COMPARING MORTALITY TRENDS VIA LEE CARTER METHOD IN THE FRAMEWORK OF MULTIDIMENSIONAL DATA ANALYSIS

Giuseppe Giordano, Maria Russolillo
Department of Statistics and Economics
University of Salerno, Italy
&
Steven Haberman
Faculty of Actuarial Science and Statistics, Cass Business School,
City University, London, UK

Keywords: Biplot, Lee-Carter Methodology, Mortality Forecasting, PCA, SVD.

Abstract

The Lee-Carter methodology 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:

$$\ln\left(m_{x,t}\right) = \alpha_x + \beta_x k_t + \varepsilon_{x,t} \qquad x = x_1,\dots,x_k;\ t = t_1,\dots,t_n \qquad [1]$$

where the subscript *x* denotes the age group and *t* indicates the calendar year. The component $\varepsilon_{x,t}$ is the error term with $E(\varepsilon_{x,t}) = 0$ and variance $\sigma_\varepsilon^2 < \infty$.

We can state the model (1) referring to the mean centred log-mortality rates:

$$\tilde{m}_{x,t} = \ln\left(m_{x,t}\right) - \alpha_x = \beta_x k_t + \varepsilon_{x,t} \qquad [2]$$

Following Lee & Carter (1992), the components β_x and k_t can be estimated according to the singular value decomposition (Eckart & Young, 1936) with suitable normality constraints. The terms $\tilde{m}_{x,t}$ are arranged in the $(x_k \times t_n)$ matrix $\tilde{\mathbf{M}}$:

$$\tilde{\mathbf{M}} = \begin{matrix} \\ x_1 \\ x_2 \\ \vdots \\ x_k \end{matrix} \begin{pmatrix} t_1 & t_2 & \cdots & t_n \\ \hline \tilde{m}_{1,1} & \tilde{m}_{1,2} & \cdots & \tilde{m}_{1,n} \\ \tilde{m}_{2,1} & \tilde{m}_{2,2} & \cdots & \tilde{m}_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{m}_{k,1} & \tilde{m}_{k,2} & \cdots & \tilde{m}_{k,n} \end{pmatrix}$$

The singular value decomposition (SVD) of the matrix $\tilde{\mathbf{M}}$ can be written as the product of three matrices which have an useful geometric interpretation. In particular, the SVD model is stated as follows:

$$\tilde{\mathbf{M}}_{x\times t} = \mathbf{S}_{x\times h}\mathbf{V}_{h\times h}\mathbf{D}'_{h\times t} \qquad \text{where } h \text{ is the rank of } \tilde{\mathbf{M}} \ (h \le \min\{x_k, t_n\}) \qquad [3]$$

and **V** is a diagonal matrix of positive singular values of $\tilde{\mathbf{M}}$. The matrices **S** and **D** hold the left and right singular vectors forming an orthogonal basis, respectively, for the rows and the columns of the matrix $\tilde{\mathbf{M}}$.

We can see the equivalence with equation (2) by rewriting the (3) in the following way:

$$\tilde{\mathbf{M}} = \sum_{\alpha=1}^{h} v_\alpha s_\alpha d'_\alpha + \mathbf{E} \qquad h \le \min\{x_k, t_n\} \qquad [4]$$

where the error term **E** in the (4) reflects the residual information not captured by the first few components of the SVD approximation. The correspondence of the two models arises by defining $\beta = vs$ and $k = d'$.

The SVD approximation allows a graphical representation in a reduced subspace of both rows and columns of the matrix $\tilde{\mathbf{M}}$. The geometric reading of such representation is carried out according to the biplot (Gabriel, 1971). The biplot is a low-dimensional display of a rectangular data matrix, where the rows and the columns are represented by points. The interpretation of the biplot is consistent with the scalar products between row and column vectors as defined in (4).

The current literature on the topic of mortality forecasts, gives emphasis mainly to the analytical role of SVD as an estimation method. So far, little attention appears to have been given, in this field, to the descriptive features of such decomposition.

In this paper we exploit both the analytical and descriptive features of the SVD in the light of principal component analysis (Hotelling, 1933). This technique aims at reducing the information held in a (*units*×*variables*) matrix of quantitative data to represent them in a reduced subspace. We propose a data analysis strategy exploiting both the analytical and the graphical properties of the Lee Carter method together with the biplot interpretation. This approach allows us to simultaneously obtain a graphical display of the mortality trends for age group, while also providing a theoretical support for the choice of the components' number to be considered in the model fitting. An interpretation of the characteristic elements of the SVD approximation and the factorial technique is given in the light of the latent demographic scenario.

The main advantage of our approach is the opportunity to graphically show the relation within the estimated *time varying* parameters *k*'s , within the *age group* coefficients β's and between these two sets. Furthermore, we are able to describe the meaningful components of the SVD and explore the information lost in the residual term. Another opportunity is to visually compare the model structure with respect to some variables of interest. For example, we can compare different countries, ethnic groups, mortality cause, etc.

In order to give an in-depth illustration of our procedure, we will analyse the Italian mortality rates from 1950 to 2000. The analysis will be carried out by comparing male and female subpopulations.

Main References

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211-218.

Gabriel, K. R. (1971). The biplot-graphic display of matrices with application to principal component analysis. *Biometrika*, **58**, 453-467.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol*. **24**, 417-441, 498-520

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

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