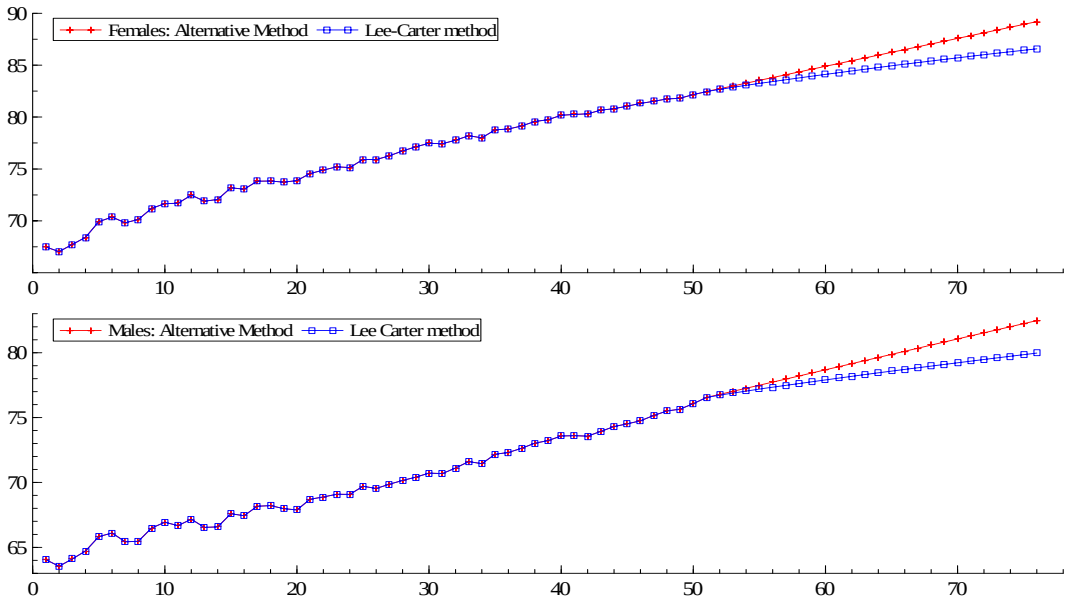

Fig. 10 Life Expectancy at birth and Forecasts

5. Conclusions

We have presented an application of the model underpinning the Lee-Carter methodology for forecasting vital rates. In particular we have focused on forecasting life expectancies on a period basis and we have compared the life expectancies forecasted under the LC model, with the time-series-based forecast. The results are interesting; the *a priori* assumption would be that they would be different, and this is what we find in our analysis. The modelling of the underlying mortality rates is a superior method in theoretical terms yet employing the alternative allow us to examine the effect of a different approach. Moreover, the difference in results is evident for both genders.

References

Box, G.E.P., Jenkins, G.M. Time Series Analysis for Forecasting and Control. San Francisco: Holden-Day, 1976.

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

Brouhns, N., Denuit, M., Van Keilegom I. (2005). Bootstrapping the Poisson Log-Bilinear model for mortality forecasting. *Scandinavian Actuarial Journal, 212-224.*

Brouhns, N., Denuit, M., Vermunt, J.K. (2002b). Measuring the Longevity risk in mortality projections. *Bulletin of the Swiss Association of Actuaries*, 105-130.

Elandt-Johnson, R.C. and Johnson, N.L. (1980). Survival Models and Data Analysis. New York: Wiley.

Francis X. Diebold. Elements of Forecasting, 3rd edition, Cincinnati: South-Western College Publishing, 2004.

Hamilton, J.D. (1994). Time Series Analysis. Princeton University Press, Princeton, N.J.

Human Mortality Database. University of California, Berkeley (USA), available at www.mortality.org

Keyfitz, N. (1977). Introduction to the Mathematics of Population with Revisions. Addison-Wesley Publishing Co., Reading, Massachusetts.

Lee, R.D., Carter, L.R. (1992). Modelling and Forecasting U.S. Mortality. *Journal of the American Statistical Association 87*, 659-671.

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

Lee, R., Miller, T. (2000). Assessing the Performance of the Lee-Carter Approach to Modeling and Forecasting Mortality, available at www.demog.berkeley.edu

Lee, R., Miller, T. (2001). Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. *Demography 38, 4,* 537-549.

Makridakis, S., Wheelwright, S.C., Hyndman, R.J. Forecasting methods and applications, third edition. John Wiley and Sons, 1998.

Renshaw, A.., Haberman, S. (2003a). Lee-Carter mortality forecasting: a parallel generalised linear modelling approach for England and Wales mortality projections. *Applied Statistics 52*, *1*, 119-137.

Wilmoth, J.R. with the assistance of Andreev, K., Bubenheim, M., Jdanov, D., Philipov, D., Shkolnikov, V. and Vachon, P. (2000). Methods Protocol for the Human Mortality Database.

APPENDICES

Appendix A: raw estimates of α_x, β_x and k_t

Estimation αx		
Age Group	**Females**	**Males**
0	-4,033699707	-3,835790179
1-4	-7,213929985	-7,10874839
5-9	-8,160779498	-7,8680444
10-14	-8,26407312	-7,813319463
15-19	-7,864148005	-6,945887141
20-24	-7,651584535	-6,717847306
25-29	-7,452283749	-6,677807228
30-34	-7,176244489	-6,52591061
35-39	-6,82668437	-6,267172662
40-44	-6,42665121	-5,857718365
45-49	-5,97721047	-5,367224209
50-54	-5,5362239	-4,85898577
55-59	-5,099981417	-4,371698852
60-64	-4,618943106	-3,908334419
65-69	-4,091446245	-3,46120974
70-74	-3,513642443	-3,004627826
75-79	-2,91609241	-2,533438599
80-84	-2,340328469	-2,054049223
85-89	-1,816543952	-1,608759955
90-94	-1,360558507	-1,204260676
95-99	-0,98275526	-0,858826013
100-104	-0,683975682	-0,571001792

Estimation βx		
Age Group	**Females**	**Males**
0	0,102499919	0,141392134
1-4	0,115756234	0,154637924
5-9	0,076369591	0,1048845
10-14	0,06054872	0,077513092
15-19	0,046862446	0,036496079
20-24	0,052411099	0,027122682
25-29	0,052634309	0,028254762
30-34	0,049035161	0,029940744
35-39	0,046391497	0,03824621
40-44	0,041574381	0,043840993
45-49	0,0371411	0,043890003
50-54	0,035471203	0,040208161
55-59	0,034728713	0,034730071
60-64	0,036185567	0,029289642
65-69	0,038141047	0,024775806
70-74	0,03928069	0,024092927
75-79	0,03702842	0,024840021
80-84	0,031747846	0,024819726
85-89	0,025296883	0,023897241
90-94	0,018481342	0,021030037
95-99	0,013483991	0,018642631
100-104	0,00892984	0,007454613

Raw kt

Year	Females	Males
1950	14,76794409	8,583743518
1951	14,89155126	9,913426279
1952	13,408219	8,394651321
1953	12,52239962	7,846443414
1954	10,62992411	6,157210056
1955	9,842540872	6,678494859
1956	10,87206043	7,63722919
1957	9,952250993	6,926523736
1958	8,333794839	5,760670293
1959	7,666052493	4,684290325
1960	7,514274804	5,502144527
1961	6,452023934	4,67497325
1962	7,350708893	5,924349376
1963	7,437428484	6,008580608
1964	5,035484367	4,149811148
1965	5,574608009	4,440954214
1966	3,765069333	3,160189717
1967	4,097046229	2,997962203
1968	4,145444986	3,741762929
1969	3,538285116	3,239875422
1970	2,803734674	2,594174948
1971	2,15831635	2,1059274
1972	1,746190265	1,687278316
1973	1,709463933	2,05402866
1974	0,043476874	0,49638264
1975	0,085074077	1,008839268
1976	-0,48958581	0,302168793
1977	-1,406965414	-0,064006487
1978	-2,454922639	-0,523079984
1979	-2,794406103	-0,942828408
1980	-1,942370504	-0,584121479
1981	-3,91343144	-1,793326397
1982	-4,657875266	-2,739795273
1983	-4,248820047	-2,475348897
1984	-5,901708476	-3,960446015
1985	-5,964310903	-4,129592882
1986	-6,714897273	-4,911413514
1987	-7,42132546	-5,417138884
1988	-7,575054175	-5,531328024
1989	-8,556732355	-6,055195481
1990	-8,365313614	-5,705874419
1991	-8,247115251	-5,245437366
1992	-8,664645546	-5,975670242
1993	-8,610286143	-6,494699583
1994	-9,306065233	-7,242041819
1995	-9,841560985	-7,062980989
1996	-10,14304994	-8,138229592
1997	-11,10687368	-9,046047702
1998	-11,81185036	-9,821817576
1999	-12,90587746	-11,02581041
2000	-13,29832396	-11,78585499

Appendix B: k_t re-estimated

Reestimated kt		
Year	**Females**	**Males**
1950	12,239065	7,127597
1951	13,261274	8,301183
1952	12,594144	7,754879
1953	12,055052	7,318683
1954	9,651698	5,478459
1955	9,163218	5,265724
1956	11,254295	7,463961
1957	9,853836	6,647819
1958	7,950729	4,876913
1959	7,034265	4,145711
1960	7,632881	5,196660
1961	6,159950	4,190676
1962	7,683678	5,772402
1963	7,480289	5,813922
1964	5,299921	4,079973
1965	6,234911	4,901331
1966	4,399276	3,478168
1967	4,684472	3,650243
1968	5,330330	4,475201
1969	4,693498	4,603788
1970	3,442278	3,068314
1971	2,802146	2,750540
1972	2,117816	2,410300
1973	2,955119	3,103914
1974	1,363817	2,088754
1975	1,885872	2,989996
1976	0,000000	2,486175
1977	0,000000	0,000000
1978	-0,767902	0,000000
1979	-1,524324	0,000000
1980	-1,043566	0,000000
1981	-1,971115	0,000000
1982	-3,159363	-0,890456
1983	-2,010931	0,000000
1984	-4,507087	-1,983047
1985	-4,416243	-2,085719
1986	-5,308517	-2,968874
1987	-6,862385	-4,323954
1988	-7,192844	-4,789860
1989	-8,570573	-5,953422
1990	-8,606526	-6,217906
1991	-9,022171	-6,417020
1992	-10,235814	-7,592074
1993	-10,703506	-7,944450
1994	-11,313666	-8,573662
1995	-12,367972	-9,549911
1996	-13,367315	-10,741046
1997	-13,494783	-11,336797
1998	-13,349057	-11,294018
1999	-14,263166	-12,552274
2000	-15,503808	-14,116502

Appendix C: Forecasted k_t

Forecasted kt		
Years	**Kt_Females**	**Kt_Males**
2001	-16,07029342	-14,54138354
2002	-16,63677847	-14,96626551
2003	-17,20326352	-15,39114747
2004	-17,76974857	-15,81602944
2005	-18,33623361	-16,2409114
2006	-18,90271866	-16,66579337
2007	-19,46920371	-17,09067533
2008	-20,03568876	-17,5155573
2009	-20,60217381	-17,94043926
2010	-21,16865886	-18,36532123
2011	-21,73514391	-18,79020319
2012	-22,30162895	-19,21508515
2013	-22,868114	-19,63996712
2014	-23,43459905	-20,06484908
2015	-24,0010841	-20,48973105
2016	-24,56756915	-20,91461301
2017	-25,1340542	-21,33949498
2018	-25,70053924	-21,76437694
2019	-26,26702429	-22,18925891
2020	-26,83350934	-22,61414087
2021	-27,39999439	-23,03902284
2022	-27,96647944	-23,4639048
2023	-28,53296449	-23,88878677
2024	-29,09944953	-24,31366873
2025	-29,66593458	-24,7385507

Appendix D: Comparison between the two different approaches

	Alternative Method	Lee-Carter method
Years	**Female(1,1,1)**	**Female(0,1,1)**
2001	82,72917824	82,67351256
2002	82,99835649	82,86201097
2003	83,26753473	83,04787938
2004	83,53671298	83,23118379
2005	83,80589122	83,41198693
2006	84,07506946	83,59034841
2007	84,34424771	83,76632497
2008	84,61342595	83,93997054
2009	84,8826042	84,11133649
2010	85,15178244	84,28047172
2011	85,42096068	84,44742281
2012	85,69013893	84,61223418
2013	85,95931717	84,77494818
2014	86,22849542	84,93560524
2015	86,49767366	85,09424397
2016	86,7668519	85,25090125
2017	87,03603015	85,40561236
2018	87,30520839	85,55841107
2019	87,57438664	85,70932971
2020	87,84356488	85,85839928
2021	88,11274312	86,00564951
2022	88,38192137	86,15110895
2023	88,65109961	86,29480504
2024	88,92027786	86,43676419
2025	89,1894561	86,57701179
	Alternative Method	**Lee Carter method**
Years	**Males (1,1,1)**	**Males (0,1,0)**
2001	76,78736631	76,74722129
2002	77,02473263	76,8981853
2003	77,26209894	77,04728176
2004	77,49946526	77,19458884
2005	77,73683157	77,34018047
2006	77,97419788	77,48412656
2007	78,2115642	77,62649319
2008	78,44893051	77,76734287
2009	78,68629683	77,90673464
2010	78,92366314	78,04472435
2011	79,16102945	78,18136473
2012	79,39839577	78,31670564
2013	79,63576208	78,45079417
2014	79,8731284	78,58367483
2015	80,11049471	78,71538963
2016	80,34786102	78,84597827
2017	80,58522734	78,97547824
2018	80,82259365	79,10392494
2019	81,05995997	79,23135178
2020	81,29732628	79,35779032
2021	81,53469259	79,48327033
2022	81,77205891	79,60781993
2023	82,00942522	79,73146564
2024	82,24679154	79,85423249
2025	82,48415785	79,9761441

Cass Business School
City of London

FACULTY OF ACTUARIAL SCIENCE AND STATISTICS

Actuarial Research Papers since 2001

Report Number	Date	Publication Title	Author
135.	February 2001.	On the Forecasting of Mortality Reduction Factors. ISBN 1 901615 56 1	Steven Haberman Arthur E. Renshaw
136.	February 2001.	Multiple State Models, Simulation and Insurer Insolvency. ISBN 1 901615 57 X	Steve Haberman Zoltan Butt Ben Rickayzen
137.	September 2001	A Cash-Flow Approach to Pension Funding. ISBN 1 901615 58 8	M. Zaki Khorasanee
138.	November 2001	Addendum to "Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving". ISBN 1 901615 59 6	Peter D. England
139.	November 2001	A Bayesian Generalised Linear Model for the Bornhuetter-Ferguson Method of Claims Reserving. ISBN 1 901615 62 6	Richard J. Verrall
140.	January 2002	Lee-Carter Mortality Forecasting, a Parallel GLM Approach, England and Wales Mortality Projections. ISBN 1 901615 63 4	Arthur E.Renshaw Steven Haberman.
141.	January 2002	Valuation of Guaranteed Annuity Conversion Options. ISBN 1 901615 64 2	Laura Ballotta Steven Haberman
142.	April 2002	Application of Frailty-Based Mortality Models to Insurance Data. ISBN 1 901615 65 0	Zoltan Butt Steven Haberman
143.	Available 2003	Optimal Premium Pricing in Motor Insurance: A Discrete Approximation.	Russell J. Gerrard Celia Glass
144.	December 2002	The Neighbourhood Health Economy. A Systematic Approach to the Examination of Health and Social Risks at Neighbourhood Level. ISBN 1 901615 66 9	Les Mayhew
145.	January 2003	The Fair Valuation Problem of Guaranteed Annuity Options : The Stochastic Mortality Environment Case. ISBN 1 901615 67 7	Laura Ballotta Steven Haberman
146.	February 2003	Modelling and Valuation of Guarantees in With-Profit and Unitised With-Profit Life Insurance Contracts. ISBN 1 901615 68 5	Steven Haberman Laura Ballotta Nan Want
147.	March 2003.	Optimal Retention Levels, Given the Joint Survival of Cedent and Reinsurer. ISBN 1 901615 69 3	Z. G. Ignatov Z.G., V.Kaishev R.S. Krachunov
148.	March 2003.	Efficient Asset Valuation Methods for Pension Plans. ISBN 1 901615707	M. Iqbal Owadally
149.	March 2003	Pension Funding and the Actuarial Assumption Concerning Investment Returns. ISBN 1 901615 71 5	M. Iqbal Owadally

150.	Available August 2004	Finite time Ruin Probabilities for Continuous Claims Severities	D. Dimitrova Z. Ignatov V. Kaishev
151.	August 2004	Application of Stochastic Methods in the Valuation of Social Security Pension Schemes. ISBN 1 901615 72 3	Subramaniam Iyer
152.	October 2003.	Guarantees in with-profit and Unitized with profit Life Insurance Contracts; Fair Valuation Problem in Presence of the Default Option[1]. ISBN 1-901615-73-1	Laura Ballotta Steven Haberman Nan Wang
153.	December 2003	Lee-Carter Mortality Forecasting Incorporating Bivariate Time Series. ISBN 1-901615-75-8	Arthur E. Renshaw Steven Haberman
154.	March 2004.	Operational Risk with Bayesian Networks Modelling. ISBN 1-901615-76-6	Robert G. Cowell Yuen Y, Khuen Richard J. Verrall
155.	March 2004.	The Income Drawdown Option: Quadratic Loss. ISBN 1 901615 7 4	Russell Gerrard Steven Haberman Bjorn Hojgarrd Elena Vigna
156.	April 2004	An International Comparison of Long-Term Care Arrangements. An Investigation into the Equity, Efficiency and sustainability of the Long-Term Care Systems in Germany, Japan, Sweden, the United Kingdom and the United States. ISBN 1 901615 78 2	Martin Karlsson Les Mayhew Robert Plumb Ben D. Rickayzen
157.	June 2004	Alternative Framework for the Fair Valuation of Participating Life Insurance Contracts. ISBN 1 901615-79-0	Laura Ballotta
158.	July 2004.	An Asset Allocation Strategy for a Risk Reserve considering both Risk and Profit. ISBN 1 901615-80-4	Nan Wang
159.	December 2004	Upper and Lower Bounds of Present Value Distributions of Life Insurance Contracts with Disability Related Benefits. ISBN 1 901615-83-9	Jaap Spreeuw
160.	January 2005	Mortality Reduction Factors Incorporating Cohort Effects. ISBN 1 90161584 7	Arthur E. Renshaw Steven Haberman
161.	February 2005	The Management of De-Cumulation Risks in a Defined Contribution Environment. ISBN 1 901615 85 5.	Russell J. Gerrard Steven Haberman Elena Vigna
162.	May 2005	The IASB Insurance Project for Life Insurance Contracts: Impart on Reserving Methods and Solvency Requirements. ISBN 1-901615 86 3.	Laura Ballotta Giorgia Esposito Steven Haberman
163.	September 2005	Asymptotic and Numerical Analysis of the Optimal Investment Strategy for an Insurer. ISBN 1-901615-88-X	Paul Emms Steven Haberman
164.	October 2005.	Modelling the Joint Distribution of Competing Risks Survival Times using Copula Functions. I SBN 1-901615-89-8	Vladimir Kaishev Dimitrina S, Dimitrova Steven Haberman
165.	November 2005.	Excess of Loss Reinsurance Under Joint Survival Optimality. ISBN1-901615-90-1	Vladimir K. Kaishev Dimitrina S. Dimitrova
166.	November 2005.	Lee-Carter Goes Risk-Neutral. An Application to the Italian Annuity Market. ISBN 1-901615-91-X	Enrico Biffis Michel Denuit

167.	November 2005	Lee-Carter Mortality Forecasting: Application to the Italian Population. ISBN 1-901615-93-6	Steven Haberman Maria Russolillo

Statistical Research Papers

Report Number	Date	Publication Title	Author
1.	December 1995.	Some Results on the Derivatives of Matrix Functions. ISBN 1 874 770 83 2	P. Sebastiani
2.	March 1996	Coherent Criteria for Optimal Experimental Design. ISBN 1 874 770 86 7	A.P. Dawid P. Sebastiani
3.	March 1996	Maximum Entropy Sampling and Optimal Bayesian Experimental Design. ISBN 1 874 770 87 5	P. Sebastiani H.P. Wynn
4.	May 1996	A Note on D-optimal Designs for a Logistic Regression Model. ISBN 1 874 770 92 1	P. Sebastiani R. Settimi
5.	August 1996	First-order Optimal Designs for Non Linear Models. ISBN 1 874 770 95 6	P. Sebastiani R. Settimi
6.	September 1996	A Business Process Approach to Maintenance: Measurement, Decision and Control. ISBN 1 874 770 96 4	Martin J. Newby
7.	September 1996.	Moments and Generating Functions for the Absorption Distribution and its Negative Binomial Analogue. ISBN 1 874 770 97 2	Martin J. Newby
8.	November 1996.	Mixture Reduction via Predictive Scores. ISBN 1 874 770 98 0	Robert G. Cowell.
9.	March 1997.	Robust Parameter Learning in Bayesian Networks with Missing Data. ISBN 1 901615 00 6	P.Sebastiani M. Ramoni
10.	March 1997.	Guidelines for Corrective Replacement Based on Low Stochastic Structure Assumptions. ISBN 1 901615 01 4.	M.J. Newby F.P.A. Coolen
11.	March 1997	Approximations for the Absorption Distribution and its Negative Binomial Analogue. ISBN 1 901615 02 2	Martin J. Newby
12.	June 1997	The Use of Exogenous Knowledge to Learn Bayesian Networks from Incomplete Databases. ISBN 1 901615 10 3	M. Ramoni P. Sebastiani
13.	June 1997	Learning Bayesian Networks from Incomplete Databases. ISBN 1 901615 11 1	M. Ramoni P.Sebastiani
14.	June 1997	Risk Based Optimal Designs. ISBN 1 901615 13 8	P.Sebastiani H.P. Wynn
15.	June 1997.	Sampling without Replacement in Junction Trees. ISBN 1 901615 14 6	Robert G. Cowell
16.	July 1997	Optimal Overhaul Intervals with Imperfect Inspection and Repair. ISBN 1 901615 15 4	Richard A. Dagg Martin J. Newby
17.	October 1997	Bayesian Experimental Design and Shannon Information. ISBN 1 901615 17 0	P. Sebastiani. H.P. Wynn
18.	November 1997.	A Characterisation of Phase Type Distributions. ISBN 1 901615 18 9	Linda C. Wolstenholme
19.	December 1997	A Comparison of Models for Probability of Detection (POD) Curves. ISBN 1 901615 21 9	Wolstenholme L.C
20.	February 1999.	Parameter Learning from Incomplete Data Using Maximum Entropy I: Principles. ISBN 1 901615 37 5	Robert G. Cowell

21.	November 1999	Parameter Learning from Incomplete Data Using Maximum Entropy II: Application to Bayesian Networks. ISBN 1 901615 40 5	Robert G. Cowell
22.	March 2001	FINEX : Forensic Identification by Network Expert Systems. ISBN 1 901615 60X	Robert G.Cowell
23.	March 2001.	Wren Learning Bayesian Networks from Data, using Conditional Independence Tests is Equivalant to a Scoring Metric ISBN 1 901615 61 8	Robert G Cowell
24.	August 2004	Automatic, Computer Aided Geometric Design of Free-Knot, Regression Splines. ISBN 1-901615-81-2	Vladimir K Kaishev, Dimitrina S.Dimitrova, Steven Haberman Richard J. Verrall
25.	December 2004	Identification and Separation of DNA Mixtures Using Peak Area Information. ISBN 1-901615-82-0	R.G.Cowell S.L.Lauritzen J Mortera,
26.	November 2005.	The Quest for a Doner : Probability Based Methods Offer Help. ISBN 1-90161592-8	P.F.Mostad T. Egeland., R.G. Cowell V. Bosnes Ø. Braaten

Faculty of Actuarial Science and Statistics

Actuarial Research Club

The support of the corporate members

CGNU Assurance
English Matthews Brockman
Government Actuary's Department

is gratefully acknowledged.

ISBN 1-901615-93-6