

Sensitivity Analysis of the Lee-Carter Model Fitting Mortality by Causes of Death (Italy, 1982-2002)

*Sensitività del modello di Lee-Carter per lo studio della mortalità per
causa (Italy, 1982-2002)*

Tiziana Barugola
Dipartimento di Statistica e Matematica “Diego de Castro” Torino
barugola@econ.unito.it

Carlo Maccheroni
Dipartimento di Statistica e Matematica “Diego de Castro” Torino
cmaccheroni@econ.unito.it

Riassunto - Le linee dell'evoluzione futura della mortalità sono delineate dai risultati attesi della ricerca bio-medica, ma i metodi sviluppati per la previsioni dei livelli di mortalità generale si mostrano problematici se generalizzati all'analogia previsione per le varie cause di morte. Obiettivo del lavoro è di evidenziare come uno dei modelli previsivi oggi più diffusi in ambito statistico-demografico, il modello di Lee-Carter, non possa essere applicato in modo acritico nell'ambito della mortalità per causa.

Keywords: Lee-Carter method, Singular Value Decomposition, Principal Component Analysis (PCA), Causes of Death

1. The Lee-Carter mortality model fitting

The general scenario of the future evolution of mortality by cause of death is already outlined by the increasing successes of biomedical research, but the rates for the various causes of death are still difficult to be predicted. The Lee-Carter method (LC, hereinafter) aims to show the trend in historical mortality data by a simple model, and seeks it by the Singular Value Decomposition (SVD) approximation method applied to the logarithms of the central death rate (m_x). It can be written as a sum of three elements: the age-specific component (a_x), the product of a time-varying parameter (k_t) with an age-specific component (b_x , that is the tendency of mortality by age), and the “error” (1). Computing the matrix $R = [r_{xt}]$ with $r_{x,t} = \ln(m_{x,t}) - \hat{a}_x$, applying to the SVD and denoting the (ordered) singular value, the left and right singular vector of $[r_{xt}]$ by s_i , $u_i(x)$, and $v_i(t)$ respectively, the LC model is:

$$\ln(m_{xt}) = a_x + b_x k_t + \varepsilon_{x,t} = a_x + s_1 u_1(x) v_1(t) + \sum_{i=2} s_i u_i(x) v_i(t) \quad (1)$$

It is clear that what was labeled as an error after decomposition is captured in the remaining element of the $[r_{xt}]$. A measure of goodness of fit can be obtained by summing all the unexplained age group variances, but also computing this ratio by age groups (in a more accurate examination of the model, see Lee-Carter, 1992, 2000).

2. The fitted mortality by causes of death

The first focus of the model is the proportion of variance explained by the first singular value. In the developed countries, for all causes of death, the first singular value

captures over 90 per cent of the mortality variation (see Tuljapurkar, 2005), but when we consider data by the causes of death this falls. In this study, we have analysed Italian population data by gender and five age groups from year 1982 to 2002, considering only a small number of causes of deaths.

In the first approach, we have applied the principal component analysis (PCA), since the LC model can be seen as a special case of PCA, to compute the cumulative percentage of variance explained by principal components. In table 1 we report cumulative percentage of variance referred to the first-third component. When we consider all causes the first principal component explains – for Italian males and females – 72.9 and 82.4 per cent of variance respectively, but if we take the causes of death into account we observe an information loss and the models fitted by the LC method show significant departures from linearity in the time component (k_t).

Table 1. Cumulative percentage of variance explained in mortality by first- third principal component for varied causes of death and gender.

Causes of death	All causes		A		B		C		D		E		F	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F
1st comp.	72.9	82.4	63.7	69.3	43.1	46.5	70.0	69.3	67.5	71.8	71.1	69.3	70.6	80.7
2nd comp.	89.4	92.6	75.5	78.6	67.2	72.0	81.5	82.1	75.6	80.5	85.5	78.8	86.2	85.4
3rd comp.	94.2	95.0	81.0	84.4	77.3	79.7	87.3	87.4	81.5	85.8	90.8	83.9	91.9	89.0

Legend: A: Neoplasms; B: Endocrine, Nutritional and Metabolic diseases; C: Diseases of the Circulatory System; D: Diseases of the respiratory system; E: Injury and poisoning; F: Other causes

An absurd trend is showed in Endocrine, Nutritional and Metabolic diseases because in this category deaths caused by AIDS are also included. But if we leave out those deaths we obtain (quite) a linear trend particularly for females, less for males, that is why we have to consider those deaths separately. Briefly, while on the whole we have observed a good fit of the different models, the measures of goodness of fit by age groups give values, which are hard to accept.

In view of those remarks and the problems inherent in forecasting mortality by causes, we are critical about the use of the LC method as a “routine” since it does not work well at the first step of parameter estimate of the mortality trend.

Table 2. Coefficient of determination by age-groups, gender and causes of death.

Age	A		B		C		D		E		F	
	M	F	M	F	M	F	M	F	M	F	M	F
50-54	0.94	0.85	0.00	0.73	0.97	0.93	0.92	0.62	0.92	0.89	0.97	0.92
70-74	0.21	0.77	0.10	0.85	0.97	0.83	0.96	0.91	0.81	0.88	0.95	0.83
80-84	0.53	0.07	0.04	0.81	0.97	0.84	0.88	0.82	0.73	0.87	0.94	0.69

References

- Lee, D. R., Carter, L. R., (1992), Modeling and Forecasting U.S. Mortality, *Journal of the American Statistical Association*, 87 (419): 659-671.
- Lee D. R., (2000), The Lee-Carter Method for Forecasting Mortality, with Various extensions and Application, *North American Actuarial journal*, 4 (1) :80-91.
- Tuljapurkar, S., (2005), Stochastic Forecasts of Mortality, Population and Pension System, *Perspective on Mortality Forecasting*, 2,65-77.