IJEBMRI
INTERNATIONAL JOURNAL

https://ijebmri.org

International Journal of Economics, Business, Management Research Intelligence (IJEBMRI)

Original Article

# Estimation of Parameters of the Lee–Carter Model for Mortality Prediction

Chinton Emmanuel

*Department of Statistics, UCC*

Received Date: 11 October2025 Revised Date: 31 October 2025 Accepted Date: 14 November 2025

## Abstract

Mortality forecasting is critical for public-health planning, pension provisioning and actuarial risk management. The widely-used Lee–Carter model expresses log-age-specific mortality rates as the sum of an age profile plus a time-varying mortality index multiplied by age-sensitivity coefficients. Estimating its parameters—age effect $a_x$, sensitivity $b_x$and time index $k_t$ is a key step before forecasting. This paper presents a comprehensive review of parameter estimation methods for the Lee–Carter model, outlines a detailed step-by-step methodology (including singular value decomposition and time-series modelling of $k_t$), discusses practical issues (identifiability, smoothing, time-series model selection) and offers recommendations for future research in contexts of limited data or changing mortality regimes.

## Keywords

Mortality forecasting, Mortality index, Identifiability issues, Time-series modeling, Public health planning, Singular Value Decomposition (SVD).

## Introduction

Forecasting mortality rates is fundamental for demographic projections, life-insurance pricing, pension reserve setting, and public policy. As populations age and longevity increases, the need for accurate mortality predictions becomes more pressing. One of the benchmark methods is the Lee–Carter model (Lee & Carter, 1992), which decomposes the log of age-specific mortality rates into three components: an average age profile, a time-varying index capturing overall mortality level, and age-specific sensitivity coefficients. The model's popularity stems from its simplicity, interpretability and ability to generate stochastic forecasts.

Despite its widespread usage, precise estimation of its parameters remains non-trivial. The choices made in estimation affect forecast accuracy, residual behaviour and policy implications. This paper aims to (1) summarize relevant literature on estimation methods, (2) provide a detailed methodological exposition of parameter estimation for the Lee–Carter model, (3) discuss pragmatic issues in applied settings, and (4) propose directions for future research.

## Literature Review

Since its introduction by Lee & Carter (1992), the model has become a standard in mortality forecasting (Lee & Carter, 1992; Lee & Miller, 2001). The original model is given by

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

Where $m_{x,t}$is the mortality rate at age $x$in year $t$; $a_x$is the age-specific average log rate; $b_x$is the sensitivity of mortality at age $x$to the time index $k_t$; and $\varepsilon_{x,t}$is the error term (Lee & Carter, 1992; see also the online demonstration by MathWorks).

Early work reviewed its performance. For example, Booth et al. (2002) compared the original Lee–Carter method with several variants across 10 developed countries, finding that although some extensions improved log-mortality forecasts, the improvement for life-expectancy forecasts was less clear. A review by Cairns et al. (in "Mortality Modelling and Forecasting: A Review of Methods") noted that the Lee–Carter two-factor model remains highly successful but pointed to gaps in uncertainty quantification and multi-population issues.

More recent literature has focused specifically on estimation enhancements: smoothing of mortality rates before fitting (Fazle Rabbi & Mazzuco, 2020) found improved accuracy by applying LASSO-type regularization. Other extensions include moving from SVD-based estimation to maximum-likelihood or GLM frameworks (Koissi & Shapiro, 2008) to better handle over-dispersion and variable age-sensitivity. Newer methodological research explores robust estimation via probabilistic principal components (Guo & Li, 2022) and hybridising with machine-learning methods.

Collectively, these works highlight two key points: (a) parameter estimation in the Lee–Carter framework critically influences forecasts; and (b) there remains active research in improving estimation especially when data are noisy, mortality patterns shift, or populations are small/heterogeneous

## A. Methodology: Estimation of the Lee–Carter Model Parameters

This section presents a detailed step-by-step methodology for estimating parameters of the Lee–Carter model—namely $a_x$, $b_x$, and $k_t$—along with commentary on practical implementation issues.

## B. Model Specification

We begin with the classic model formulation:

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

for ages $x = x_1, \dots, x_X$and years $t = t_1, \dots, t_T$. Here $m_{x,t}$may represent age-specific central death rates or other standard mortality rates for age group $x$in year $t$(Lee & Carter, 1992; MathWorks example).

Because the model is non-identifiable without constraints (for instance one could add a constant to $k_t$and subtract it from $a_x$, or scale $b_x$and divide $k_t$accordingly), standard constraints are typically imposed:

$$\sum_x b_x = 1, \sum_t k_t = 0.$$

These constraints make the solution unique and interpretable (MathWorks example).

## C. Estimation via Singular Value Decomposition (SVD)

The seminal estimation approach recommended by Lee & Carter (1992) is via singular value decomposition of a centred log-mortality matrix. The steps are as follows:

1. Compute $a_x$:

$$a_x = \frac{1}{T}\sum_{t=1}^{T} \ln(\, m_{x,t}), \forall x.$$

This is the average over time of the log mortality rate for each age.

2. Centre the log-rates:

$$M_{x,t} = \ln(m_{x,t}) - a_x.$$

3. Apply SVD:

Let the matrix $M = [M_{x,t}]$(of dimension $X \times T$). Compute

$$M = U\,\Sigma\, V^{\top}.$$

Choose the first singular vector (largest singular value) corresponding to the dominant mortality dimension. Then set

$$b_x \propto U_{x,1}, k_t \propto \Sigma_{1,1}\, V_{t,1}.$$

(See the Wikipedia description of algorithmic steps.)

4. Impose constraints and scale:
After the preliminary $b_x, k_t$are extracted, they are rescaled to satisfy the identifiability constraints $\sum_x b_x = 1$and $\sum_t k_t = 0$.Typically one divides by $\sum_x b_x$to force the first constraint, and then adjusts $k_t$to sum to zero.

5. Estimate residual error:
Fit the equation

$$\widehat{m}_{x,t} = \exp(a_x + b_x\, k_t)$$

and compute residuals $\varepsilon_{x,t} = \ln(m_{x,t}) - (a_x + b_x\, k_t)$. Check distribution and age/time-patterns of residuals for model adequacy (see Fazle Rabbi & Mazzuco, 2020).

Forecasting $k_t$:
Having estimated the historical series $\{k_t\}_{t=1}^{T}$, one fits a univariate time-series model (commonly ARIMA) and projects into future years: $k_{T+h}$for $h = 1,2,\ldots,H$. Then forecasts of age-specific mortality are:

$$\ln(\widehat{m}_{x,T+h}) = a_x + b_x\, k_{T+h}.$$

## D. Alternative Estimation: GLM / Maximum Likelihood

While SVD is convenient and computationally efficient, alternative estimation methods frame the Lee–Carter model in a statistical inference context (Renshaw & Haberman, 2003; Koissi & Shapiro, 2008). For example:

- Specify death counts $D_{x,t}$and exposures $E_{x,t}$. Then model
$$D_{x,t} \sim \text{Poisson}(E_{x,t} m_{x,t}), \ln(m_{x,t}) = a_x + b_x\, k_t.$$
- Estimate parameters $a_x, b_x, k_t$by maximum likelihood under identifiability constraints (Koissi & Shapiro, 2008).
- This framework can handle heteroscedasticity, over-dispersion (negative-binomial extensions) and provides likelihood-based inference for parameter uncertainty (Neves et al., 2018).

## E. Practical Implementation Issues

In applying the Lee–Carter estimation procedure, practitioners should be aware of several common issues:

- Data smoothing: Raw age-specific mortality rates may be noisy, especially at older ages or for small populations. Smoothing (e.g., using splines or LASSO-type regularization) before estimation can improve parameter stability and forecast accuracy (Fazle Rabbi & Mazzuco, 2020).
- Age-sensitivity $b_x$invariance assumption: The original model assumes that $b_x$is fixed over time (i.e., the age-pattern of mortality change remains constant). Empirical work shows this may not hold (Koissi & Shapiro, 2008).
- Time-series model selection for $k_t$: While a random-walk with drift or ARIMA (0,1,0) model is standard, checking residual autocorrelation, structural breaks or non-stationarity is recommended. (See Guo & Li, 2022).
- Forecast uncertainty: Generating predictive intervals often requires simulation of the $k_t$process and propagation of parameter uncertainty (see Lee & Miller, 2001).
- Small populations / limited time series: In contexts with few years of data or small populations (e.g., developing countries), the SVD estimates can become unstable; methods that borrow strength across ages or populations, or that apply regularization, may produce more reliable estimates (see review by Cairns et al., 2015).

## F. Summary of Estimation Steps

In applied work the following procedure may be followed:

- Prepare dataset: age-specific mortality rates (or death counts + exposures) for $x = 1 \ldots X$ages, $t = 1 \ldots T$years.
- (Optional) Smooth the data if necessary (e.g., spline, LASSO) to reduce noise.
- Compute $a_x = \frac{1}{T}\sum_t \ln(m_{x,t})$.
- Centre the log-rates: $M_{x,t} = \ln(m_{x,t}) - a_x$.

- Perform SVD on $M$. Extract first singular vectors for $b_x$, $k_t$.
- Scale $b_x$, $k_t$ to satisfy $\sum_x b_x = 1$ and $\sum_t k_t = 0$.
- Compute fitted rates $\hat{m}_{x,t} = \exp(a_x + b_x\, k_t)$; examine residuals $\ln(m_{x,t}) - (a_x + b_x\, k_t)$ for age/time patterns.
- Fit a univariate time-series model (e.g., ARIMA) to $\{k_t\}$; forecast future values $k_{T+h}$.
- Calculate future mortality forecasts: $\ln(\hat{m}_{x,T+h}) = a_x + b_x\, k_{T+h}$.
- (Optional) Use simulation (e.g., Monte Carlo) to propagate uncertainty in $k_t$ and (if possible) parameter estimates $a_x, b_x$.
- Document assumptions and limitations: fixed age-pattern $b_x$, forecasting horizon, data quality, smoothing effects.

## Recommendations for Future Research

Given the methodological review and detailed estimation exposition, the following areas offer promising directions for forward-looking research:

- Adaptation for data-scarce or small-population contexts: Many low- and middle-income countries have short time-series or incomplete age-specific mortality data. Research into robust estimation in such contexts (e.g., hierarchical Lee–Carter frameworks, pooling across populations) is needed.
- Time-varying age-sensitivity coefficients $b_{x,t}$: Since mortality decline patterns differ by age over time, models allowing $b_x$ to vary with $t$ may improve long-term forecasts.
- Hybrid modelling with machine-learning techniques: Integrating the interpretability of the Lee–Carter model with flexible machine-learning approaches (e.g., ensemble methods, neural networks) may enhance forecasting in non-standard mortality regimes (Novkaniza et al., 2024; hybrid studies).
- Structural break and pandemic effects: With events like COVID-19 altering mortality patterns, extensions of the Lee–Carter model to accommodate jumps or changing drift in $k_t$ merit further work (see Bayesian jump modelling).
- Improved uncertainty quantification: While forecasting intervals are often produced, further work is needed on long-term uncertainty, parameter estimation error, and model selection uncertainty, especially for policy relevant horizons (50+ years).
- Application to cause-specific mortality: While the model is typically used for all-cause mortality, applying and adapting it for cause-specific death rates may improve understanding of epidemiological transition dynamics.

## Conclusion

The Lee–Carter model remains one of the most influential and widely applied approaches for mortality forecasting. Its decomposition into age-, time- and sensitivity-components provides transparency, interpretability and a clear forecasting pipeline. Accurate estimation of its parameters $a_x$, $b_x$ and $k_t$ is central to its effectiveness. While the original SVD approach remains a solid starting point, practitioners must attend to data smoothing, identifiability constraints, residual diagnostics, and time-series modelling of $k_t$. Alternative approaches (e.g., GLM/MLE) and recent methodological advances (robust estimation, machine-learning hybrids) offer enhancements. For researchers and practitioners working in settings with limited data, changing mortality regimes or unique age-patterns, the recommended future directions offer fruitful paths. Ultimately, careful estimation, validation and transparent reporting of assumptions will enhance the reliability and policy-relevance of mortality forecasts.

## References

[1] Booth, H., Hyndman, R. J., Tickle, L., & de Jong, P. (2002). Lee–Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research*, 15(9), 289-314.

[2] Fazle Rabbi, A. M., & Mazzuco, S. (2020). Mortality forecasting with the Lee–Carter method: adjusting for smoothing and lifespan disparity. *European Journal of Population*, 37(1), 97-120.

[3] Guo, Y., & Li, J. S.-H. (2022). Robust parameter estimation for the Lee–Carter family: a probabilistic principal component approach.

[4] Koissi, M.-C., & Shapiro, A. F. (2008). The Lee–Carter model under the condition of variables age-specific parameters. *Proceedings of the 43rd Actuarial Research Conference*.

[5] Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*.

[6] Li, N., & Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. *Demography*, 42(3), 575-594. (as cited in more recent literature)

[7] Neves, C., et al. (2018). Five different distributions for the Lee–Carter model of mortality.
[8] "Mortality modelling and forecasting: a review of methods." (2011). *Annals of Actuarial Science*, (abstract online).
[9] Novkaniza, F., Athalla Syach, M., Novita, M. (2024). Modified Lee–Carter model using extreme value theory for forecasting mortality rates amidst extreme events during COVID-19 era in Indonesia. *Communications in Mathematical Biology and Neuroscience*, 2024:76.