

贝叶斯推断与深度神经网络在死亡率预测中的应用与对比研究

张 钰, 肖鸿民*

西北师范大学数学与统计学院, 甘肃 兰州

收稿日期: 2025年2月18日; 录用日期: 2025年3月11日; 发布日期: 2025年3月25日

摘 要

随着社会的不断发展和死亡率的下降, 死亡率数据分析及未来预测一直是精算师、人口学家和统计学家的研究重点。经典的单因子Lee-Carter模型通过对数转换将死亡率分解为年龄和时间项, 尽管形式简单, 但在处理复杂死亡率模式时表现有限。为此, 我们采用了神经网络DNN模型, 通过多层神经网络捕捉数据中的非线性关系, 期望有所改善。同时, 在经典Heligman-Pollard模型的基础上, 引入了动态分量和随机游走结构, 放宽了参数平稳性假设, 从而能更精准地描述不同年龄段死亡率的变化。我们还使用GMRF模型, 通过高效捕捉跨年龄组和跨年份的死亡率相关性, 并通过基于梯度的MCMC采样器提高采样效率, 以期减少误差。最后, 结合均方根误差(RMSE)、平均绝对误差(MAE)和平均绝对百分比误差(MAPE)等评价指标, 对不同模型的结果进行了比较分析。结果表明, GMRF模型在预测精度方面表现最佳, Heligman-Pollard模型适用于描述死亡率的动态变化, 而DNN模型则适合于大规模数据且需捕捉非线性关系的场景。

关键词

死亡率预测, Heligman-Pollard模型, DNN, GMRF模型

Application and Comparative Study of Bayesian Inference and Deep Neural Networks in Mortality Rate Prediction

Yu Zhang, Hongmin Xiao*

College of Mathematics and Statistics, Northwest Normal University, Lanzhou Gansu

Received: Feb. 18th, 2025; accepted: Mar. 11th, 2025; published: Mar. 25th, 2025

*通讯作者。

文章引用: 张钰, 肖鸿民. 贝叶斯推断与深度神经网络在死亡率预测中的应用与对比研究[J]. 统计学与应用, 2025, 14(3): 194-204. DOI: 10.12677/sa.2025.143071

Abstract

With society's development and declining mortality rates, mortality data analysis and forecasting have been key research areas for actuaries, demographers, and statisticians. The Lee-Carter model decomposes mortality into age and time components using a logarithmic transformation. While simple, it has limitations in handling complex mortality patterns. To improve, we used a neural network (DNN) model to capture nonlinear relationships in the data. Additionally, we enhanced the Heligman-Pollard model with dynamic components and random walk structure, allowing more accurate description of age-specific mortality changes. We also use GMRF model to capture the mortality correlation across age groups and years efficiently, and improve the sampling efficiency through gradient-based MCMC sampler in order to reduce errors. Finally, we compared the model's using metrics like RMSE, MAE, and MAPE. The results show that the GMRF model performs best in accuracy, the Heligman-Pollard model excels at describing dynamic mortality changes, and the DNN model is best for large datasets with nonlinear relationships.

Keywords

Mortality Prediction, Heligman-Pollard Model, DNN, GMRF Model

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

死亡率数据分析和未来预测一直是精算师、人口学家和统计学家的研究重点。虽然静态问题较为简单,但动态问题更加复杂,至今只有部分解决方案。现有的死亡率预测模型各异,通常通过选择一个适合的模型,基于拟合优度或其他因素进行未来预测,不仅估计死亡率,还能给出预测区间以体现不确定性。图1展示了1960至2020年英国威尔士的死亡率数据。

国际上各个国家的对数死亡率的平均水平随着年龄的增长呈现出一种特定的趋势,这表明不同国家在随年龄增长的变化模式上却展现出显著的共性。进一步分析,这种共性揭示了一个重要现象:随着年龄的增长,不同国家的死亡率变化遵循着相似规律。在初始阶段,由于医疗保健、营养状况和生活环境的改善,死亡率可能呈现下降趋势。然而,随着年龄的增长,人体机能逐渐衰退,慢性疾病和衰老相关疾病的发病率增加,导致死亡率重新上升。这一发现不仅深化了我们对全球人口死亡率变化规律的理解,也强调了死亡率预测在全球范围内的重要性和应用价值。因此,死亡率预测不仅是一个统计学问题,更是一项对社会福祉和全球健康具有重要意义的工作。

在1992年, Li 和 Lee 提出了影响力深远的单因子 Lee-Carter 模型。这个模型简洁而经典,仅包含年龄项和时间项,被认为是最早的单因子动态死亡率模型之一,并考虑将死亡率的变化分解成随着时间而变化的时间因子和不随着时间而变化的年龄因子,进而提出了 Lee-Carter 模型。在文中运用美国死亡率数据对其进行拟合与预测,发现效果较好[1]。在此基础上持续研究, Li 和 Lee 发现 Lee-Carter 死亡率建模和预测方法在长时间序列数据的情况下运行良好,对于数据缺失间隔不均匀的情况,采用带漂移项的随机游走模型来进行预测[2]。黄顺林和王晓军利用考虑出生年效应的 Lee-Carter 模型,对中国男性死亡率数据进行了精准拟合。相比其他模型,其残差分析显示拟合效果更佳。预测表明,中国男性预期寿命

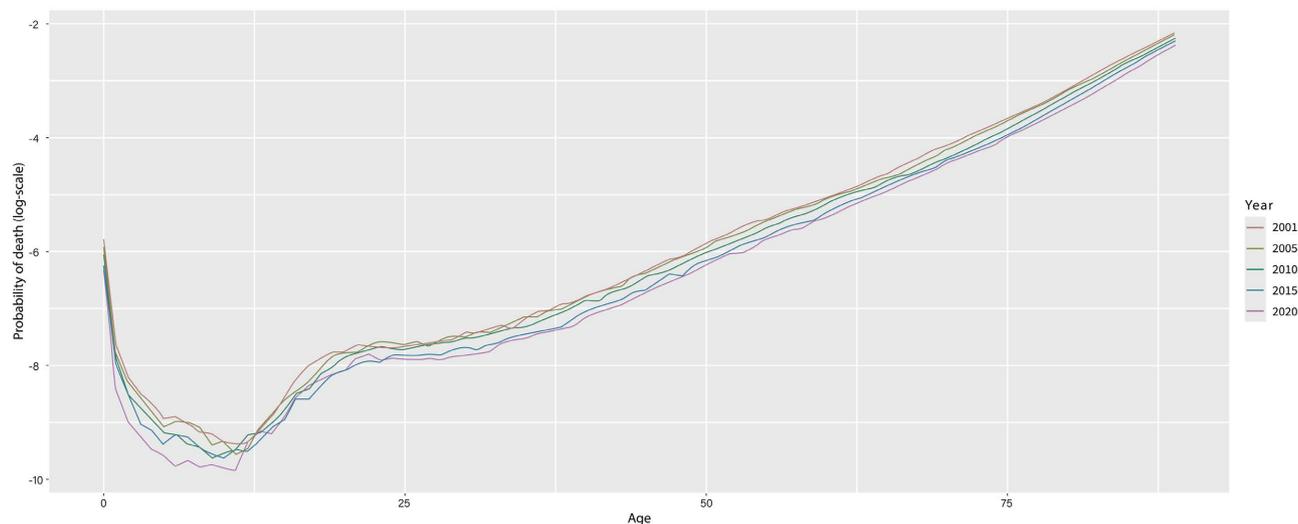


Figure 1. Diagram of mortality rate data for the UK

图 1. 英国的死亡率数据

将持续上升,但增幅逐步放缓[3]。李志生和刘恒甲在 1992~2007 年中国人口分年龄组死亡率数据基础上,对 Lee-Carter 模型参数拟合采用四种不同方法,即对奇异值分解法(SVD)、最小二乘法(OLS)、加权最小二乘法(WLS)和极大似然法(MLE)的拟合结果和预测能力进行了比较分析。

Dellaportas 等在 2001 年探索了使用加权最小二乘法估计 Heligman-Pollard 模型的方法,并提出使用马尔科夫链蒙特卡洛(MCMC)进行贝叶斯推理[4]。随着 MCMC 技术的进步和应用范围的扩大,越来越多的研究开始采用贝叶斯方法进行死亡率建模。Angelos Alexopoulos 等在 2017 年提出了一种基于 Heligman-Pollard 模型公式的动态模型[5]。Christian Evan Chandra 等 2022 年结合使用截断的 HP 模型和 Makeham 模型并采用 Metropolis-Hastings 算法进行采样,利用贝叶斯方法估计死亡率模型的参数,这种方法允许模型在面对数据限制时整合先验信息并提供更稳健的估计[6]。

近年来,神经网络及其深度学习扩展——深度神经网络(DNN),因其在复杂模式识别和数据拟合方面的优势,逐渐成为死亡率预测的一个重要工具。Donatien Hainaut 提出了一种预测和模拟人类死亡率的神经网络方法,与有和没有群体效应的 Lee-Carter 模型相比,神经方法具有极好的预测能力[7]。此外,Andrea Nigri 等提出了一个深度神经网络(DNN)模型,通过利用类似于人口学间接估计技术的深度学习算法,从观测或预测的预期寿命中推导出特定年龄死亡率[8]。

2. 方法与模型

2.1. 经典死亡率模型

Lee 和 Carter (1992)提出了著名的 Lee-Carter 模型,这是一种简洁而经典的单因子模型,包含年龄项和时间项。作为最早的单因子动态死亡率模型之一, Lee-Carter 模型采用了对数转换,形成了一个关于年龄和时间的双线性函数,即:

$$\ln m_{x,t} = \alpha_x + \beta_x + \varepsilon_{x,t} \quad (1)$$

其中, $m_{x,t}$ 为 t 时刻年龄为 x 岁的人的中心死亡率, α_x 为不同年龄段死亡率对数变动的均值; β_x 为不同年龄段死亡率对数变动的趋势; k_t 为时间因素变量; $\varepsilon_{x,t}$ 是一个误差项(白噪声)。同样的还可以附加下述约束条件以保证参数估计结果的唯一性: $\sum_t k_t = 0$, $\sum_x \beta_x = 1$ 。

该模型根据观察到的对数死亡率可通过奇异值分解法(SVD)、最小二乘法(OLS)、加权最小二乘法(WLS)、极大似然法(MLE)拟合参数。

2.2. 深度神经网络 DNN 模型

深度神经网络(DNN)是一种由多层神经元组成的人工神经网络,能够自动从数据中提取特征并学习复杂模式。其结构通常包括输入层、多个隐藏层和输出层。每层神经元接收前一层输出,通过加权求和及非线性激活将信息传递至下一层。隐藏层的数量决定了网络的“深度”,使其能够捕捉数据中的深层特征。DNN 的目标是对于给定的训练数据 $\{(x_1, y_1), \dots, (x_n, y_n)\}$, 找到一个函数 $f(\cdot)$, 使得在遵循未知分布 $P(x, y)$ 的测试数据上预期误差最小化:

$$\int L(y, f(x))P(x, y)dx dy \quad (2)$$

其中, $L(y, f(x))$ 是用于衡量给定 x 相对于实际值 y 预测误差的损失函数。本文应用的 DNN 模型图如图 2 所示, 指定了在一个通用时间 t_i 出生的生命期望, 一个特定年龄段的死亡率向量, 其结构如图所示, 其中输入是出生时生命期望随时间 $t \in T$ 的向量 $e_0 = (e_{0,t_1}, e_{0,t_2}, \dots, e_{0,t_n})$

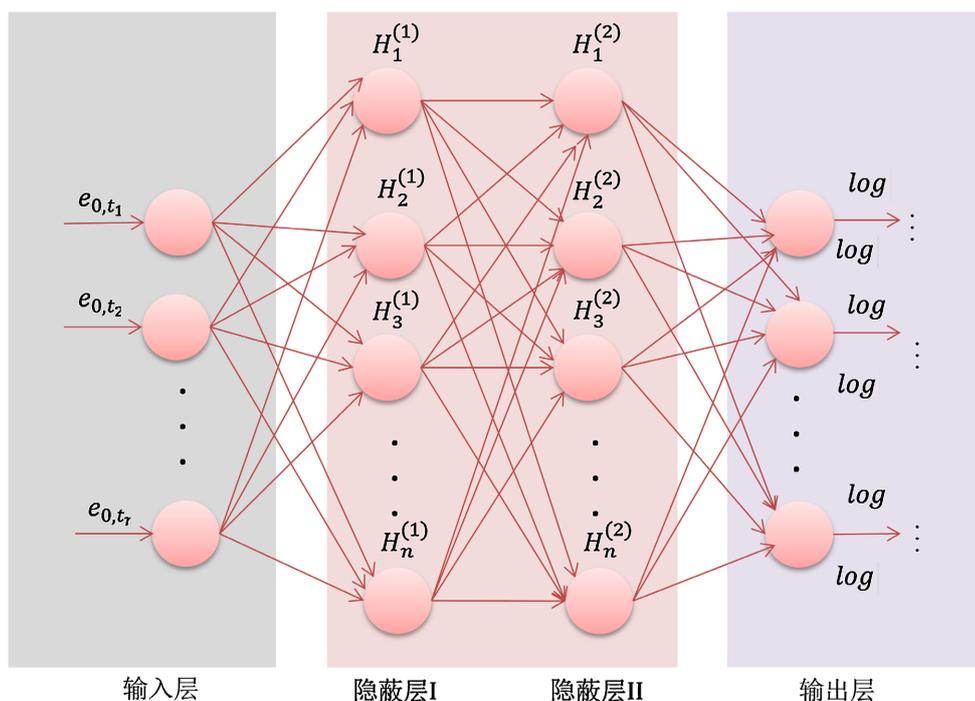


Figure 2. Diagram of DNN model

图 2. DNN 模型图

DNN 模型采用三层架构,包括输入层、隐藏层和输出层。输入是随时间 e_0 的预期寿命的向量,输出为每个年龄对应的对数死亡率。隐藏层 $H^{(k)}$ 的计算公式为:

$$H^{(k)} = f^{(k)}(W^{(k)}H^{(k-1)} + b^{(k)}) \quad (3)$$

其中, $f^{(k)}$ 为激活函数(使用由 $f(z) = \max(z, 0)$ 定义的整流线性单元(ReLU)函数), $W^{(k)}$ 为权重矩阵, $H^{(k-1)}$ 为前一层的输出, $b^{(k)}$ 为偏置项。

输入层接收预期寿命向量 e_0 , 经过隐藏层逐层传递, 最终输出对数死亡率 $M = \log(m_{a,t})$, 其中 $a \in A$ 表示年龄, $t \in T$ 表示年份。其中设 $M = \log(m_{a,t})$ ($a \in A, t \in T$) 为一个死亡率矩阵, 行代表年龄, 列代表日历年。其中 $K=3$ (输入层、隐藏层和输出层), 对于一个包含输入层、隐藏层和输出层的三层标准架构, 关系如下:

$$M = f^{(3)} \left(w^{(3)} f^{(2)} \left(W^{(2)} f^{(1)} \left(W^{(1)} e_0 + b^{(1)} \right) + b^{(2)} \right) + b^{(3)} \right) \quad (4)$$

DNN 目标通过训练数据 $\{(x_1, y_1), \dots, (x_n, y_n)\}$ 学习一个函数使得预测输出 \hat{M} 与实际输出 M 之间的误差最小化。误差通过在多种数据集上表现优异的均方误差(MSE)损失函数衡量:

$$MSE = \frac{1}{n} \sum_{i=1}^n (M_i - \hat{M}_i)^2 \quad (5)$$

采用梯度下降法优化模型参数, 通过反向传播算法计算损失函数关于权重和偏置的梯度, 按照以下公式来更新参数:

$$\omega_{n,r}^{(k)} \leftarrow \omega_{n,r}^{(k)} - \eta \nabla MSE \quad (6)$$

其中而 η 是学习率, 决定了参数更新的步幅。 ∇MSE 表示损失函数关于权重的梯度。通过这种方式, 神经网络能够在训练过程中逐渐减小误差, 提升模型的预测性能。

2.3. Heligman-Pollard 动态模型

Heligman 和 Pollard 在 1980 年提出的 Heligman-Pollard 模型, 被广泛用于死亡率分析, 包括参数化、分解和预测。该模型特别在描述不同年龄段的死亡率模式方面表现出独到之处, 为精确反映特别是老年人群的死亡率变动, 该模型通过引入灵活参数设置, 其复合公式有效捕捉了死亡率在不同年龄段的变化特征, 公式如下:

$$\frac{P_z}{1-p_z} = A^{(Z+B)^C} + D \exp \left[-E \{ \log(Z) - \log(F) \}^2 \right] + GH^z \quad (7)$$

p_z 是年龄 Z 的死亡率概率, 公式由三部分组成, 根据 Alexopoulos 的设置, 每个参数的值的逻辑范围如下: $A \in (0,1)$, $B \in (0,1)$, $C \in (0,1)$, $D \in (0,1)$, $E \in (0,\infty)$, $G \in (0,1)$, $H \in (0,\infty)$ 。而在时间序列或空间建模中, 潜在状态指的是数据生成过程中的隐藏变量, 不易直接观察, 而需要通过模型推断。在死亡率建模中, 潜在状态可能表现在年龄或时间维度上参数的动态变化趋势, 对 Heligman-Pollard 模型的八个参数引入动态分量, 假设这些参数作为随机游走参数演化, 从而放宽了死亡率曲线的平稳性假设。

为定义动态版本, 设假设模型的八个参数作为随机游走参数演化, 放宽死亡率曲线的平稳性假设。为定义动态版本, 设,

$$\varphi(t) = (\tilde{A}_t, \tilde{B}_t, \tilde{C}_t, \tilde{D}_t, \tilde{E}_t, \tilde{F}_t, \tilde{G}_t, \tilde{H}_t)' \quad (8)$$

$\varphi(t)$ 为模型参数在时间 t 的潜在状态, 其中 $\varphi(t) \in R^8$, 元素通过变换从原始的参数变量中获得。其中 t 表示某一年, T 表示有数据的过去年份数量。死亡的概率假设由 Heligman-Pollard 模型给出:

$$\frac{P_z}{1-p_z} = A_t^{(Z+B_t)^{C_t}} + D_t \exp \left[-E_t \{ \log(Z) - \log(F_t) \}^2 \right] + G_t H_t^z \quad (9)$$

其中, $Z = 0,1,\dots,\omega$, $t = 1,2,\dots,T$, ω 表示所采取数据集中最年长的年龄。将上述等式右边记作 $K(Z, \varphi_t)$, 则有

$$\frac{P_{zt}}{1-P_{zt}} = K(Z, \varphi_t) \quad (10)$$

死亡率概率 p_{zt} 及其补集 $1-p_{zt}$, 于是有

$$P_{zt} = \frac{K(Z, \varphi_t)}{1+K(Z, \varphi_t)} \quad (11)$$

似然函数是针对所有时间点 t 和所有年龄组 z 的数据构建的。对于每个 t 和 z , 是观测到的个体数, 而 d_{zt} 是该组潜在的总人数。每个 z 和 t 的贡献是:

$$\binom{n_{zt}}{d_{zt}} [K(Z, \varphi_t)]^{d_{zt}} [1+K(Z, \varphi_t)]^{-n_{zt}} \quad (12)$$

这里的二项式系数 $\binom{n_{zt}}{d_{zt}}$ 表示从 d_{zt} 个个体中选择 n_{zt} 个存活个体的不同方式的数目。 $K(Z, \varphi_t)^{d_{zt}}$ 表示所有选择的 d_{zt} 个存活个体的概率乘积。 $[1-K(Z, \varphi_t)]^{-n_{zt}}$ 是所有可能的存活与死亡组合的总概率进行归一化。将所有年龄组和时间点的贡献乘在一起, 形成整体的似然函数:

$$\gamma(d|\varphi) = \prod_{t=1}^T \prod_{z=0}^{\omega} \binom{n_{zt}}{d_{zt}} [K(Z, \varphi_t)]^{d_{zt}} [1+K(Z, \varphi_t)]^{-n_{zt}} \quad (13)$$

其中, d 是一个向量, 其元素为 d_{zt} , 且 $Z = 0, 1, \dots, \omega$ 和 $t = 1, 2, \dots, T$ 。为了对潜在状态进行动态建模, 假设潜在状态 φ_t 遵循随机游走结构,

$$\gamma(\varphi_t | \varphi_{t-1}, \mu, \Sigma, \eta, \xi) = \phi_8(\varphi_t; \varphi_{t-1} + \mu, \Sigma, \eta, \xi) \quad (14)$$

其中 $t = 1, 2, \dots, T$, 这个公式表示的是一个概率分布 $\gamma(\varphi_{it})$, 其中 φ_{it} 是向量 φ_t 的第 i 个元素, 当 φ_{it} 属于区间 $[\eta_i, \xi_i]$ 时, $\gamma(\varphi_{it})$ 与 1 成正比; 当 φ_{it} 不属于区间 $[\eta_i, \xi_i]$ 时, $\gamma(\varphi_{it})$ 为 0。随机游走过程通过(14)定义, 对 Heligman-Pollard 模型的参数引入了大量的先验结构, 同时放宽了这些参数在时间上的平稳性假设。本文固定了向量 η 和 ξ 的元素, 向量 η 具体为:

$$\eta = (-10.61, -10.61, -5.99, -11.29, -25.33, -\infty, -17.5, -1.39)' \quad (15)$$

向量 ξ 具体为:

$$\xi = (-2.75, -0.2, 2.2, -3.48, 4.09, 2.64, -3.48, 0.18)' \quad (16)$$

对于随机游走过程的漂移 μ , 假设 $\gamma(\mu) = \phi_8(\mu, 0, M^{-1})$, 其中 M 是一个 8×8 的对角矩阵, 其对角线元素为 0.001, 对于协方差矩阵 Σ , 采取逆 Wishart 分布: $\Sigma | \alpha \sim IW(\nu, \Psi)$, ν 是自由度, Ψ 是一个 $p \times p$ 的正定矩阵, 其密度函数为:

$$p(\Sigma) = \frac{|\Psi|^{(\nu-p-1)/2} \exp\left(-\frac{1}{2} \text{tr}(\Psi \Sigma^{-1})\right)}{2^{(\nu p)/2} \Gamma_p\left(\frac{\nu}{2}\right) |\Sigma|^{(\nu+p+1)/2}} \quad (17)$$

假设 Ψ 是一个对角矩阵, 其对角元素为 $1/\alpha_1, \dots, 1/\alpha_8$, 其中 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_8)$ 。对于每一个参数 α_i 都独立同分布(IID)地从逆伽马分布抽取,

$$\alpha_1 \sim IG\left(\frac{1}{2}, \frac{1}{l^2}\right) \quad (18)$$

设置 $l = 10^5$, 这种逆伽马先验分布结构意味着 Σ 对角线上地标准差 σ_i 服从半 $t(\nu, l)$ 分布, 选 $\nu = 2$,

用 $\theta = (\Sigma, \mu, \alpha)$ 表示模型参数, 模型后验分布为:

$$\gamma(\varphi, \theta | d, \eta, \xi) \propto \gamma(\theta) \gamma(d | \varphi) \prod_{t=2}^T \phi_8(\varphi_t; \varphi_{t-1} + \mu, \Sigma, \eta, \xi) \quad (19)$$

提出的模型是一个基于潜在高斯模型与 φ 和超参数 θ 的分布的框架。为了从后验分布中获得样本, 利用 Metropolis-within-Gibbs 采样器, 交替从 $\gamma(\theta | d)$ 和 $\gamma(\theta)$ 中采样, 条件分布所需样本从 $\gamma(\theta | d)$ 采样。从 $\gamma(\theta | d)$ 采样通过 7 个 Metropolis-Hastings 步骤更新每个 φ 来执行。

2.4. GMRF 动态模型

马尔可夫随机场描述一组变量 (x_1, \dots, x_n) 的联合分布, 该分布是通过其全条件分布 $\pi(x_i | x_{-i})$ 来确定的, 其中 x_{-i} 表示除 x_i 之外的所有变量集合。当这些条件分布是高斯分布时, 这种马尔可夫随机场被称为高斯马尔可夫随机场(GMRF), 可通过内在高斯马尔可夫随机场模型模拟死亡率。基于似然函数公式(13)相关模型模拟死亡率, 将每年各年龄组的死亡概率 p_{zt} 转换成变量 $x_{zt} = \log \left[\frac{p_{zt}}{1 - p_{zt}} \right]$, 对所有 $z = 0, \dots, \omega$ 和 $t = 1, \dots, T$ 。其中,

$$x_t = (x_{0t}, \dots, x_{\omega t})' \quad (20)$$

并设

$$x = (x'_1, \dots, x'_T) \quad (21)$$

这为一个 $(\omega+1)T$ 维向量, 包含了所有时间点和所有年龄组的数据, 想象成一个有 $(\omega+1) \times T$ 个节点的网格, 其中 z 表示行(年龄组), t 表示列(年份)。对于向量 x , 假设遵循一个 $(\omega+1)T$ 维高斯分布时, 均值为 $\mu(b_1 1_{\omega+1}, 2b_1 1_{\omega+1}, \dots, T b_1 1_{\omega+1})$, 其中 $1_{\omega+1}$ 是全 1 的 $(\omega+1)$ 维向量, 精度矩阵为:

$$Q = \tau(\rho_{age, R_{\omega+1}} \otimes I_T + \rho_{year, 1_{\omega+1}} \otimes R_T) \quad (22)$$

这里, $R_{\omega+1}$ 是一个 $(\omega+1) \times (\omega+1)$ 矩阵, 其中 R_{ij} 满足: 若 $i=1$ 和 $j=1$ 时 $R_{ij}=1$; 若 $i=\omega+1$ 和 $j=\omega+1$ 时 $R_{ij}=1$; $i=j$ 且 $i, j \neq 1$ 且 $i, j \neq \omega+1$ 时 $R_{ij}=2$; 若 $|i-j|=1$ 时, $R_{ij}=-1$; 其余情况 $R_{ij}=0$ 。 x 本质是一个高斯马尔可夫随机场, 矩阵 Q 是奇异的。对于 $Z=0, 1, \dots, \omega-1$ 和 $t=1, 2, \dots, T-1$, 随机变量 x_{zt} 的完全条件分布是正态分布, 其均值为:

$$\frac{1}{4} \{ \rho_{age} (x_{z-1,t} + x_{z+1,t}) + \rho_{year} (x_{z,t-1} + x_{z,t+1}) \}$$

方差为 $\frac{1}{4\tau}$ 。参数 ρ_{age} 和 ρ_{year} 分别控制跨年龄组和跨年份的死亡概率的关联性, 捕捉了年龄纬度和日历时间纬度的相关性, 被期望为不同值, 为保证能够唯一确定模型的参数, 假设 $\rho_{age} + \rho_{year} = 2$ 。

2.5. 贝叶斯推断

此模型的似然函数由等式(13)右侧各项的乘积给出:

$$\gamma(d | x) = \prod_{t=1}^T \prod_{z=0}^{\omega} \binom{n_{zt}}{d_{zt}} p_{zt}^{d_{zt}} [1 - p_{zt}]^{n_{zt} - d_{zt}} \quad (23)$$

其中 d 是一个 $(\omega+1)T$ 维向量, 元素为 d_{zt} 。模型参数表示为 $\theta = (b, \rho_{age}, \tau)$, 构建 MCMC 算法用于从联合后验分布中采样, 其密度为:

$$\gamma(\theta, x | d) \propto \gamma(\theta) \gamma(x | \theta) \gamma(d | x) \quad (24)$$

其中, $\gamma(\theta)$ 是参数的先验分布的密度, $\gamma(x | \theta)$ 是均值为 μ 精度矩阵为 Q 的 $(\omega+1)T$ 变量高斯分布的密度。模型参数 b 、 ρ_{age} 和 τ 的完全条件分布中采样。对于 ρ_{age} , 采用随机游走的 Metropolis-Hastings 步骤。 $\gamma(x | \theta, d)$ 中采样, 使用基于梯度的辅助 MCMC 采样器首先构建辅助变量 $u_k \in \mathbb{R}^{(\omega+1)T}$ 是从高斯分布中抽取的:

$$N \left[x + \left(\frac{\delta}{2} \right) \nabla \log \{ \gamma(d | x, \theta) \}, \frac{\delta}{2} I_{(\omega+1)T} \right] \quad (25)$$

计算状态 x 和 θ 下对数似然的梯度 $\nabla \log \gamma(d | \theta, x)$, 有

$$q(x_{prop} | x) \propto \left(\frac{\delta}{2} \right)^{(\omega+1)T} e^{-\frac{1}{\delta} \|x_{prop} - u_x\|^2} \gamma(x_{prop} | \theta) \quad (26)$$

从(26)提取新的值 x_{prop} 中, 并且选择 x_{prop} 以 Metropolis-Hastings 接受概率 $\min(1, \alpha)$ 被接受, 其中

$$\alpha = \frac{\gamma(d | x_{prop}, \theta)}{\gamma(d | x, \theta)} \exp \{ f(u_x, x_{prop}) - f(u_x, x) \} \quad (27)$$

且

$$f(u_x, x) = \left(u_x - x - \left(\frac{\delta}{4} \right) \nabla \log \gamma(d | \theta, x) \right)^T \nabla \log \gamma(d | \theta, x) \quad (28)$$

调整参数 δ 以实现 50% 到 60% 的接受率。对于每个 z 和 k , 目标为预测在未来时间点 $t = T+1, T+2, \dots, T+k$ 的死亡概率 p_z , 通过向量 $x^* = (x_{T+1}, \dots, x_{T+k})$ 表达。所需的预测密度为:

$$\gamma(x^* | d) = \int \gamma(x^* | \theta, x) \gamma(\theta | x, d) \gamma(x | d) dx d \theta \quad (29)$$

3. 实证分析

3.1. 数据分析及预处理

在本研究中, 从人类死亡数据库中选取了智利和日本的死亡率数据作为研究对象。这些国家代表了不同的经济和社会发展水平, 能够提供具有广泛适用性的分析结果。从公开的死亡率数据库中提取了 1995 年至 2020 年的死亡率数据。该数据涵盖了 0 到 89 岁年龄段的死亡率, 并按年份和年龄段进行分类, 以确保每个年龄组的死亡率数据充分且具有代表性。

为了后续模型训练和评估提供可靠的基础以确保数据的质量和一致性, 首先, 数据从人类死亡率数据库中读取数据。接着, 对“年龄”进行处理, 使用正则表达式移除可能存在的“+”符号, 并将该列转换为整数类型, 以确保年龄数据的格式统一。随后, 对数据进行筛选, 仅保留“年份”列在 1995 年至 2020 年之间且“年龄”列小于等于 89 的数据。这一步骤的目的是排除异常值或不符合要求的数据, 确保数据的合理性和有效性。接下来, 处理“死亡率”列中的异常值。由于后续步骤需要对死亡率取对数, 而死亡率为 0 时会导致对数计算无意义, 因此将死亡率为 0 的值替换为一个非常小的正数。然后, 对“死亡率”进行对数变换, 生成新的“对数死亡率”列。这一步骤的目的是将偏态分布的数据转换为接近正态分布的形式, 使其更适合线性模型的假设。为了消除不同特征之间的量纲差异, 数据被进一步缩放。通过最小-最大缩放方法, 将“年份”和“年龄”列的值缩放到 [0, 1] 的范围内。最后, 数据被分为训练集和测试集。这一步骤的目的是确保模型能够在未见过的数据上进行评估, 从而验证其泛化能力。

图 3、图 4 分别是 Heligman-Pollard 模型和 GMRF 在预测 2018 年日本死亡率中的表现。

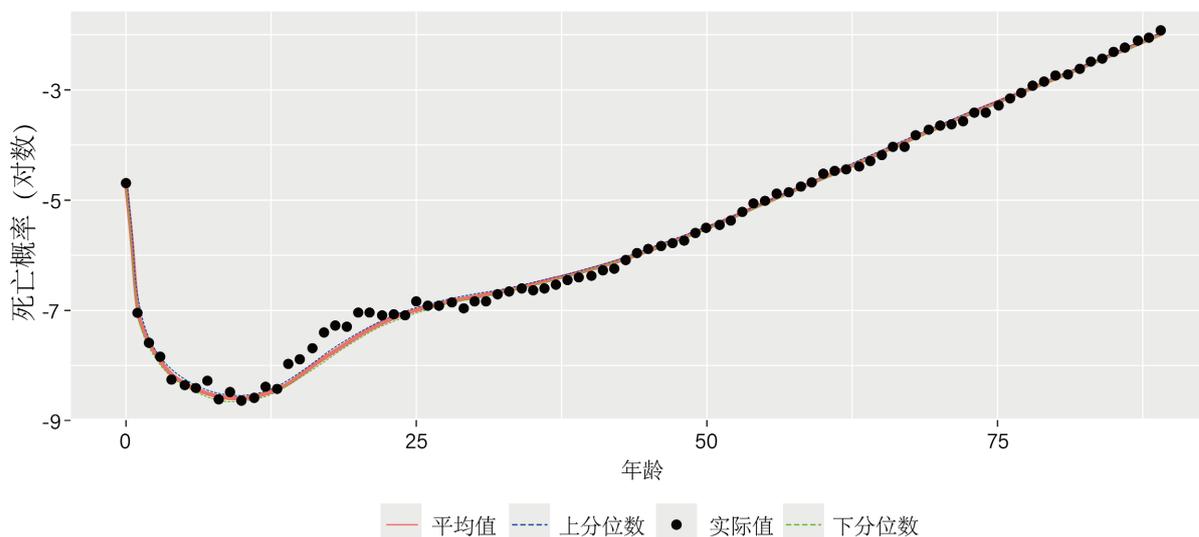


Figure 3. Diagram of HP model result

图 3. HP 模型结果图

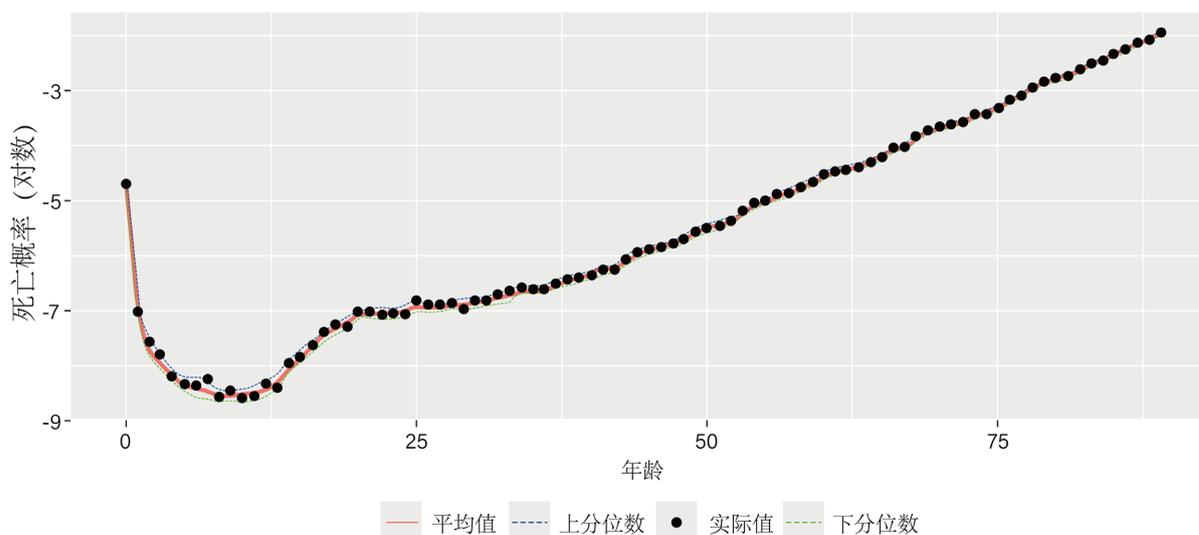


Figure 4. Diagram of GMRF model result

图 4. GMRF 模型结果图

为了比较四种模型的效果, 最终选择使用智利的预测数据进行对比分析。在经典的 LC 模型中, 我们分别对全年龄组(0~89 岁)和老年组(40~89 岁)进行了预测。而在 DNN 模型中, 数据按照年份划分, 2001 年之前的数据用于训练集, 2001 年及之后的数据则作为测试集。

3.2. 模型评估

本文采用的评价指标有均方根误差(RMSE)、平均绝对误差(MAE)和平均绝对百分比误差(MAPE), 表达式为:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (26)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (27)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (28)$$

其中 y_i 为真实值, \hat{y}_i 为预测值, n 为样本总数。

3.3. 模型对比分析

利用测试集对组合模型进行测试, 将建立之后不断调优的动态模型和神经网络模型得出最终的预测结果, 对比结果如表 1 所示。

Table 1. Model parameter settings

表 1. 模型对比结果

	RMSE	MAE	MAPE (%)
LC	0.1687	0.1306	13.0862
LC(局部)	0.0054	0.0023	10.6682
DNN	0.0150	0.0088	17.224
HP	0.0014	0.0006	8.4098
GMRF	0.0006	0.0002	3.8572

4. 结论

Lee-Carter 模型单因子假设限制了其对复杂死亡率模式的捕捉能力。实证结果显示其预测误差较高 (RMSE: 0.1687, MAE: 0.1306, MAPE: 13.0862), 局部数据表现有所改善 (RMSE: 0.0054, MAE: 0.0023, MAPE: 10.6682) 但仍未达到最优。DNN 模型通过多层神经元结构捕捉死亡率数据中的非线性关系, 采用梯度下降法和反向传播算法优化参数。尽管其理论拟合能力强, 但受数据量和超参数调优的影响, 实证预测误差 (RMSE: 0.0150, MAE: 0.0088, MAPE: 17.224) 高于 Heligman-Pollard 和 GMRF 模型。Heligman-Pollard 模型通过引入动态分量和随机游走结构, 放宽了参数平稳性假设以更好地捕捉死亡率的动态变化。采用贝叶斯推断和 Metropolis-within-Gibbs 采样器, 其预测精度较高 (RMSE: 0.0014, MAE: 0.0006, MAPE: 8.4098), 在描述不同年龄段死亡率模式方面表现优异。GMRF 模型基于梯度的辅助 MCMC 采样器提高了采样效率, 预测误差最低 (RMSE: 0.0006, MAE: 0.0002, MAPE: 3.8572), 能够有效捕捉时空依赖性, 为死亡率预测提供了可靠框架。

总得来说, GMRF 模型预测精度最优, 适用于高精度预测和复杂时空依赖性场景; Heligman-Pollard 模型适合描述死亡率动态变化; DNN 模型适用于数据充足且需捕捉非线性关系的场景, 但需注意超参数调优; Lee-Carter 模型简洁且解释性强, 但在复杂数据上表现有限。未来改进方向可结合 GMRF 的动态建模能力和 DNN 的非线性拟合能力开发混合模型, 或引入社会经济因素等外部变量, 可能进一步提升预测精度。

基金项目

国家自然科学基金项目 (NO.12061066); 甘肃省自然科学基金项目 (NO. 20JR5RA528)。

参考文献

- [1] Lee, R.D. and Carter, L.R. (1992) Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical*

Association, **87**, 659-671. <https://doi.org/10.2307/2290201>

- [2] Li, N., Lee, R. and Tuljapurkar, S. (2004) Using the Lee-Carter Method to Forecast Mortality for Populations with Limited Data. *International Statistical Review*, **72**, 19-36. <https://doi.org/10.1111/j.1751-5823.2004.tb00221.x>
- [3] 黄顺林, 王晓军. 加入出生年效应的死亡率预测及其在年金系数估计中的应用[J]. 统计与信息论坛, 2010(5): 81-86.
- [4] Dellaportas, P., Smith, A.F.M. and Stavropoulos, P. (2001) Bayesian Analysis of Mortality Data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **164**, 275-291. <https://doi.org/10.1111/1467-985x.00202>
- [5] Alexopoulos, A., Dellaportas, P. and Forster, J.J. (2018) Bayesian Forecasting of Mortality Rates by Using Latent Gaussian Models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **182**, 689-711. <https://doi.org/10.1111/rssa.12422>
- [6] Chandra, C.E., Abdullah, S. and Devila, S. (2022) Estimating Indonesian Complete Life Table and Fair Annual Pure Premium Range from Abridged Life Table with Bayesian Method and Bootstrapping. *International Journal on Advanced Science, Engineering and Information Technology*, **12**, 2226-2236. <https://doi.org/10.18517/ijaseit.12.6.14947>
- [7] Hainaut, D. (2018) A Neural-Network Analyzer for Mortality forecast. *ASTIN Bulletin*, **48**, 481-508. <https://doi.org/10.1017/asb.2017.45>
- [8] Nigri, A., Levantesi, S. and Aburto, J.M. (2022) Leveraging Deep Neural Networks to Estimate Age-Specific Mortality from Life Expectancy at Birth. *Demographic Research*, **47**, 199-232. <https://doi.org/10.4054/demres.2022.47.8>