Bayesian mortality forecasting with a Conway-Maxwell-Poisson specification

joint work with Jackie Siaw Tze Wong, University of Essex

Midwest Actuarial Research Seminars
Purdue University, West Lafayette
28 February 2026

Emiliano A. Valdez
University of Connecticut

Mortality forecasting: what it is and ongoing relevance

What it is. Mortality forecasting uses past data to predict future death rates by age, sex, and year. It also measures uncertainty in the forecasts.

Why it still matters.

- *Pensions and social security*: Life expectancy affects costs and funding.
- *Insurance and annuities*: Needed for pricing and risk management.
- *Public policy*: Helps plan healthcare and public spending.
- *Economic planning*: Population aging affects the economy.

Data and notation

- We use male death data and the corresponding exposures for England and Wales, extracted from the Human Mortality Database (HMD).
  - Observed deaths: d_{xt} for age x and calendar year t
  - Exposure-to-risk: e_{xt} (central exposure)
  - Underlying mortality rate: μ_{xt} (central rate)
  - Ages: $\{x_1, \ldots, x_A\} = \{0, 1, \ldots, 99\}$ with $A = 100$
  - Years: $\{t_1, \ldots, t_T\} = \{1961, \ldots, 2002\}$ with $T = 42$
  - Holdout validation: years 2003-2021
- Each cell (x, t) represents one age-year combination. There are 100×42 total cells.
- The death count model "typically" assumes that

$$\mathrm{E}[D_{xt}] = e_{xt}\mu_{xt}$$

Standard Poisson

- Let D_{xt} be the random variable that represents the number of deaths for age x in year t.

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

- Structural assumptions for μ_{xt}:
  - Lee-Carter: (Lee and Carter (1992))

$$\log \mu_{xt} = \alpha_x + \beta_x \kappa_t \ ,$$

with constraints for model identifiability: $\sum_x \beta_x = 1$ and $\sum_t \kappa_t = 0$.

  - Lee-Carter with cohorts: (Renshaw and Haberman (1992))

$$\log \mu_{xt} = \alpha_x + \beta_x \kappa_t + \gamma_c \ ,$$

with additional constraints: $\sum_c \gamma_c = \sum_c c\gamma_c = \sum_c c^2\gamma_c = 0$.

- Poisson–LC (Poisson Lee-Carter); Poisson–LCC (Poisson Lee-Carter with cohorts

Interpretation

- The LC model has the following components:
  - Age effect (α_x): the average mortality level (log-central death rate) at age x over all years. It shows the basic shape of mortality by age. Mortality is usually high at young and old ages, and low in early adulthood.
  - Period effect (κ_t): the overall time trend in mortality, appropriately modulated by β_x.
  - Age-Period sensitivity (β_x): shows how strongly mortality at age x reacts to changes in κ_t.
- The LCC model adds the component:
  - cohort effect γ_c: the effect of being born in year c. It reflects shared experiences of a generation that affect mortality over life.
- Limitation: The Poisson model imposes $Var[D_{xt}] = \mathrm{E}[D_{xt}]$, which is often not true with real data.

Overdispersion: evidence via Pearson residuals

- Many studies report overdispersion (extra Poisson variation), e.g., Brillinger (1986), Dean and Lawless (1989).
  - We check overdispersion using the squared Pearson residual:

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

  where $\hat{\mu}_{xt}$ is the MLE of the underlying mortality rate. Specifically,

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

Heatmap for testing overdispersion

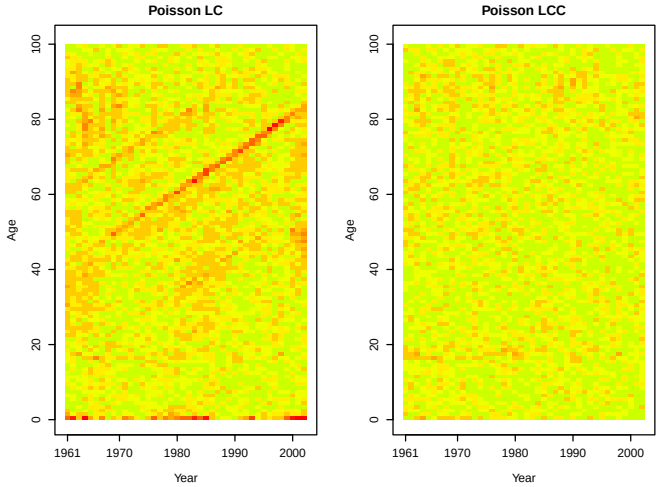


Figure 1: Heat maps of r_{xt}^2 for the Poisson–LC (left) and Poisson–LCC (right) models. Green/yellow rectangular cells indicate areas with good fit; while orange/red-colored cells indicate areas with significantly poor fit.

Statistical test results

- Data: England and Wales males, 1961–2002
- Under a correct Poisson model (with mild conditions), each r_{xt}^2 has an approximate χ_1^2 distribution, asymptotically.
- If the model is correct, only about 5% of cells ($AT \times 0.05 = 210$) should have poor fit. We define poor fit as $r_{xt}^2 > 3.84$, where 3.84 is the 95$^{\text{th}}$ precentile of χ_1^2.
- Observed number of poor-fit cells:
  - Poisson–LC model: 1128 cells (about 27%)
  - Poisson–LCC model: 474 cells (about 11%)
- Additionally, we perform goodness of fit test, using overall Pearson deviance $r^2 = \sum_{x,t} r_{xt}^2$
  - Poisson–LC model: 16709.85
  - Poisson–LCC model: 6628.97
  - Compare these to χ_{df}^2 = 4107.51 where $df = (A-1)(T-2) = 3960$
- Both values are much larger than the benchmark $\Rightarrow$ poor model fit.

Proposed solution: Conway–Maxwell–Poisson (CMP) distribution

- CMP extends the Poisson model by adding a dispersion parameter $\nu \geq 0$.
- ν is learned from data (using Bayesian), which enables for explicit inference about dispersion.
- Retains Poisson as a special case ($\nu = 1$) and includes geometric/Bernoulli as limiting cases.
- We use CMP to model death counts within LC and LCC mortality models.

How CMP compares to common alternatives:

- Poisson: Var $=$ Mean
- Neg Bin: Var $=$ Mean$(1 + \text{Mean}/\varphi)$
- CMP: Var $\approx \frac{1}{\nu}(\text{Mean} + 1/2 - 1/((2\nu))$

Using our empirical mortality data: CMP improves residual calibration and out-of-sample scores versus Poisson and Negative Binomial (NB).

Conway-Maxwell-Poisson (CMP) Distribution

The count r.v. Y follows a CMP distribution if its pmf has the form

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

where

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

- No explicit expression for normalizing constant $Z(\lambda, \nu)$; has to be numerically evaluated.
- Parameters: $\lambda > 0$ (rate-like), $\nu \geq 0$ (dispersion)
- Standard Poisson is recovered when $\nu = 1$.
- Conway and Maxwell (1962)
- Computational note: $Z(\lambda, \nu)$ requires truncation/approximation; implemented in R package COMPoissonReg

How ν controls dispersion

- When ν and λ are not too small, approximately

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

- Recall that for Poisson, $\frac{\mathbb{P}[Y=y-1|\lambda]}{\mathbb{P}[Y=y|\lambda]} = \frac{y}{\lambda}$. For CMP distribution, we have

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

  - When $\nu = 1$, no dispersion as in ordinary Poisson.
  - When $\nu < 1$, overdispersion; the rate of decay decreases less than Poisson and has a longer tail.
  - When $\nu > 1$, underdispersion; the rate of decay increases more in a nonlinear function, thus shortening the tail of the distribution.

- Issue overdispersion in the Poisson assumption has been explored by Wong et al. (2018).
- We propose

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

- Suppose we write $\lambda_{xt}=(e_{xt}\mu_{xt}+\frac{1}{2}-\frac{1}{2\nu})^{\nu}$ for simplicity, the corresponding PMF is

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

- This slightly unconventional parameterization ensures that the expected deaths equals the product of exposures and mortality rates, i.e.

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

- It may be sufficient to specify $D_{xt}|\mu_{xt}, \nu \sim \text{CMP}\ ((e_{xt}\mu_{xt})^{\nu}, \nu)$ in order to achieve $\mathbb{E}[D_{xt}] \approx e_{xt}\mu_{xt}$, particularly since ν is typically not too small.

- The implied variance is

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

- We too consider the Lee–Carter (LC) and Lee–Carter with cohorts (LCC) structural models for μ_{xt}, with resulting models as **CMP–LC** and **CMP–LCC**.

Bayesian estimation

- The benefits of a Bayesian framework are well documented (e.g., Wong et al., 2018).
- Here, the dispersion parameter ν is treated as an unknown random variable and estimated directly from the data.
- This allows inference on dispersion without fixing ν in advance and provides a coherent framework for uncertainty quantification, including overdispersion and parameter uncertainty.
- Full details on priors, posterior derivation, and the MCMC algorithm are given in the paper.

Estimates for the LC models

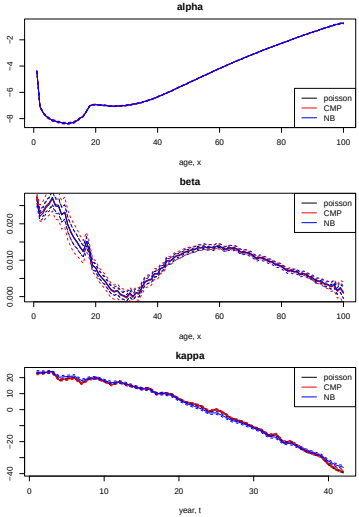


Figure 2: Plot of the medians (solid lines) and 95% credible intervals (dashed lines) of the estimated model parameters under the Poisson–LC, CMP–LC, and NB–LC models.

Estimates for the LCC models

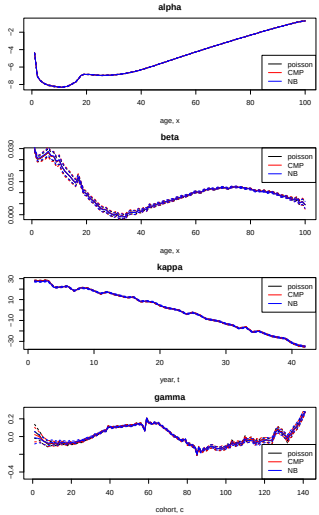
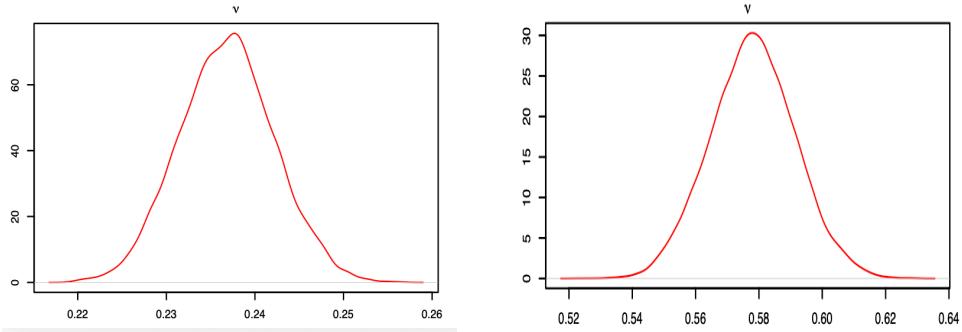


Figure 3: Plot of the medians (solid lines) and 95% credible intervals (dashed lines) of the estimated model parameters under the Poisson–LCC, CMP–LCC, and NB–LCC models.

Results: Distribution of dispersion parameter



(a) LC model **(b)** LCC model

Figure 4: Density plots of estimated dispersion parameters for the LC model (left panel) and the LCC model (right panel). Estimated dispersion (posterior mean): CMP–LC ($\hat{\nu} = 0.237$); CMP–LCC ($\hat{\nu} = 0.578$).

Residual heatmaps for the LC models

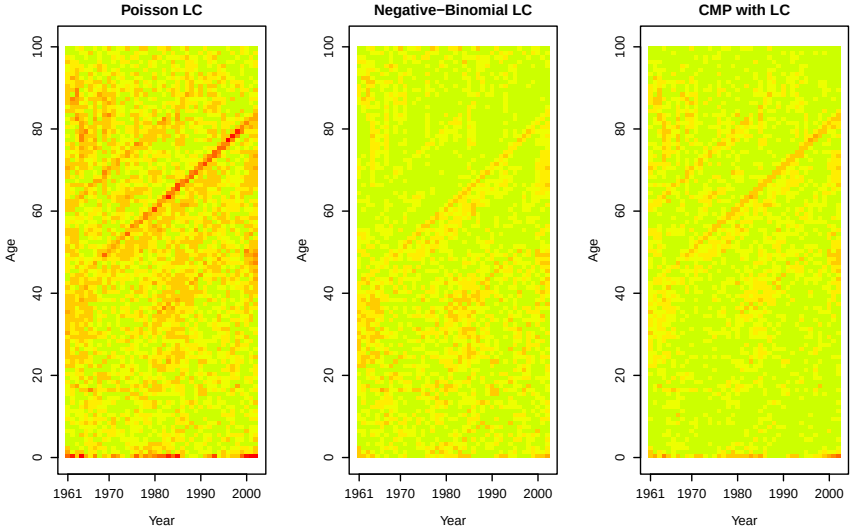


Figure 5: Heatmap of r_{xt}^2 for the LC models.

Residual heatmaps for the LCC models

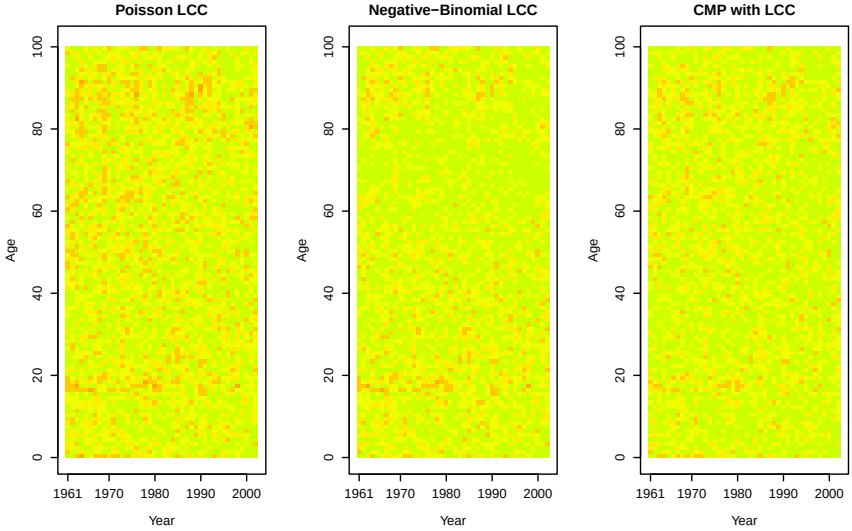


Figure 6: Heatmap of r_{xt}^2 for the LCC models.

Model validation

Table 1: The estimated total model residuals r^2 under each of the model specifications. Also included in parentheses are the percentages of cells with poor fit, i.e. $r^2_{xt} > 3.84$. Models that generate less than 5% poor fit cells are good.

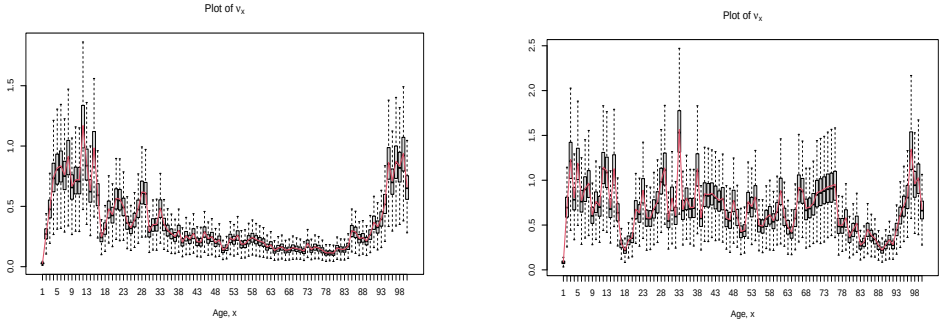
Model	Poisson	NB	CMP
LC	16709.85	4392.80	**3970.60**
	(26.86%)	(5.62%)	(**4.40**%)
LCC	6628.97	4006.45	**3850.75**
	(11.29%)	(4.71%)	(**4.33**%)

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**

Logarithmic scores (LOGS); Continuous Ranked Probability Scores (CRPS); Dawid-Sebastiani scores (DSS)

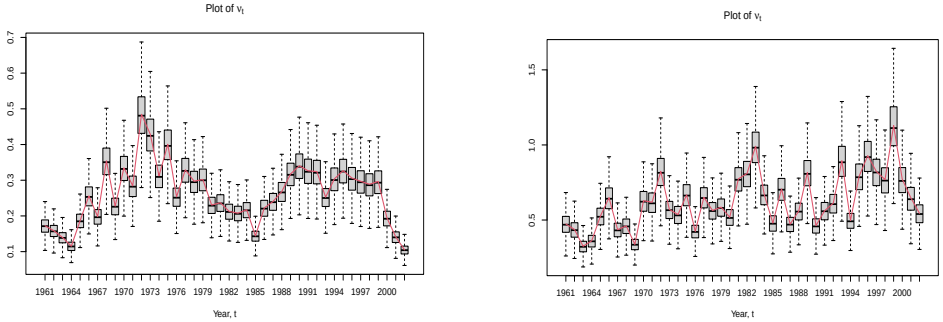
Dispersion varies with age x



(a) LC model **(b)** LCC model

Figure 7: Boxplots of the estimated age-specific dispersion parameters ν_x ($x = 1, \ldots, A$) under the CMP–LC (top) and CMP–LCC (bottom) models. Posterior means are included in red.

Dispersion varies with year t



(a) LC model **(b)** LCC model

Figure 8: Boxplots of the estimated year-specific dispersion parameters ν_t $(t = 1, \ldots, T)$ under the CMP–LC (top) and CMP–LCC (bottom) models. Posterior means are included in red.

Key takeaways

- The CMP distribution flexibly captures under-, equi-, and overdispersion through the parameter ν.
- For England and Wales male data, CMP achieves superior residual calibration and out-of-sample performance compared with Poisson and NB models.
- Incorporating cohort effects (LCC) reduces spurious overdispersion and improves trend projections.

Future research directions

- Formal model selection for alternative time-series specifications of κ_t and γ_c.
- Allow dispersion to vary across age or time (e.g., ν_x or ν_t).
- Apply the CMP framework to other populations, gender, and stress periods.
- Incorporate explicit pandemic or shock components when appropriate.
- Assess the financial implications of dispersion in mortality modelling.

Selected references

- Sellers, K.F. (2023). *The Conway–Maxwell–Poisson Distribution*. Cambridge University Press: United Kingdom.
- Conway, R. and W. Maxwell (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*. 12(2): 132-136.
- Dean, C. and J. Lawless (1989). Tests for detecting overdispersion in Poisson regression models. *Journal of the American Statistical Association*. 84(406): 467-472.
- Sellers, K.F., T. Lotze, and A. Raim (2023). COMPOissonReg: Conway–Maxwell–Poisson (COM-Poisson) Regression. *CRAN*. R package version 0.8.1.
- Wong, J.S.T., J.J. Forster, and P.W.F. Smith (2018). Bayesian mortality forecasting with overdispersion. *Insurance: Mathematics and Economics*. 83:206-221.
- Wong, J.S.T. and E.A. Valdez (2026). Bayesian mortality forecasting with a Conway–Maxwell–Poisson specification. Available at `https://arxiv.org/abs/2601.01686`.

- Thank you -