Bayesian Mortality Modeling with Pandemics

A Vanishing Jump Approach

Based on Goes, J., Barigou, K., & Leucht, A. (2025)

Ziye Lin - ACTSC 991

Background

The LC model proposed by Lee and Carter (1992) has been successful in modeling mortality over time:

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t} \ , \ \varepsilon_{x,t} \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$$

Limitation: The LC model assumes that mortality rates evolve smoothly over time

- No sudden changes or shocks

Not a serious issue, (maybe) $\longrightarrow$ until COVID-19 hit the world.

A Prior Approach — Liu and Li (2015)

Motivation:
We need a model that can accommodate unexpected jumps in mortality rates, such as those caused by pandemics and wars.

A Prior Approach — Liu and Li (2015):
Liu and Li (2015) extended the classical LC model to include an extra jump term:

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + N_t J_{x,t} + \varepsilon_{x,t}$$

- α_x, β_x, κ_t, $\varepsilon_{x,t}$ have the same meanings as in the classical LC model
- N_t (jump occurrence) are i.i.d. Bernoulli random variables with parameter p denoting the probability of a mortality jump in a calendar year
- $J_{x,t}$ is the effect of a mortality jump occurred in year t on age group x

A Prior Approach — Liu and Li (2015)

Liu and Li (2015)'s Model:

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + N_t J_{x,t} + \varepsilon_{x,t}$$

Three variations were proposed for this model: $J0$, $J1$, $J2$.

The $J1$ variation (focus of Goes et al., 2025):

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \beta_x^{(J)} N_t Y_t + \varepsilon_{x,t} \ ,$$

- Y_t are i.i.d. normal random variables measuring the severity of the mortality jump at calendar year t
- $\beta_x^{(J)}$ are fixed quantities for the age patterns associated with the jump effects

A Prior Approach — Liu and Li (2015)

Liu and Li (2015)'s Model (J1):

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \beta_x^{(J)} N_t Y_t + \varepsilon_{x,t} \ ,$$

Weaknesses:

- Different jumps have the same age patterns
- Yearly jumps are independent

New Methodology: Serial Dependence Structures

Goes et al. (2025) proposed the following formulation:

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \beta_x^{(J)} J_t + \varepsilon_{x,t} \ ,$$
$$\kappa_t = \kappa_{t-1} + d + \zeta_t \ , \ \zeta_t \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_\zeta^2) \ .$$

To capture the serial dependence of J_t, Goes et al. proposed two options:

1. Autoregressive structure (AR model)
2. Moving average structure (MA model)

Option 1: Autoregressive Structure

The AR model:

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \beta_x^{(J)} J_t + \varepsilon_{x,t} \ ,$$

where

$$J_t = a J_{t-1} + N_t Y_t$$

N_t, Y_t have the same meanings as in the model of Liu and Li (2015).

The serial dependence is controlled by $a \in [0, 1)$.

Option 2: Moving Average Structure

The MA model (with order $\mathcal{Q} = 1$):

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \beta_x^{(J)} J_t + \varepsilon_{x,t} \ ,$$

where

$$J_t = N_t Y_t + b N_{t-1} Y_{t-1}$$

N_t, Y_t have the same meanings as in the model of Liu and Li (2015).

The serial dependence is controlled by $b \in [0, 1)$.

Two Options for Serial Dependence of J_t

The AR model:

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \beta_x^{(J)} J_t + \varepsilon_{x,t} \ ,$$
$$J_t = a J_{t-1} + N_t Y_t \ , \ a \in [0,1) \ .$$

The MA model (with order $\mathcal{Q} = 1$):

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \beta_x^{(J)} J_t + \varepsilon_{x,t} \ ,$$
$$J_t = N_t Y_t + b N_{t-1} Y_{t-1} \ , \ b \in [0,1) \ .$$

Estimation Procedure

Estimation of mortality models generally has two routes (Haberman and Renshaw, 2012):

1. Based on the central death rates: $\ln(m_{x,t})$ — *Route I*
2. Based on the mortality improvement: $Z_{x,t} := \ln(m_{x,t+1}) - \ln(m_{x,t})$ — *Route II*

Mitchel et al. (2013) and Wong et al. (2023) provided some discussion on and comparison between *Route I* and *Route II*. They concluded that *Route II* are generally better than *Route I*.

Goes et al. (2025) proceed with *Route II*:

Model the mortality improvements directly: $Z_{x,t} := \ln(m_{x,t+1}) - \ln(m_{x,t})$

Estimation: Route II

Mortality Improvement: $Z_{x,t} := \ln(m_{x,t+1}) - \ln(m_{x,t})$

By Goes et al.'s formulation,

$$Z_{x,t} = \beta_x(\kappa_{t+1} - \kappa_t) + \beta_x^{(J)}(J_{t+1} - J_t) + \vartheta_{x,t} \ ,$$

where $\vartheta_{x,t} = \varepsilon_{x,t+1} - \varepsilon_{x,t} \overset{i.i.d.}{\sim} \mathcal{N}(0, 2\sigma_\varepsilon^2)$.

It can be further rewritten as

$$\begin{aligned} Z_{x,t} &= \beta_x \Delta\kappa_{t+1} + \beta_x^{(J)} \Delta J_{t+1} + \vartheta_{x,} \\ &= \beta_x(d + \zeta_{t+1}) + \beta_x^{(J)} \Delta J_{t+1} + \vartheta_{x,} \ , \end{aligned}$$

where $\Delta J_t = J_{t+1} - J_t$, $\Delta\kappa_t = \kappa_{t+1} - \kappa_t$.

Identifiability Constraints

Similarly to a standard LC model, Goes et al.'s model suffers from nonidentifiability issues.

Let $\boldsymbol{B}_x = \left(\beta_x, \beta_x^{(J)}\right)^T$, $\boldsymbol{K}_t = (\kappa_t, J_t)^T$, then

$$\begin{aligned}\ln(m_{x,t}) &= \alpha_x + \beta_x \kappa_t + \beta_x^{(J)} J_t + \varepsilon_{x,t} \\ &= \alpha_x + \boldsymbol{B}_x^T \boldsymbol{K}_t + \varepsilon_{x,t} \ .\end{aligned}$$

Let there be an invertible matrix $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$ and a matrix $\boldsymbol{D} \in \mathbb{R}^{2\times 1}$, the following sets of parameters will result in the same fitted mortality rates (Hunt and Blake, 2020):

$$\begin{aligned}\left\{\tilde{\alpha},\ \tilde{\boldsymbol{B}}_x,\ \tilde{\boldsymbol{K}}_t\right\} &= \left\{\alpha_x,\ \boldsymbol{A}^{-1}\boldsymbol{B}_x,\ \boldsymbol{A}\boldsymbol{K}_t\right\} \ ; \\ \left\{\tilde{\alpha},\ \tilde{\boldsymbol{B}}_x,\ \tilde{\boldsymbol{K}}_t\right\} &= \left\{\alpha_x - \boldsymbol{D}^T\boldsymbol{B}_x,\ \boldsymbol{B}_x,\ \boldsymbol{K}_t + \boldsymbol{D}\right\} \ .\end{aligned}$$

Since $\boldsymbol{A}$ is 2×2, $\boldsymbol{D}$ is 2×1, there are six free parameters in total.

Identifiability Constraints

Nonidentifiability:

$$\left\{\tilde{\alpha},\ \tilde{\boldsymbol{B}}_x,\ \tilde{\boldsymbol{K}}_t\right\} = \left\{\alpha_x,\ \boldsymbol{A}^{-1}\boldsymbol{B}_x,\ \boldsymbol{A}\boldsymbol{K}_t\right\}\ ;$$
$$\left\{\tilde{\alpha},\ \tilde{\boldsymbol{B}}_x,\ \tilde{\boldsymbol{K}}_t\right\} = \left\{\alpha_x - \boldsymbol{D}^T\boldsymbol{B}_x,\ \boldsymbol{B}_x,\ \boldsymbol{K}_t + \boldsymbol{D}\right\}\ .$$

However, recall that the mortality improvements are given by

$$Z_{x,t} = \beta_x(\kappa_{t+1} - \kappa_t) + \beta_x^{(J)}(J_{t+1} - J_t) + \vartheta_{x,t}\ .$$

There is no $\alpha_x \Rightarrow \boldsymbol{D} = \boldsymbol{0}_{2\times 1}$.

The amount of identifiability constraints reduced to four $\Rightarrow$ the four entries in $\boldsymbol{A}$.

Identifiability Constraints

To identify β_x, $\beta_x^{(J)}$, $\Delta\kappa_t$, ΔJ_t, and d, the following constraints are needed:

$$\sum_{x=1}^{A} \beta_x = 1 \;\; ; \;\; \sum_{x=1}^{A} \beta_x^{(J)} = 1 \;\; ; \Delta J_2 = 0 \;\; ; \;\; \zeta_2 = 0 \;\; .$$

However, for N_t, a (vanishing effect in AR model), and b (vanishing effect in MA model), additional constraints might be needed depending on the assumed structure for J_t. Further details and proofs given in Goes et al. (2025).

More discussion on identifiability constraints in LC-type models is provided in Hunt and Blake (2020).

Estimation: Priors (Summary)

$$Z_{x,t} = \beta_x(d + \xi_{t+1}) + \beta_x^{(J)}\Delta J_t + \vartheta_{x,t} \ .$$

To estimate the parameters from above:

- $\Delta\kappa_t$: a normal prior $\mathcal{N}(d, \xi_t)$
- Y_t : a half-normal prior $\mathcal{N}^+(\mu_Y, \sigma_Y^2)$
- N_t : a Bernoulli prior $Bern(p)$
- $(\beta_1, \ldots, \beta_A)$ and $(\beta_1^{(J)}, \ldots, \beta_A^{(J)})$: multivariate Dirichlet priors
- a (in AR model) and b (in MA model): truncated normal on $[0, 1)$ with mean 0 and standard deviation 0.4

To choose the hyperparameters:

- d (drift): a normal prior (to ensure that $P(d \neq 0) = 1$)
- p (jump probability): $Beta(1, 20)$ (a subjective choice)
- μ_Y (in the prior of Y_t): a half-normal prior (ensure positive value on Y_t)
- all standard deviations have half-normal priors

(Goes et al., 2025)

Estimation: NIMBLE Package in R

NIMBLE package in R is used by Goes et al. to estimate the parameters:

- Allows for the selection of multiple samplers, including the famous MCMC and Hamiltonian Monte Carlo (HMC)
- Can choose a different sampler for each parameter
- Assess the convergence of all model parameters via *the split-$\hat{R}$ statistic*, *bulk effect sample size (Bulk-ESS)*, and tail effective sample size (Tail-ESS), computed by `rstan` package in R (Stan Development Team, 2023; Vehtari et al., 2021)
- A list of samplers for each parameter is provided in Goes et al. (2025)

Mortality Forecasting

Using the definition of $Z_{x,t}$, the mortality rate at time $t+1$ can be calculated as

$$\ln(m_{x,t+1}) = \ln(m_{x,t}) + Z_{x,t} \ .$$

Suppose we want to predict mortality rates for the next H years starting from time $t = T$, we need to predict future improvements $Z_{x,T+1}, Z_{x,T+2}, \ldots, Z_{x,T+H}$, and apply the above equation recursively.

Mortality Forecasting: General Procedure

The procedure for predicting $Z_{x,T+1}, Z_{x,T+2}, \ldots, Z_{x,T+H}$ (with the AR model) is as follows.

for $s = 1, \ldots, S$:

Step 1: Generate new values of $N_{T+1}^{(s)}$ by first drawing a value of $p^{(s)}$ from the posterior distribution and then sampling $N_{T+1}^{(s)}$ from a Bernoulli distribution with parameter $p^{(s)}$.

Step 2: Generate new values of $J_{T+1}^{(s)}$. Start by drawing $\mu_Y^{(s)}$ and $\sigma_Y^{(s)}$ from the posterior distribution. Then sample a new value of $Y_{T+1}^{(s)}$ from a normal distribution with mean $\mu_Y^{(s)}$ and standard deviation $\sigma_Y^{(s)}$. Afterwards, draw $a^{(s)}$ and $J_T^{(s)}$ from the posterior distribution. Use the newly generated $N_{T+1}^{(s)}$ from Step 1 to compute a future value of $J_{T+1}^{(s)}$.

Step 3: Generate a new error term $\varepsilon_{x,T+1}^{(s)}$ by sampling from a normal distribution with mean 0 and standard deviation $\sigma_r^{(s)}$.

Step 4: Obtain the s-th posterior draw for the remaining parameters and substitute all values into equation (7) to generate $Z_{x,T+1}^{(s)}$.

Figure: Goes et al. (2025)'s procedure for mortality forecasting

Repeated these steps for $T+1, \ldots, T+H$ to generate $\ln(m_{x,T+1}), \ldots, \ln(m_{x,T+H})$.

Key Empirical Findings

Model Comparison with Liu and Li (2015)'s model

Table 1. Sources of mortality data

Country	Year	Source
Counts of death		
Poland	1990–2023	Eurostat
Spain	1990–2023	Eurostat
U.S.	1990–2021	HMD
	2022–2023	CDC
Population estimate		
Poland	1990–2020	Eurostat
	2021–2023	Statisics Poland
Spain	1990–2023	Eurostat
U.S.	1990–2021	HMD
	2022–2023	US Census Bureau

Key Empirical Findings

Model Comparison with Liu and Li (2015)'s model

- Compare performance using Watanabe-Akaike information criterion (WAIC; Watanabe, 2010) and leave-one-out cross-validation (LOO-CV) on the deviance scale in Bayesian setting
- WAIC and LOO-CV calculated using the `loo` package in R
- Lower values of WAIC and LOO-CV deviance indicate better fit

Table 2. In-sample fit comparison of the Liu–Li and own models on COVID-19 data for multiple countries. Bold value denotes best of the column

Model	U.S.	Spain	Poland
WAIC			
AR	−1463.47	−1193.73	−1174.92
MA	**−1465.79**	**−1194.63**	**−1181.99**
Liu–Li	−1462.45	−1187.28	−1178.66
LOO-CV			
AR	−1455.42	−1188.21	−1166.10
MA	**−1458.00**	**−1188.66**	**−1176.04**
Liu–Li	−1453.39	−1180.85	−1170.29

Key Empirical Findings

Out-of-sample performance using world war data

Mortality Data: England and Wales from 1900 to 2000 from the HMD website

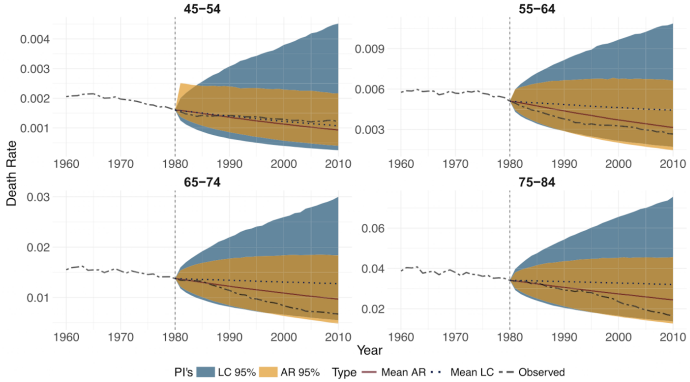

Figure 6. Thirty-year ahead forecasts of death rates for different age groups of England and Wales including prediction intervals (PI) using both a Lee–Carter (LC) and AR model. Dashed black line denotes actual observed values, while dotted line denotes the mean forecast (mean LC) of the LC model and the solid line the mean forecast (mean AR) of own model.

A Possible Multi-Population Extension

Goes et al. (2025) proposed a possible extension of their models for multi-population mortality modeling:

$$\ln(m_{x,t,c}) = \alpha_{x,c} + \beta_{x,c}\kappa_{t,c} + \beta^{(J)}_{x,c} J_{t,c} + \varepsilon_{x,t,c} \ , \ c = 1, \ldots, C \ .$$

The set of index $c \in \{1, 2, \ldots, C\}$ denotes the countries; $\boldsymbol{\kappa}_t = (\kappa_{t,1}, \ldots, \kappa_{t,C})^T$ can be modeled as a multivariate random walk with drift.

Similarly, the AR structure will be

$$J_{t,c} = a_c J_{t-1,c} + N_t Y_{t,c} \ .$$

Concluding Remark

"[The] shock of a pandemic's mortality can trigger a compensatory response. There is an argument that a pandemic accelerates the demise of those already in poor health...it is believed that many of those who die during a pandemic would have died anyway in the near future, resulting in a slight decrease in the mortality rates among survivors. This contradicts our assumption of a pandemic effect that slowly vanished over time, making our model unsuitable for this type of scenario."

(Goes et al., 2025; page 23)

"Mortality Dips" (?)

References I


Perry de Valpine, Daniel Turek, Christopher J. Paciorek, Clifford Anderson-Bergman, Duncan Temple Lang, and Rastislav Bodik.
Programming with models: Writing statistical algorithms for general model structures with nimble.
Journal of Computational and Graphical Statistics, 26(2):403–413, 2017.


Julius Goes, Karim Barigou, and Anne Leucht.
Bayesian mortality modelling with pandemics: a vanishing jump approach.
Journal of the Royal Statistical Society Series C: Applied Statistics, 2025;, qlaf018.


Steven Haberman and Arthur E. Renshaw.
Parametric mortality improvement rate modelling and projecting.
Insurance: Mathematics and Economics, 50(3):309–333, 2012.

References II

Andrew Hunt and David Blake.
Identifiability in age/period mortality models.
Annals of Actuarial Science, 14(2):461–499, 2020.

Ronald D. Lee and Lawrence R. Carter.
Modeling and forecasting U.S. mortality.
Journal of the American Statistical Association, 87(419):659–671, 1992.

Yanxin Liu and Johnny Siu-Hang Li.
The age pattern of transitory mortality jumps and its impact on the pricing of catastrophic mortality bonds.
Insurance: Mathematics and Economics, 64:135–150, 2015.

References III

Daniel Mitchell, Patrick Brockett, Rafael Mendoza-Arriaga, and Kumar Muthuraman.

Modeling and forecasting mortality rates.
Insurance: Mathematics and Economics, 52(2):275–285, 2013.

Stan Development Team.
Stan modeling language user's guide and reference manual, version 2.32.
`https://mc-stan.org`, 2025.
Accessed July 23, 2025.

Daniel Turek, Perry deValpine, and Christopher J. Paciorek.
nimbleHMC: Hamiltonian Monte Carlo and Other Gradient-Based MCMC Sampling Algorithms for nimble, 2024.
R package version 0.2.3.

References IV


Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner.
Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of mcmc.
Bayesian Analysis, 16(2):667–718, 2021.


Sumio Watanabe.
Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory.
Journal of Machine Learning Research, 11(116):3571–3594, 2010.


Jackie S T Wong, Jonathan J Forster, and Peter W F Smith.
Bayesian model comparison for mortality forecasting.
Journal of the Royal Statistical Society Series C: Applied Statistics, 72(3):566–586, 03 2023.