Lee-Carter models: the wider context

Rob J Hyndman 15 December 2022

Basellini, Camarda & Booth (2022) provide a very useful review of the Lee-Carter method and its most important extensions, in the context of forecasting mortality rates. Bringing together this rich literature in a convenient overview will be extremely helpful to researchers navigating the literature and identifying the most appropriate variant for tackling a range of mortality modelling problems. I'd like to take a broader perspective, and highlight the value of this modelling framework for data *other* than mortality rates, and to point out some of the connections between the Lee-Carter model and several large modelling classes that are widely used in other areas of application.

Other age-specific applications

First, within demography, analogous models have been used for fertility, net migration, and other series indexed by age. If we let the observation at time t and age x be denoted by $y_{x,t}$, then we can write

$$y_{x,t} = a_x + \sum_{j=1}^{J} b_{x,j} k_{t,j} + \varepsilon_{x,t}, \tag{1}$$

where $\sum_x b_{x,j} = 1$ and $\sum_j k_{t,j} = 0$ for all j. When $y_{x,t} = \log(m_{x,t})$, $m_{x,t}$ denotes the mortality rate for age x and time t, and $J = 1$, we have the original Lee-Carter model. Lee-Carter mortality forecasts are obtained when SVD is used to estimate the components, and $k_{t,j}$ follows a random walk with drift.

Lee (1993) applied the same model to fertility data, by setting $y_{x,t}$ equal to the fertility rate for age x at time t, and $J = 1$. Later Hyndman & Ullah (2007) used a similar model for fertility, where $y_{x,t}$ denotes smoothed log fertility rates, $J > 1$, and the various $k_{t,j}$ series were modelled using univariate ARIMA models. A Bayesian variation of the model was used by Wiśniowski et al. (2015) for fertility, mortality, immigration and emigration data.

Hyndman & Booth (2008) applied (1) to smoothed log mortality rates, smoothed log fertility rates, and smoothed net migration rates, in order to construct stochastic age-specific population forecasts. Hyndman, Zeng & Shang (2021) used a similar idea, but combined it with the coherent product-ratio models of Hyndman, Booth & Yasmeen (2013) to study the effect of pension age on the old-age dependency ratio.

In fact, whenever we have age-specific data, a similar model can be used. For example, Mason & Miller (2018) used the model for age-specific health care consumption, Url, Hyndman & Dokumentov (2016) used it to model age-specific labour market participation rates, and Yap et al. (2018) used it to model age-specific hospital admission rates.

Lee-Carter as a dynamic factor model

The Lee-Carter model can be considered a special case of some larger model classes that have been well-studied by statisticians and econometricians. For example, it is a special case of a dynamic factor model, as pointed out by French & O'Hare (2013), and can be used for *any* multivariate time series, where x denotes the individual series. A dynamic factor model can be written as (Stock & Watson, 2016)

$$\begin{aligned}\boldsymbol{y}_t &= \boldsymbol{a} + \boldsymbol{B}(L)\boldsymbol{k}_t + \boldsymbol{e}_t,\\ \boldsymbol{k}_t &= \boldsymbol{d} + \boldsymbol{C}(L)\boldsymbol{k}_{t-1} + \boldsymbol{\eta}_t,\end{aligned}$$

where $\boldsymbol{y}_t$ is an N-vector of the observed time series at time t, $\boldsymbol{k}_t$ is a J-vector of unobserved factors, $\boldsymbol{B}(L)$ and $\boldsymbol{C}(L)$ are lag polynomial matrices of size $N \times J$ and $J \times J$ respectively, and $\boldsymbol{e}_t$ and $\boldsymbol{\eta}_t$ are multivariate white noise errors that are also uncorrelated with each other. The specific Lee-Carter model is obtained when $J = 1$, $\boldsymbol{B}(L) = [b_{1,1}, \dots, b_{N,1}]'$, and $\boldsymbol{C}(L) = 1$. There is a rich literature on dynamic factor models, largely in the context of economic and finance data, that could profitably be applied to mortality and other demographic data.

Lee Carter as a functional data model

When x is a continuous variable (such as age), the model can be considered within the framework of functional data models (Hyndman & Ullah, 2007), and then (1) is most naturally expressed as

$$y_t(x) = a(x) + \sum_{j=1}^{J} b_j(x)k_{t,j} + e_t(x).$$

Within this model class, developments have included the use of dynamic functional principal components, in contrast to the static components obtained using SVD (Gao, Shang & Yang, 2019), and the use of functional partial least squares rather than principal components (Hyndman & Shang, 2009). A combination of the functional data approach with dynamic factor models was proposed by Hays, Shen & Huang (2012).

Generalized linear time series factor models

Brouhns, Denuit & Vermunt (2002) showed that it is helpful to embed a Lee-Carter type model within a Generalized Linear Modelling (GLM) framework to account for the count nature of deaths. Once we realise that a Lee-Carter model is a special case of a functional time series factor model, it becomes clear that it will also be useful to consider these within a more general GLM-like framework. Early approaches along these lines (but not in the context of mortality data) include Shen & Huang (2008) and Clark & Wells (2022). An application of such models to mortality data is provided by Li, Huang & Shen (2018).

Where to from here?

I have only touched on the vast literature available for these model families, each of which contains the Lee-Carter model as a special case. Surely there is much to be learned from their applications across many fields, and in the methodological and theoretical developments that have appeared outside of the mortality modelling literature.

References

Basellini, U, CG Camarda & H Booth (2022). Thirty years on: A review of the Lee-Carter method for forecasting mortality. *International J Forecasting*. in press.

Brouhns, N, M Denuit & JK Vermunt (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance, Mathematics & Economics* **31**(3), 373–393.

Clark, NJ & K Wells (2022). Dynamic generalised additive models (DGAMs) for forecasting discrete ecological time series. *Methods in Ecology and Evolution*. in press.

French, D & C O'Hare (2013). A dynamic factor approach to mortality modeling. *J Forecasting* **32**(7), 587–599.

Gao, Y, HL Shang & Y Yang (2019). High-dimensional functional time series forecasting: An application to age-specific mortality rates. *Journal of Multivariate Analysis* **170**, 232–243.

Hays, S, H Shen & JZ Huang (2012). Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics* **6**(3), 870–894.

Hyndman, RJ & H Booth (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International J Forecasting* **24**(3), 323–342.

Hyndman, RJ, H Booth & F Yasmeen (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography* **50**(1), 261–283.

Hyndman, RJ & HL Shang (2009). Forecasting functional time series (with discussion). *Journal of the Korean Statistical Society* **38**(3), 199–221.

Hyndman, RJ & S Ullah (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis* **51**(10), 4942–4956.

Hyndman, RJ, Y Zeng & HL Shang (2021). Forecasting the old-age dependency ratio to determine a sustainable pension age. *Australian & New Zealand J Statistics* **63**(2), 241–256.

Lee, RD (1993). Modeling and forecasting the time series of U.S. fertility: age distribution, range, and ultimate level. *International J Forecasting* **9**, 187–202.

Li, G, JZ Huang & H Shen (2018). Exponential family functional data analysis via a low-rank model. *Biometrics* **74**(4), 1301–1310.

Mason, CN & T Miller (2018). International projections of age specific healthcare consumption: 2015–2060. *J Economics of Ageing* **12**, 202–217.

Shen, H & JZ Huang (2008). Interday forecasting and intraday updating of call center arrivals. *Manufacturing and Service Operations Management* **10**(3), 391–410.

Stock, JH & MW Watson (2016). "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics". In: *Handbook of macroeconomics*. Vol. 2. Elsevier. Chap. 8, pp.415–525.

Url, T, RJ Hyndman & A Dokumentov (2016). *Long-term forecasts of age-specific participation rates with functional data models*. Working paper 03/16. Department of Econometrics & Business Statistics, Monash University. robjhyndman.com/publications/participation-rates/.

Wiśniowski, A, PWF Smith, J Bijak, J Raymer & JJ Forster (2015). Bayesian population forecasting: Extending the Lee-Carter method. *Demography* **52**(3), 1035–1059.

Yap, SZZ, SM Zahari, Z Derasit & S Shariff (2018). Comparing methods for Lee–Carter parameter's estimation for predicting hospital admission rates. In: *Proceedings of the Second International Conference on the Future of ASEAN (ICoFA) 2017–Volume 2*. Springer, pp.361–372.