Journal of the Royal Statistical Society Series A: Statistics in Society, 2026, 00, 1–23
https://doi.org/10.1093/jrsssa/qnag089
Published: 15 July 2026
Original article

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society Series A Statistics in Society
OXFORD

Bayesian mortality forecasting with a Conway–Maxwell–Poisson specification

Jackie Siaw Tze Wong[1], **Emiliano A. Valdez**[2]

[1]School of Mathematics, Statistics, and Actuarial Science, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK
[2]Department of Mathematics, University of Connecticut, Storrs, CT 06269-1009, USA
Address for correspondence: Jackie Siaw Tze Wong, School of Mathematics, Statistics, and Actuarial Science, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK. Email: jw19203@essex.ac.uk

Abstract

This paper presents a novel approach to stochastic mortality modelling by using the Conway–Maxwell–Poisson (CMP) distribution to model death counts within the Lee–Carter framework, both with and without cohort effects. Unlike standard Poisson or negative binomial distributions, the CMP is a more adaptable choice because it can account for different levels of variability in the data, a feature known as dispersion. Specifically, it can handle data that are underdispersed (less variable than expected), equidispersed (as variable as expected), and overdispersed (more variable than expected). We develop a Bayesian formulation that treats the dispersion level as an unknown parameter, using a Gamma prior to enable a robust and coherent integration of the parameter, process, and distributional uncertainty. The model is calibrated using Markov chain Monte Carlo (MCMC) methods, with model performance evaluated using standard statistical criteria such as residual analysis and scoring rules. An empirical study using England and Wales male mortality data shows that our CMP-based models provide a better fit for both existing data and future predictions compared to traditional Poisson and negative binomial models, particularly when the data exhibit overdispersion. Finally, we conduct a sensitivity analysis with respect to prior specification to assess robustness.

Keywords Bayesian estimation, Conway–Maxwell–Poisson (CMP), Lee–Carter, mortality forecasting, overdispersion, uncertainty calibration

1 Introduction

Mortality forecasting is vital for demographers, actuaries, and policymakers because it provides insights into future population dynamics, the sustainability of retirement and social security systems, and the solvency of insurance companies. A major research advance came with the Lee–Carter (LC) model (Lee & Carter, 1992), which introduced a simple yet powerful stochastic framework to decompose and extrapolate age-specific death rates. Basellini et al. (2023) and Booth and Tickle (2008) provide comprehensive surveys of subsequent developments. While foundational models such as LC often assume that death counts follow simple distributions, like the Poisson, an important area of modern research focuses on properly modelling dispersion, the unexpected noise in the death data that the main model structure does not typically capture. It is crucial to account for overdispersion, and failure to accurately model this extra variability results in model's prediction intervals that are

Received: January 1, 2026. **Revised:** May 22, 2026. **Accepted:** July 1, 2026
© The Royal Statistical Society 2026.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

too narrow. This, in turn, leads to an underestimation of longevity risk and may result in significant financial consequences.

There is a breadth of literature that investigates the incorporation of overdispersion in mortality/longevity modelling. Renshaw and Haberman (2005) proposed the use of the quasi-Poisson likelihood approach to quantify overdispersion by introducing a single variance parameter, increasing the flexibility of their model specification, leading to several options of error distribution. Delwarde et al. (2007) presented a mixture of Poisson and gamma distributions that results in a negative binomial LC model to account for overdispersion. In similar settings, Li et al. (2009) introduced an age-specific latent variable to represent mortality variations, which upon marginalization, also leads to the negative-binomial LC model. Wong et al. (2018) presented two variants of the overdispersion models, including negative-binomial and Poisson log-normal specifications, to quantify the overall heterogeneity in mortality data. More recently, Rakhmawan et al. (2024) incorporate the Conway–Maxwell–Poisson (CMP) distribution into a Generalized Autoregressive Score framework, enabling dynamic modelling of both the mean and dispersion via score-driven updates.

In this paper, we present three main contributions. First, we introduce the use of CMP distribution as a flexible approach for modelling death counts in the presence of dispersion. Originally proposed in Conway and Maxwell (1962), the CMP model allows us to capture under-, equi-, and over-dispersion in mortality data and provides us an opportunity to illustrate its versatility and practical relevance for demographic and actuarial applications.

Secondly, we develop a Bayesian formulation of the proposed model in which the dispersion is treated as an unknown but estimable parameter and assigned a Gamma prior. This specification enables explicit inference on both the nature and level of dispersion, while coherently integrating multiple sources of uncertainty within a unified probabilistic framework. The advantages of the Bayesian approach are well documented in the literature, such as Cairns et al. (2011) and Wong et al. (2018). It is worth noting that Bayesian framework allows us to avoid the two-stage estimation procedure which is prevalent in frequentist mortality modelling framework (see for example, Czado et al., 2005; Wong et al., 2018). In addition, there is an emerging body of recent work applying Bayesian methods in mortality and longevity modelling, e.g. Diana et al. (2025), Goes et al. (2025), Shi et al. (2024). Bayesian implementation of the CMP distribution has not, to the best of our knowledge, previously been applied in the context of mortality modelling.

Finally, we demonstrate the proposed methodology using England and Wales male death data. The empirical results are encouraging and show clear improvements over the conventional Poisson and negative binomial specifications.

Although Rakhmawan et al. (2024) also consider the use of CMP distribution for modelling death counts, there are several key distinctions between our approaches. Firstly, we embed the CMP model with the LC rate model, both with and without cohort effects, where we demonstrate that the inclusion of cohort parameters further enhances the calibration of dispersion. Moreover, we propose an alternative parametrization of the CMP model that exhibits a mean-matching property tailored to mortality analysis (which will be detailed in Section 2.2). Finally, we develop a novel Bayesian implementation of the CMP modelling framework, simultaneously facilitating data-driven estimation of the dispersion parameter and coherent integration of various sources of uncertainty.

The rest of the paper is organized as follows. The remainder of Section 1 introduces the data, the notation, and the Poisson specification for modelling death counts. Section 2 presents the CMP distribution as a flexible alternative to handle overdispersion. Section 3 outlines the Bayesian implementation, including the prior specification and the corresponding estimation algorithm. Empirical results based on England and Wales male mortality data are reported in Section 4. Finally, we conclude with a sensitivity analysis on the prior specification and some closing remarks.

1.1 Data and notation

Let d_{xt} denote the observed number of deaths for age group x in year t, where $x = \{x_1, \ldots, x_A\}$ and $t = \{t_1, \ldots, t_T\}$ represent a set of A different age groups and T different years, respectively. Suppose

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

further that e_{xt} and μ_{xt} respectively denote the corresponding central exposure to risk and central mortality rate.

The data chosen for illustrative purposes are male death data and the corresponding exposures of England and Wales, extracted from the Human Mortality Database (HMD).[1] They are classified by single year of age from 0 to 99, and years ranging from 1961 to 2002. Hence, here we have $\{x_1, \ldots, x_A\} = \{0, \ldots, 99\}$ and $\{t_1, \ldots, t_T\} = \{1961, \ldots, 2002\}$ with $A = 100$ and $T = 42$. The data for years 2003-2021 were reserved as a validation set (see Section 4).

1.2 Poisson distribution for modelling death counts

The Poisson distribution is commonly employed to model death counts (see, for example, McCullagh, 2019 or Selvin, 2004). More recent applications include Hayat and Higgins (2014), Tawiah et al. (2020), and Zhang and Ma (2023) etc. Let D_{xt} be the random variable denoting the number of deaths of age x year t, then

$$D_{xt}|\mu_{xt} \sim \text{Poisson}(e_{xt}\mu_{xt}), \tag{1}$$

with the corresponding probability mass function (PMF)

$$\mathbb{P}[D_{xt} = d_{xt}|\mu_{xt}] = \frac{(e_{xt}\mu_{xt})^{d_{xt}} \exp(-e_{xt}\mu_{xt})}{d_{xt}!} \propto \mu_{xt}^{d_{xt}} \exp(-e_{xt}\mu_{xt}). \tag{2}$$

There are several structural assumptions for μ_{xt}, each defining different mortality rate models in the literature. For example, specifying the popular Lee–Carter model (Lee & Carter, 1992) gives:

$$\begin{cases} D_{xt}|\mu_{xt} \sim \text{Poisson}(e_{xt}\mu_{xt}), \\ \log \mu_{xt} = \alpha_x + \beta_x \kappa_t, \end{cases} \tag{3}$$

where the following constraints are adopted for model identifiability,

$$\sum_x \beta_x = 1 \quad \text{and} \quad \sum_t \kappa_t = 0. \tag{4}$$

This yields the **Poisson Lee–Carter (Poisson–LC)** as proposed by Brouhns et al. (2002). The parameters can be interpreted as follows:

- α_x : is the average of the logarithm of death rates over time (i.e. $\alpha_x = \frac{\sum_t \log \mu_{xt}}{T}$).
- β_x : is the age-specific pattern of mortality improvement, quantifying the sensitivity of mortality at each age to overall changes.
- κ_t : represents the overall temporal trend in mortality, appropriately modulated by β_x.

In this paper, we also consider the inclusion of cohort effects for comparison purposes. Following the formulation of Renshaw and Haberman (2005),

$$\begin{cases} D_{xt}|\mu_{xt} \sim \text{Poisson}(e_{xt}\mu_{xt}), \\ \log \mu_{xt} = \alpha_x + \beta_x \kappa_t + \gamma_c, \end{cases} \tag{5}$$

where γ_c are cohort parameters with $c = t - x \in \{1, \ldots, C\}$ denoting the cohort index, which represent cohorts born in $\{1861, \ldots, 2001\}$, and $C = A + T - 1$. We will follow the suggestion of Wong et al. (2023) in imposing, together with (4), the following set of constraints[2]:

$$\sum_c \gamma_c = \sum_c c\gamma_c = \sum_c c^2\gamma_c = 0, \tag{6}$$

We refer to this model as the **Poisson Lee–Carter with cohorts (Poisson–LCC)** model.

[1] See http://www.mortality.org.

[2] We note that the constraints are different from both Renshaw and Haberman (2005) and Hunt and Villegas (2015), but this choice leads to stable computations with good convergence properties.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

1.3 Overdispersion

The Poisson model imposes the mean-variance equality ($\mathbb{E}[D_{xt}] = \text{Var}[D_{xt}] = e_{xt}\mu_{xt}$), a restrictive assumption that rarely holds in practice. It implies homogeneity within each age-period cell, i.e. individuals of the same age and year are assumed to share identical mortality risks. This is unrealistic, as mortality is also influenced by factors such as smoking, income, ethnicity, and genetic background Brown (2003), leading to additional variation across individuals, commonly termed overdispersion.

The issue of overdispersion due to extra-Poisson variability has been noted by several articles, e.g. Brillinger (1986), Breslow (1984), Hinde and Demétrio (1998), Hougaard et al. (1997), Dean and and (1989), and Dean (1992). Here, we illustrate this phenomenon using the square of Pearson residuals given as:

$$r_{xt}^2 = \frac{(d_{xt} - \mathbb{E}[D_{xt}])^2}{\text{Var}[D_{xt}]}\bigg|_{\mu_{xt}=\hat{\mu}_{xt}} = \frac{(d_{xt} - e_{xt}\hat{\mu}_{xt})^2}{e_{xt}\hat{\mu}_{xt}}, \tag{7}$$

where $\hat{\mu}_{xt}$ is the maximum likelihood estimate (MLE) of the underlying mortality rate. Specifically,

$$\hat{\mu}_{xt} = \begin{cases} \exp\left(\hat{\alpha}_x + \hat{\beta}_x\hat{\kappa}_t\right), & \text{for Poisson–LC,} \\ \exp\left(\hat{\alpha}_x + \hat{\beta}_x\hat{\kappa}_t + \hat{\gamma}_c\right), & \text{for Poisson–LCC.} \end{cases}$$

where $\hat{\alpha}_x$, $\hat{\beta}_x$, $\hat{\kappa}_t$, $\hat{\gamma}_c$ are the MLE of the parameters. A colour-coded heat map of r_{xt}^2 can then be constructed to visualize the lack of fit of both the Poisson–LC and Poisson–LCC models for our mortality data, as depicted in Figure 1.

Under the null hypothesis that the Poisson–LC model is the true underlying model (and some mild conditions), each r_{xt}^2 has an approximate chi-squared distribution with degrees of freedom one (χ_1^2) asymptotically. Ideally, we should expect only around 5% of the rectangular cells ($AT \times 0.05 = 210$) to have poor fit (which we define as $r_{xt}^2 > 3.84$, where 3.84 is the 95th precentile of χ_1^2). However, it is evident from Figure 1 that the heat map for the Poisson–LC model is scattered with more than the expected number of orange/red cells (1128, about 27%), and is especially obvious for infants and cohorts born around the year 1920 (diagonal lines of orange/red cells), suggesting the model inadequacy in accounting for additional variations in the data. Even after the inclusion of cohort effects as in the Poisson–LCC model, the heat map remains scattered with orange/red cells (474, about 11%), indicating a poor overall fit.

Additionally, we can perform the Pearson's chi-squared overall goodness of fit test. In particular, the model deviance computed as the sum of r_{xt}^2,

$$r^2 = \sum_{x,t} r_{xt}^2 = \sum_{x,t} \frac{(d_{xt} - e_{xt}\hat{\mu}_{xt})^2}{e_{xt}\hat{\mu}_{xt}},$$

giving values of 16,709.85 and 6,628.97 for the Poisson–LC and Poisson–LCC models respectively. Again, under the null hypothesis that this model is a good fit to the data, r^2 should follow an approximate chi-squared distribution with degrees of freedom (df) given as $df = (A-1)(T-2) = 3,960$ (see Renshaw & Haberman, 2005). Since the model deviances of 16,709.85 and 6,628.97 are substantially larger than the critical value of the conventional chi-squared statistics (i.e. the 95th percentile of χ_{df}^2 is 4,107.51), this clearly suggests that both Poisson–LC and Poisson–LCC models do not provide satisfactory fit to the data.

As noted by Wong et al. (2018), ignoring overdispersion leads to underestimated variance, causing overfitting and yielding overly narrow forecasts. Allowing for overdispersion produces more flexible fits, better captures heterogeneity, and generates more reliable prediction intervals (see Section 4.2). Moreover, incorporating cohort effects further improves calibration between data and uncertainty, reducing projection biases while maintaining appropriate interval widths. In the following sections, we demonstrate how the CMP distribution addresses the extra-Poisson variation in death counts.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

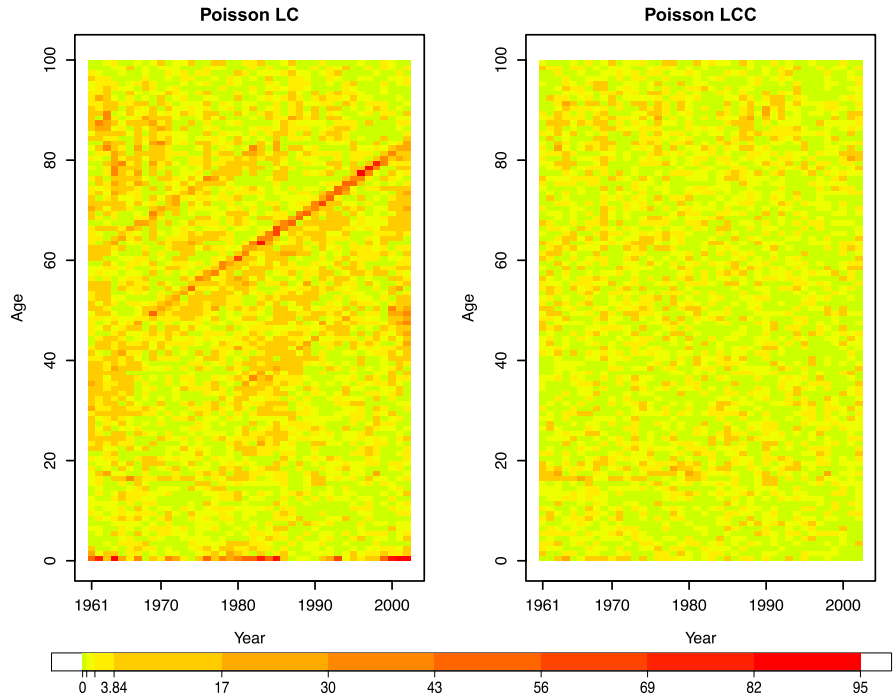

Figure 1. Heat maps of r_{xt}^2 for the Poisson–LC (left) and Poisson–LCC (right) models. Green/yellow rectangular cells indicate areas with good fit; while orange/red-colored cells indicate areas with significantly poor fit.

2 The Conway–Maxwell–Poisson (CMP) distribution

The CMP distribution was originally proposed by Conway and Maxwell (1962) as a generalization of the Poisson distribution. The PMF of a random variable Y following a CMP distribution is given by:

$$\mathbb{P}[Y = y|\lambda, \nu] = \frac{1}{Z(\lambda, \nu)} \frac{\lambda^y}{(y!)^\nu}, \quad \text{for } y \in \{0, 1, 2, \ldots\},$$

where

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$$

is the normalizing constant that ensures the PMF sums to 1. The model parameters satisfy $\lambda > 0$ and $\nu \geq 0$. This is commonly denoted by $Y \sim \text{CMP}\ (\lambda, \nu)$.

2.1 Some properties

The key strength of this distribution is its ability to model count data exhibiting both overdispersion and underdispersion through the dispersion parameter ν. We note that the moments of $Y \sim \text{CMP}\ (\lambda, \nu)$ are given by

$$\mathbb{E}[Y] = \sum_{j=0}^{\infty} \frac{j\lambda^j}{(j!)^\nu Z(\lambda, \nu)}, \quad \text{and} \quad \text{Var}[Y] = \sum_{j=0}^{\infty} \frac{j^2\lambda^j}{(j!)^\nu Z(\lambda, \nu)} - (\mathbb{E}[Y])^2.$$

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

Or more specifically, when ν and λ are not too small, we have

$$\mathbb{E}[Y] \approx \lambda^{1/\nu} + \frac{1}{2\nu} - \frac{1}{2} \quad \text{and} \quad \text{Var}[Y] \approx \frac{\lambda^{1/\nu}}{\nu}, \tag{8}$$

approximately (see for example Chanialidis et al., 2018). This allows for $\text{Var}[Y] > \mathbb{E}[Y]$ (overdispersion; when $\nu < 1$), $\text{Var}[Y] < \mathbb{E}[Y]$ (underdispersion; when $\nu > 1$) or $\text{Var}[Y] = \mathbb{E}[Y]$ (equidispersion; when $\nu = 1$) depending on the value of ν.

Also note that the ratio of subsequent probabilities

$$\frac{\mathbb{P}[Y = y-1|\lambda, \nu]}{\mathbb{P}[Y = y|\lambda, \nu]} = \frac{y^\nu}{\lambda},$$

indicating that the decay rate follows one of the cases below:

- $\nu < 1$: the rate of decay is slower than linear decrease, meaning the distribution has heavier/longer tail (overdispersion).
- $\nu = 1$: the rate of decay is linear, which becomes the ordinary Poisson distribution (equidispersion).
- $\nu > 1$: the rate of decay is quicker than linear decrease, meaning the distribution has lighter/shorter tail (underdispersion).
- $\nu = 0$: this is a special case where the CMP distribution reduces to a geometric distribution with probability of $1 - \lambda$ where $0 < \lambda < 1$.
- $\nu \to \infty$: this is a special case where the CMP distribution reduces to a Bernoulli distribution with probability of success $\frac{\lambda}{1+\lambda}$.

Computations involving the CMP distribution can be carried out using the R package *COMPoissonReg* (Sellers et al., 2023). Further details on the CMP distribution are provided in Daly and Gaunt (2016).

2.2 CMP distribution for modelling death counts

This issue of overdispersion in the Poisson assumption has been explored by Wong et al. (2018) in the context of mortality modelling/forecasting, where two extensions to the above specification were proposed. In this work, we present the use of the CMP distribution as an alternative, which introduces a general dispersion parameter to relax the stringent assumptions of the Poisson model. That is, we propose

$$D_{xt}|\mu_{xt}, \nu \sim \text{CMP}\left((e_{xt}\mu_{xt} + \frac{1}{2} - \frac{1}{2\nu})^\nu, \nu\right), \tag{9}$$

where ν is the dispersion parameter to be estimated. Suppose we write $\lambda_{xt} = \left(e_{xt}\mu_{xt} + \frac{1}{2} - \frac{1}{2\nu}\right)^\nu$ for simplicity, the corresponding PMF is

$$\mathbb{P}[D_{xt} = d_{xt}|\mu_{xt}, \nu] = \frac{1}{Z(\lambda_{xt}, \nu)} \frac{\lambda_{xt}^{d_{xt}}}{(d_{xt}!)^\nu},$$

where

$$Z(\lambda_{xt}, \nu) = \sum_{j=0}^{\infty} \frac{\lambda_{xt}^j}{(j!)^\nu}.$$

Note that this (somewhat unconventional) formulation is designed to ensure that the expected number of deaths coincides with the product of exposures and mortality rates, i.e.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

$$\begin{aligned}\mathbb{E}[D_{xt}] &\approx ((e_{xt}\mu_{xt} + \frac{1}{2} - \frac{1}{2\nu})^{\nu})^{1/\nu} + \frac{1}{2\nu} - \frac{1}{2} \\ &= e_{xt}\mu_{xt} + \frac{1}{2} - \frac{1}{2\nu} + \frac{1}{2\nu} - \frac{1}{2} \\ &= e_{xt}\mu_{xt},\end{aligned} \tag{10}$$

as implied from (8).

We remark that it may be sufficient to specify $D_{xt}|\mu_{xt}, \nu \sim \text{CMP}\left((e_{xt}\mu_{xt})^{\nu}, \nu\right)$ in order to achieve $\mathbb{E}[D_{xt}] \approx e_{xt}\mu_{xt}$, particularly since ν is typically not too small. We also note that Rakhmawan et al. (2024) employed an alternative formulation, specifying $\lambda_{xt} = e_{xt}\mu_{xt}$, and demonstrated its applicability within their framework. Nevertheless, we proceed with the formulation in (9) as it offers a more accurate adjustment (closer alignment with the conditional mean) without introducing significant additional complexity.

The specification also implies that the variance of the number of deaths, expressed in terms of mean, is given by

$$\begin{aligned}\text{Var}[D_{xt}] &\approx \frac{\left(\left(e_{xt}\mu_{xt} + \frac{1}{2} - \frac{1}{2\nu}\right)^{\nu}\right)^{1/\nu}}{\nu} \\ &= \frac{e_{xt}\mu_{xt} + \frac{1}{2} - \frac{1}{2\nu}}{\nu} \\ &= \frac{\mathbb{E}[D_{xt}] + \frac{1}{2} - \frac{1}{2\nu}}{\nu}.\end{aligned} \tag{11}$$

So ν governs the level of overdispersion, where the term $1/\nu$ can also be (approximately) interpreted as the multiplicative increase in the variance of D_{xt} relative to its mean. For instance, if $\nu < 1$, we have $\text{Var}[D_{xt}|\mu_{xt}, \nu] > \mathbb{E}[D_{xt}|\mu_{xt}, \nu]$ which is the case for overdispersion, and vice versa. And when $\nu = 1$, $\text{Var}[D_{xt}|\mu_{xt}, \nu] = \mathbb{E}[D_{xt}|\mu_{xt}, \nu]$ indicating the case of equidispersion.

Similar to before, we consider two structural models for μ_{xt},

$$\log \mu_{xt} = \begin{cases} \alpha_x + \beta_x \kappa_t, \\ \alpha_x + \beta_x \kappa_t + \gamma_c, \end{cases} \tag{12}$$

coupled with the use of (9). We term the former the **CMP–LC** model, and the latter as the **CMP–LCC** model.

Using the property of independence and the PMF above, we also observe that the full data likelihood is

$$f(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \nu) = \prod_{x,t} \mathbb{P}[D_{xt} = d_{xt}|\mu_{xt}, \nu] = \prod_{x,t} \frac{1}{Z(\lambda_{xt}, \nu)} \frac{(\lambda_{xt})^{d_{xt}}}{(d_{xt}!)^{\nu}}. \tag{13}$$

3 Bayesian estimation

The motivation for considering a Bayesian framework has been extensively discussed in the literature (e.g. Wong et al., 2018) . In our setting, this approach allows the dispersion parameter ν to be treated as an unknown random variable to be estimated naturally and directly from the data. This data-driven specification enables inference on the level of dispersion without the need to prespecify a fixed value of ν. This also forms a coherent framework for uncertainty calibration, including that due to overdispersion and other sources of uncertainty, such as parameter uncertainty through prior distributions. In the following subsections, we describe the prior specification, derive the joint posterior distributions, and outline our proposed MCMC sampling procedure.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

3.1 Prior specification

In principle, the specified prior distributions should reflect our prior knowledge about the model parameters. But in most applications, we do not have prior domain knowledge, so Bayesian methods allow us to learn about all parameters (including ν) from the data. In this paper, we will predominantly follow the specification suggested by Wong et al. (2023), except for ν where we explore various specifications.

First, we specify

$$\begin{aligned}
\boldsymbol{\alpha} &\sim N(-5 \cdot \mathbf{1}_A, 4 \cdot I_A), \\
\boldsymbol{\beta} &\sim N\left(\frac{1}{A} \cdot \mathbf{1}_{A-1}, 0.005 \cdot \left(I_{A-1} - \frac{1}{A} \cdot J_{A-1}\right)\right), \\
\beta_1 &= 1 - \sum_{x=2}^{A} \beta_x,
\end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_A)^\top$, $\boldsymbol{\beta} = (\beta_2, \ldots, \beta_A)^\top$, $\mathbf{1}_A$ is a length–A vector of ones, I_A and J_A are respectively the identity matrix and a matrix of ones with dimension $A \times A$.

Following the formulation of Czado et al. (2005), we impose for κ_t

$$\left.\begin{aligned}
&\kappa_t - \eta_t = \rho(\kappa_{t-1} - \eta_{t-1}) + \epsilon_t, \text{ for } t = 2, 3, \ldots, T, \\
&\kappa_1 = \eta_1 + \epsilon_1,
\end{aligned}\right\} \tag{14}$$

where $\eta_t = \psi_1 + \psi_2 t$ denotes the linear drift and $\epsilon_t \overset{\text{ind}}{\sim} N(0, \sigma_\kappa^2)$. Then, depending on the value of ρ, this forms either a first-order autoregressive (AR(1)) model ($|\rho| < 1$) or a random walk model ($\rho = 1$) for κ_t. The model in (14) can be expressed multivariately (along with the constraint) as

$$\left.\begin{aligned}
&\boldsymbol{\kappa} | \rho, \boldsymbol{\psi}, \sigma_\kappa^2 \sim N(\boldsymbol{\mu}_\kappa, \sigma_\kappa^2 V), \\
&\kappa_1 = -\sum\nolimits_{t=2}^{T} \kappa_t,
\end{aligned}\right\} \tag{15}$$

where $\boldsymbol{\kappa} = (\kappa_2, \ldots, \kappa_T)^\top$, $\boldsymbol{\mu}_\kappa = (I_{T-1} - B_{21} B_{11}^{-1} \mathbf{1}_{T-1}^\top) \cdot X\boldsymbol{\psi}$, $X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 2 & 3 & \cdots & T \end{pmatrix}^\top$, $\boldsymbol{\psi} = (\psi_1, \psi_2)^\top$, expressions for B and V are provided in online supplementary material, Appendix A. For other hyperparameters,

$$\begin{aligned}
\frac{\rho + 1}{2} &\sim \text{Beta}(3, 2), \text{ where } \rho \in (-1, 1), \\
\sigma_\kappa^{-2} &\sim \text{Gamma}(1, 0.0001), \\
\boldsymbol{\psi} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2000 & 0 \\ 0 & 2 \end{pmatrix}\right),
\end{aligned}$$

For the cohort parameters γ_c, we use an ARIMA(1,1,0) as adopted by Villegas et al. (2018), i.e.

$$\left.\begin{aligned}
&(\gamma_c - \gamma_{c-1}) = \rho_\gamma(\gamma_{c-1} - \gamma_{c-2}) + \epsilon_c^\gamma, \text{ for } c = 3, \ldots, C, \\
&\gamma_2 - \gamma_1 = \frac{1}{\sqrt{1 - \rho_\gamma^2}} \epsilon_2^\gamma, \\
&\gamma_1 = 100\epsilon_1^\gamma,
\end{aligned}\right\} \tag{16}$$

where $\epsilon_c^\gamma \overset{\text{ind}}{\sim} N(0, \sigma_\gamma^2)$ for $c = 1, \ldots, C$, and ρ_γ and σ_γ^2 are hyperparameters. Applying the constraints $\sum_c \gamma_c = \sum_c c\gamma_c = \sum_c c^2 \gamma_c = 0$ on (16), we write

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

$$\boldsymbol{\gamma} \sim N(\mathbf{0}, \sigma_\gamma^2 W_\gamma), \tag{17}$$

where $\boldsymbol{\gamma} = (\gamma_2, \ldots, \gamma_{71}, \gamma_{73}, \ldots, \gamma_{C-1})^\top$, and W_γ is provided in online supplementary material, Appendix A. The remaining cohort components $\{\gamma_1, \gamma_{72}, \gamma_C\}$ can then be derived as follows

$$\begin{aligned}
\gamma_1 &= \frac{1}{71 \times (C-1)} \sum_{c \neq 1,72,C} (c-72)(C-c)\gamma_c, \\
\gamma_{72} &= -\frac{1}{69 \times 71} \sum_{c \neq 1,72,C} (C-c)(c-1)\gamma_c, \\
\gamma_C &= \frac{1}{69 \times (C-1)} \sum_{c \neq 1,72,C} (c-1)(72-c)\gamma_c.
\end{aligned}$$

As recommended by Wong et al. (2023), we set priors

$$\rho_\gamma \sim N(0, 1) \quad \text{and} \quad \sigma_\gamma \sim \text{Uniform}(0.1).$$

For the dispersion parameter ν, our initial strategy is to specify (improper) flat prior, i.e.

$$\pi(\nu) \propto 1, \quad \text{for } \nu > 0 \tag{18}$$

so that results derived from our Bayesian analysis are directly comparable with those from a frequentist framework. This is in the spirit of data-driven inferences, where we allow the data to 'speak for itself'. This also serves as an indication if there is sufficient information in the data to estimate ν. We remark that the use of improper flat priors is common and convenient, which typically leads to proper posteriors, as pointed out by Gelfand and Sahu (1999).

It is also of interest to determine the influence of prior specification on the results. We could explore different choices of $\pi(\nu)$. We focus on priors of the form

$$\pi(\nu) \sim \text{Gamma}(a, b),$$

where a and b are hyperparameters to be specified as follows:

- $a = 1$ and $b = 100$.
- $a = 1$ and $b = 10$.
- $a = 1$ and $b = 1$.
- $a = 1$ and $b = 0.1$.
- $a = 1$ and $b = 0.01$.

From top to bottom, the prior variance of ν increases from 1×10^{-4} to 1×10^4, indicating that the priors become increasingly vague or non-informative. A sensitivity analysis with respect to these values of a and b will be presented later in Section 4.

3.2 Posterior distributions

First, denote $\theta = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\psi}, \sigma_\kappa^2, \rho, \nu\}$ as the full set of parameter for the CMP–LC model, the joint posterior distribution is then

$$\begin{aligned}
\pi(\theta|\boldsymbol{d}) &\propto f(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \nu) \cdot \pi(\theta) \\
&= f(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \nu) \cdot \pi(\boldsymbol{\alpha}) \cdot \pi(\boldsymbol{\beta}) \cdot \pi(\boldsymbol{\kappa}|\boldsymbol{\psi}, \sigma_\kappa^2, \rho) \cdot \pi(\boldsymbol{\psi}) \cdot \pi(\sigma_\kappa^2) \cdot \pi(\rho) \cdot \pi(\nu),
\end{aligned}$$

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

where $f(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \nu)$ is the data likelihood as given in (13), and the rest are prior distributions of all other parameters as described in Section 3.1. Similarly, for the CMP–LCC model, $\theta = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\psi}, \sigma_\kappa^2, \rho, \boldsymbol{\gamma}, \sigma_\gamma^2, \rho_\gamma, \nu\}$, giving the joint posterior

$$\pi(\theta|\boldsymbol{d}) \propto f(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \nu) \cdot \pi(\boldsymbol{\alpha}) \cdot \pi(\boldsymbol{\beta}) \cdot \pi(\boldsymbol{\kappa}|\boldsymbol{\psi}, \sigma_\kappa^2, \rho) \cdot \pi(\boldsymbol{\psi}) \cdot \pi(\sigma_\kappa^2) \cdot \pi(\rho) \cdot \pi(\boldsymbol{\gamma}|\sigma_\gamma^2, \rho_\gamma) \cdot \pi(\sigma_\gamma^2) \cdot \pi(\rho_\gamma) \cdot \pi(\nu).$$

3.3 The MCMC updating scheme

We apply the variable-at-a-time random walk Metropolis-Hastings (MH) algorithm as described in O'Hagan and Forster (2004). For a description of the MCMC updating schemes of all other parameters please refer to Appendix B in the online supplementary materials. Here we describe in detail the MCMC updating step for the dispersion parameter, ν.

Note that the conditional posterior distribution of ν is proportional to the likelihood function multiplied by their prior distributions, i.e.

$$\pi(\nu|\boldsymbol{d}, \theta_{-\nu}) \propto f(\boldsymbol{d}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \nu) \times \pi(\nu), \tag{19}$$

where $\theta_{-\nu} = \theta \backslash \{\nu\} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\psi}, \sigma_\kappa^2, \rho\}$ is the full set of parameters excluding ν. MCMC updating scheme is then applied iteratively to sample from (19), which can be combined with other posterior samples to form samples of the joint posterior distribution, $\pi(\theta|\boldsymbol{d})$.

To avoid proposing negative values, we perform the updating scheme on a log scale, i.e.

$$\log(\nu^*) \sim N(\log(\nu^{(i)}), \sigma_\nu^2),$$

where $\nu^{(i)}$ is the current iterate, σ_ν^2 is the proposal variance chosen so that the resulting acceptance rate is between 0.15-0.45 (Roberts & Rosenthal, 2001). From our pilot runs, setting $\sigma_\nu^2 = 0.005$ leads to an acceptance rate of about 0.40.

The acceptance ratio for ν^* is

$$\alpha(\nu^*|\theta_{-\nu}^{(i)}, \nu^{(i)}) = \min\left\{1, \frac{f(\boldsymbol{d}|\boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)}, \boldsymbol{\kappa}^{(i)}, \nu^*) \times \pi(\nu^*)}{f(\boldsymbol{d}|\boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)}, \boldsymbol{\kappa}^{(i)}, \nu^{(i)}) \times \pi(\nu^{(i)})}\right\}. \tag{20}$$

If the flat prior in (18) is used, then the expression in (20) reduces to simply a ratio of likelihoods.

3.4 Initialization and convergence diagnostics

For the initialization of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}$, we use the MLE obtained by iteratively fitting the Poisson–LCC model using the *glm* function in R. The initial values of σ_κ^2 and ρ are obtained by fitting an AR(1) with linear drift model on $\boldsymbol{\kappa}$ (using the *arima* function within the *forecast* package in R). The cohort-related parameters, σ_γ and ρ_γ, are initialized by fitting an ARIMA (1,1,0) model to the MLE of $\boldsymbol{\gamma}$. Finally, $\boldsymbol{\psi}$ is initialized as $(0, 0)^\top$, while the dispersion parameter is initialized as $\nu = 0.5$. Note that the initialization is proposed on the basis of values close to the MLE to speed up convergence, but it should have little effect on the estimation procedure.

We apply a burn-in phase of 1,000 iterations to mitigate the effect of initialization. In addition, we apply 50th posterior sample thinning (collecting one realization every 50 iterations) to each of the parameters to reduce the autocorrelations of these series. After discarding burn-in iterations and applying thinning, we obtain a sample of size 10,000 for each of the parameters under both the CMP–LC and CMP–LCC models.

Before making any inferential comparisons, trace plots and auto-correlation plots (see for example Lunn et al., 2013, Chapter 3) can be used as diagnostic tools for detecting anomalies in the MCMC

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

generated posterior samples. By referring to online supplementary material, Appendix C, the trace plots of some of the randomly selected parameters emerge as if convergence has been attained, with proper mixing and no apparent anomaly. The sample auto-correlations also appear to decay fairly quickly after applying thinning, except perhaps κ_t, which are relatively more correlated. A few commonly used convergence diagnostics such as those proposed by Gelman and Rubin (1992) and Geweke (1991), all of which are presented in the Supplementary Materials. Specifically, the Gelman's statistic, potential scale reduction factor (PSRF) computed from the posterior samples under both the CMP–LC and CMP–LCC are all very close to 1, indicating that there is no convergence issue. The Z statistics computed using the approach by Geweke (1991) also satisfy the convergence criterion.

In summary, the MCMC generated posterior samples seem to be well-behaved and, thus, are ready to be used to perform subsequent computations for accurate inferences to be drawn.

4 Results

When presenting our results, we also include, where appropriate, those obtained under the Poisson and negative binomial specifications for D_{xt} using Bayesian methods, for comparison purposes (see Czado et al., 2005 and Wong et al., 2018, respectively). The Poisson specification is as described in Section 1.2 and the negative binomial (NB) as follows,

$$D_{xt} \sim \text{Neg-Bin}\left(\phi, \frac{\phi}{e_{xt}\mu_{xt} + \phi}\right), \tag{21}$$

where ϕ is the overdispersion parameter. Under the NB specification, $\mathbb{E}[D_{xt}] = e_{xt}\mu_{xt}$ and $\text{Var}[D_{xt}] = \mathbb{E}[D_{xt}] \times (1 + \mathbb{E}[D_{xt}]/\phi) > \mathbb{E}[D_{xt}]$. The comparison is intended to highlight firstly the significance of accounting for overdispersion, and secondly the difference between two variants of incorporating overdispersion (i.e. between the CMP and NB distributions).

4.1 Estimated parameters

Figures 2 and 3 depict the fitted values (posterior medians) of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$, accompanied by the associated 95% credible intervals under respectively the LC and LCC models (for CMP, Poisson, and NB specifications). Evidently, the fitted values under these models are rather similar, with the overdispersion models producing slightly wider credible intervals. This is the general feature of models which account for overdispersion, where the responses (D_{xt}) are allowed to have more variabilities due to the extra flexibility offered by the model likelihood, permitting the parameters to be more volatile, and hence, the wider credible intervals. Some exceptions observed for $\boldsymbol{\beta}$ in the CMP–LC model, where the CMP specification yields credible intervals that are noticeably wider. $\boldsymbol{\kappa}$ for the NB–LC model also appear to demonstrate more linearity than the CMP–LC model.

The kernel posterior densities of other parameters for the LC and LCC models are illustrated in Figures 4 and 5. The posteriors are mostly similar between the CMP and Poisson specifications. Between the CMP and NB specifications, they are rather similar (visually) except for σ_κ^2, ψ_1, and ρ. In particular, the estimated σ_κ^2 are smaller, while the estimated ψ_1 and ρ are larger for the NB specification. Interestingly, the posteriors of ρ are uni-modal and have values close to one, contrary to the findings of Wong et al. (2018) for England and Wales female death data where ρ demonstrated bimodal posteriors.

The estimated mean dispersion parameter ν under the CMP–LC and CMP–LCC models are respectively around 0.237 and 0.578. Recall from (11) that $1/\nu$ measures the relative (multiplicative) increase in the variance of D_{xt} with respect to its mean. Substituting the posterior means into the expression gives $\frac{1}{\nu} \approx 4.22$[3] and $\frac{1}{\nu} \approx 1.73$, indicating that the level overdispersion implied by the CMP–LCC is lower than that by the CMP–LC model. This finding is consistent with that by Wong et al. (2023), where incorporating cohort effects allows for better calibration between data signals and uncertainties, helping to prevent misclassification of uncaptured data signals as overdispersion.

[3] This means that $\text{Var}[D_{xt}]$ is approximately 4.22 times $\mathbb{E}[D_{xt}]$.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

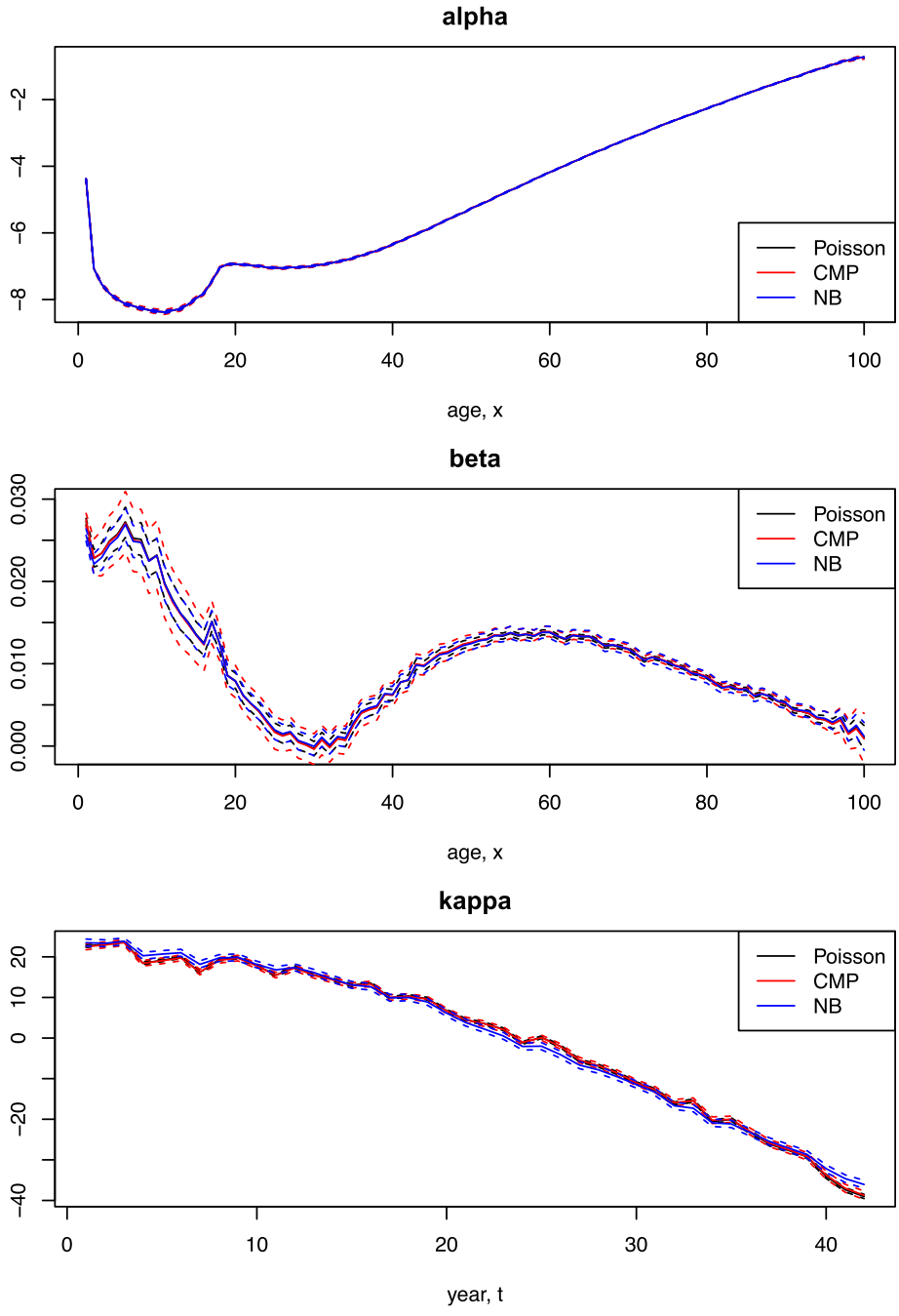

Figure 2. Plot of the medians (solid lines) and 95% credible intervals (dashed lines) of the estimated model parameters under the Poisson–LC, CMP–LC, and NB–LC models.

4.2 Fitted and projected crude mortality rates

We also assess the performance of the models using crude mortality rates (observable) rather than the underlying mortality rates, μ_{xt} (unobservable). To obtain the fitted crude mortality rates, the fitted number of deaths, D^F_{xt}, is first generated through Equation 9, where joint posterior samples of α_x,

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

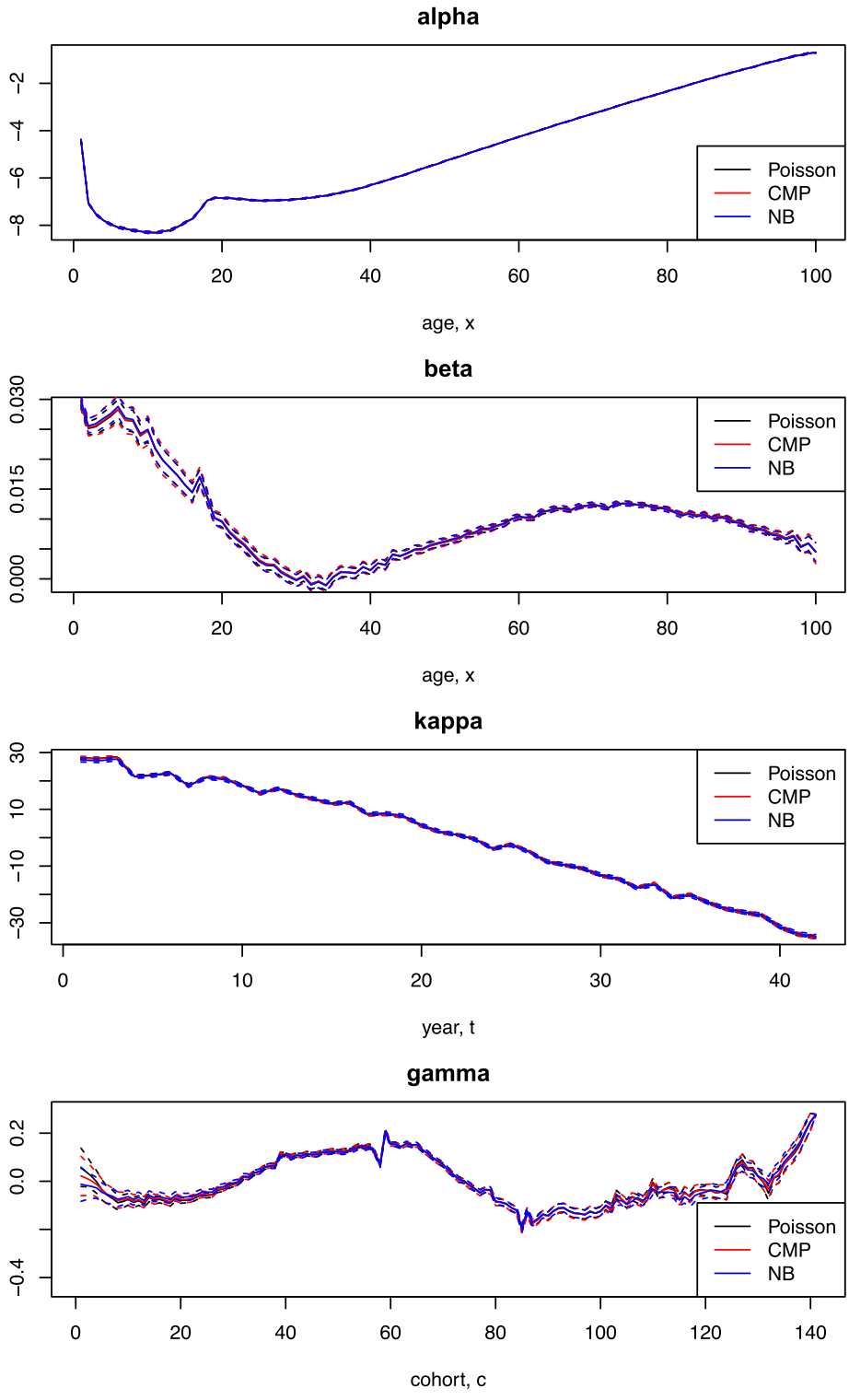

Figure 3. Plot of the medians (solid lines) and 95% credible intervals (dashed lines) of the estimated model parameters under the Poisson–LCC, CMP–LCC, and NB–LCC models.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

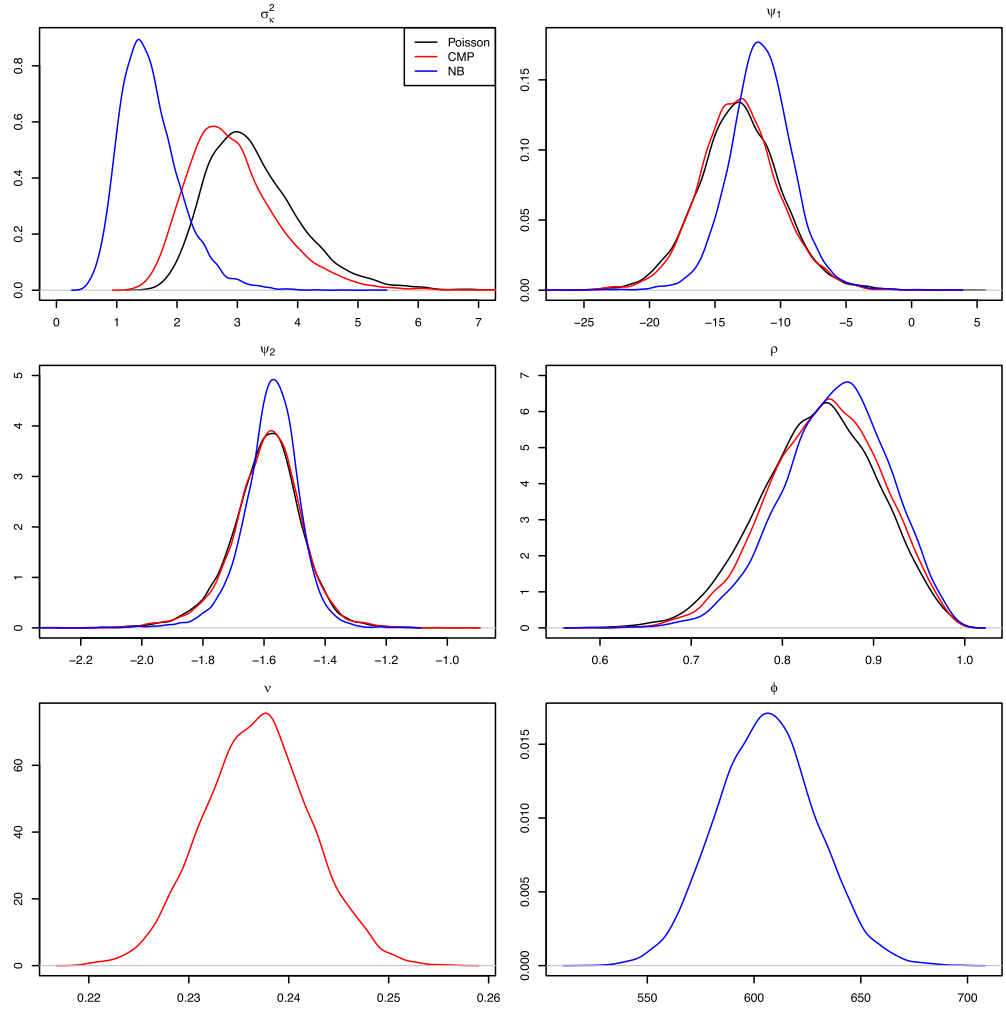

Figure 4. Density plots of estimated parameters for the LC models.

$\beta_x, \kappa_t, \gamma_c$, and ν are substituted as appropriate. The crude fitted mortality rate for each $x = 1, \ldots, A$ and $t = 1, \ldots, T$ is then computed as $\hat{\mu}_{xt} = \frac{D_{xt}^F}{e_{xt}}$.

For mortality projection, we note that the posterior predictive distribution of 1-year ahead number of deaths for each age (with age parameters held fixed), under the LCC model for instance, is

$$f(D_{x\,T+1}|\boldsymbol{d}) = \int f(D_{x\,T+1}|\alpha_x, \beta_x, \kappa_{T+1}, \gamma_{T+1-x}, \nu) \times f(\kappa_{T+1}|\kappa_T, \boldsymbol{\psi}, \rho, \sigma_\kappa^2) \times f(\gamma_{T+1-x}|\gamma_{T-x}, \gamma_{T-x-1}, \rho_\gamma, \sigma_\gamma^2) \times \pi(\alpha_x, \beta_x, \kappa_T, \boldsymbol{\psi}, \rho, \sigma_\kappa^2, \gamma_{T+1-x}, \rho_\gamma, \sigma_\gamma^2, \nu|\boldsymbol{d}) \mathrm{d}\alpha_x \ldots \mathrm{d}\phi, \tag{22}$$

where $\pi(\alpha_x, \beta_x, \kappa_T, \boldsymbol{\psi}, \rho, \sigma_\kappa^2, \gamma_{T+1-x}, \rho_\gamma, \sigma_\gamma^2, \nu|\boldsymbol{d})$ is the joint posterior distribution. This suggests the following generation procedure for the projected crude mortality rates:

(i) Generate κ_{T+1} from Equation 14 where joint posterior samples of $(\boldsymbol{\psi}, \rho, \kappa_T, \sigma_\kappa^2)$ are appropriately substituted.
(ii) Generate γ_T from Equation 16 where joint posterior samples of $(\gamma_{T-1}, \gamma_{T-2}, \rho_\gamma, \sigma_\gamma^2)$ are substituted. Other 'observed' cohort components $(\gamma_{T-99}, \ldots, \gamma_{T-3})$ are also available from posterior samples.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

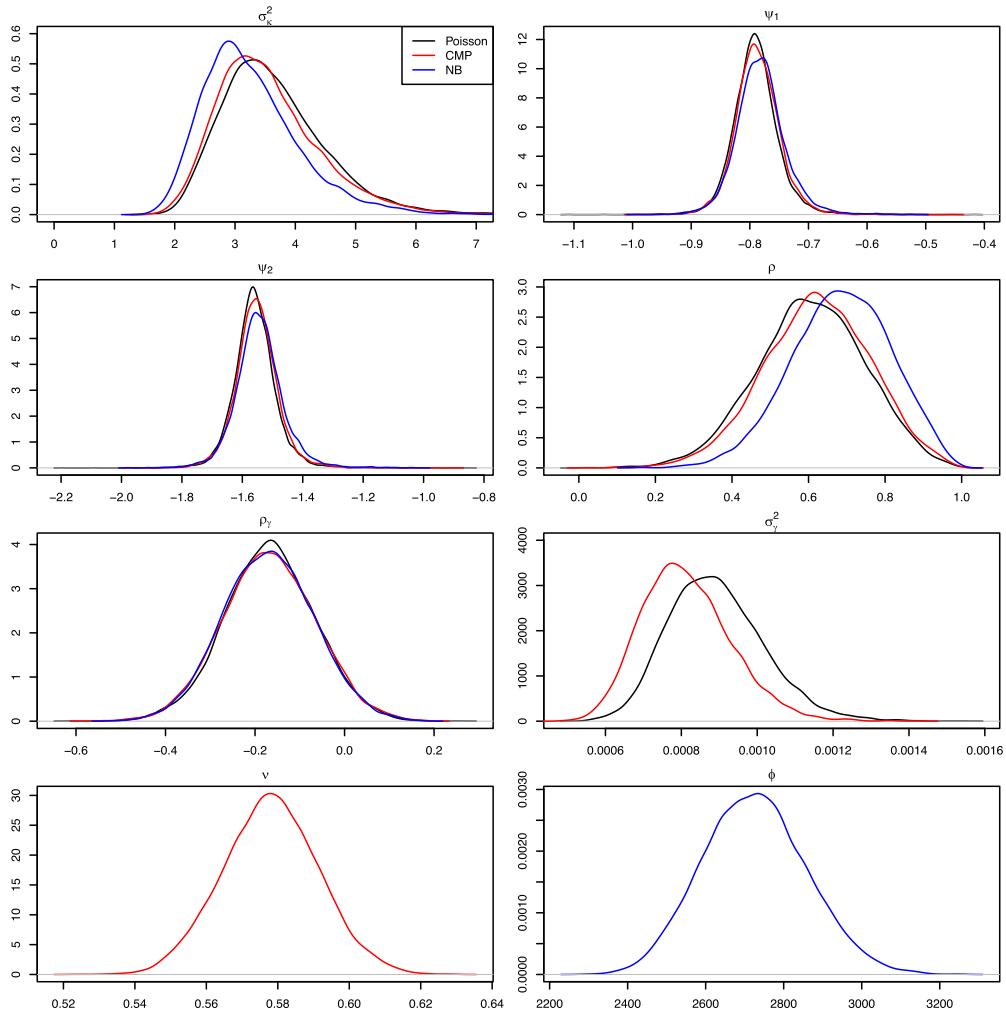

Figure 5. Density plots of estimated parameters for the LCC models.

(iii) Generate for each x the forecasted number of deaths using (9), i.e.

$$D^F_{xT+1} \sim \text{CMP}\left((e_{xT+1}\exp(\alpha_x + \beta_x\kappa_{T+1} + \gamma_{T+1-x}) + \frac{1}{2} - \frac{1}{2\nu})^\nu, \nu\right),$$

where e_{xT+1} are holdout central exposed to risks, joint posterior samples of (α_x, β_x, ν) are substituted, κ_{T+1} and γ_{T+1-x} are respectively from steps (i) and (ii).

(iv) Compute for each x the projected crude mortality rates as $\hat{\mu}_{xT+1} = \frac{D^F_{xT+1}}{e_{xT+1}}$.

By analogy, h-year ahead projections can be obtained by recursive implementation of the above algorithm. Having generated samples for $\hat{\mu}_{xt}$ for $x = 1, \ldots, A$ and $t = 1, \ldots, T + h$, sample percentiles can be constructed to form median and 95% intervals.

Under the Poisson and NB specifications, we generate from $D^F_{xt} \sim \text{Poisson}(e_{xt}\mu_{xt})$ and $D^F_{xt} \sim \text{Neg} - \text{Bin}\left(\phi, \frac{\phi}{e_{xt}\mu_{xt}+\phi}\right)$, using posterior samples of the relevant parameters, and then compute the fitted crude mortality rates as $\hat{\mu}_{xt} = \frac{D^F_{xt}}{e_{xt}}$.

Note that the above generation procedure of fitted and projected mortality rates builds in uncertainties due to random variations (Poisson, CMP, and NB), parameter, and forecast uncertainties.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

This allows us to better visualize the overall performances of each distributional specification for D_{xt}, taking into account various sources of uncertainties.

Figure 6 shows the fitted and projected crude death rates for several chosen ages when $h = 19$. Also included are the observed and holdout crude mortality rates. Generally speaking, both the CMP and

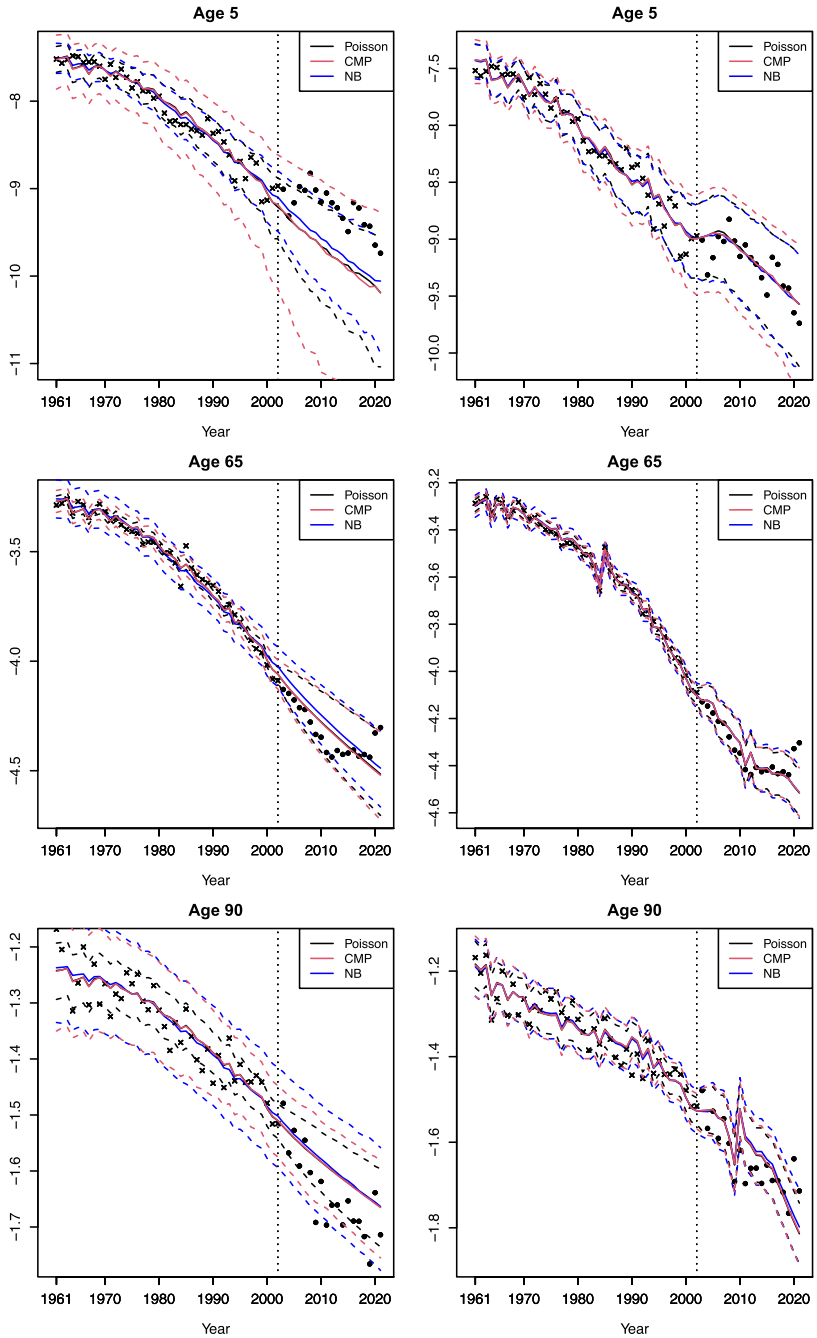

Figure 6. Plots of observed and holdout log crude death rates, fitted log crude death rates and the associated 19-year ahead projection of the crude log death rates for several chosen ages under the LC (left panels) and LCC (right panels) rate models, accompanied by 95% credible intervals in dashed lines.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

NB specifications yield mortality forecasts with good properties, with close adherence to the holdout crude rates and 95% credible intervals that have high coverage rates. By contrast, the Poisson specification yields credible intervals that are overly narrow, highlighting the importance of accounting for overdispersion. Comparing between the CMP and NB specifications, the CMP specification produces wider 95% credible intervals for earlier ages, while the reverse is true for the NB case where it generates wider intervals for older ages.

Incorporating cohort effects generally leads to a better description of the underlying mortality trend for both fitted and projected rates. In particular, the 95% prediction intervals of the cohort models (all specifications) offer realistic coverage rates for including the holdout rates, despite having narrower widths. Models without cohort effects generally fail to capture (and hence project) the mortality trends, leading to unrealistic median projections and unnecessarily wide intervals. Simultaneous inclusion of overdispersion and cohort components facilitates the calibration between signals and errors, reducing the bias in the mortality projection with adequately wide intervals to characterize uncertainty. This is consistent with the findings of Wong et al. (2023). Comparing the various specifications, again the Poisson specification gives low coverages due to overly narrow intervals, while CMP and NB specifications yield sufficiently wide intervals with high coverage. Note that the two most recent mortality rates were observed to be unusually high due to Covid-19 effects, which we do not address here.

4.3 Model assessments

Heatmaps of residuals by age and time are presented in Figures 7 and 8 to demonstrate in-sample goodness of fit. In general, both the CMP and NB specifications yield residuals that are well-behaved, indicating that the error distributions are well-calibrated, hence, model adequacy. By contrast, the Poisson specification clearly led to poor distribution of residuals, highlighting its inadequacy. Comparing the CMP and NB specifications, the NB appears to capture variations more effectively in ages 60–80 but struggles at younger ages, whereas the CMP performs better at younger ages. Overall, the residual patterns suggest that the CMP specification provides superior fit.

Table 1 reports the total model residuals r^2, as defined in Section 1.3, for assessing overall goodness of fit. Evidently, the CMP specification outperforms its counterparts by producing smaller overall deviance.

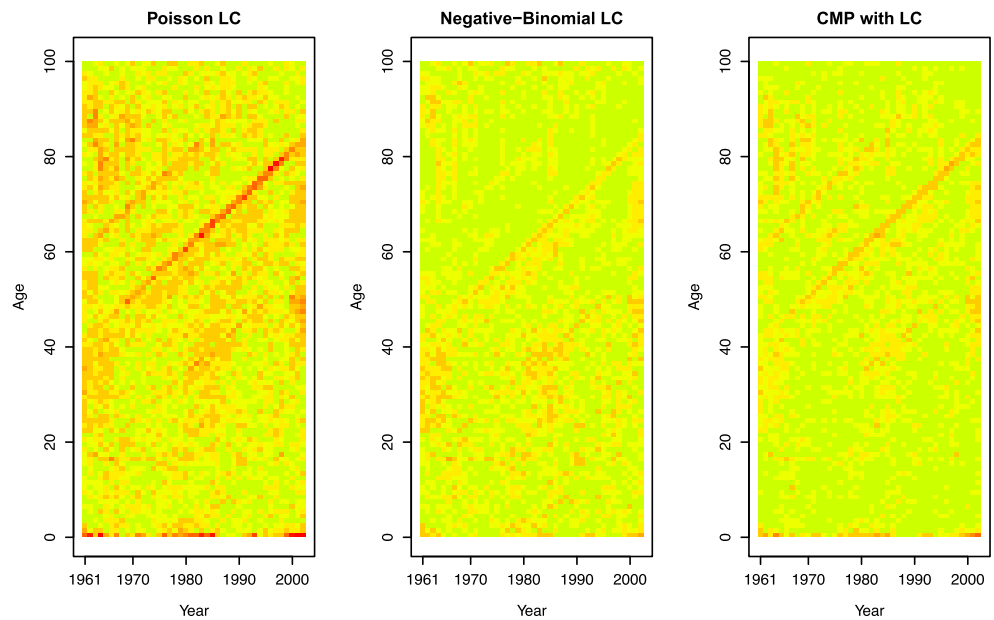

Figure 7. Heatmap of r_{xt}^2 for the LC models.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

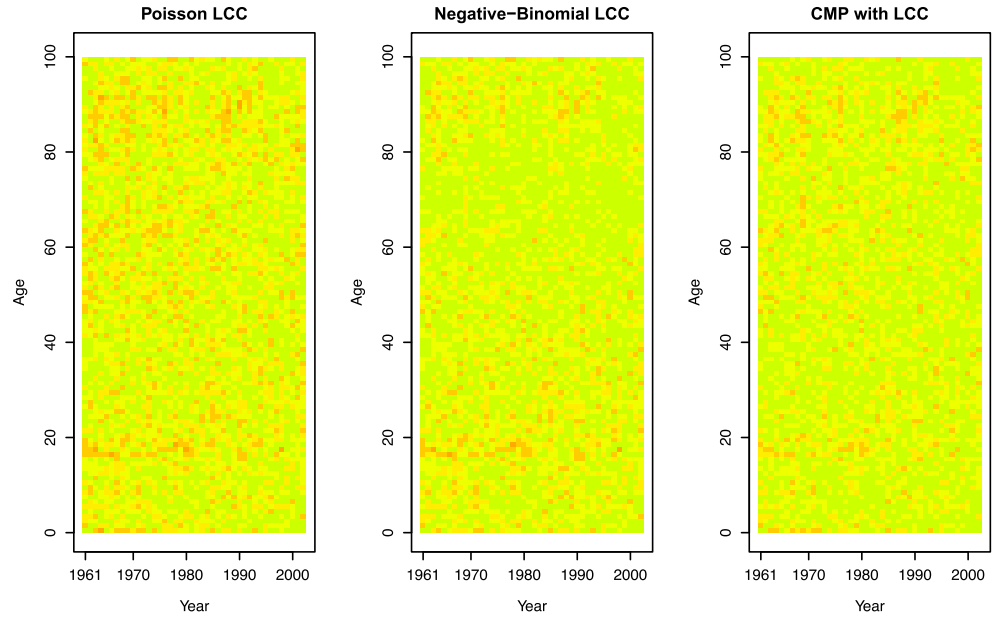

Figure 8. Heatmap of r_{xt}^2 for the LCC models.

Table 1. The estimated total model residuals r^2 under each of the model specifications

Model	Poisson	NB	CMP
LC	16, 709.85 (26.86%)	4, 392.80 (5.62%)	**3**, **970.60** (**4.40%**)
LCC	6, 628.97 (11.29%)	4, 006.45 (4.71%)	**3**, **850.75** (**4.33%**)

Note. Also included in parentheses are the percentages of cells with poor fit, i.e. $r_{xt}^2 > 3.84$. Models that generate less than 5% poor fit cells are good.

Table 2. Various scores for assessing out-of-sample predictive performances under each of the model specifications (the smaller the better)

Model	Metric	Poisson	NB	CMP
LC	LOGS	10.82	9.05	**7.20**
	CRPS	235.62	254.97	**227.60**
	DSS	15.37	13.91	**12.71**
LCC	LOGS	15.57	13.86	**8.95**
	CRPS	153.95	153.32	**151.85**
	DSS	15.34	14.32	**13.29**

The out-of-sample predictive performances of the models are compared by assessing their logarithmic scores (LOGS), continuous ranked probability scores (CRPS), and Dawid-Sebastiani scores (DSS), see for example Krüger et al. (2021). These measures are adopted as evaluation metrics as they jointly assess the calibration and sharpness of the full posterior predictive distribution, which is essential in our

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

context where modelling the dispersion is a key objective. The metrics can be derived from the MCMC posterior samples of D_{xt}^{F} (for $x = 1, \ldots, A$ and $t = 2003, \ldots, 2021$) and computed using the R package *scoringRules* (Jordan et al., 2019). See online supplementary material, Appendix D for more details about the computation of the metrics. The results, as presented in Table 2, indicate that the CMP specifications for death counts generated better out-of-sample forecasts than their counterparts.

4.4 Sensitivity analysis

Finally, we examine the influence of prior specification on the estimation of ν and assess the sensitivity of the results to the hyperparameters a and b. For illustration, this analysis is performed on the CMP–LC model only. Figure 9 displays the posterior distributions of ν under various Gamma prior specifications, as described in Section 3.1.

We first observe that the posterior distributions of ν under the priors Gamma(1, 0.01), Gamma(1, 0.1), Gamma(1, 1), and Gamma(1, 10) closely resemble that obtained using a flat prior. This is unsurprisingly expected, as these priors have variances substantially larger than the posterior variance of ν (approximately 3×10^{-5}), and therefore exert minimal influence on the posterior estimation. In these cases, the estimation of ν is largely driven by the data.

An exception arises when the prior $\nu \sim$ Gamma(1, 100) is specified. This prior has a mean of 0.01 and a variance of 1×10^{-4}, which is of comparable magnitude to the variance implied by the data. As a result, the prior is sufficiently informative to influence the Bayesian learning process, shifting the posterior distribution of ν slightly towards the prior mean of 0.01. This reflects the increased certainty assigned to the prior, relative to other specifications, resulting in a more pronounced impact on the Bayesian estimation.

In summary, as the variance of the Gamma prior decreases (i.e. for larger values of b), the prior contributes a more prominent role in the estimation of ν.

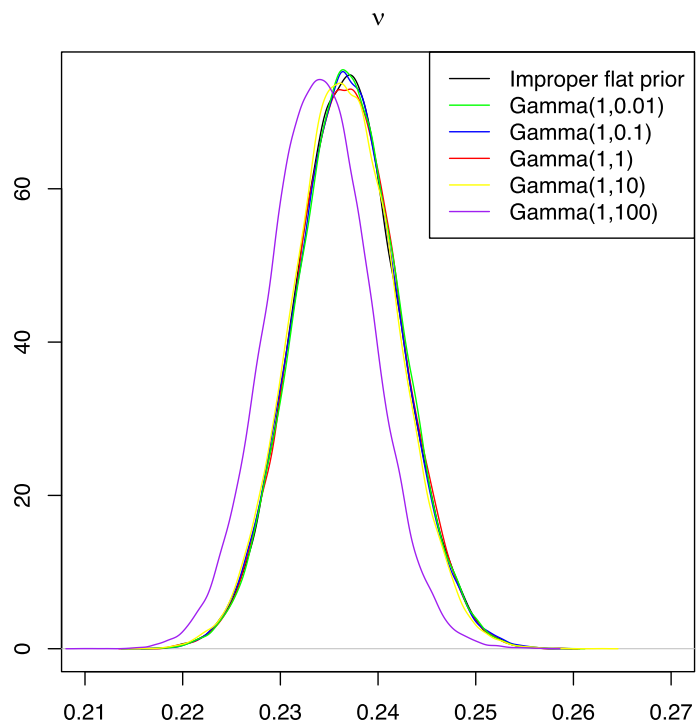

Figure 9. Kernel density plots of the posteriors of ν for the CMP–LC model under various prior specifications.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

5 Conclusion

In this paper, we introduce the CMP distribution as a flexible framework for modelling death counts. The distribution is embedded within the LC and LCC mortality rate models and estimated using a Bayesian approach. The proposed framework highlights the ability of the CMP distribution to accommodate underdispersion, equidispersion, and overdispersion through an explicit dispersion parameter, ν, whose value can be inferred directly from the data. By doing so, this approach allows all relevant sources of uncertainty to be coherently incorporated, particularly in settings where overdispersion is present. To facilitate Bayesian inference, we adopt a non-informative gamma prior for the dispersion parameter and develop a corresponding MCMC algorithm to generate posterior samples.

The proposed methodology is illustrated using England and Wales male mortality data across the entire age range. Empirical results demonstrate that the CMP-based specification led to superior in-sample and out-of-sample performances, as assessed through residual diagnostics and a range of proper scoring rules, when compared with commonly used alternative distributions, such as the Poisson and NB models. These findings underscore the importance of combining flexible distributional assumptions with appropriate structural components in mortality modelling. In particular, incorporating cohort effects alongside a more flexible distribution that accommodates dispersion leads to improved calibration between systematic mortality trends and random variation. A sensitivity analysis was also conducted to assess the impact of prior choices on Bayesian estimation, and the results indicate under the proposed specification appears robust.

Several additional avenues for future research emerge from our work. First, the time-series dynamics of the period and cohort effects, κ_t and γ_c, were specified using standard formulations from the literature. Extending this framework to allow for formal model selection among alternative time-series specifications may further enhance model stability, particularly when applied to other mortality datasets, although the gains may be modest for the data considered here. Second, the dispersion parameter ν was assumed to be constant across ages and periods, capturing only a single overall level of dispersion. This may be considered restrictive, as dispersion could vary systematically across the mortality surface. Allowing for age-specific or time-specific dispersion parameters, such as ν_x or ν_t, represents a promising direction for future research. Furthermore, it would be of interest to explore and formally compare alternative mathematical formulations of λ_{xt}, as discussed in Section 2.2.

Overall, our study demonstrates that integrating the CMP distribution into established mortality models provides a powerful and flexible approach for analysing death counts. By explicitly modelling dispersion within a coherent Bayesian framework, the proposed methodology delivers improved model fit, richer inference, and a solid foundation for further methodological and applied work in mortality modelling and projections.

Acknowledgments

The authors greatly appreciate the constructive comments from the Editor, Associate Editor, and the two anonymous reviewers, which have helped improve the quality of the paper. The authors also thank the audience with their feedback at the 2026 Midwest Actuarial Research Seminars on 28 February 2026 at Purdue University, where this paper was presented.

Conflicts of interest

The authors declare that they have no conflicts of interest with respect to the publication of this article. No financial or personal relationships that could have influenced the work have been disclosed.

Funding

This research received no specific grant from any funding agency in the public, commercial, or nonprofit sectors.

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

Data availability

Data and code supporting the findings of this study are available from the following GitHub repository: https://github.com/jstw1g09/JRSSA-CMP-LC-LCC.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series A*.

References

Basellini, U., Camarda, C. G., & Booth, H. (2023). Thirty years on: A review of the Lee-Carter method for forecasting mortality. *International Journal of Forecasting*, *39*(3), 1033–1049. https://doi.org/10.1016/j.ijforecast.2022.11.002

Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, *3*(1–2), 3–43. https://doi.org/10.1017/S1748499500000440

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Journal of the Royal Statistical Society: Series C, Applied Statistics*, *33*(1), 38–44. https://doi.org/10.2307/2347661

Brillinger, D. R. (1986). A biometrics invited paper with discussion: The natural variability of vital rates and associated statistics. *Biometrics*, *42*(4), 693–734. https://doi.org/10.2307/2530689

Brouhns, N., Denuit, M., & Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, *31*(3), 373–393. https://doi.org/10.1016/S0167-6687(02)00185-3

Brown, J. R. (2003). Redistribution and insurance: Mandatory annuitization with mortality heterogeneity. *The Journal of Risk and Insurance*, *70*(1), 17–41. https://doi.org/10.1111/1539-6975.00045

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G., & Khalaf-Allah, M. (2011). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin: The Journal of the IAA*, *41*(1), 29–59. https://doi.org/10.2143/AST.41.1.2084385

Chanialidis, C., Evers, L., Neocleous, T., & Nobile, A. (2018). Efficient Bayesian inference for COM-Poisson regression models. *Statistics and Computing*, *28*(3), 595–608. https://doi.org/10.1007/s11222-017-9750-x

Conway, R., & Maxwell, W. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, *12*(2), 132–136.

Czado, C., Delwarde, A., & Denuit, M. (2005). Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, *36*(3), 260–284. https://doi.org/10.1016/j.insmatheco.2005.01.001

Daly, F., & Gaunt, R. E. (2016). The Conway-Maxwell-Poisson distribution: Distributional theory and approximation. *ALEA, Latin American Journal of Probability and Mathematical Statistics*, *13*, 635–658. https://doi.org/10.30757/ALEA.v13-25

Dean, C., & and, J. F. L. (1989). Tests for detecting overdispersion in Poisson regression models. *Journal of the American Statistical Association*, *84*(406), 467–472. https://doi.org/10.1080/01621459.1989.10478792

Dean, C. B. (1992). Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association*, *87*(418), 451–457. https://doi.org/10.1080/01621459.1992.10475225

Delwarde, A., Denuit, M., & Partrat, C. (2007). Negative binomial version of the Lee–Carter model for mortality forecasting. *Applied Stochastic Models in Business and Industry*, *23*(5), 385–401. https://doi.org/10.1002/asmb.679

Diana, A., Wong, J. S. T., & Pittea, A. (2025). A unified Bayesian framework for mortality model selection. *Annals of Actuarial Science*, *19*(3), 462–481. https://doi.org/10.1017/S1748499525100146

Gelfand, A. E., & Sahu, S. K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, *94*(445), 247–253. https://doi.org/10.1080/01621459.1999.10473840

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–511. https://doi.org/10.1214/ss/1177011136

Geweke, J. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Technical report, Federal Reserve Bank of Minneapolis.

Goes, J., Barigou, K., & Leucht, A. (2025). Bayesian mortality modelling with pandemics: A vanishing jump approach. *Journal of the Royal Statistical Society: Series C, Applied Statistics*, *74*(4), 1150–1182. https://doi.org/10.1093/jrsssc/qlaf018

Hayat, M. J., & Higgins, M. (2014). Understanding Poisson regression. *Journal of Nursing Education*, *53*(4), 207–215. https://doi.org/10.3928/01484834-20140325-04

Hinde, J., & Demétrio, C. G. (1998). Overdispersion: Models and estimation. *Computational Statistics & Data Analysis*, *27*(2), 151–170. https://doi.org/10.1016/S0167-9473(98)00007-3

Hougaard, P., Lee, M.-L. T., & Whitmore, G. A. (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics*, *53*(4), 1225–1238. https://doi.org/10.2307/2533492

Hunt, A., & Villegas, A. M. (2015). Robustness and convergence in the Lee–Carter model with cohort effects. *Insurance: Mathematics and Economics*, *64*, 186–202. https://doi.org/10.1016/j.insmatheco.2015.05.004

Jordan, A., Krüger, F., & Lerch, S. (2019). Evaluating probabilistic forecasts with scoringRules. *Journal of Statistical Software*, *90*(12), 1–37. https://doi.org/10.18637/jss.v090.i12

Krüger, F., Lerch, S., Thorarinsdottir, T., & Gneiting, T. (2021). Predictive inference based on Markov chain Monte Carlo output. *International Statistical Review*, *89*(2), 274–301. https://doi.org/10.1111/insr.12405

Lee, R. D., & Carter, L. R. (1992). Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association*, *87*(419), 659–671. https://doi.org/10.2307/2290201

Li, J. S.-H., Hardy, M. R., & Tan, K. S. (2009). Uncertainty in mortality forecasting: An extension to the classical Lee-Carter approach. *ASTIN Bulletin: The Journal of the IAA*, *39*(1), 137–164. https://doi.org/10.2143/AST.39.1.2038060

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2013). *The BUGS book: A practical introduction to Bayesian analysis*. Chapman and Hall/CRC.

McCullagh, P. (2019). *Generalized linear models*. Routledge.

O'Hagan, A., & Forster, J. (2004). *Kendall's advanced theory of statistics* (3rd ed., Vol. *2B*). Kendall's Library of Statistics.

Rakhmawan, S. A., Mahmood, T., Abbas, N., & Riaz, M. (2024). Unifying mortality forecasting model: An investigation of the COM–Poisson distribution in the GAS model for improved projections. *Lifetime Data Analysis*, *30*(4), 800–826. https://doi.org/10.1007/s10985-024-09634-x

Renshaw, A. E., & Haberman, H. (2005). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, *38*(3), 556–570. https://doi.org/10.1016/j.insmatheco.2005.12.001

Roberts, G. O., & Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, *16*(4), 351–367. https://doi.org/10.1214/ss/1015346320

Sellers, K. F., Lotze, T., & Raim, A. (2023). *COMPoissonReg: Conway–Maxwell Poisson (COM-Poisson) regression*. CRAN. R package version 0.8.1.

Selvin, S. (2004). *Statistical analysis of epidemiologic data* (Vol. *35*). Oxford University Press.

Shi, J., Shi, Y., Wang, P., & Zhu, D. (2024). Multi-population mortality modelling: A Bayesian hierarchical approach. *ASTIN Bulletin: The Journal of the IAA*, *54*(1), 46–74. https://doi.org/10.1017/asb.2023.29

Tawiah, K., Iddi, S., & Lotsi, A. (2020). On zero-inflated hierarchical Poisson models with application to maternal mortality data. *International Journal of Mathematics and Mathematical Sciences*, *2020*(1), 1407320. https://doi.org/10.1155/2020/1407320

Villegas, A. M., Kaishev, V. K., & Millossovich, P. (2018). StMoMo: An R package for stochastic mortality modeling. *Journal of Statistical Software*, *84*(3), 1–38. https://doi.org/10.18637/jss.v084.i01

Wong, J. S. T., Forster, J. J., & Smith, P. W. F. (2018). Bayesian mortality forecasting with overdispersion. *Insurance: Mathematics and Economics*, *83*(2018), 206–221. https://doi.org/10.1016/j.insmatheco.2017.09.023

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026

Wong, J. S. T., Forster, J. J., & Smith, P. W. F. (2023). Bayesian model comparison for mortality forecasting. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *72*(3), 566–586. https://doi.org/10.1093/jrsssc/qlad021

Zhang, X., & Ma, R. (2023). Forecasting waved daily COVID-19 death count series with a novel combination of segmented Poisson model and ARIMA models. *Journal of Applied Statistics*, *50*(11-12), 2561–2574. https://doi.org/10.1080/02664763.2021.1976119

Downloaded from academic.oup.com/jrsssa/advance-article/doi/10.1093/jrsssa/qnag089/8734935 by guest on 18 November 2026