DRAFT updated 28 March 2004

EVALUATION OF THE VARIANTS OF THE LEE-CARTER METHOD OF FORECASTING MORTALITY: A MULTI-COUNTRY COMPARISON

Heather Booth,[1] Australian National University
Leonie Tickle, Macquarie University
Len Smith, Australian National University

Introduction

As Keyfitz observed in 1981, one might have thought that population forecasters would be obsessed with eagerness to see how well they have done in the past, and that users would demand reports on the error of current forecasts; but 'no such obsession or demand is to be seen' (Keyfitz 1981:580). Population futures have always been a central concern of demographers and those who use their work, from the studies of Malthus in the eighteenth century to those of Pearl and Reed in the twentieth. But, perhaps as a result of the failure of these and other authors to correctly foretell the demographic future, demographers for most of the last century retreated into non-committal scenario-building projections: demographic forecasting based on formal statistical methods has developed only in the last two decades. Actuaries similarly are centrally concerned with the future survival of current lives, but again formal statistical methods of mortality forecasting are a comparatively recent development.

The publication of the Lee-Carter method (Lee and Carter 1992) marked the beginning of a new era of interest in mortality forecasting. Since then several other methods have been developed, but the Lee-Carter method is still regarded as among the best currently available and is now widely used. On the basis of this method, Tuljapurkar, Li and Boe (2000) claimed to have identified a universal pattern of constant rates of mortality decline in the world's most developed countries, with rates of decline higher than those incorporated in official projections, leading to higher forecast levels of life expectancy.

The Lee-Carter method uses matrix decomposition to reduce annual age specific death rates to a time-dependent index of level of mortality, and a set of time-independent parameters which modify the overall level at particular ages. It uses standard time series methods to model and forecast the level index over time. As with time-series-based forecasting in general, the philosophy of the Lee-Carter approach is that the past is the best guide to the future. Thus accurate modelling of past trends is an essential basis for forecasting future levels of mortality, and accurate modelling of the past variability of mortality is an essential basis for estimating the uncertainty of the forecast. In this context, the timescale becomes a central issue, in two ways: first, *how much* of the past provides the best guide to *how much* of the future? Second, how dependent is the answer

[1] Correspondence to heather.booth@anu.edu.au

on the specific time in history when the forecast is made? Even in the context of a statistical forecast, judgment will be required to answer these questions, but the objective is to minimise the role of judgement and maximise the role of formal theory on the one hand, and formal evaluation on the other.

Since the publication of the Lee-Carter method, enhancements have been proposed by Lee and Miller (2001) and Booth, Maindonald and Smith (2002). These address the choice of fitting period, the method for the adjustment of the time parameter and the choice of jump-off rates. The three variants of the Lee-Carter method have not been comprehensively evaluated. This paper presents the results of an evaluation of the Lee-Carter, Lee-Miller and Booth-Maindonald-Smith variants based on data by sex for ten countries. The evaluation involves fitting the different variants to data up to 1985, forecasting for the period since that date, and comparing the forecasts with actual mortality in that period. This evaluation is not explicitly concerned with modelling the variability of forecasts themselves; explicit modelling of variability forms the basis of an alternative approach used by Keyfitz (1981), Keilman (1997) and others.

The Three Variants

The Lee-Carter method

The Lee-Carter method of mortality forecasting combines a demographic model of mortality with time-series methods of forecasting. The method is generally interpreted as making use of the longest available time series of data. The Lee-Carter model of mortality is

$$\ln m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t} \qquad \text{….(1)}$$

where $m_{x,t}$ is the central death rate at age *x* in year *t,* k_t is an index of the level of mortality at time *t,* a_x is a general pattern of mortality by age, b_x is the relative speed of change at each age, and $\varepsilon_{x,t}$ is the residual at age *x* and time *t*. The a_x are calculated as the average of $\ln m_{x,t}$ over time, and the b_x and k_t are estimated by singular value decomposition.

The second-stage estimation involves adjusting k_t by refitting to total observed deaths. This adjustment gives greater weight to ages at which deaths are high, thereby partly counterbalancing the equalising effect of using logratres in the Lee-Carter model. k_t is then extrapolated using the time series model

$$k_t = k_{t-1} + d + e_t \qquad \text{….(2)}$$

where *d* is constant annual change in k_t, and e_t are uncorrelated errors. The combination of the standard errors in *d* and e_t represents the uncertainty associated with a one-year forecast. This is used to produce probabilistic prediction intervals for the forecast values of k_t. Forecast age-specific death rates are obtained using extrapolated k_t and fixed a_x and b_x. In this case, the jump-off rates (i.e. the rates in the last year of the fitting period or jump-off year) are fitted rates.

It should be noted that the Lee-Carter method does not prescribe the linear time series model of a random walk with drift for all situations. However, this model has been found to be the most appropriate in almost all cases; even where a different model was indicated, the more complex model was found to give results which were only marginally different to the random walk with drift (Lee and Miller 2001). Further, Tuljapurkar et al. (2000) found that the decline in mortality was constant, i.e. k_t was linear, for the G7 countries, reinforcing the use of a random walk with drift as an integral part of the Lee-Carter method.

The Lee-Miller variant

The Lee-Miller variant differs from the original Lee-Carter method in three ways:
1. the fitting period is reduced to commence in 1950;
2. the adjustment of k_t involves fitting to e(0) in year t;
3. the jump-off rates are taken to be the actual rates in the jump-off year.

In their evaluation of the Lee-Carter method, Lee and Miller (2001) noted that for US data the Lee-Carter model did not perform particularly well when using the fitting period 1900-1989 to forecast the period 1990-1997. The main source of error was the mismatch between fitted rates for the last year of the fitting period or jump-off year (1989) and actual rates in that year; this jump-off error or bias amounted to 0.6 years in life expectancy for males and females combined (Lee and Miller 2001 p.539). Jump-off bias was avoided by setting a_x equal to the actual rates in the jump-off year, equivalent to constraining the model such that k_t passes through zero in the jump-off year.

It was also noted that the pattern of change in mortality was not fixed over time, as the Lee-Carter model assumes. Based on different age patterns of change (or b_x patterns) for 1900-1950 and 1950-1995, Lee and Miller (2001) adopted 1950 as the first year of the fitting period. This 'simple and satisfactory solution' (Lee and Miller 2001 p.545) to changing age patterns of change had been adopted by Tuljapurkar et al. (2000).

The adjustment of k_t by fitting to e(0) was adopted to avoid the use of population data as required for fitting to D_t (Lee and Miller 2001).

The Booth-Maindonald-Smith variant

The Booth-Maindonald-Smith variant also differs from the Lee-Carter method in three ways:
1. the fitting period is chosen based on statistical goodness-of-fit criteria under the assumption of linear k_t;
2. the adjustment of k_t involves fitting to the age distribution of deaths, $D_{x,t}$;
3. the jump-off rates are taken to be the fitted rates based on this fitting methodology.

Booth, Maindonald and Smith (2002) fitted the Lee-Carter model to Australian data for 1907-1999 and found that the 'universal pattern' (Tuljapurkar et al. 2000) of constant mortality decline as represented by linear k_t did not hold. In addition, problems were encountered in meeting the assumption of constant b_x in the underlying Lee-Carter model. Taking linearity in k_t as a starting point, the Booth-Maindonald-Smith variant seeks to maximise the fit of the overall model by restricting the fitting period, which also results in the assumption of constant b_x being better met. The choice of fitting period is based on the ratio of the mean deviances of the fit of the underlying Lee-Carter model and of the overall linear fit: this ratio is computed for all fitting periods (that is for all years marking the start of the period, which always ends in the same year) and the period for which this ratio is substantially smaller than that for periods starting in previous years is chosen.

The procedure for the adjustment of k_t was modified. Rather than fit to total deaths, D_t, the Booth-Maindonald-Smith variant fits to the age distribution of deaths, $D_{x,t}$, using the Poisson distribution to model the death process and the deviance statistic to measure goodness of fit (Booth, Maindonald and Smith 2002). The jump-off rates are taken to be the fitted rates under this adjustment.

Data

The data for this study are taken from the Human Mortality Database (<http://www.mortality.org> or <http://www.humanmortality.de>) and the Australian Demographic DataBank (Australian Centre for Population Research). Ten countries were selected giving 20 sex-specific populations for analysis. The selected countries are those with reliable data series commencing in 1941 or earlier. It was desirable to use only countries for which the available time series of data commenced somewhat earlier than 1950 in order to maintain the full and consistent comparison of the three variants. Lee and Carter (1992) used US data for the full period available, 1900-1989.[2] Therefore this multi-country analysis uses data for the period commencing in 1900 where possible. Though for some countries the data extend back to the nineteenth century, these were truncated at 1900. The Lee-Carter method could be interpreted as using all available data for the fitting period, but the use of pre-1900 data would both reduce comparability of methods across countries and necessitate a time series model with a non-linear trend which falls outside the scope of both applications to date and the current analysis. The selected countries are shown in Table 1 along with the dates used to define the fitting periods.

The data consist of central death rates and mid-year populations by sex and single years of age to 110 years (except Australia to 100 years). For this analysis, data at older ages (age 90 and above) were grouped in order to avoid problems associated with erratic rates at these ages.

[2] The US data by single years of age in the Human Mortality Database commence only in 1959; thus the US is not included in this study. Lee and Carter used data by five-year age groups.

Table 1: Countries and years defining fitting and forecasting periods

Country	Startyear				Endyear
	LC	LM	BMS [m]	BMS [f]	
Australia	1921	1950	1968	1970	2000
Canada	1921	1950	1974	1976	1996
Denmark	1921	1950	1968	1967	2000
England and Wales	1900	1950	1968	1972	1998
Finland	1941	1950	1971	1971	2000
France	1900	1950	1971	1969	1997
Italy	1906	1950	1968	1968	1999
Norway	1900	1950	1969	1963	2000
Sweden	1900	1950	1976	1969	2001
Switzerland	1900	1950	1962	1962	2001

Note: The fitting period is defined by startyear to 1985; the forecasting period is defined by 1986 to endyear.

Methods and measures

The three variants were fitted to periods ending in 1985 and used to forecast death rates from 1986 to the last year of available data (1996 to 2001, depending on the country). The variants are evaluated by comparing point forecast log death rates with actual log death rates.

Forecasting error is measured in terms of absolute error (| forecast – actual |) and error (forecast – actual). These are averaged over the relevant number of forecast years to produce the mean year absolute error and mean year error, which are indexed by age. They are also averaged over age to produce the mean age absolute error and mean age error, which are indexed by year. Averaging over both year and age produces two single indices of overall error: overall absolute error and overall error. These overall measures are averaged across countries to produce average overall absolute error and average overall error.

In addition to these errors in log death rates, the error in life expectancy is examined. Error in life expectancy denotes the error in life expectancy by forecast year, and mean error in life expectancy denotes the error in life expectancy averaged over years. These measures are averaged across countries to produce average error in life expectancy and average mean error in life expectancy.

Components of error are identified by comparing results based on relevant combinations of fitting period, adjustment method and jump-off rates.

Uncertainty in the forecasts is derived from the standard error of k_t in the time series model (equation 2). Two components of uncertainty can be distinguished: uncertainty due to innovation, in other words e_t, and uncertainty in the drift. From equation (1), the standard error of $\ln m_{x,t}$ is equal to the standard error of k_t multiplied by the constant b_x.

(Note that Lee and Carter (1992, p 670) found the standard errors of a_x and b_x to become less significant over forecast time in comparison to the standard error of k_t and that by 10 years into the forecast of US mortality 98 per cent of the standard error of e(0) was accounted for by uncertainty in k_t.) As $\ln m_{x,t}$ are on a common scale for given age, it is possible to compare standard errors in $\ln m_{x,t}$ between variants, sex and countries for given ages, and hence for derived statistics covering the entire age range. Such comparisons are only slightly affected by the differing levels of mortality among the 20 populations and different fitting periods of the three variants.

Each $\ln m_{x,t}$ is a stochastic process determined by the stochastic process k_t. Hence, ignoring error terms, $\varepsilon_{x,t}$, the variations in $\ln m_{x,t}$ are perfectly correlated. This means that the prediction interval for life expectancy and other life table functions can be derived directly from the prediction interval for k_t without having to worry about the cancellation of errors. The approach adopted to comparing uncertainty in life expectancy is to take their 95 per cent prediction intervals. These are asymmetric due to the log transform and the transformation involved in the life table. Again, these prediction intervals are roughly comparable among populations.

In what follows, the three variants are referred to as LC, LM and BMS. The corresponding three fitting periods are referred to by "long", "1950" and "short", reflecting the variable length in the LC and BMS variants and fixed length in LM. The three adjustment methods are referred to by "D_t", "e(0)" and "$D_{x,t}$".

Findings

Comparison of variants

Findings based on overall absolute errors for the 20 populations considered (Table 2) show that point estimates from LM and BMS are superior to those from LC. Relative to LC (Table 3), most overall absolute errors are in the range 30 to 70 per cent. Of the three variants, BMS has the lowest error for 15 of the 20 populations, as well as the lowest average error for both females and males.

An additional finding is that LC consistently underestimates mortality, especially for females, as indicated by the negative average overall error (Table 4). LM and BMS do not show a marked tendency to over- or under-estimate, and have overall errors closer to zero.

Table 2: Overall absolute error by sex, variant and country

Country	Female			Male		
	LC	LM	BMS	LC	LM	BMS
Australia	0.306	0.149	0.120	0.485	0.178	0.136
Canada	0.242	0.094	0.107	0.296	0.097	0.105
Denmark	0.307	0.238	0.215	0.184	0.217	0.190
England and Wales	0.272	0.114	0.095	0.384	0.132	0.107
Finland	0.667	0.276	0.265	0.559	0.207	0.193
France	0.360	0.100	0.093	0.361	0.123	0.118
Italy	0.355	0.152	0.151	0.258	0.177	0.189
Norway	0.733	0.190	0.180	0.217	0.201	0.178
Sweden	0.708	0.189	0.192	0.254	0.212	0.177
Switzerland	0.529	0.208	0.186	0.266	0.191	0.172
Average	**0.448**	**0.171**	**0.160**	**0.326**	**0.174**	**0.156**

Note: Overall absolute error is the mean over age and year of the absolute error in log death rates.

Table 3: Overall absolute error relative to LC by sex, variant and country

Country	Female			Male		
	LC	LM	BMS	LC	LM	BMS
Australia	1.00	0.49	0.39	1.00	0.37	0.28
Canada	1.00	0.39	0.44	1.00	0.33	0.35
Denmark	1.00	0.77	0.70	1.00	1.18	1.04
England and Wales	1.00	0.42	0.35	1.00	0.34	0.28
Finland	1.00	0.41	0.40	1.00	0.37	0.34
France	1.00	0.28	0.26	1.00	0.34	0.33
Italy	1.00	0.43	0.42	1.00	0.69	0.73
Norway	1.00	0.26	0.25	1.00	0.93	0.82
Sweden	1.00	0.27	0.27	1.00	0.83	0.70
Switzerland	1.00	0.39	0.35	1.00	0.72	0.65
Average	**1.00**	**0.38**	**0.36**	**1.00**	**0.53**	**0.48**

Note: Overall absolute error is the mean over age and year of the absolute error in log death rates. The country average is the ratio of the average errors.

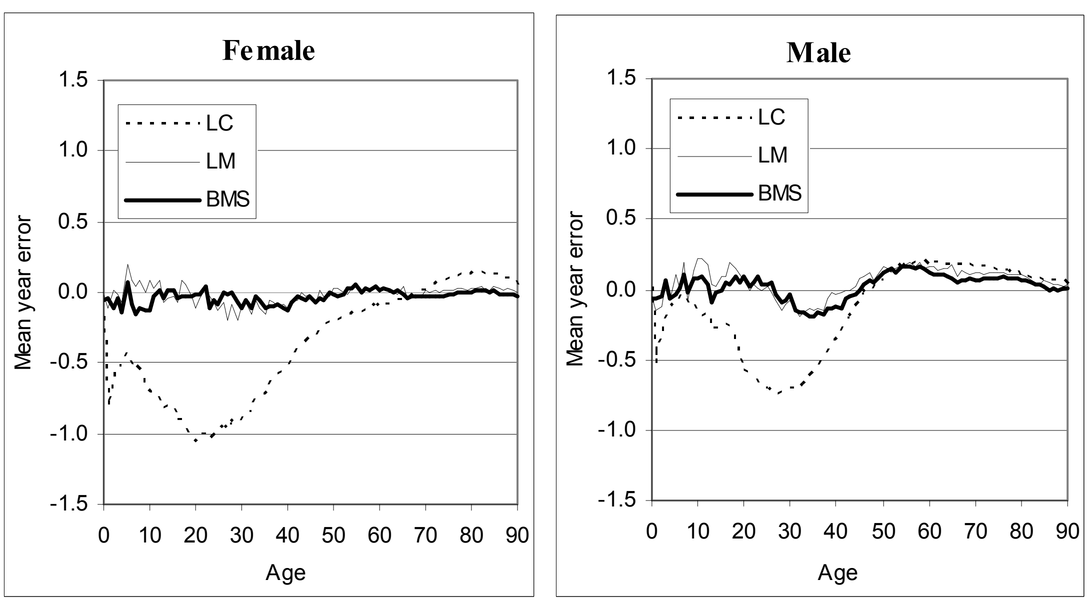
Table 4: Overall error by sex, variant and country

Country	Female			Male		
	LC	LM	BMS	LC	LM	BMS
Australia	-0.170	0.052	0.004	-0.256	0.102	0.042
Canada	-0.221	-0.022	-0.058	-0.146	0.026	-0.055
Denmark	-0.231	0.040	0.038	0.074	0.133	0.111
England and Wales	-0.221	0.002	0.016	-0.247	0.035	0.030
Finland	-0.614	-0.167	-0.168	-0.410	0.004	-0.044
France	-0.272	0.020	0.030	-0.208	0.064	0.058
Italy	-0.265	-0.059	-0.083	-0.078	-0.001	-0.032
Norway	-0.675	-0.001	-0.041	0.121	0.094	0.106
Sweden	-0.644	-0.006	-0.047	-0.094	0.075	-0.006
Switzerland	-0.461	-0.015	-0.020	-0.121	0.027	0.029
Average	**-0.377**	**-0.016**	**-0.033**	**-0.136**	**0.056**	**0.024**

Note: Overall error is the mean over age and year of the error in log death rates.

Across the age range, patterns of error in the log death rates are similar across different countries so an average of all countries is shown in Figure 1. Errors are small and show no consistent age pattern for LM and BMS apart from a tendency to overestimate for males at ages 45+. The LC method produces large negative mean year errors at the younger ages, particularly for females, and small positive mean year errors at the older ages. This is due to the fact that the longer LC fitting period produces estimates of b_x that do not reflect the age pattern of change in the forecasting period. The dominance of the large negative errors at the younger ages accounts for the overall underestimation observed for LC in Table 4.

Figure 1: Mean year error by age, by sex and variant, averaged across countries



Note: Mean year error is the mean over years of error in log death rates.

Although LC underestimates overall mortality when measuring error in log death rates, this does not necessarily translate into an overestimate of life expectancy, due to the variation in the magnitude and sign of errors over age. Table 5 shows that LC and, to a lesser extent, BMS do overestimate female life expectancy. All variants underestimate male life expectancy, due to the overstatement of mortality at the older ages observed in Figure 1. For this measure, in contrast to the overall absolute error measure, LC results do not appear to be substantially inferior to those for LM and BMS: this is because the large negative and positive errors in different parts of the age range partly cancel.

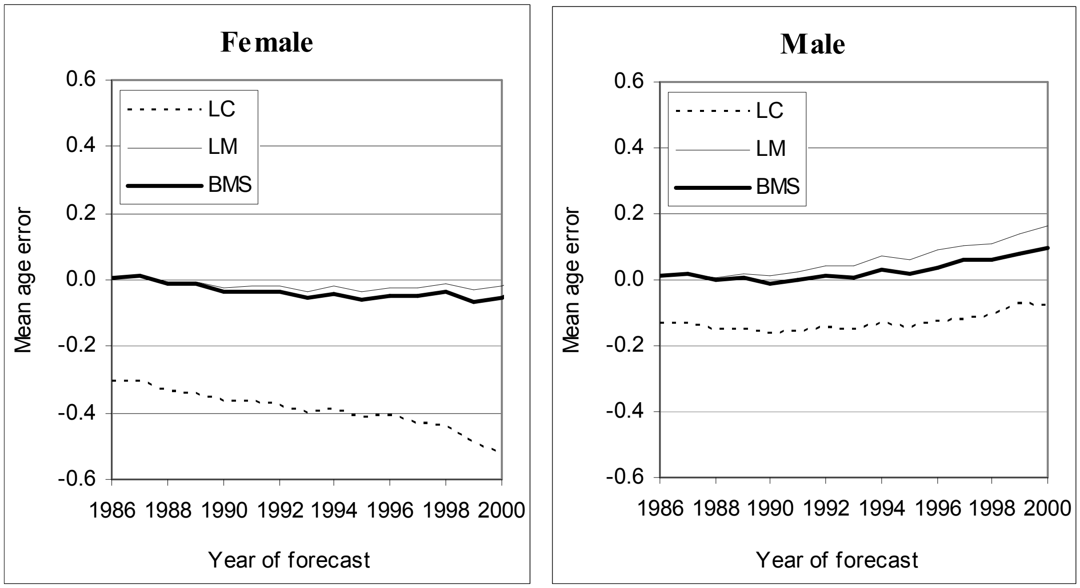
Table 5: Mean error in life expectancy by sex, variant and country

Country	Female			Male		
	LC	LM	BMS	LC	LM	BMS
Australia	-0.70	-0.77	-0.11	-1.06	-1.53	-0.59
Canada	0.33	0.30	0.32	-0.49	-0.53	0.18
Denmark	1.26	0.49	0.40	-0.73	-1.10	-1.18
England and Wales	0.04	-0.46	-0.42	-0.48	-0.97	-0.78
Finland	0.79	0.48	0.82	0.14	-0.61	-0.11
France	-0.32	-0.43	-0.29	-0.40	-0.88	-0.76
Italy	-0.65	-0.52	-0.26	-1.24	-1.06	-0.74
Norway	0.90	0.10	0.44	-1.33	-1.50	-1.13
Sweden	0.61	0.12	0.15	-0.73	-1.33	-0.64
Switzerland	0.76	0.33	0.57	-0.03	-0.46	-0.32
Average	**0.30**	**-0.04**	**0.16**	**-0.64**	**-1.00**	**-0.61**

Note: Mean error in life expectancy is error averaged over years.

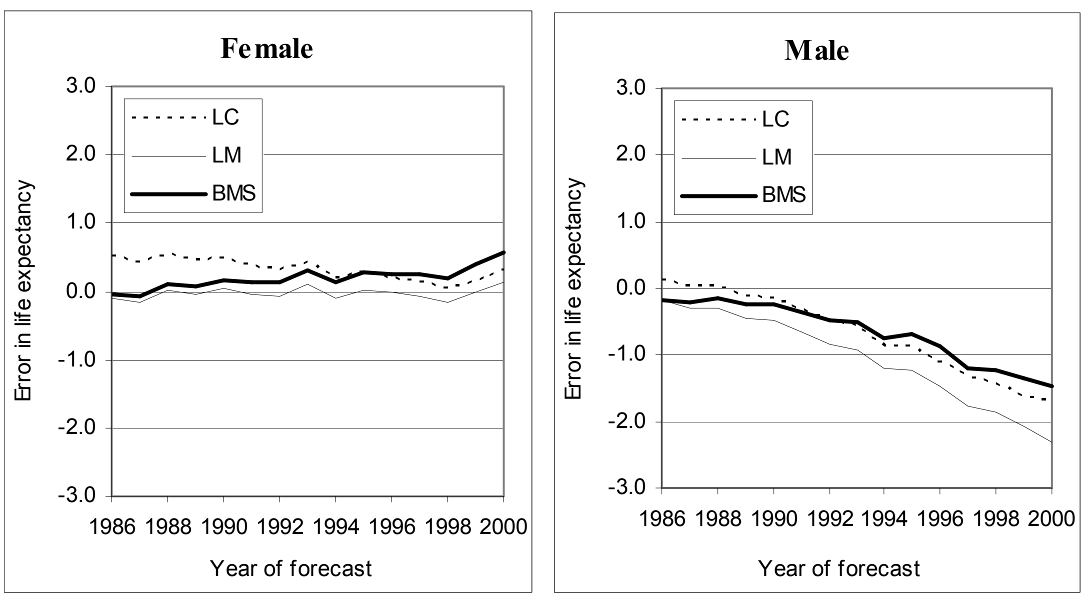
Over time, the pattern of mean age error varies by population. However, some consistent trends can be discerned. LC errors almost always start with a sizeable negative in the first year and remain negative throughout the entire duration of the forecast. BMS errors are typically very close to zero in the first year of the forecast, indicating minimal jump-off error. Both LM and BMS usually remain reasonably close to zero for the duration of the forecast (though this is not the case for all populations). Figure 2 averages across countries, but gives a sense of the general pattern. All three variants give a reasonably good approximation of life expectancy for females over the forecast period (Figure 3). The 1950 startyear for LM gives the most appropriate rate of improvement in life expectancy: LC gives a slightly too gradual improvement (and also has a sizeable positive jump-off bias) and BMS is slightly too steep. The rate of improvement in male life expectancy is underestimated by all three variants: the shorter fitting period for BMS gives the best results except in the very early years.

Figure 2: Mean age error by forecast year, by sex and variant, averaged across countries



Note: Averages from 1996 to 2000 include a decreasing number of populations (see Table 1). Since only two countries have data to 2001, results have been shown to 2000 only.

Figure 3: Error in life expectancy by forecast year, by sex and variant, averaged across countries

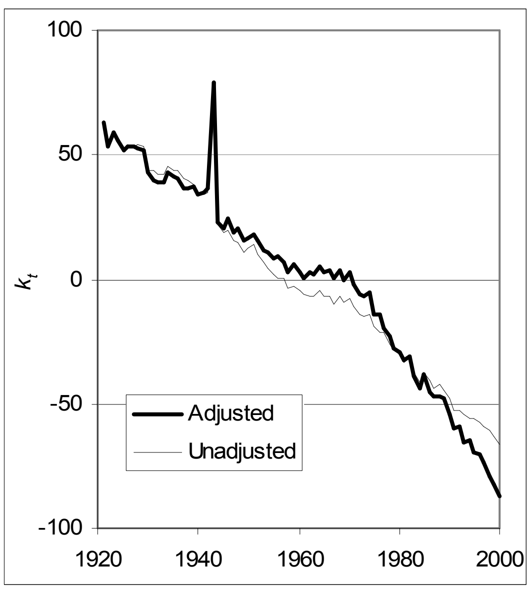


Decomposition of error

Error can be decomposed into three components corresponding to the source of error: fitting period, method of adjustment and jump-off rates. These are inter-related. The choice of method of adjustment is independent of the two other components. Choice of fitting period (or startyear) is independent of other considerations for LC and LM, but for BMS is dependent on the shape of fitted k_t, which in turn is influenced to a small extent by the method of adjustment particularly where deviations from linearity occur (see Figure 4). For LC and BMS, jump-off bias is dependent on both fitting period and method of adjustment (see below).

These error components are examined in order of dependence. The most dependent component, jump-off error is discussed first. After its removal, the effect of fitting period is discussed. Finally the net effect of adjustment method is considered.

Figure 4: k_t and adjusted k_t for Australia, both sexes combined, 1921-2000



Jump-off bias

Jump-off bias derives from the error in the fit of the underlying Lee-Carter model after adjustment of k_t. In terms of log death rates (see equation 1), it is equal to $\varepsilon_{x,t}$ in the jump-off year (i.e. 1985). For the LM variant, this error is zero because actual rates are used as jump-off rates. In the LC and BMS variants, the size of $\varepsilon_{x,t}$ in the jump-off year is determined by the goodness of fit of the Lee-Carter model which is dependent on the fitting period (or startyear) as well as the adjustment of k_t.

Jump-off bias is thus a fixed quantity in terms of log death rates and death rates. Added to actual rates it results in a fixed bias in life expectancy in the jump-off year. However, the size of this bias in life expectancy will not remain constant over the forecast years because of the effects of entropy of the life table and of the combination of (positive and negative values of) b_x and jump-off bias in log death rates.

Figure 5 shows size of jump-off bias in life expectancy for LC and BMS for the 20 populations. The bias is much greater for LC than BMS, and greater for females than males. For females, LC bias is as large as 1.19 years (for Norway), whereas BMS bias is at most 0.15 years. While in most cases, jump-off bias for the LC variant is positive, it is less consistent in direction for BMS.

Figure 5: Jump-off bias by sex and country, LC and BMS variants

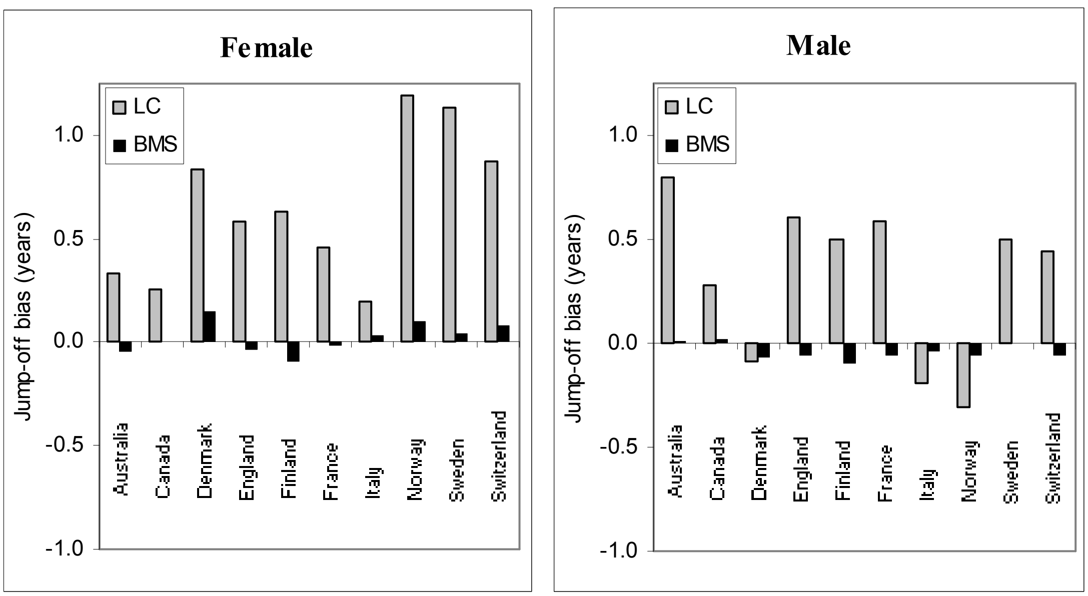


Table 6 shows the effect of jump-off bias on overall absolute error for the LC and BMS variants. The magnitude of the average effect of bias is greatest for LC: the effect for BMS is only 4-7 per cent of that for LC. Jump-off bias also accounts for a greater proportion of average overall error in the LC variant than in the BMS variant: for LC, jump-off bias accounts for 57 per cent of overall error for females and 42 per cent for males compared with 11 and 4 per cent respectively for the BMS variant. Unlike LC, however, in the case of BMS jump-off bias serves to reduce average overall error, indicating that for absolute error, at least, fitted rates provide a better jump-off point than actual rates.

Table 6: Magnitude and relative size of effect of jump-off bias on overall absolute error by sex, variant and country

Country	Female				Male			
	Bias		% of total error		Bias		% of total error	
	LC	BMS	LC	BMS	LC	BMS	LC	BMS
Australia	0.139	-0.021	45.4	-17.1	0.283	-0.004	58.5	-3.2
Canada	0.132	-0.012	54.5	-11.6	0.184	0.002	62.3	1.9
Denmark	0.085	-0.031	27.8	-14.2	-0.009	-0.028	-4.8	-14.6
England and Wales	0.145	-0.018	53.4	-18.9	0.220	-0.008	57.2	-7.4
Finland	0.308	-0.036	46.2	-13.7	0.222	-0.015	39.7	-7.7
France	0.247	0.002	68.7	1.8	0.225	0.012	62.4	10.5
Italy	0.184	0.008	51.7	5.0	0.078	0.023	30.4	12.4
Norway	0.507	-0.023	69.2	-13.0	0.032	0.000	14.8	-0.2
Sweden	0.495	-0.023	69.9	-12.1	0.062	-0.021	24.3	-12.0
Switzerland	0.300	-0.025	56.7	-13.7	0.071	-0.017	26.7	-9.9
Average	**0.254**	**-0.018**	**56.8**	**-11.3**	**0.137**	**-0.006**	**42.0**	**-3.6**

Country comparison shows that there is general consistency within variant in the direction of the effect of jump-off bias, especially for LC and females, but considerable variation in magnitude. For males, the effect of bias for LC ranges from –0.009 for Denmark to 0.283 for Australia, while for BMS the range is –0.028 for Denmark to 0.023 for Italy. Corresponding ranges for females are 0.085 for Denmark to 0.507 for Norway, and –0.036 for Finland to 0.008 for Italy. On average bias is greater for females than males. For LC, there is no pattern between the sexes, but for BMS there is a tendency for the sexes to follow similar patterns across countries.

Fitting period

The effect of different fitting periods is essentially measuring the effect of different trends in k_t. Average overall absolute errors net of jump-off bias are shown in Table 7. The marginal effects due to fitting period are small in comparison with jump-off bias for LC but are commensurate with jump-off bias for BMS. It is seen that average overall absolute error is greatest for the long fitting period; in other words, reducing the fitting period results in reduced error. Whether the 1950 or short fitting period is most advantageous is unclear: for females, error is smallest for the fitting period starting in 1950 while for males error is smallest for the short fitting period.

Table 7: Average overall absolute error net of jump-off error by sex, fitting period and method of adjustment

Fitting period	Female			Male		
	Adjustment			Adjustment		
	D_t	e(0)	$D_{x,t}$	D_t	e(0)	$D_{x,t}$
Long	0.194	0.189	0.190	0.189	0.187	0.186
1950	0.171	0.171	0.171	0.175	0.174	0.174
Short	0.180	0.178	0.179	0.167	0.162	0.162

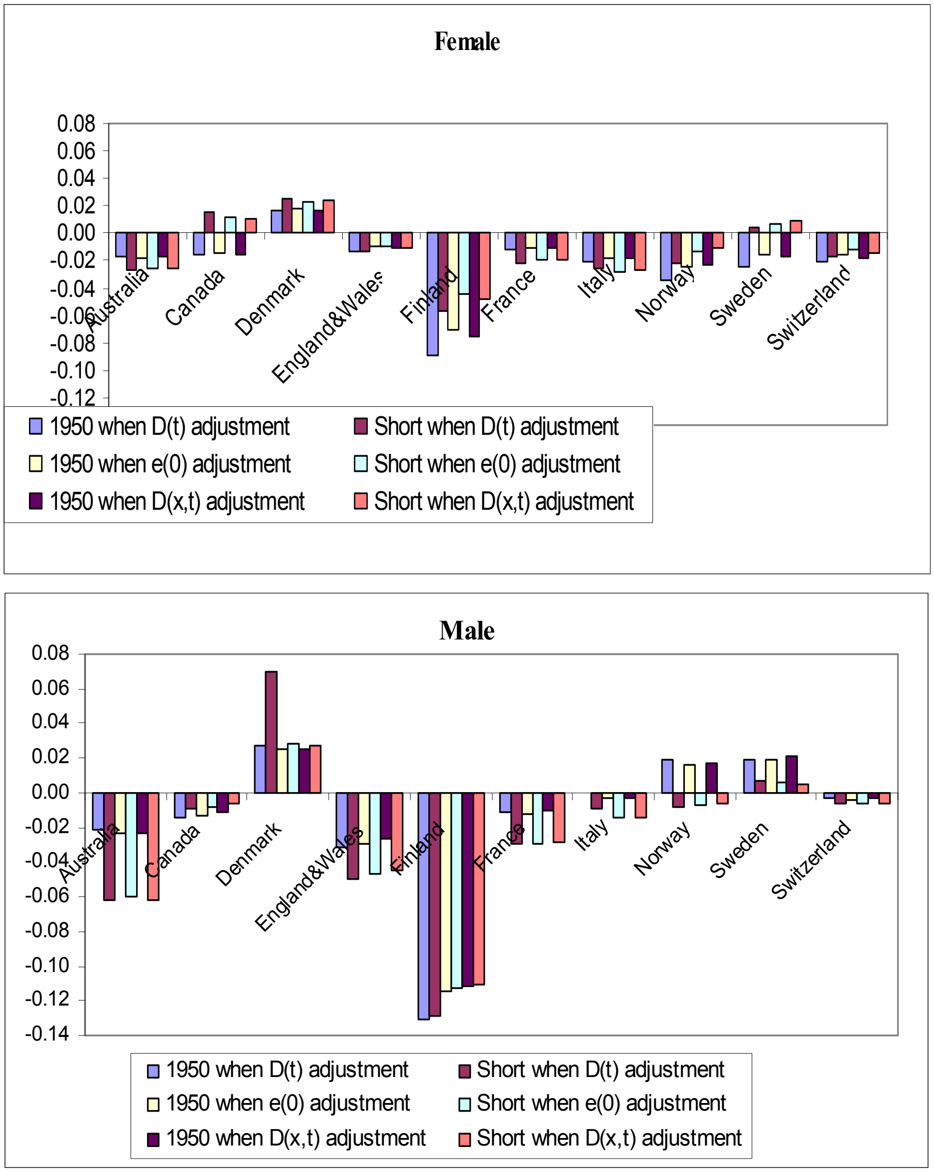
The short period would be expected to produce smaller errors if the short term trend were a better guide than the longer term trend to the future. Thus, use of the short period might be expected to result in smaller errors than use of 1950. The fact that this is not the case for females (in absolute error terms) suggests that the post-1985 trend differs from that in the 10-20 years prior to 1985.

Country comparisons shown in Figure 6 show that the reduction in overall absolute error due to the use of a fitting period starting in 1950 or a short fitting period is fairly consistent across countries. Of the 20 populations, 14 show reductions in error due to reductions in length of fitting period, and in nine of these the short fitting period gives greater reductions in error. Denmark is a notable exception for both sexes, with both increases in error and greater increases for the short fitting period than the 1950 fitting period. It should be noted, however, that there were difficulties in using the method for Denmark with the chosen fitting and forecasting periods due to the erratic nature of k_t in the later part of the fitting period, stemming from a poor fit of the base model. For Swedish males, the reduced period also resulted in greater error; in this case the 1950 fitting period produced the greater effect. In a further three cases, inconsistent results were obtained: for Canada and Sweden females, the 1950 fitting period improved forecast accuracy, while the short fitting period resulted in greater error; and for Norway males the reverse was true.

Adjustment method

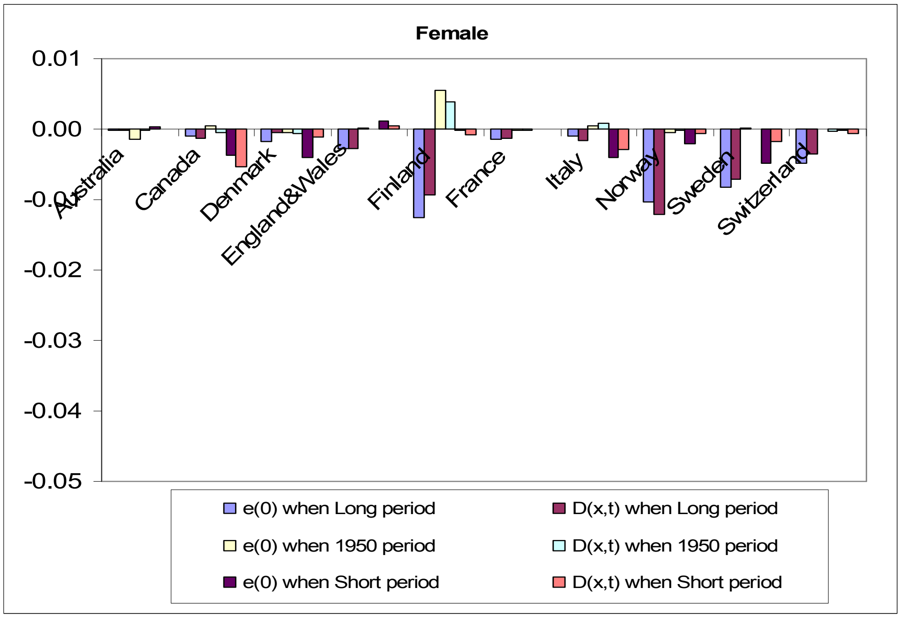
Table 7 also allows comparison by method of adjustment. For the average overall absolute error measure, the effect of adjustment method is small compared with the effect of fitting period and jump-off bias, and in some cases is extremely marginal. For females, adjustment by e(0) produces least error for long and short periods while there is no difference for the fitting period starting in 1950. For males, adjustment by $D_{x,t}$ is marginally superior for the long fitting period. It can be concluded that adjustment method makes virtually no difference to the overall absolute error. This raises the question as to whether any adjustment improves the forecast over using unadjusted k_t.

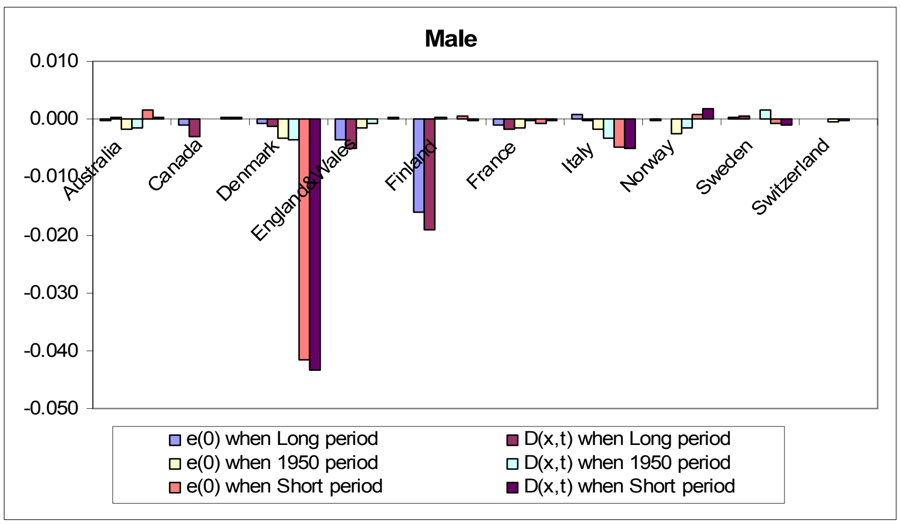
Figure 6: Effect of fitting period, marginal to LC, on overall absolute error by sex, country and method of adjustment



Country comparisons again indicate a high degree of consistency in the direction of effect, but considerable variation among countries. There is no consistency in patterns between the sexes. Taking all comparisons across the 20 populations and three fitting periods, the D_t adjustment gives the lowest error in 13 cases, the e(0) the lowest error in 23 cases and the $D_{x,t}$ the lowest error in 24.

Figure 7: Effect of adjustment method by sex, country and fitting period





Uncertainty

Uncertainty is examined by comparing the 95 per cent prediction intervals for life expectancy. Table 8 shows lower and upper interval width in 1996, the latest year for which data are available for all countries, for LM and BMS relative to LC. On average across countries, the intervals are reduced in width for both LM and BMS, despite the smaller number of observations. For females, the intervals are reduced by 20-40 per cent with the LM reduction being about 10 percentage points greater than the BMS. For males, the reduction is as great as 60-70 per cent with the BMS reduction being greater. Despite average reductions in uncertainty, BMS resulted in up to 36 per cent greater uncertainty for females in Australia and Finland, while LM failed to reduce uncertainty in two populations. The country pattern in uncertainty is similar in males and females.

In general, the smaller the prediction interval the better the method. However, if a variant were to over-fit the data, it would artificially reduce the standard error.[3] This is potentially more of a problem for BMS than LC and LM because of the selection of a shorter period in order to maximise the linear fit. However, as Table 8 shows, there is no consistent difference between LM and BMS in the amount of reduction in prediction width and the differences between these two variants are not great. Hence, either both BMS and LM over-fit the data or this is not an important issue. The latter view seems likely.

Table 8: Width of lower and upper prediction intervals for forecast life expectancy in 1996 relative to LC intervals by sex, variant and country

Country	Female				Male			
	Lower interval		Upper interval		Lower interval		Upper interval	
	LM	BMS	LM	BMS	LM	BMS	LM	BMS
Australia	0.95	1.31	1.00	1.36	0.27	0.35	0.36	0.46
Canada	0.79	0.79	0.81	0.81	0.88	0.65	0.91	0.69
Denmark	0.68	0.66	0.72	0.67	1.01	0.07	0.63	0.05
England and Wales	0.50	0.53	0.56	0.59	0.48	0.48	0.55	0.55
Finland	0.87	1.17	0.90	1.25	0.45	0.50	0.53	0.59
France	0.41	0.43	0.45	0.47	0.17	0.12	0.21	0.15
Italy	0.40	0.46	0.44	0.50	0.30	0.26	0.33	0.30
Norway	0.88	0.87	0.95	0.97	0.48	0.39	0.31	0.45
Sweden	0.63	0.77	0.69	0.82	0.37	0.47	0.42	0.55
Switzerland	0.58	0.66	0.62	0.71	0.39	0.43	0.42	0.47
Average	**0.61**	**0.70**	**0.66**	**0.76**	**0.39**	**0.32**	**0.40**	**0.38**

The only true test of the forecast uncertainty is out-of-sample forecast accuracy. Examination of whether actual life expectancy falls within the prediction interval for all years shows that for females there are only nine instances where this is not the case. These all occurred for the lower prediction interval of the LC variant in the first five years of the forecast: Denmark in 1986, Norway in 1986 to 1990, Sweden in 1986 and 1988 and Switzerland in 1986. For males, a different pattern occurs. First the LC prediction interval almost always contains actual life expectancy (except for Canada) while for LM and BMS there are a significant number of exceptions involving almost all countries. These all occur for the upper prediction interval and are mostly in the later years of the forecast. These are due to the unprecedented rapid decline in mortality that occurred among males during the forecast period in the countries concerned. In total,

[3] It is also possible to underestimate the standard error due to model mis-specification.

BMS has fewer occurrences than LM and the magnitude of excess over prediction interval is smaller for BMS.

While for females, these results may indicate that the prediction intervals for LM and BMS could be too wide, for LC and all three variants for males the combination of jump-off bias and error in drift preclude determination of the accuracy of the prediction intervals.

Conclusion

It has been shown that the LM and BMS variants are superior to LC in both forecast accuracy and width of prediction interval. The decomposition of error has demonstrated that jump-off bias is a substantial source of error for LC. In addition, the LC adjustment by fitting to D_t has been shown to be marginally inferior to the other two adjustment methods. It has also been shown that BMS performs better than LM in terms of overall absolute error in 15 out of the 20 populations considered. However in overall error terms, LM is superior for females and BMS superior for males. The prediction intervals for LM and BMS are similar and their evaluation inconclusive.

These results are limited to the forecasting period adopted. Further research is needed to determine whether they may be more widely generalised to other forecasting periods, particularly in the post-WW2 era. The consistency within sex indicates that it is likely that they may be generalised to other developed countries.

References

Booth, H., J. Maindonald and L. Smith (2002). "Applying Lee-Carter under conditions of variable mortality decline." *Population Studies* 56(3): 325-336.

Keilman, N. (1997) “Ex-post errors in official population forecasts in industrialized countries”, Journal of Official Statistics, 13(3): 245-277.

Keyfitz, N (1981) “The limits on population forecasting.” *Populations and Development Review* **7**(4): 579-593.

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

Lee, R. D. and L. R. Carter (1992). "Modeling and forecasting U.S. mortality." *Journal of the_American Statistical Association* 87(419): 659-671.

Tuljapurkar, S., N. Li and C. Boe (2000) “A universal pattern of mortality decline in the G7 countries”, *Nature* 405: 789-792.