Supplementary materials for Bayesian model comparison for mortality forecasting

Supplementary materials caption:

- Appendix A: Expressions for $\boldsymbol{V}$, $\boldsymbol{W}$, and $\boldsymbol{W}_\gamma$.
- Appendix B: Deriving the priors for the API model by moments-matching.
- Appendix C: MCMC Scheme for the API Model.
- Appendix D: MCMC Scheme for the LC Model.
- Appendix E: Computation of marginal likelihoods.
- Appendix F: MH/Gibbs updating for $\boldsymbol{\gamma}'$, σ_γ^2, and ρ_γ.

Figures caption:

- Figure C.1: A plot of proposal variances (left) and the corresponding acceptance rates (right) of the MCMC algorithms for α_x and β_x under the API model.

Appendix A Expressions for $\boldsymbol{V}$, $\boldsymbol{W}$, and $\boldsymbol{W}_\gamma$

(a) $\boldsymbol{V} = \boldsymbol{B}^{22} - \boldsymbol{B}^{21}(\boldsymbol{B}^{11})^{-1}\boldsymbol{B}^{12}$, $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top$ and is partitioned such that

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}^{11}_{1\times 1} & \boldsymbol{B}^{12}_{1\times(T-1)} \\ \boldsymbol{B}^{21}_{(T-1)\times 1} & \boldsymbol{B}^{22}_{(T-1)\times(T-1)} \end{pmatrix}, \boldsymbol{Q} = (\boldsymbol{I}_T - \boldsymbol{P})^\top(\boldsymbol{I}_T - \boldsymbol{P}), \boldsymbol{P} = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ \rho & 0 & & & \vdots \\ 0 & \rho & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \rho & 0 \end{pmatrix}_{T\times T},$$

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{T\times T}.$$

(b) $\boldsymbol{W} = \boldsymbol{E}^{22} - \boldsymbol{E}^{21}(\boldsymbol{E}^{11})^{-1}\boldsymbol{E}^{12}$, $\boldsymbol{E} = \boldsymbol{D}\boldsymbol{Q}^{-1}\boldsymbol{D}^\top$ and is partitioned such that

$$\boldsymbol{E} = \begin{pmatrix} \boldsymbol{E}^{11}_{2\times 2} & \boldsymbol{E}^{12}_{2\times(T-2)} \\ \boldsymbol{E}^{21}_{(T-2)\times 2} & \boldsymbol{E}^{22}_{(T-2)\times(T-2)} \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & \cdots & T-1 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{T\times T},$$

and $\boldsymbol{Q}$ is as above.

(c) $\boldsymbol{W}_\gamma = [\boldsymbol{E}^{22}_\gamma - \boldsymbol{E}^{21}_\gamma(\boldsymbol{E}^{11}_\gamma)^{-1}\boldsymbol{E}^{12}_\gamma]$, $\boldsymbol{E}_\gamma = \boldsymbol{D}_\gamma\boldsymbol{Q}_\gamma^{-1}\boldsymbol{D}_\gamma^\top$ which is partitioned such that

$$\boldsymbol{E}_\gamma = \begin{pmatrix} \boldsymbol{E}^{11}_{\gamma\ 3\times 3} & \boldsymbol{E}^{12}_{\gamma\ 3\times(C-3)} \\ \boldsymbol{E}^{21}_{\gamma\ (C-3)\times 3} & \boldsymbol{E}^{22}_{\gamma\ (C-3)\times(C-3)} \end{pmatrix},$$

$$\boldsymbol{D}_\gamma = \begin{array}{r} \text{row } 1 \\ \text{row } 2 \\ \text{row } 3 \\ \text{row } 4 \\ \text{row } 5 \\ \vdots \\ \text{row } 73 \\ \text{row } 74 \\ \vdots \\ \text{row } C \end{array} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & \cdots & \cdots & \cdots & \cdots & 1 \\ 1 & 2 & 3 & 4 & 5 & \cdots & \cdots & \cdots & \cdots & C \\ 1^2 & 2^2 & 3^2 & 4^2 & 5^2 & \cdots & \cdots & \cdots & \cdots & C^2 \\ 0 & 1 & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & & & & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \ddots & \ddots & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 1 & 0 \end{pmatrix}_{C\times C},$$

$\boldsymbol{Q}_\gamma = \boldsymbol{R}_\gamma^\top\boldsymbol{R}_\gamma$, with

$$\boldsymbol{R}_\gamma = \begin{pmatrix} 1/100 & 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ -\sqrt{1-\rho_\gamma^2} & \sqrt{1-\rho_\gamma^2} & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ \rho_\gamma & -(1+\rho_\gamma) & 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & & & \vdots \\ 0 & \cdots & 0 & \rho_\gamma & -(1+\rho_\gamma) & 1 & 0 & \cdots & 0 \end{pmatrix}_{C\times C}.$$

Appendix B Deriving the priors for the API model by moments-matching

In what follows, we will be applying the Law of iterated expectation repeatedly. Using the parameter correspondence relationship in Section 2.3, we require

$$\mathbb{E}[\alpha_x^{\mathrm{API}}] = \mathbb{E}[\alpha_x^{\mathrm{LC}}] + \mathbb{E}[\beta_x^{\mathrm{LC}}]\mathbb{E}[\psi_1^{\mathrm{LC}}] = -5,$$

and

$$\mathrm{Var}[\alpha_x^{\mathrm{API}}] \quad = \quad \mathrm{Var}[\alpha_x^{\mathrm{LC}}] + \mathrm{Var}[\psi_1^{\mathrm{LC}}] \times (\mathbb{E}[\beta_x^{\mathrm{LC}}])^2 + \mathrm{Var}[\beta_x^{\mathrm{LC}}] \times \mathbb{E}[(\psi_1^{\mathrm{LC}})^2] = 14.$$

Hence, the prior $\alpha_x^{\mathrm{API}} \overset{\mathrm{ind}}{\sim} N(-5, 14)$. Similarly, we require

$$\mathbb{E}[\beta_x^{\mathrm{API}}] = \mathbb{E}[\beta_x^{\mathrm{LC}}]\mathbb{E}[\psi_2^{\mathrm{LC}}] = 0,$$

and

$$\mathrm{Var}[\beta_x^{\mathrm{API}}] \quad = \quad \mathrm{Var}[\psi_2^{\mathrm{LC}}] \times (\mathbb{E}[\beta_x^{\mathrm{LC}}])^2 + \mathrm{Var}[\beta_x^{\mathrm{LC}}] \times \mathbb{E}[(\psi_2^{\mathrm{LC}})^2] = 0.01.$$

Hence, the prior $\beta_x^{API} \overset{\mathrm{ind}}{\sim} N(0, 0.01)$. Finally, ignoring the constraints, we require

$$\mathbb{E}[\epsilon_t'^{\mathrm{API}}] = \mathbb{E}[\beta_x^{\mathrm{LC}}]\mathbb{E}[\epsilon_t'^{\mathrm{LC}}] = 0 \times 0 = 0,$$

and

$$\begin{aligned}
\mathrm{Var}[\epsilon_t'^{\mathrm{API}}|(\sigma_\kappa^{\mathrm{API}})^2] &= \mathrm{Var}[\beta_x^{\mathrm{LC}}\epsilon_t'^{\mathrm{LC}}|(\sigma_\kappa^{\mathrm{LC}})^2] \\
&= \mathrm{Var}[\epsilon_t'^{\mathrm{LC}}|(\sigma_\kappa^{\mathrm{LC}})^2] \times (\mathbb{E}[\beta_x^{\mathrm{LC}}])^2 + \mathbb{E}[(\epsilon_t'^{\mathrm{LC}})^2|(\sigma_\kappa^{\mathrm{LC}})^2] \times \mathrm{Var}[\beta_x^{\mathrm{LC}}] \\
&= 0.005 \times (\mathrm{Var}[\epsilon_t'^{\mathrm{LC}}|(\sigma_\kappa^{\mathrm{LC}})^2] + (\mathbb{E}[\epsilon_t'^{\mathrm{LC}}|(\sigma_\kappa^{\mathrm{LC}})^2])^2). \qquad (1)
\end{aligned}$$

Knowing that

$$\epsilon_t'^{\mathrm{LC}}|(\sigma_\kappa^{\mathrm{LC}})^2, \rho^{\mathrm{LC}} \sim N\left(0, \frac{(\sigma_\kappa^{\mathrm{LC}})^2}{1-(\rho^{\mathrm{LC}})^2}\right),$$

we marginalise over ρ^{LC} to obtain

$$\mathbb{E}[\epsilon_t'^{\mathrm{LC}}|(\sigma_\kappa^{\mathrm{LC}})^2] = \mathbb{E}[\mathbb{E}[\epsilon_t'^{\mathrm{LC}}|(\sigma_\kappa^{\mathrm{LC}})^2, \rho^{\mathrm{LC}}]] = \mathbb{E}[0] = 0,$$

and

$$\mathrm{Var}[\epsilon_t'^{\mathrm{LC}}|(\sigma_\kappa^{\mathrm{LC}})^2] \quad = \quad \mathbb{E}[\mathrm{Var}[\epsilon_t'^{\mathrm{LC}}|(\sigma_\kappa^{\mathrm{LC}})^2, \rho^{\mathrm{LC}}]] + \mathrm{Var}[\mathbb{E}[\epsilon_t'^{\mathrm{LC}}|(\sigma_\kappa^{\mathrm{LC}})^2, \rho^{\mathrm{LC}}]] = (\sigma_\kappa^{\mathrm{LC}})^2 \times \mathbb{E}\left[\frac{1}{1-(\rho^{\mathrm{LC}})^2}\right].$$

Since

$$\mathbb{E}\left[\frac{1}{1-(\rho^{\mathrm{LC}})^2}\right] = \int_{-1}^{1} \frac{1}{1-(\rho^{\mathrm{LC}})^2} \times \frac{3}{4}(1+\rho^{\mathrm{LC}})^2(1-\rho^{\mathrm{LC}})\mathrm{d}\rho^{\mathrm{LC}} = \frac{3}{2},$$

we have

$$\mathrm{Var}[\epsilon_t'^{\mathrm{LC}}|(\sigma_\kappa^{\mathrm{LC}})^2] = \frac{3}{2}(\sigma_\kappa^{\mathrm{LC}})^2.$$

Suppose also that

$$\frac{\rho^{\mathrm{API}}+1}{2} \sim \mathrm{Beta}(3,2), \text{ for } \rho^{\mathrm{API}} \in (-1,1),$$

a similar calculation as above gives

$$\mathrm{Var}[\epsilon_t'^{\mathrm{API}}|(\sigma_\kappa^{\mathrm{API}})^2] = \frac{3}{2}(\sigma_\kappa^{\mathrm{API}})^2.$$

Substituting all these back into Equation (1), we require $(\sigma_\kappa^{\mathrm{API}})^2 = 0.005(\sigma_\kappa^{\mathrm{LC}})^2$. Hence, the prior $(\sigma_\kappa^{API})^{-2} \sim \mathrm{Gamma}(0.1, 5 \times 10^{-7})$.

Appendix C MCMC Scheme for the API Model

We drop superscripts hereon for simplicity. Denoting $f(\cdot|rest)$ as the density of $\cdot$ conditional on the rest of the parameters, the conditional posterior densities of the parameters are given as

i.
$$f(\alpha_x|rest) \propto \frac{\exp(\sum_t d_{xt}\alpha_x - 2.5^{-1} \times |\alpha_x + 5|)}{\prod_t [e_{xt}\exp(\alpha_x + \beta_x t + \kappa_t) + \phi]^{d_{xt}+\phi}}.$$

ii.
$$f(\beta_x|rest) \propto \frac{\exp(\sum_t d_{xt}\beta_x t - 0.03^{-1} \times |\beta_x|)}{\prod_t [e_{xt}\exp(\alpha_x + \beta_x t + \kappa_t) + \phi]^{d_{xt}+\phi}}.$$

iii.
$$f(\boldsymbol{\kappa}_{-1,2}|rest) \propto \frac{\exp\left(\sum_{x,t} d_{xt}\kappa_t - (4\sigma_\kappa^2)^{-1}\boldsymbol{\kappa}_{-1,2}^\top \boldsymbol{W}^{-1}\boldsymbol{\kappa}_{-1,2}\right)}{\prod_{x,t}[e_{xt}\exp(\alpha_x + \beta_x t + \kappa_t) + \phi]^{d_{xt}+\phi}}.$$

iv.
$$f(\sigma_\kappa^2|rest) \propto (\sigma_\kappa^2)^{-\frac{T-1}{2}} \times \exp\left(-\lambda\sigma_\kappa^2 - \frac{1}{4\sigma_\kappa^2}\boldsymbol{\kappa}_{-1,2}^\top \boldsymbol{W}^{-1}\boldsymbol{\kappa}_{-1,2}\right).$$

v.
$$\lambda|rest \sim \text{Gamma}\left(2, 2.5 \times 10^{-7} + \sigma_\kappa^2\right).$$

vi.
$$f(\rho|rest) \propto |\boldsymbol{W}|^{-\frac{1}{2}} \exp\left(-\frac{1}{4\sigma_\kappa^2}\boldsymbol{\kappa}_{-1,2}^\top \boldsymbol{W}^{-1}\boldsymbol{\kappa}_{-1,2}\right) \times (1+\rho)^2(1-\rho) \times I_{(-1,)}(\rho),$$

where $I_{(-1,1)}(\rho)$ is the indicator function. This corresponds to fitting a stationary AR(1) model on κ_t. Manually setting $\rho = 1$ within the MCMC gives the random walk model.

vii.
$$f(\phi|rest) \propto \prod_{x,t}\left[\frac{\Gamma(d_{xt}+\phi)}{(e_{xt}\exp(\alpha_x + \beta_x t + \kappa_t) + \phi)^{d_{xt}+\phi}}\right]\frac{1}{[\Gamma(\phi)]^{AT}}\phi^{a_\phi - 1}\exp(-b_\phi\phi).$$

For most parameters (except λ), random walk MH algorithm is applied, where proposal variances are numerically determined using pilot runs to achieve certain acceptance rates (controlled to lie within $[0.15, 0.45]$). For example, the proposal variances of α_x and β_x numerically determined using the automated search algorithm described in Wong et al., 2018) are illustrated in Figure C.1.

The constraints $\sum_t \kappa_t = \sum_t t\kappa_t = 0$ induce (time-varying) correlation among κ_t. Therefore, we update $\boldsymbol{\kappa}_{-1,2}$ in one single block to avoid slow convergence rate in the MH updating scheme, i.e. $\boldsymbol{\kappa}_{-1,2}^* \sim N(\boldsymbol{\kappa}_{-1,2}^{i-1}, \boldsymbol{\Sigma}_\kappa)$, where $\boldsymbol{\kappa}_{-1,2}^{i-1}$ is the current iterate and $\boldsymbol{\Sigma}_\kappa$ is the proposal variance. As motivated by Roberts and Rosenthal (2001), we choose

$$\boldsymbol{\Sigma}_\kappa = \frac{2.38^2}{39} \times (-\boldsymbol{H}_\kappa^{\text{mode}})^{-1},$$

where $\boldsymbol{H}_\kappa^{\text{mode}}$ is the sub-matrix of the Hessian matrix of the joint posterior distribution corresponding to $\boldsymbol{\kappa}_{-1,2}$, evaluated at the joint posterior mode (found by numerical optimisation). The matrix $\boldsymbol{H}_\kappa$ is such that its ij^{th} element is given as follows: for $i \neq j$,

$$\begin{aligned}[\boldsymbol{H}_\kappa]_{ij} &= -i \times j\sum_{x=1}^{A}\frac{(d_{x1}+\phi)\phi e_{x1}\exp(\alpha_x + \beta_x t_1 + \kappa_1)}{[e_{x1}\exp(\alpha_x + \beta_x t_1 + \kappa_1) + \phi]^2} \\ &\quad -(i+1)(j+1)\sum_{x=1}^{A}\frac{(d_{x2}+\phi)\phi e_{x2}\exp(\alpha_x + \beta_x t_2 + \kappa_2)}{[e_{x2}\exp(\alpha_x + \beta_x t_2 + \kappa_2) + \phi]^2} - \frac{1}{2\sigma_\kappa^2}[\boldsymbol{W}^{-1}]_{ij},\end{aligned}$$

where $[\boldsymbol{W}^{-1}]_{ij}$ is the ij^{th} element of $\boldsymbol{W}^{-1}$; while for $i = j$,

$$
\begin{aligned}
[\boldsymbol{H}_\kappa]_{ij} &= -\sum_{x=1}^{A} \frac{(d_{x\,i+2} + \phi)\phi e_{x\,i+2} \exp(\alpha_x + \beta_x t_{i+2} + \kappa_{i+2})}{[e_{x\,i+2} \exp(\alpha_x + \beta_x t_{i+2} + \kappa_{i+2}) + \phi]^2} \\
&\quad -i^2 \sum_{x=1}^{A} \frac{(d_{x1} + \phi)\phi e_{x1} \exp(\alpha_x + \beta_x t_1 + \kappa_1)}{[e_{x1} \exp(\alpha_x + \beta_x t_1 + \kappa_1) + \phi]^2} \\
&\quad -(i+1)^2 \sum_{x=1}^{A} \frac{(d_{x2} + \phi)\phi e_{x2} \exp(\alpha_x + \beta_x t_2 + \kappa_2)}{[e_{x2} \exp(\alpha_x + \beta_x t_2 + \kappa_2) + \phi]^2} - \frac{1}{2\sigma_\kappa^2}[\boldsymbol{W}^{-1}]_{ii}.
\end{aligned}
$$

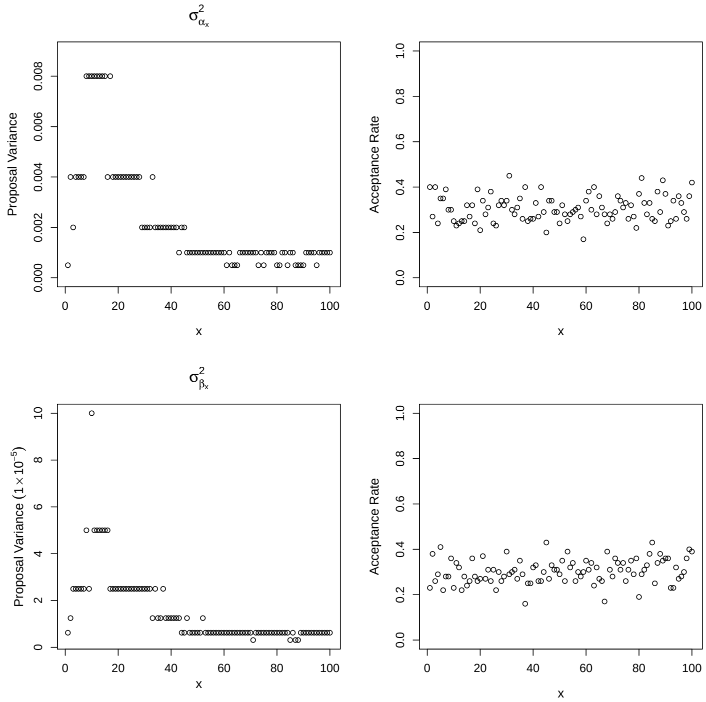

Fig. C.1. A plot of proposal variances (left) and the corresponding acceptance rates (right) of the MCMC algorithms for α_x and β_x under the API model.

Appendix D MCMC Scheme for the LC Model

The condition posterior densities of $(\sigma_\kappa)^2$, ρ, and $\boldsymbol{\kappa}_{-1}$ are

i. $\sigma_\kappa^{-2}|rest \sim \text{Gamma}\left(1 + \frac{T-1}{2}, 0.0001 + \frac{1}{2}(\boldsymbol{\kappa}_{-1} - \boldsymbol{\mu}_\kappa)^\top (\boldsymbol{V})^{-1}(\boldsymbol{\kappa}_{-1} - \boldsymbol{\mu}_\kappa)\right)$.

ii. $f(\rho|rest) \propto |\boldsymbol{V}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2(\sigma_\kappa)^2}(\boldsymbol{\kappa}_{-1} - \boldsymbol{\mu}_\kappa)^\top (\boldsymbol{V})^{-1}(\boldsymbol{\kappa}_{-1} - \boldsymbol{\mu}_\kappa)\right] \times (1+\rho)^2(1-\rho)I_{(-1,1)}(\rho)$.

iii.

$$
f(\boldsymbol{\kappa}_{-1}|rest) \propto \frac{\exp(\sum_{x,t} d_{xt}\beta_x\kappa_t)}{\prod_{x,t}[e_{xt}\exp(\alpha_x + \beta_x\kappa_t) + \phi]^{d_{xt}+\phi}} \times \exp\left(-\frac{1}{2\sigma_\kappa^2}(\boldsymbol{\kappa}_{-1} - \boldsymbol{\mu}_\kappa)^\top (\boldsymbol{V})^{-1}(\boldsymbol{\kappa}_{-1} - \boldsymbol{\mu}_\kappa)\right).
$$

After some tuning, the proposal variance matrix we suggest for updating $\boldsymbol{\kappa}_{-1}$ is $\frac{6.38^2}{39} \times \boldsymbol{G}$, where $\boldsymbol{G}$ is the sub-matrix of $[\boldsymbol{H}^{\text{LC}}]^{-1}$ corresponding to $\boldsymbol{\kappa}_{-1}$ evaluated at the MLE of the likelihood function of the LC model, and $\boldsymbol{H}^{\text{LC}}$ is the Hessian matrix of the likelihood function. In practice, $\boldsymbol{H}^{\text{LC}}$ can be estimated by fitting the LC model in the frequentist framework, then extracting the corresponding covariance matrix estimated (for example, using the function "glm" iteratively in R).

Appendix E Computation of marginal likelihoods

For a set of models $M \in M^S$ under consideration with parameters $\boldsymbol{\theta}_M$, the marginal likelihood (ML) of model M, $f_M(\boldsymbol{d})$, is given by

$$f_M(\boldsymbol{d}) \;=\; \int f_M(\boldsymbol{d}|\boldsymbol{\theta}_M) f_M(\boldsymbol{\theta}_M) \mathrm{d}\boldsymbol{\theta}_M, \tag{2}$$

where $f_M(\boldsymbol{d}|\boldsymbol{\theta}_M)$ is the negative-binomial likelihood given as

$$f_M(\boldsymbol{d}|\boldsymbol{\theta}_M) \;=\; \frac{\phi^{AT\phi}\exp(\sum_{x,t} d_{xt} M_{xt})}{[\Gamma(\phi)]^{AT}} \times \prod_{x,t}\left[\frac{e_{xt}^{d_{xt}}\Gamma(d_{xt}+\phi)}{\Gamma(d_{xt}+1)[e_{xt}\exp(M_{xt})+\phi]^{d_{xt}+\phi}}\right],$$

and $f_M(\boldsymbol{\theta}_M)$ is the density of the joint prior distribution under model M.

For the model $M =$ API-AR1, we set (suppressing superscripts) $\boldsymbol{\theta}_M = \{(\boldsymbol{\alpha})^\top, (\boldsymbol{\beta})^\top, (\boldsymbol{\kappa}_{-1,2})^\top, \rho, \log(\sigma_\kappa)^2, \log(\lambda), \log\phi\}^\top$, where $(\sigma_\kappa)^2$, λ and ϕ are log-transformed so that the parameter spaces span $(-\infty, \infty)$. The joint prior density is given by

$$\begin{aligned} f_M(\boldsymbol{\theta}_M) \;=\; & 0.2^{100} \times \exp\left(-\frac{2}{5}\sum_{x=1}^{100}|\alpha_x + 5|\right) \times 0.06^{-100} \times \exp\left(-\frac{100}{3}\sum_{x=1}^{100}|\beta_x|\right) \times (2\pi)^{-\frac{T-2}{2}} \\ & \times(\sigma_\kappa^2)^{-\frac{T-1}{2}}|\boldsymbol{W}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma_\kappa^2}\boldsymbol{\kappa}_{-1,2}^\top \boldsymbol{W}^{-1}\boldsymbol{\kappa}_{-1,2}\right) \times \frac{3}{4}(1+\rho)^2(1-\rho) \times I_{(-1,1)}(\rho) \\ & \times\lambda\sigma_\kappa^2\exp(-\lambda\sigma_\kappa^2) \times 2.5\times 10^{-7} \times \lambda e^{-2.5\times10^{-7}\times\lambda} \times \frac{0.05^{25}}{\Gamma(25)}\phi^{25}e^{-0.05\phi}. \end{aligned} \tag{3}$$

Similarly, for the model $M =$ LC-AR1, we set $\boldsymbol{\theta}_M = \{(\boldsymbol{\alpha})^\top, (\boldsymbol{\beta}_{-1})^\top, (\boldsymbol{\kappa}_{-1})^\top, \rho, \boldsymbol{\psi}, \log(\sigma_\kappa)^2, \log\phi\}^\top$, where $(\sigma_\kappa)^2$ and ϕ are log-transformed. The joint prior density is given by

$$\begin{aligned} f_M(\boldsymbol{\theta}_M) \;=\; & (8\pi)^{-50}\exp\left[-\frac{\sum_{x=1}^{100}(\alpha_x+5)^2}{8}\right] \times (2\pi)^{-\frac{A-1}{2}}\left|0.005\times(\boldsymbol{I}_{A-1} - \frac{1}{A}\boldsymbol{J}_{A-1})\right|^{-\frac{1}{2}} \\ & \times\exp\left[-100(\boldsymbol{\beta}_{-1} - \frac{1}{A}\boldsymbol{1}_{A-1})^\top(\boldsymbol{I}_{A-1} - \frac{1}{A}\boldsymbol{J}_{A-1})^{-1}(\boldsymbol{\beta}_{-1} - \frac{1}{A}\boldsymbol{1}_{A-1})\right] \\ & \times(2\pi)^{-\frac{T-1}{2}}(\sigma_\kappa^2)^{-\frac{T-1}{2}}|\boldsymbol{V}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2\sigma_\kappa^2}(\boldsymbol{\kappa}_{-1} - \boldsymbol{\mu}_\kappa)^\top\boldsymbol{V}^{-1}(\boldsymbol{\kappa}_{-1} - \boldsymbol{\mu}_\kappa)\right] \times \frac{3}{4}(1+\rho)^2(1-\rho) \\ & \times I_{(-1,1)}(\rho) \times (2\pi)^{-1} \times 4000^{-\frac{1}{2}} \times \exp\left[-\frac{1}{8000}\times\boldsymbol{\psi}^\top\begin{pmatrix} 2 & 0 \\ 0 & 2000\end{pmatrix}\boldsymbol{\psi}\right] \\ & \times 0.0001\sigma_\kappa^{-2}e^{-0.0001\sigma_\kappa^{-2}} \times \frac{0.05^{25}}{\Gamma(25)}\phi^{25}e^{-0.05\phi}. \end{aligned} \tag{4}$$

For $M =$ APCI-AR1 and $M =$ LCC-AR1, the expressions of $f_M(\boldsymbol{\theta}_M)$ are respectively given by adding the following term to (3) and (4),

$$(2\pi)^{-\frac{C-3}{2}}(\sigma_\gamma^2)^{-\frac{C-3}{2}}|\boldsymbol{W}_{\boldsymbol{\gamma}}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma_\gamma^2}(\boldsymbol{\gamma}')^\top\boldsymbol{W}_\gamma^{-1}\boldsymbol{\gamma}'\right) \times 10 \times I_{(0,0.1)}(\sigma_\gamma) \times (2\pi)^{-\frac{1}{2}} \times \exp(-\frac{1}{2}\rho_\gamma^2).$$

For the API-RW, LC-RW, APCI-RW, LCC-RW models, simply set ρ to equate one.

The optimal iterative formula for bridge sampling by Meng and Wong (1996) is used,

$$\hat{f}_M^{(t+1)}(\boldsymbol{d}) = \frac{\frac{1}{N_2}\sum_{i=1}^{N_2}\left[\frac{\tilde{l}_i}{N_1\tilde{l}_i + N_2\hat{f}_M^{(t)}(\boldsymbol{d})}\right]}{\frac{1}{N_1}\sum_{i=1}^{N_1}\left[\frac{1}{N_1 l_i + N_2\hat{f}_M^{(t)}(\boldsymbol{d})}\right]}, \tag{5}$$

where $\hat{f}_M^{(t)}(\boldsymbol{d})$ is the t^{th} iteration of the estimator, $l_i = \frac{f_M(\boldsymbol{d}|\boldsymbol{\theta}_M^i)f_M(\boldsymbol{\theta}_M^i)}{g_M(\boldsymbol{\theta}_M^i)}$, $\tilde{l}_i = \frac{f_M(\boldsymbol{d}|\tilde{\boldsymbol{\theta}}_M^i)f_M(\tilde{\boldsymbol{\theta}}_M^i)}{g_M(\tilde{\boldsymbol{\theta}}_M^i)}$, $\{\boldsymbol{\theta}_M^i\}_{i=1}^{N_1}$ is a sample of size N_1 from the posterior distribution with density $f_M(\boldsymbol{\theta}_M|\boldsymbol{d})$ available from our MCMC output, and $\{\tilde{\boldsymbol{\theta}}_M^i\}_{i=1}^{N_2}$ is a sample of size N_2 from g_M (see below). The bridge sampling algorithm we implement is as follows:

(a) Compute the sample mean, $\bar{\theta}_M$, and covariance matrix, σ_M^2, from the first half of the posterior samples, $\{\boldsymbol{\theta}_M^i\}_{i=1}^{5000}$.

(b) Simulate a sample, $\{\tilde{\boldsymbol{\theta}}_M^i\}_{i=1}^{10000}$, from g_M, which is the density of $N(\bar{\theta}_M, \sigma_M^2)$.

(c) Starting with an initial guess, iterate Equation (5) using $\{\boldsymbol{\theta}_M^i\}_{i=5001}^{10000}$ and $\{\tilde{\boldsymbol{\theta}}_M^i\}_{i=1}^{10000}$ until convergence to form $\hat{f}_M^1(\boldsymbol{d})$, where $N_1 = 5000$, $N_2 = 10000$.

(d) Repeat steps 1-3, but this time using second half of the posterior sample for moments-matching, and first half for evaluating Equation (5), forming $\hat{f}_M^2(\boldsymbol{d})$.

(e) The estimated marginal likelihood for model M is then $\hat{f}_M(\boldsymbol{d}) = \frac{\hat{f}_M^1(\boldsymbol{d}) + \hat{f}_M^2(\boldsymbol{d})}{2}$.

Appendix F MH updating for γ', σ_γ^2, and ρ_γ

This section applies to both the APCI and LCC models. The conditional posterior densities of $\boldsymbol{\gamma}'$, σ_γ, and ρ_γ are

i. $f(\boldsymbol{\gamma}'|rest) \propto \frac{\exp(\sum_{x,t} d_{xt}\gamma_c)}{\prod_{x,t}[e_{xt}\exp(M_{xt})+\phi]^{d_{xt}+\phi}} \times \exp\left(-\frac{1}{2\sigma_\gamma^2}(\boldsymbol{\gamma}')^\top \boldsymbol{W}_\gamma^{-1}\boldsymbol{\gamma}'\right)$.

ii. $f(\sigma_\gamma|rest) \propto (\sigma_\gamma)^{-(C-3)} \times \exp\left(-\frac{1}{2\sigma_\gamma^2}(\boldsymbol{\gamma}')^\top (\boldsymbol{W}_\gamma)^{-1}\boldsymbol{\gamma}'\right) \times I_{(0,0.1)}(\sigma_\gamma)$.

iii. $f(\rho_\gamma|rest) \propto |\boldsymbol{W}_\gamma|^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2\sigma_\gamma^2}(\boldsymbol{\gamma}')^\top (\boldsymbol{W}_\gamma)^{-1}\boldsymbol{\gamma}' - \frac{1}{2}\rho_\gamma^2\right)$.

$\boldsymbol{\gamma}'$ is updated using random walk MH in a single block. Using the proposal variance matrix $\boldsymbol{\Sigma}_\gamma = \frac{2.38^2}{138} \times \boldsymbol{H}_\gamma^M$, where $\boldsymbol{H}_\gamma^M$ is the sub-matrix of $[\boldsymbol{H}^M]^{-1}$ corresponding to $\boldsymbol{\gamma}'$, evaluated at the MLE of the likelihood function of model M, and $\boldsymbol{H}^M$ is the Hessian matrix of the likelihood function of model M.

References

Meng, X.-L. and W. H. Wong (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica,* **6**(4), 831–860.

Roberts, G. O. and J. S. Rosenthal (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science,* **16**(4), 351–367.

Wong, J. S. T., J. J. Forster, and P. W. F. Smith (2018). Bayesian mortality forecasting with overdispersion. *Insurance: Mathematics and Economics,* **83**(2018), 206–221.