Evaluating the Lee-Carter Method to Forecasting Spanish Life-Expectancy

Gabriel Pons Rotger and Montse Guillén Estany

4/02/2002

Abstract

We evaluate Lee-Carter method to forecast Spanish Life-Expectancy, and we find that this method is a very good tool to forecast the natural life expectancy and a reilable method to produce global long-run forecast of life expectancy corrected for accidental mortality.

1 INTRODUCTION

Many economic agents must decide according to some forecast of life-expectancy, and the consequences of these decisions are of crucial importance. The objective of this study is to asses the performance of the Lee-Carter (LC) model (Lee and Carter, 1992), that combines a demographic and time series model, to forecast life expectancy of Spanish males and females disaggregated by age.

Different from other studies which apply LC method to forecast life tables (see Lee and Carter, 1992; Carter, 2001; or Brouhns and Denuit, 2002), given our database¹, we forecast the more predictable part of life-expectancy considering only natural mortality, that is mortality for other causes than accidents, and then correct the point and interval forecast for the non precitable part, accident mortality. In this way, we improve the forecasting ability of the method as is shown in Subsection 4.3

2 SPANISH LIFE-EXPECTANCY

2.1 Historic Time Series

Although we are interested in forecast Spanish life expectancy for every age, we can observe a common pattern in Life Expectancy at birth. This variable is plotted for period 1930-2000 in Figure 1. As seen in the graphic, with the exception of the Civil War period where the life-expectancy decreased, life expectancy for males and females has increased from 48.38 years to 75.3 years for males and from 51.6 to 82.5 years for females. Two characteristics emerge from this figure, firstly, and as all the developed countries, there is a deceleration in life expectancy gains for both males and females, and secondly the gap between the higher females' life expectancy respect male's life expectancy has increased during all the period. The difference of 3.22 years favourable to females in 1930 has increased to 7.2 year in 2000.

2.2 Available Data

In terms of our forecasting objective, the long-term life expectancy presents a smoothed non-linear trend that a quantitative forecasting method is able to capture. However, and as seen below we have a short span database comprising only 1976-1993, with some non-regularities that convey non useful information for the future we want to forecast. Hence, we have to be very careful to forecast future trend in the life expectancy from our information set, and try not to include some irregularities in the predictive function.

The available data ranges from 1976 to 1993 and is disaggregated by sex and by age. The raw variables are: number of annual deaths at certain age x ($d_{x,t}$) and

¹See Section 2.





number of annual population at this age $(v_{x,t})$. From these variables the following life table variables are obtained:

- death probability at age x $(q_{x,t} = d_{x,t}/v_{x,t});$
- survival from age x to x + 1 probability $(p_{x,t} = 1 q_{x,t})$;
- survival up to age x probability $(_0p_{x,t} = p_{0,t} \times p_{1,t} \times ... \times p_{x-1,t});$
- survivals at age x of the life table $(s_{x,t} = 100, 000 \times_0 p_{x,t});$
- death at age x of the life table $(d_{x,t}^* = q_{x,t} \times s_{x,t});$
- life expectancy at age x $(e_{x,t} = (0.5 \times d_{x,t}^* + \dots + (t+0.5)d_{x+t,t}^* + \dots)/s_x);$

Figure 2 plots estimated life expectancy at birth with our data. It is shown how in this period life expectancy increases and also how the gap between females and males increases according with the historical time series. Comparing our estimation of life expectancy with the historical time series we detect a slightly downward bias in our estimation. For the two comparable periods, 1980 and 1990, we obtain $e_{0,80}^M =$ 72.41, $e_{0,80}^F = 78.37$, $e_{0,90}^M = 73.08$, and $e_{0,90}^F = 80.13$; while the historical data are $e_{0,80}^M = 72.52$, $e_{0,80}^F = 78.61$, $e_{0,90}^M = 73.4$, and $e_{0,90}^F = 80.49$, a difference that is not very significative.

Although it is difficult to appreciate from the plot life expectancy has a quite irregular component due the accident mortality. This can be appreciated if we plot the difference between life expectancy and the life expectancy computed excluding mortality for accidents.



It is seen how the reduction in life expectancy if we consider accidental mortality is around one year and a half for males (Figure 3) and half year for females (Figure 4). This reduction increases considerably for the period 1985-90 for both males and females and then returns to its normal pattern. Given this irregular behaviour we will model and forecast separately the natural life expectancy, that is the life expectancy without considering accidental mortality, and then we will correct our forecasts in order to consider the global life expectancy.

Figure 3: Life Expectancy Reduction because Accidents, Males



Figure 5 plots natural life expectancy at birth construct without accidental mortality. Although difficult to appreciate for this scale of values, these time series present a smoother trending behaviour and for this reason an easier to forecast trend.



Figure 4: Life Expectancy Reduction because Accidents, Females

Figure 5: Natural Life Expectancy at Birth, - Females, - - Males



As seen in the graphic, and as detected for the global life expectancy, the gap between natural life expectancy at birth for females and males has increased along the period. At 1976 $e_0^F = 77.4$ and $e_0^M = 72.6$, more than 4 years of difference and at the end of the sampling period 1993 $e_0^F = 81.27$ and $e_0^M = 75.06$, that is more than 6 years of difference.

3 LEE-CARTER METHOD

3.1 Mortality Rates. A more forecastable time series than Life Expectancy

While life expectancy presents a general non linear trend behaviour, Lee and Carter (1992) highlight that mortality rates $m_{x,t}$ for age-groups present a quite linear decreasing trend and therefore they are easier to forecast than life-expectancy is. This linearity

is also present in the Spanish death rates. Figure 6 shows the linearity of the trend component of death rates time series for the ages 0 and 75, while it is also shown the more irregular behaviour for death rates of 25 years old population, for both males and females. If we observe the death rates over ages for the extremes of the sampling period, we detect the reduction of these death rates for practically all ages with the important exception of death rates associated to ages between 23 and 37 for males² which for some years the death rates have increased instead of following the long-run trend of decreasing. This non decreasing trend of death rates corresponding to young males is due the short span of the sample and it was also observed in the paper of Lee and Carter for the years 60's-80's. Thus it seem a rather general behaviour. Obviously such a irregular behaviour is difficult to fit with our model that is designed to capture more long-run behaviour. Anyway, given the very low death rates associated to these ages the poor fit and forecast of young males death rates will have a very small impact on the life expectancy forecasts as it will be seen in the results.



Figure 6: Actual Death Rates, Selected Ages, – Males, - - Females

3.2 Lee-Carter Model of Mortality Rates

We assume that the annual death rates disaggregated by sex for every age can be represented by the following model:

$$\ln m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t}, \\ k_t \tilde{I}(1),$$

where a_x is an age-specific constant that captures the shape across age of the mortality; b_x is another age-specific constant that reflects the specific time dependence of the

 $^{^2\}mathrm{It}$ is also detected for females of 29 years old.



Figure 7: Actual Death Rates, Males, Selected Years, -1976 - - 1993

death rate; k_t is a mortality index, a common stochastic trend that represents the common long-run decreasing trending behaviour of the death rates, and $\varepsilon_{x,t}$ is age-specific historical influences not captured by the model.

Because we do not have the average population $pop_{x,t}$, we obtain the death rates $m_{x,t} (= d_{x,t}/pop_{x,t})$ indirectly from the death probability using the following formula:

$$m_{x,t} = \frac{q_{x,t}}{1 - q_{x,t}/2}.$$

We estimate the parameters a_x , b_x , and k_t by least squares using the single value decomposition imposing a normalization such that the constant term across time is the sample mean of the logs of the death rates

$$a_x = T^{-1} \Sigma_{t=76}^{93} \ln m_{xt}.$$

Then an initial approximate SVD estimation of the mortality index is obtained by summing across all the ages the deviations of the logs of death rates:

$$k_t^1 = \sum_{x \in X} (\ln m_{xt} - a_x),$$

and the slopes b_x are computed regressing the deviations with the initial estimation of the mortality index

$$b_x \mid \text{LS} (\ln m_{xt} - a_x) \text{ on } k_t^1.$$

We avoid to use raw data on death rates for oldest olds, that in this paper consider those ages between 89 and 100. Instead, we estimate these death rates with the Lee-Carter model. For this purpose, using the information for the ages between 75 and 89 we fit a quadratic trend depending on age to the parameters a_x and then forecast a_x . The slope is maintained fixed for oldest olds $b_x = b_{89}$ for 90 < x .100.



Figure 8: Actual Death Rates, Females, Selected Years, -1976 - - 1993

Table 1: Modelling the a parameters, Sample: Ages 75-89

	Ι	Regressand:	a_x			
	Μ	ales	Fen	Females		
Regressors	Coef	t-stat.	Coef.	t-stat.		
α	-3.02832	-654.715	-3.64352	-437.1260		
x	0.10974	82.49171	0.14942	62.33037		
x^2	-0.00038	-4.633364	-0.00115	-7.884734		
R^2		0.999		0.999		

Thus, the age-specific parameters for all ages estimated and forecasted are listed in the tables 2 and 3. As seen in the table 2 the slope b_x takes negatives values for males and ages 23-37 reflecting that mortality at these ages tends to increase when decrease for the remainder of ages. For the females, a negative slope is only detected for age 29.

We reestimate the mortality index such that the actual total number of deads

$$d_t = \sum_{x \in X} d_{x,t}$$

is equal to the fitted total number of deaths

$$\widehat{d}_t = \sum_{x \in X} e^{a_x + b_x k_t} pop_{x,t},$$

where the average population is derived from

$$pop_{x,t} = \frac{d_{x,t}}{m_{x,t}}$$

	Table 2: Fitted and Forecasted a and b, Males							
$Ag\epsilon$	a_x	b_x	Age	a_x	b_x	Age	a_x	b_x
0	-4.6564	0.0477	34	-6.9740	-0.0068	68	-3.6144	0.0134
1	-6.3495	0.0412	35	-6.8976	-0.0057	69	-3.5236	0.0143
2	-7.5843	0.0380	36	-6.8237	-0.0034	70	-3.4071	0.0142
3	-7.9817	0.0382	37	-6.7368	-0.0060	71	-3.3186	0.0156
4	-8.2504	0.0334	38	-6.6475	0.0002	72	-3.2174	0.0168
5	-8.3935	0.0390	39	-6.5347	0.0009	73	-3.1047	0.0145
6	-8.4821	0.0248	40	-6.4317	0.0032	74	-3.0138	0.0145
7	-8.5122	0.0249	41	-6.3446	0.0046	75	-2.9125	0.0143
8	-8.6232	0.0376	42	-6.2242	0.0062	76	-2.8081	0.0165
9	-8.7057	0.0382	43	-6.1086	0.0050	77	-2.7067	0.0151
10	-8.7400	0.0336	44	-6.0183	0.0089	78	-2.6043	0.0150
11	-8.7663	0.0212	45	-5.8921	0.0071	79	-2.4942	0.0152
12	-8.7133	0.0293	46	-5.7876	0.0054	80	-2.3775	0.0155
13	-8.6329	0.0244	47	-5.6889	0.0103	81	-2.2837	0.0152
14	-8.5760	0.0167	48	-5.5831	0.0068	82	-2.1700	0.0133
15	-8.4507	0.0210	49	-5.4761	0.0077	83	-2.0674	0.0130
16	-8.3255	0.0255	50	-5.3470	0.0107	84	-1.9649	0.0144
17	-8.2029	0.0196	51	-5.2738	0.0115	85	-1.8659	0.0121
18	-8.1312	0.0162	52	-5.1697	0.0105	86	-1.7683	0.0111
19	-8.0934	0.0109	53	-5.0540	0.0108	87	-1.6625	0.0080
20	-7.9990	0.0133	54	-4.9484	0.0113	88	-1.5631	0.0053
21	-7.9799	0.0078	55	-4.8615	0.0118	89	-1.4715	0.0061
22	-7.8572	0.0030	56	-4.7722	0.0137	90	-1.3684	0.0061
23	-7.7828	-0.0151	57	-4.6694	0.0123	91	-1.2710	0.0061
24	-7.7172	-0.0073	58	-4.5743	0.0136	92	-1.1744	0.0061
25	-7.6265	-0.0161	59	-4.4651	0.0125	93	-1.0785	0.0061
26	-7.5314	-0.0178	60	-4.3592	0.0136	94	-0.9834	0.0061
27	-7.4469	-0.0178	61	-4.2851	0.0111	95	-0.8890	0.0061
28	-7.3738	-0.0151	62	-4.1845	0.0118	96	-0.7954	0.0061
29	-7.3021	-0.0164	63	-4.0923	0.0117	97	-0.7025	0.0061
30	-7.2140	-0.0200	64	-3.9934	0.0105	98	-0.6104	0.0061
31	-7.1828	-0.0121	65	-3.8960	0.0114	99	-0.5190	0.0061
32	-7.0775	-0.0084	66	-3.8085	0.0128	100	-0.4284	0.0061
33	-7.0870	-0.0120	67	-3.7119	0.0137			

Table 2: Fitted and Forecasted a and b, Males

Age	a_x	b_x	Age	a_x	b_x	Age	a_x	b_x
0	-4.8831	0.0252	34	-7.5586	0.0068	68	-4.3706	0.0132
1	-6.5205	0.0220	35	-7.4430	0.0062	69	-4.2538	0.0128
2	-7.6623	0.0217	36	-7.3731	0.0078	70	-4.1168	0.0136
3	-8.0924	0.0189	37	-7.2976	0.0070	71	-4.0101	0.0125
4	-8.3969	0.0181	38	-7.1992	0.0104	72	-3.8729	0.0123
5	-8.5898	0.0146	39	-7.1397	0.0104	73	-3.7462	0.0132
6	-8.6821	0.0188	40	-7.0193	0.0103	74	-3.6155	0.0132
7	-8.7702	0.0064	41	-6.9649	0.0102	75	-3.4814	0.0131
8	-8.8681	0.0242	42	-6.8486	0.0073	76	-3.3446	0.0135
9	-8.9059	0.0172	43	-6.7512	0.0067	77	-3.2167	0.0127
10	-8.8745	0.0148	44	-6.6608	0.0090	78	-3.0699	0.0128
11	-8.9921	0.0246	45	-6.5700	0.0098	79	-2.9393	0.0117
12	-8.9010	0.0143	46	-6.4934	0.0118	80	-2.7875	0.0124
13	-8.8278	0.0184	47	-6.3961	0.0102	81	-2.6591	0.0109
14	-8.8764	0.0120	48	-6.2808	0.0088	82	-2.5243	0.0105
15	-8.7032	0.0119	49	-6.2017	0.0097	83	-2.3882	0.0096
16	-8.7253	0.0131	50	-6.0805	0.0112	84	-2.2544	0.0099
17	-8.6972	0.0094	51	-6.0091	0.0116	85	-2.1306	0.0088
18	-8.7221	0.0075	52	-5.9247	0.0107	86	-2.0080	0.0080
19	-8.6559	0.0116	53	-5.8375	0.0102	87	-1.8908	0.0067
20	-8.4760	0.0083	54	-5.7525	0.0109	88	-1.7778	0.0054
21	-8.5198	0.0117	55	-5.6815	0.0116	89	-1.6743	0.0041
22	-8.4501	0.0049	56	-5.5843	0.0124	90	-1.5469	0.0041
23	-8.4107	0.0093	57	-5.4935	0.0119	91	-1.4354	0.0041
24	-8.2833	0.0071	58	-5.3963	0.0130	92	-1.3262	0.0041
25	-8.1571	0.0056	59	-5.3082	0.0126	93	-1.2192	0.0041
26	-8.1160	0.0065	60	-5.1822	0.0148	94	-1.1146	0.0041
27	-8.0626	0.0036	61	-5.1162	0.0122	95	-1.0123	0.0041
28	-7.9472	0.0000	62	-4.9970	0.0126	96	-0.9123	0.0041
29	-7.8850	-0.0019	63	-4.9025	0.0115	97	-0.8146	0.0041
30	-7.8070	0.0022	64	-4.8069	0.0116	98	-0.7191	0.0041
31	-7.7292	0.0063	65	-4.7032	0.0119	99	-0.6260	0.0041
32	-7.6989	0.0055	66	-4.6036	0.0129	100	-0.5352	0.0041
33	-7.6178	0.0071	67	-4.4858	0.0123			

Table 3: Fitted and Forecasted a and b, Females

Then by searching in the neighbourhood of the values of the initial mortality index we find the final mortality index k_t^2 that equals

$$k_t^2 \left| \qquad d_t = \widehat{d}_t.$$

Figure 9 plots the final mortality index k_t^2 for males and females and table 4 displays the values of the index. Both indexes decline with time at a quite constant rate, with the female mortality index declining at a higher rate. This reflects the increase in the gap of the natural life expectancy between females and males. The graphic shows a practically linear pattern of the indexes with the exception of years 1983 1985 for females and 1985 and 1987 for males where the index of mortality presents small jumps.



Figures 10 to 12 shows the actual and fitted death rates for the selected ages 0, 25 and 75. It is shown how the model fits quite well except for the young group for which a different and more irregular behaviour have been detected. On the other hand, the fit of the model is similar for different years (Figures 13 and 14).

4 FORECASTING LIFE EXPECTANCY

We model the forecast index by a time series model and then we produce point and interval forecasts for the period 1994-2000. We evaluate the forecasting ability of our model by to ways. Firstly, we restimate the model for a reduced sample that spans from 1976 to 1990 and evaluate the performance of the point forecast of natural life expectancy by computing weighted measures with the prediction errors. Then we

Voor	Malog	Formelog
rear	males	remates
1976	12.9775	22.4530
1977	10.0490	18.6172
1978	8.8070	16.2450
1979	8.0240	11.3262
1980	4.6510	8.0450
1981	4.1570	6.7165
1982	0.7610	1.1752
1983	0.9730	3.7450
1984	-0.7680	-1.0380
1985	1.4060	0.2280
1986	-2.7225	-2.8290
1987	-4.6935	-7.1440
1988	-4.7965	-8.0168
1989	-5.3042	-10.2290
1990	-5.1565	-11.0130
1991	-5.5200	-12.3960
1992	-7.9880	-17.7760
1993	-8.0170	-17.9160

Table 4: Index of Mortality, 1976-1993

Figure 10: Actual and Fitted 0-Death Rates, - Actual, - - Fitted





Figure 11: Actual and Fitted 25-Death Rates, - Actual, - - Fitted

Figure 12: Actual and Fitted 75-Death Rates, - Actual, - - Fitted





Figure 13: Actual and Fitted 1976-Death Rates, - Actual, - - Fitted

Figure 14: Actual and Fitted 1993-Death Rates, - Actual, - - Fitted



evaluate the performance of our forecast of global life expectancy with the recent available figures of life expectancy of the Statistics National Institute (INE) for 2000. The Lee-Carter model succeeds in both evaluations, and therefore we believe that this methodology is very convenient to forecast for the unknown future.

4.1 Modelling the Mortality Index

We have detected unit roots in both mortality index, and also in the individual death rates time series. Thus we will use as in Lee and Carter (1992) or as in other works a stochastic trend to model the trend behaviour of the mortality index. Given the short span and also the presence of some outliers for the index the estimated models to forecast the mortality index are

$$\Delta k_t^M = -1.849 - 0.390 \Delta k_t^M + 3.344 d85_t + \hat{u}_t^M, \\ \hat{\sigma}^M = 1.306$$

for males and

$$\Delta k_t^F = -3.795 - 0.601 \Delta k_t^F + \hat{u}_t^F, \\ \hat{\sigma}^F = 1.998.$$

4.2 Evaluating Natural Life Expectancy Forecast Performance

In order to evaluate the short-run performance of our point forecast of the natural life expectancy we restimate the model of the mortality index for the period 1976-1990. The we point forecast the natural life expectancy for males and females for the period 1991-1993. The forecasts are quite good as is reflected in the following measures. For males we have

$$WRSMFE = 0.156$$
$$WMAFE = 0.156$$
$$WMAPFE = 0.47\%,$$

and for females we have

$$WRSMFE = 0.097$$
$$WMAFE = 0.095$$
$$WMAPFE = 0.27\%.$$

Where these weighted measures are defined as

$$WMAPFE = \sum_{x \in X} MAPFE_x \times \omega_x,$$

where

$$error_{x,T}(h) = e_{x,T+h} - \hat{e}_{x,T}(h)$$

with forecast origin T = 1990 and forecast horizon h = 1, 2, 3, and

$$RMAPFE_{x} = \frac{\sum_{h=1}^{3} error_{x,T}(h)^{2}}{3}$$
$$MAFE_{x} = \frac{\sum_{h=1}^{3} |error_{x,T}(h)|}{3},$$
$$MAPFE_{x} = \frac{\sum_{h=1}^{3} |\frac{error_{x,T}(h)}{e_{x,T+h}}|}{3},$$
$$\omega_{x} = \frac{\overline{pop}_{x}}{\sum_{x \in X} \overline{pop}_{x}},$$
$$\overline{pop}_{x} = \frac{pop_{x,91} + pop_{x,92} + pop_{x,93}}{3}.$$

Hence, from all these measures we can conclude that the Lee-Carter model produces excellent forecasts of natural life expectancy. If we compare the quality of forecasts for males and females, we observe how female life expectancy with aWMAPFE = 0.27% is better forecasted than male life expectancy with WMAPFE = 0.47%.

4.3 Evaluating Global Life Expectancy Forecast Performance

Our objective is to forecast life Expectancy considering all mortality causes. Thus, we must include in some way the increase of mortality produced for non-natural causes, that is the accidental mortality. Given its irregular behaviour as shown in Figures 3 and 4 we do not model it but include it through some robust central measure. Specifically we adopt the median of the difference between Natural and Global Life Expectancy, and consider this median as the reduction of life expectancy because accidents. We will modify all point and interval forecast of the natural life expectancy adding this factor.

4.3.1 Life-Expectancy Reduction Because Accidents

4.3.2 Point and Interval Forecast of Life Expectancy

Age	Males	Females	Age	Males	Females	Age	Males	Females
0	1.47	0.52	34	0.68	0.27	68	0.14	0.10
1	1.46	0.51	35	0.66	0.27	69	0.14	0.10
2	1.44	0.50	36	0.64	0.26	70	0.13	0.09
3	1.43	0.49	37	0.62	0.26	71	0.12	0.09
4	1.42	0.49	38	0.59	0.25	72	0.12	0.08
5	1.41	0.48	39	0.57	0.25	73	0.11	0.08
6	1.40	0.47	40	0.55	0.25	74	0.11	0.07
7	1.39	0.47	41	0.53	0.24	75	0.10	0.07
8	1.39	0.46	42	0.50	0.24	76	0.09	0.06
9	1.38	0.46	43	0.48	0.23	77	0.09	0.06
10	1.37	0.45	44	0.47	0.23	78	0.08	0.06
11	1.37	0.45	45	0.45	0.22	79	0.08	0.05
12	1.36	0.44	46	0.43	0.22	80	0.07	0.05
13	1.35	0.44	47	0.41	0.21	81	0.07	0.04
14	1.34	0.43	48	0.39	0.21	82	0.06	0.04
15	1.32	0.43	49	0.37	0.20	83	0.05	0.04
16	1.31	0.42	50	0.36	0.20	84	0.05	0.03
17	1.28	0.41	51	0.34	0.19	85	0.04	0.03
18	1.24	0.40	52	0.32	0.19	86	0.03	0.02
19	1.20	0.39	53	0.31	0.18	87	0.02	0.02
20	1.15	0.38	54	0.29	0.18	88	0.01	0.01
21	1.11	0.37	55	0.28	0.17	89	0.00	0.00
22	1.07	0.36	56	0.27	0.16	90	0.00	0.00
23	1.02	0.35	57	0.25	0.16	91	0.00	0.00
24	0.99	0.34	58	0.24	0.15	92	0.00	0.00
25	0.95	0.33	59	0.23	0.15	93	0.00	0.00
26	0.91	0.32	60	0.22	0.14	94	0.00	0.00
27	0.88	0.32	61	0.20	0.14	95	0.00	0.00
28	0.84	0.31	62	0.19	0.13	96	0.00	0.00
29	0.81	0.30	63	0.18	0.13	97	0.00	0.00
30	0.78	0.30	64	0.17	0.12	98	0.00	0.00
31	0.75	0.29	65	0.17	0.12	99	0.00	0.00
32	0.73	0.29	66	0.16	0.11	100	0.00	0.00
33	0.70	0.28	67	0.15	0.11			

Table 5: Life Expectancy Reduction Factor

	[Point]	INE
Males	74.31	74.91	75.49	75.3
Females	81.73	82.32	82.90	82.5