

Article

A New Class of Counting Distributions Embedded in the Lee–Carter Model for Mortality Projections: A Bayesian Approach

Yaser Awad ¹, Shaul K. Bar-Lev ² and Udi Makov ^{3,*}

¹ Sakhnin Academic College, Galilee St. 100, Sakhnin 30810, Israel; awad_y@netvision.net.il

² Faculty of Industrial Engineering and Technology Management, HIT—Holon Institute of Technology, 52 Golomb Street, Holon 5810201, Israel; shaulb@hit.ac.il

³ Actuarial Research Center, University of Haifa, 199 Aba Khoushy Ave. Mount Carmel, Haifa 3498838, Israel

* Correspondence: udimakov@gmail.com

Abstract: The Lee–Carter model, the dominant mortality projection modeling in the literature, was criticized for its homoscedastic error assumption. This was corrected in extensions to the model based on the assumption that the number of deaths follows Poisson or negative binomial distributions. We propose a new class of families of counting distributions, namely, the ABM class, which belongs to a wider class of natural exponential families. This class is characterized by its variance functions and contains the Poisson and the negative binomial distributions as special cases, offering an infinite class of additional counting distributions to be considered. We are guided by the principle that the choice of distribution should be made from a pool of distributions as large as possible. To this end, and following a data mining approach, a training set of historical mortality data of the population could be modeled using the ABM’s rich choice of distributions, and the chosen distribution should be the one that proved to offer superior projection results on a test set of mortality data. As an alternative to parameter estimation via the singular value decomposition used in the classical Lee–Carter model, we adopted Bayesian estimation, harnessing the Markov Chain Monte Carlo methodology. A numerical study demonstrates that when fitting mortality data using this new class of distributions, while traditional distributions may provide desirable projections for some populations, for others, alternative distributions within the ABM class can potentially produce superior results for the entire population or particular age groups, such as the oldest-old.

Keywords: Lee–Carter; counting distributions; mortality projections; natural exponential family



Citation: Awad, Yaser, Shaul K. Bar-Lev, and Udi Makov. 2022. A New Class of Counting Distributions Embedded in the Lee–Carter Model for Mortality Projections: A Bayesian Approach. *Risks* 10: 111. <https://doi.org/10.3390/risks10060111>

Academic Editors: Ermanno Pitacco and Annamaria Olivieri

Received: 24 March 2022

Accepted: 18 May 2022

Published: 27 May 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The seminal paper by Lee and Carter (1992) (LC) introduced a model which is one of the most well-known and widely applied models for forecasting mortality rates. Within this model, the time series of the log mortality rates, $\ln m_{xt}$, of each age is described by an age-specific intercept α_x plus a common trend k_t for all age groups multiplied by an age-specific coefficient β_x ,

$$\ln m_{xt} = \alpha_x + \beta_x k_t + \varepsilon_{xt}.$$

The error term ε_{xt} is assumed to be distributed with a mean 0 and variance σ_ε^2 , reflecting influences missed by the model. The age and time-specific mortality rate m_{xt} is calculated as (D_{xt}/E_{xt}) , where D_{xt} denotes the number of deaths in a population at age x , $x = 1, 2, \dots, P$, and at time t , $t = 1, 2, \dots, T$, and E_{xt} is the exposure to the risk of death. To ensure the identifiability of model parameters, constraints are imposed such that the sum of β_x over age is 1 and the sum of k_t over time is 0. To forecast mortality rates into the future, a simple random walk with drift is proposed for k_t :

$$k_t = k_{t-1} + \theta + w_t.$$

The homoscedastic error assumption of the Lee–Carter model was criticized for its limiting impact on predictions (Brouhns et al. 2002; Danesi et al. 2015; Idrizi 2018). This led to the introduction of the Poisson log-bilinear LC-type model (Brouhns et al. 2002), which, in contrast, is intrinsically heteroscedastic, namely:

$$D_{xt} \sim \text{Poisson}(\mu_{xt}), \mu_{xt} = E_{xt}m_{xt}.$$

Here, the number of dead is directly modelled by a Poisson distribution, whose parameter is estimated by maximum likelihood estimation (MLE). This alternative approach gained momentum and alternative discrete distributions were proposed. In particular, a binomial distribution was proposed by Wang and Lu (2005) and a negative binomial distribution was suggested by Delwarde et al. (2007) and Renshaw and Haberman (2008). See Azman and Pathmanathan (2022) for further discussion of these distributions within the GLM framework.

The Lee–Carter model has been widely used for many purposes (Shair et al. 2018), such as forecasting mortality reduction factors, assessing the adequacy of retirement income, population projections and the projection of mortality trends for the oldest-old (older than 80, and in some sources 85). This age group is of considerable interest for policymakers as it is destined to grow as a proportion of the entire population and can outstrip existing infrastructures' capacity (Buettner 2002). This is a fairly recent phenomenon. In Canada, for instance (Legare et al. 2015), the 21st century brought about the most significant gain in life expectancy at age 85 (7.79% for women and 9.93% for men). Clearly, policies need to be devised that can meet people's special needs in what is called the fourth age (Baltes and Smith 2003), and accurate mortality projection for this age group is a must. We shall, therefore, focus on the adequacy of potential underlying discrete distribution functions to produce accurate mortality projections using the Lee–Carter model for this age group. This will be discussed in Section 4.

In essence, while the above-cited papers relied on popular and commonly used discrete distributions, one cannot say that one particular distribution is universally superior. Indeed, it is entirely plausible to assume that there is a winning distribution for any given population or even for a specific population at a certain age range. Ideally, one should consider a rich class of family of counting distributions, much richer than the two already suggested, and use the data to pick the most suitable distribution for the population under study. This paper proposes an infinitely countable set of families of counting distributions, where the Poisson, negative binomial and Abel families of distributions are special cases. Our aim is to study this family, incorporate it into the framework of the LC model and use real data to seek the most suitable distribution for mortality projection. While there is little doubt that the distributions discussed above could prove adequate for specific populations or age groups, other distributions within the suggested family could have the upper hand.

The paper is organized as follows. Section 2 presents the new class of counting distributions. Section 3 is devoted to the new class and its Bayesian framework. Section 4 (divided into Section 4.1: Methods and Section 4.2: Results) reports a numerical study in which superior members of this class are chosen for mortality projections of the oldest-old in three populations. Finally, Section 5 offers a discussion.

2. A New Class of Counting Distributions on the Set of Nonnegative Integers \mathbb{N}_0

The new class of families of counting distributions on the non-negative integers belongs to a wider class of natural exponential families (NEFs), characterized by their variance functions (VFs). In order to comprehend this class we decompose this section into subsections. We first present some preliminaries on NEFs and their associated VFs. We then introduce a class of NEFs having polynomial structure and then suggest the new class of families of counting distributions, named ABM, first introduced by Awad et al. (2016), where the class was defined and its usefulness for mortality projections was preliminary sketched. Furthermore, such a class has been investigated by Bar-Lev and Ridder (2021a, 2021b) from

a classical frequency approach and has shown superiority with respect to various metrics or goodness-of-fit tests for different count datasets (for further details see item 6 in Section 2.3).

2.1. NEFs—Some Preliminaries

The following preliminaries are mainly taken from Letac and Mora (1990) and are briefly presented here for completeness.

Let ν be a non-Dirac positive Radon measure on \mathbb{R} , and $L(\theta) = \int e^{\theta x} \nu(dx)$ its Laplace transform. Assuming that $\Theta = \text{int}\{\theta \in \mathbb{R} : L(\theta) < \infty\} \neq \emptyset$, then the NEF generated by ν is defined by the probability distributions

$$\mathcal{F} = \left\{ F_\theta : F_\theta(\nu(dx)) = e^{\theta x - \kappa(\theta)} \nu(dx), \theta \in \Theta \right\}, \quad (1)$$

where $\kappa(\theta) = \log L(\theta)$, the cumulant transform of ν , is strictly convex and real analytic on Θ . If X_θ represents a r.v. having distribution F_θ of the form given in (1) then the expectation and variance of X_θ are given, respectively, by $E(X_\theta) \doteq m = \kappa'(\theta)$ and $V(X_\theta) = \kappa''(\theta)$ where $m = \kappa'(\theta)$ is strictly monotone and thus its inverse, say, $\theta = \psi(m)$, $m \in M = \kappa'(\Theta)$ is well defined. The set M of all means of (1) is called the mean parameter space of \mathcal{F} . The variance of F_θ can be expressed in terms of m by $V(m) = \kappa''(\theta) = \kappa''(\psi(m))$. The pair (V, M) is called the VF of \mathcal{F} and it uniquely determines \mathcal{F} within the class of NEFs. For example, (m, \mathbb{R}^+) and (m^2, \mathbb{R}^+) are, respectively, the VFs of the Poisson and exponential NEFs and are uniquely determined by them.

2.2. The Mean Value Parametrization of NEFs

As indicated above, the VF of an NEF \mathcal{F} uniquely determines F within the class of NEFs. Let (V, M) be a given VF of an NEF \mathcal{F} generated by ν . Then, simple calculations show both $\theta = \psi(m)$ and the cumulant transform $\kappa(\theta) = \kappa(\psi(m))$ of ν can be expressed in terms of m as:

$$\theta = \psi(m) = \int \frac{dm}{V(m)} + c_1, \quad \kappa(\theta) = \kappa(\psi(m)) = \int \frac{m}{V(m)} dm + c_2, \quad (2)$$

where one needs to determine the constants c_1 and c_2 so that F_θ , $\theta \in \Theta$, constitutes a probability distribution (not an easy task). Accordingly, a mean value parametrization of an NEF \mathcal{F} generated by a measure ν is given by:

$$\mathcal{F} = \left\{ \exp\{\psi(m)x - \kappa(\psi(m))\} \nu(dx), m \in M \right\}. \quad (3)$$

Such a representation of \mathcal{F} is more natural as it is expressed in terms of the mean m rather than a somewhat artificial parameter θ . A comprehensive description of NEFs in terms of their mean value representation is reviewed in Bar-Lev and Kokonendji (2017).

Remark 1. *The task of computing the constants c_1 and c_2 is not simple and might be rather cumbersome. However, from a Bayesian perspective, when (3) is used as a prior distribution on m , then in the calculation of the respective posterior distribution, such constants are cancelled out (as the likelihood function is the only relevant component). As this paper is concerned with a Bayesian framework, one can assume without any loss of generality that $c_1 = c_2 = 0$. Henceforth, we indeed assume so.*

2.3. Polynomial VFs of Counting NEFs Supported on the Set of Nonnegative Integers \mathbb{N}_0

The innovative and breakthrough Proposition 4.4 of Letac and Mora (1990, p. 13) provided conditions under which a given VF (V, M) is associated with a counting NEF F supported on the set of non-negative integers \mathbb{N}_0 , i.e., where all members of F are composed of counting distributions on \mathbb{N}_0 . They provided general examples of two classes of VFs which fulfill the premises of their Proposition 4.4 and thus their associated NEFs'

distributions are supported on the non-negative integers. One of these two classes has the form:

$$V(m) = m \prod_{i=1}^k \left(1 + \frac{m}{p_i}\right), p_i > 0, i = 1, \dots, k, k \in \mathbb{N}_0, M = \mathbb{R}^+, \text{ where } \prod_{i=1}^0 \doteq 1. \quad (4)$$

They proved that such VFs constitute counting NEFs supported on \mathbb{N}_0 , namely, counting distributions with non-negative integer support. Moreover, their Proposition 4.4 enables to compute (at least theoretically and numerically) the corresponding measure ν (we skip details as they are irrelevant for our Bayesian framework analysis). Note that the two special cases of (4) with $k = 0$ and $k = 1$ correspond, respectively, to the Poisson and negative binomial NEFs. However, the general setting (4) for $k \geq 3$ does not allow an explicit calculation of $\theta = \psi(m)$ and $k(\theta) = k(\psi(m))$ in (2), implying that the mean value parametrization of the corresponding NEFs in the form (3) is not explicitly expressible in terms of m and thus becomes useless for any practical consideration.

2.4. A New Class of Polynomial VFs—The ABM NEFs

As we already noted, the fact that a given pair (V, M) is known to be a VF of some NEF does not necessarily enable the construction of the corresponding mean value parametrization (3), as in most cases the integrals for $\psi(m)$ and $k(\psi(m))$ in (2) are not explicitly expressible analytically in closed forms, and indeed, this is the situation for the class (4) in its general form. Consequently, one needs to search for subclasses of (4) for which the integrals in (2) can be computed explicitly. One such special subclass takes the above point into consideration. Indeed, by taking in (4) the special case where

$$p_1 = p_2 = \dots = p_k,$$

and denoting

$$p_2 \doteq k \in \mathbb{N}_0,$$

we obtain a subclass of (4) with VFs with the form:

$$(V, M) = m \left(1 + \frac{m}{p_1}\right)^{p_2}, \mathbb{R}^+, p_1 > 0, p_2 \in \mathbb{N}_0. \quad (5)$$

As (5) is a subclass of (4) and (4) satisfies the premises of Proposition 4.4 of Letac and Mora (1990) it follows that the subclass (5) are VFs associated with counting NEFs supported on the non-negative integers.

The subclass of VFs in (5) (hereafter called the ABM class) was first introduced by Awad et al. (2016) who showed that the corresponding $\psi(m)$ and $k(\psi(m))$ (calculated from (2)) have, as opposed to the general form in (4), the following closed forms (the exact proof details appear in Bar-Lev and Kokonendji 2017):

$$\theta = \psi(m) = \ln \frac{m}{p_1 + m} + \sum_{i=1}^{p_2-1} \frac{1}{i} \frac{p_1^i}{(p_1 + m)^i} + c_1, \text{ where } \sum_{i=1}^0 = 0,$$

and

$$\kappa(\psi(m)) = -\frac{p_1^{p_2}}{(p_2 - 1)(m + p_1)^{p_2-1}} + c_2.$$

Thus, its mean value parametrization is given by the probability distribution:

$$F(m, \nu(dx)) = \exp \left\{ x \left[\ln \frac{m}{p_1 + m} + \sum_{i=1}^{p_2-1} \frac{1}{i} \frac{p_1^i}{(p_1 + m)^i} + c_1 \right] + \frac{p_1^{p_2}}{(p_2 - 1)(m + p_1)^{p_2-1}} + c_2 \right\}, m \in \mathbb{R}^+, p_1 > 0, p_2 \in \mathbb{N} \quad (6)$$

where hereafter we denote this probability distribution by $ABM(p_1, p_2)$, where p_1 is a positive real number and p_2 is a non-negative integer. (For a classical frequency approach, the constants c_1 and c_2 have been computed by [Bar-Lev and Ridder 2021b](#)). However, as noted above, for a Bayesian framework they are cancelled out when computing the posterior distribution and thus can be taken to be $c_1 = c_2 = 0$ without any loss of generality).

Note that the ABM class of VFs $\left\{m\left(1 + \frac{m}{p_1}\right)^{p_2}, p_1 > 0\right\}_{p_2 \in \mathbb{N}_0}$, or alternatively, the corresponding class $\{ABM(p_1, p_2)\}_{p_2 \in \mathbb{N}_0}$ of NEFs is composed of an infinitely countable set of families of counting NEFs supported on the non-negative integers. As special cases, this class contains the Poisson NEF ($p_2 = 0$), the negative binomial NEF ($p_2 = 1$) and the Abel NEF ($p_2 = 2$), (c.f., [Letac and Mora 1990](#), p. 31; [Bar-Lev and Ridder 2019](#), for applications to car accident claims of a Swedish insurance company dataset).

Summarizing, this ABM NEF has the following features:

1. It is a class of counting distributions supported on the non-negative integers;
2. It is overdispersed as $V(m)/m > 1$;
3. It allows a mean value parameterization in a closed form;
4. It is infinitely divisible, which allows the construction of an exponential dispersion model (EDM) with dispersion parameter space equal to \mathbb{R}^+ . EDMs are used to describe the error distribution in generalized linear models (see [Jorgensen 1987, 1997](#));
5. p_1 is an unknown parameter to be estimated (see next section). $p_2 \in \mathbb{N}_0$ is a parameter governing the particular model within the ABM class and is considered to be a decision variable (note that different values of p_2 determine different ABM NEFs). Accordingly, for given national datasets (i.e., those of US, Ireland and Ukraine), the goal will be to locate that value of p_2 , which minimizes a respective RMSE (see in the sequel). However, due to the rather cumbersome and intractable structure of the ABM probabilities (or likelihood) in (6) and the fact that the larger the p_2 , the larger the number of elements in the summands appearing in (6), no analytic solution for an optimal p_2 is feasible at all for achieving such a goal. Consequently, only numerical search algorithms are plausible. The search starts with $p_2 = 0$ (the Poisson NEF), $p_2 = 1$ (the negative binomial NEF), $p_2 = 2$ (the Abel NEF) and so on;
6. As already noted, the ABM class $\{ABM(p_1, p_2)\}_{p_2 \in \mathbb{N}_0}$ is composed of infinitely countable set of families of counting NEFs supported on the non-negative integers and thus can also be used to model real datasets by employing the classical frequency approach (and not only Bayesian). Indeed, the ABM class has been compared in [Bar-Lev and Ridder \(2021a, 2021b\)](#) with other common counting probability models (such as Poisson-inverse Gaussian distribution, new logarithmic distribution, an exponentiated discrete Lindley distribution) for various real count datasets stemming from automobile insurance claims, marketing, biometry, health, and social sciences (none of which is related to mortality projections). Members of the ABM counting class have shown superiority with respect to various metrics for goodness-of-fit tests (chi-squared test, Akaike information criterion (AIC), root-mean-square error (RMSE) and Kullback–Leibler divergence (KL)), and provided a much better fit for each of the datasets considered (more details can be found in [Bar-Lev and Ridder 2021b](#)).

3. ABM Based LC Model and its Bayesian Framework

As an alternative to parameter estimation via the singular value decomposition used in the classical LC model or the MLE in the cases discussed above, we adopt the Bayesian approach which offers advantages succinctly expressed in [Antonio et al. \(2015\)](#): a. The calibration and forecast steps are combined, which leads to more consistent estimates of the period effects; b. The Bayesian approach provides a natural framework for incorporating parameter uncertainty in mortality forecasts, which is relevant—for example—in the new insurance regulatory framework of Solvency II. The Bayesian approach allows adequate handling of small populations and missing data. Like [Czado et al. \(2005\)](#) and [Pedroza \(2006\)](#), we harness the power of the Markov Chain Monte Carlo (MCMC) methodology to

estimate the model parameters and execute mortality projection. We note that the interest in Bayesian solutions in the context of mortality projections has recently gained momentum (Ellison et al. 2020; Graziani 2020; Hilton et al. 2019; Hunt and Blake 2020; Kogure et al. 2019; Liu et al. 2020; Njenga and Sherris 2020; Wong et al. 2018).

Suppose the number of deaths D_{xt} in a population at age x and time t is distributed as follow:

$$D_{xt} \sim ABM(p_1, p_2)(\mu_{xt}), \mu_{xt} = E_{xt}m_{xt}, m_{xt} = \exp(\alpha_x + \beta_x k_t),$$

where

$$\alpha = (\alpha_{x_{\min}}, \dots, \alpha_{x_{\max}})', \beta = (\beta_{x_{\min}}, \dots, \beta_{x_{\max}})', k = (k_{t_{\min}}, \dots, k_{t_{\max}}).$$

Bayesian estimation of the unknown parameters α, β, k and p_1 are based on the joint posterior distribution function of α, β, k and p_1 given (E_{xt}, D_{xt}) , when $x = x_{\min}, x_{\min} + 1, \dots, x_{\max}, t = t_{\min}, t_{\min} + 1, \dots, t_{\max}$. The first step in the Bayesian estimation is to determine the prior probability functions for these parameters.

The prior distribution for k_t and θ

Let $k_t = k_{t-1} + \theta + w_t$, and let $w_t \sim N(0, \sigma_w^2)$ and hence $k_t \sim N(\theta, t\sigma_w^2)$. we assume $\sigma_w^{-2} \sim \text{gamma}(a_k, b_k)$ and $\theta \sim N(\theta_0, \sigma_\theta^2)$. The hyper-parameters θ_0, a_k, b_k and σ_θ^2 are arbitrary initial values.

The prior distribution for β_x

We assume $\beta_x \sim N(0, \sigma_\beta^2) \forall x$, where $\sigma_\beta^{-2} \sim \text{gamma}(a_\beta, b_\beta)$. The hyper-parameters a_β, b_β are arbitrary initial values.

The prior distribution for α_x

We suppose that the prior distribution of $\alpha_x \sim N(\alpha_{0x}, \sigma_\alpha^2) \forall x$, where $\sigma_\alpha^{-2} \sim \text{gamma}(a_\alpha, b_\alpha)$. The hyper-parameters α_{0x}, a_α and b_α are arbitrary initial values.

The prior distribution for p_1

We let $p_1 \sim \text{gamma}(a_{p_1}, b_{p_1})$. The hyper-parameters a_{p_1}, b_{p_1} are arbitrary initial values.

MH (Metropolis–Hastings) Algorithm for Estimating the Parameters α, β, k and p_1

Suppose the D_{xt} 's are independent random variables, which are distributed as (6) and $g(\Xi)$ is the joint prior distribution of the unknown parameters $\Xi = (\alpha, \beta, k, p_1)'$. Then, the posterior distribution of Ξ , given all available data $D = \{d_{xt}\}$ and p_2 , can be represented as follows:

$$f(\Xi | D, p_2) \propto \prod_x \prod_t \exp \left[d_{xt} \left(\ln \frac{E_{xt}m_{xt}}{p_1 + E_{xt}m_{xt}} + \sum_{i=1}^{p_2-1} \frac{1}{i} \frac{p_1^i}{(p_1 + E_{xt}m_{xt})^i} \right) + \frac{p_1^{p_2}}{(p_2-1)(p_1 + E_{xt}m_{xt})^{p_2-1}} \right] \times g(\Xi).$$

See Appendix A for the marginal posterior distributions of α, β, k and p_1 . We now describe the estimation of α, β, k and p_1 using the MH, conditioned on the data and all other parameters at their respective iterations. The superscript denotes the iteration number of the parameter of interest.

Estimation of k_t using the MH algorithm

Let the marginal posterior distribution of k_t be $f(k_t | D, \alpha, \beta, k_{-t}, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1)$. The estimation of k_t , is achieved by the following steps, where.

1. Draw k_t^* from the proposal density function $N(k_t^{(i)}, \sigma_t^2)$, such that σ_t^2 is assumed known;
2. Calculate the following probability:

$$\Psi(k_t^{(i)}, k_t^*) = \min \left(1, \frac{f(k_t^* | D, \alpha, \beta, k_{-t}^{(i)}, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1)}{f(k_t^{(i)} | D, \alpha, \beta, k_{-t}^{(i)}, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1)} \right),$$

where $k_{-t} = (k_{t \min}, \dots, k_{t-1}, k_{t+1}, \dots, k_{t \max})'$;

3. Draw a value u from uniform probability function in range $U(0, 1)$ and decide in accordance with the following formula:

$$\left\{ \begin{array}{l} \text{if } u \leq \Psi(k_t^{(i)}, k_t^*) \text{ then } k_t^{(i+1)} = k_t^* \\ \text{if } u > \Psi(k_t^{(i)}, k_t^*) \text{ then } k_t^{(i+1)} = k_t^{(i)}; \end{array} \right\}$$

4. Going over all values of t , we have:

$$k^{(i+1)} = (k_{t \min}^{(i+1)}, \dots, k_t^{(i+1)}, k_{t+1}^{(i)}, \dots, k_{t \max}^{(i)})'$$

5. Transforming $k^{(i+1)}$ and $\alpha^{(i)}$ to assure identifiably:

$$k^{(i+1)} - \bar{k} \rightarrow k^{(i+1)}, \alpha^{(i)} + \beta^{(i)}\bar{k} \rightarrow \alpha^{(i)},$$

where

$$\bar{k} = \frac{1}{T} \left(\sum_{j \leq t} k_j^{(i+1)} + \sum_{j > t} k_j^{(i)} \right);$$

6. Repeat steps 1 to 5.

Estimation of β_x using MH algorithm

Let the marginal posterior distribution of β_x be $f(\beta_x | D, \alpha, \beta_{-x}, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1)$. The estimation of β_x is achieved by the following steps.

1. Draw β_x^* from the proposal density function $N(\beta_x^{(i)}, \sigma_\beta^2)$, such that σ_β^2 is assumed to be known;
2. Calculate the following probability:

$$\Psi(\beta_x^{(i)}, \beta_x^*) = \min \left(1, \frac{f(\beta_x^* | D, \alpha, \beta_{-x}^{(i)}, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1)}{f(\beta_x^{(i)} | D, \alpha, \beta_{-x}^{(i)}, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1)} \right),$$

where

$$\beta_{-x} = (\beta_{x \min}, \dots, \beta_{x-1}, \beta_{x+1}, \dots, \beta_{x \max})'$$

3. Draw a value u from uniform probability function in range $U(0, 1)$ and decide in accordance with the following formula:

$$\left\{ \begin{array}{l} \text{if } u \leq \Psi(\beta_x^{(i)}, \beta_x^*) \text{ then } \beta_x^{(i+1)} = \beta_x^* \\ \text{if } u > \Psi(\beta_x^{(i)}, \beta_x^*) \text{ then } \beta_x^{(i+1)} = \beta_x^{(i)}; \end{array} \right\}$$

4. Going over all values of x , we have:

$$\beta^{(i+1)} = (\beta_{x \min}^{(i+1)}, \dots, \beta_x^{(i+1)}, \beta_{x+1}^{(i)}, \dots, \beta_{x \max}^{(i)})'$$

5. Transforming $k^{(i+1)}$ and $\beta^{(i+1)}$ to assure identifiably:

$$\frac{\beta^{(i+1)}}{\beta_{sum}} \rightarrow \beta^{(i+1)}, k^{(i+1)} \times \beta_{sum} \rightarrow k^{(i+1)},$$

where

$$\beta_{sum} = \left(\sum_{j \leq x} \beta_j^{(i+1)} + \sum_{j > x} \beta_j^{(i)} \right);$$

6. Repeat steps 1 to 5.

Estimation of α_x using MH algorithm

Let the marginal posterior distribution of α_x be $f(\alpha_x | D, \alpha_{-x}, \beta, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1)$. The estimation of α_x , is achieved by the following steps;

1. Draw α_x^* from the proposal density function $N(\alpha_x^{(i)}, \sigma_\alpha^2)$, such that σ_α^2 is assumed known;
2. Calculate the following probability:

$$\Psi(\alpha_x^{(i)}, \alpha_x^*) = \min \left(1, \frac{f(\alpha_x^* | D, \alpha_{-x}^{(i)}, \beta, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1)}{f(\alpha_x^{(i)} | D, \alpha_{-x}^{(i)}, \beta, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1)} \right),$$

where

$$\alpha_{-t} = (\alpha_{x_{min}}, \dots, \alpha_{x-1}, \alpha_{x+1}, \dots, \alpha_{x_{max}})';$$

3. Draw a value u from uniform probability function in range $U(0, 1)$ and decide in accordance with the following formula:

$$\left\{ \begin{array}{l} \text{if } u \leq \Psi(\alpha_x^{(i)}, \alpha_x^*) \text{ then } \alpha_x^{(i+1)} = \alpha_x^* \\ \text{if } u > \Psi(\alpha_x^{(i)}, \alpha_x^*) \text{ then } \alpha_x^{(i+1)} = \alpha_x^{(i)} \end{array} \right\}.$$

4. Receiving $\alpha^{(i+1)}$ in $(i + 1)$ th iteration as follows:

$$\alpha_x^{(i+1)} = (\alpha_{x_{min}}^{(i+1)}, \dots, \alpha_x^{(i+1)}, \alpha_{x+1}^{(i)}, \dots, \alpha_{x_{max}}^{(i)});$$

5. Repeat steps 1 to 4.

Estimation of p_1 using MH algorithm

Let the marginal posterior distribution of p_1 be $f(p_1 | D, \alpha, \beta, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2)$, proportional to the product of the likelihood (6) and the gamma prior distribution of p_1 . The estimation of p_1 is achieved by the following steps;

1. Draw p_1^* from the probability function $gamma(\alpha_{p_1}, b_{p_1})$, such that α_{p_1} and b_{p_1} are hyperparameters and are assumed known;
2. Calculate the following probability:

$$\Psi(p_1^{(i)}, p_1^*) = \min \left(1, \frac{f(p_1^* | D, \alpha, \beta, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2)}{f(p_1^{(i)} | D, \alpha, \beta, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2)} \right);$$

3. Draw a value u from uniform probability function in range $U(0,1)$ and decide in accordance with the following formula:

$$\left\{ \begin{array}{l} \text{if } u \leq \Psi(p_1^{(i)}, p_1^*) \text{ then } p_1^{(i+1)} = p_1^* \\ \text{if } u > \Psi(p_1^{(i)}, p_1^*) \text{ then } p_1^{(i+1)} = p_1^{(i)} \end{array} \right\}.$$

4. Then receiving $p^{(i+1)}$ in $(i + 1)$ th iteration;
5. Repeat steps 1 to 4.

Estimation of $\theta, \sigma_\alpha^2, \sigma_\beta^2$ and σ_w^2 using the Gibbs sampler

The Gibbs sampler can be used for estimating $\theta, \sigma_\alpha^2, \sigma_\beta^2$ and σ_w^2 since the marginal posterior distribution of these parameters can be written explicitly (See: [Czado et al. 2005](#)). The following are the marginal posterior sampling distributions of each of these parameters, conditioned on the data and all other parameters at their respective iterations.

1. Sampling θ :

The posterior probability function of the parameter θ , is presented as follows:

$$f(\theta | D, \alpha, \beta, k, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1) = f(\theta | k, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1) \propto f(k | \theta, \sigma_w^2) f(\theta).$$

The prior probability function of the parameter θ is $N(\theta_0, \sigma_\theta^2)$, and the hyper-parameters θ_0 and σ_θ^2 are set by the user, hence the posterior probability function of the parameter is:

$$(\theta | k, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2) \sim N\left(\frac{\theta_0 \sigma_w^2}{T \sigma_\theta^2 + \sigma_w^2}, (T \sigma_\theta^2 + \sigma_w^2)^{-1}\right).$$

2. Sampling σ_α^2 :

The posterior probability function of the parameter σ_α^2 is presented as follows:

$$f(\sigma_\alpha^2 | D, \alpha, \beta, k, \theta, \sigma_\beta^2, \sigma_w^2, p_1) \propto f(\alpha | \sigma_\alpha^2) f(\sigma_\alpha^2).$$

The prior probability function of the parameter σ_α^2 such that $\sigma_\alpha^{-2} \sim \text{gamma}(a_\alpha, b_\alpha)$, and the hyper-parameters a_α and b_α are set by the user, so the posterior probability function of the parameter is:

$$(\sigma_\alpha^{-2} | D, \alpha, \beta, k, \theta, \sigma_\beta^2, \sigma_w^2, p_1) \sim \text{gamma}\left(a_\alpha + \frac{x_{\max}}{2}, b_\alpha + \frac{1}{2} \sum_{x=x_{\min}}^{x_{\max}} (\alpha_x - \bar{\alpha})^2\right),$$

where

$$\bar{\alpha} = \frac{1}{x_{\max}} \sum_{x=x_{\min}}^{x_{\max}} \alpha_x.$$

3. Sampling σ_β^2 :

The posterior probability function of the parameter σ_β^2 is presented as follows:

$$f(\sigma_\beta^2 | D, \alpha, \beta, k, \theta, \sigma_\alpha^2, \sigma_w^2, p_1) \propto f(\beta | \sigma_\beta^2) f(\sigma_\beta^2).$$

The prior probability function of the parameter σ_β^2 such that $\sigma_\beta^{-2} \sim \text{gamma}(a_\beta, b_\beta)$, and the hyper-parameters a_β and b_β are set by the user, so the posterior probability function of the parameter is:

$$(\sigma_\beta^{-2} | D, \alpha, \beta, k, \theta, \sigma_\alpha^2, \sigma_w^2, p_1) \sim \text{gamma}\left(a_\beta + \frac{x_{\max}}{2}, b_\beta + \frac{1}{2} \sum_{x=x_{\min}}^{x_{\max}} (\beta_x)^2\right).$$

4. Sampling σ_w^2 :

The posterior probability function of the parameter σ_w^2 , is presented as follows:

$$f(\sigma_w^2 | D, \alpha, \beta, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, p_1) \propto f(k | \sigma_w^2) f(\sigma_w^2).$$

The prior probability function of the parameter σ_w^2 such that $\sigma_w^{-2} \sim \text{gamma}(a_k, b_k)$, and the hyper-parameters a_k and b_k are set by the user, so the posterior probability function of the parameter is:

$$(\sigma_w^{-2} | D, \alpha, \beta, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, p_1) \sim \text{gamma} \left(a_k + \frac{T}{2}, b_k + \frac{1}{2} \sum_{t=t_{\min}}^{t_{\max}} (k_t - k_{t-1} - \theta)^2 \right).$$

4. Numerical Experiment

4.1. Methods

To test the adequacy of the ABM class, we analyzed mortality data of men in Ireland, Ukraine and the USA, downloaded from the database of Human Mortality Database (<https://www.mortality.org> accessed on 8 March 2013). The data contain the number of dead and the size of the population exposed to risk by age and year; Ireland's and the USA's data are for 1950–2007, and Ukraine's data are for 1959–2009. This analysis aims to examine sixteen models within the ABM class, $p_2 = 0, \dots, 15$, with a particular emphasis on forecasting the mortality of the oldest-old (see below an argumentation for the restriction of the values of p_2 to $\{0, \dots, 15\}$). These models also include the Poisson and negative binomial models which feature widely in the literature, for which $p_1 = 0$ and $p_2 = 1$, respectively. Adopting a data mining approach, the models were fitted using training sets and were examined using test sets. The training sets contained data up to 2000 and the test sets, aimed at monitoring the quality of predictions, contained data from 2001 to 2007 for Ireland and the USA, and from 2001 to 2009 for Ukraine. Predictions are carried out with the estimated parameters, $\ln[m_{x,t+s}] = \hat{\alpha}_x + \hat{\beta} * \hat{k}_{t+s}$, $s = 1, 2, \dots$, where model performance (using the test sets) was checked using the root of the mean squared errors (RMSE), which was calculated in two ways:

1. Predicting mortality rates (μ) by age. In other words, after model parameters were estimated, mortality rates were predicted for a given age across years. For instance, predicting mortality rates for those age 70 was carried out over the years beyond 2000;
2. Predicting mortality rates (μ) by cohort. In other words, after model parameters were estimated, mortality rates were predicted for a cohort that was at a particular age at the beginning of the test period. For example, predicted mortality rates in 2001–2007 for a cohort aged 70 in 2001.

For every member of the ABM class (controlled by p_2), the Markov chains used to obtain posterior distributions/parameter estimates comprised 4000 iterations with the first 1000 considered a burn-in period. Convergence was established using graphical means and a sensitivity analysis ascertained that the choice of arbitrary initial hyper-parameters did not affect the final outcomes. We report the outcomes for the Poisson distribution ($p_2 = 0$) and the negative binomial distribution ($p_2 = 1$). In addition, we examined the ABM members for which $p_2 \in \{2, \dots, 15\}$. We limited our reporting to $p_2 \in \{2, \dots, 15\}$ since, for the data under study, increasing p_2 beyond 15 (our study explored all models up to $p_2 = 50$) did not alter our findings of the optimal p_2 for varying ages and resulted in much larger RMSE than those found up to $p_2 = 15$. In practice, we analyzed all 16 models within the p_2 range and, for each, we estimated p_1 as well as all other unknown parameters. Finally, we reported graphically the RMSE for the Poisson, negative binomial and for the ABM member which produced the minimal RMSE for various ages.

4.2. Results

Figure 1a,b shows Ireland’s mortality projections by age and cohort, respectively. A superior model for an age range is the one which produced the smallest RMSE. It is evident that the Poisson performs best for most ages above 70, with the negative binomial lagging behind. However, at a very old age, the Poisson diverges and the best ABM member to be chosen instead is $ABM(\cdot, p_2 = 3)$. A similar result is shown for the USA (Figure 2a,b), except that the best ABM for the very old is $ABM(\cdot, p_2 = 4)$. A different picture emerges when we focus on Ukraine’s mortality projections by age and cohort (Figure 3a,b, respectively). While the Poisson and negative binomial perform well (with Poisson being better), for ages above 96 (by cohort) or 104 (by age), the negative binomial drifts away, leaving $ABM(\cdot, p_2 = 10)$ to be the winner of the ABM class. Clearly, the recommended projection policy for Ukraine is to use the Poisson for most ages but to rely on $ABM(\cdot, p_2 = 10)$ for very old ages. Naturally, other countries with their specific national datasets may yield different ABM models (i.e., different p_2 ’s) for mortality projections.

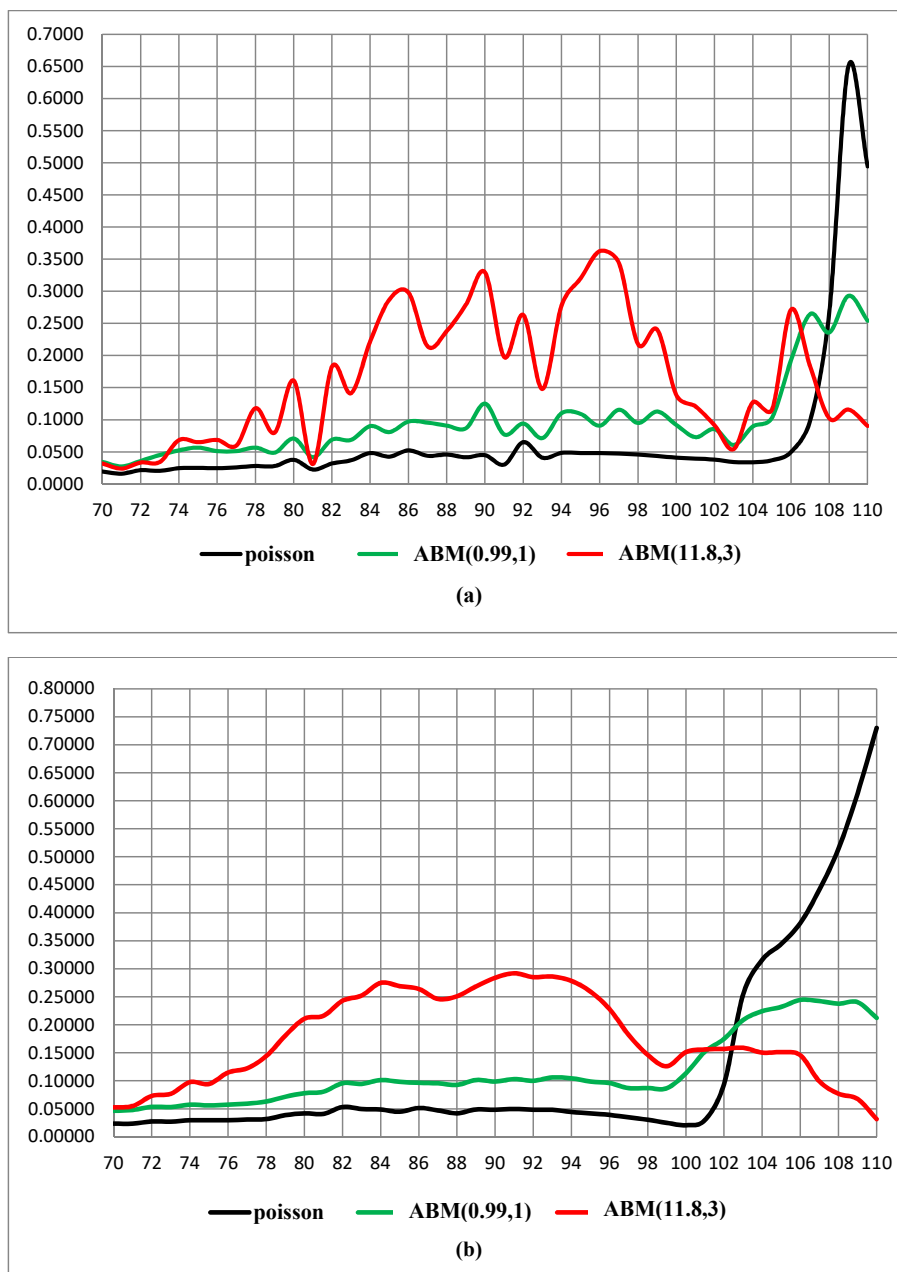


Figure 1. RMSE for projecting mortality rate, Ireland, 2001–2007. (a) by age; (b) by cohort.

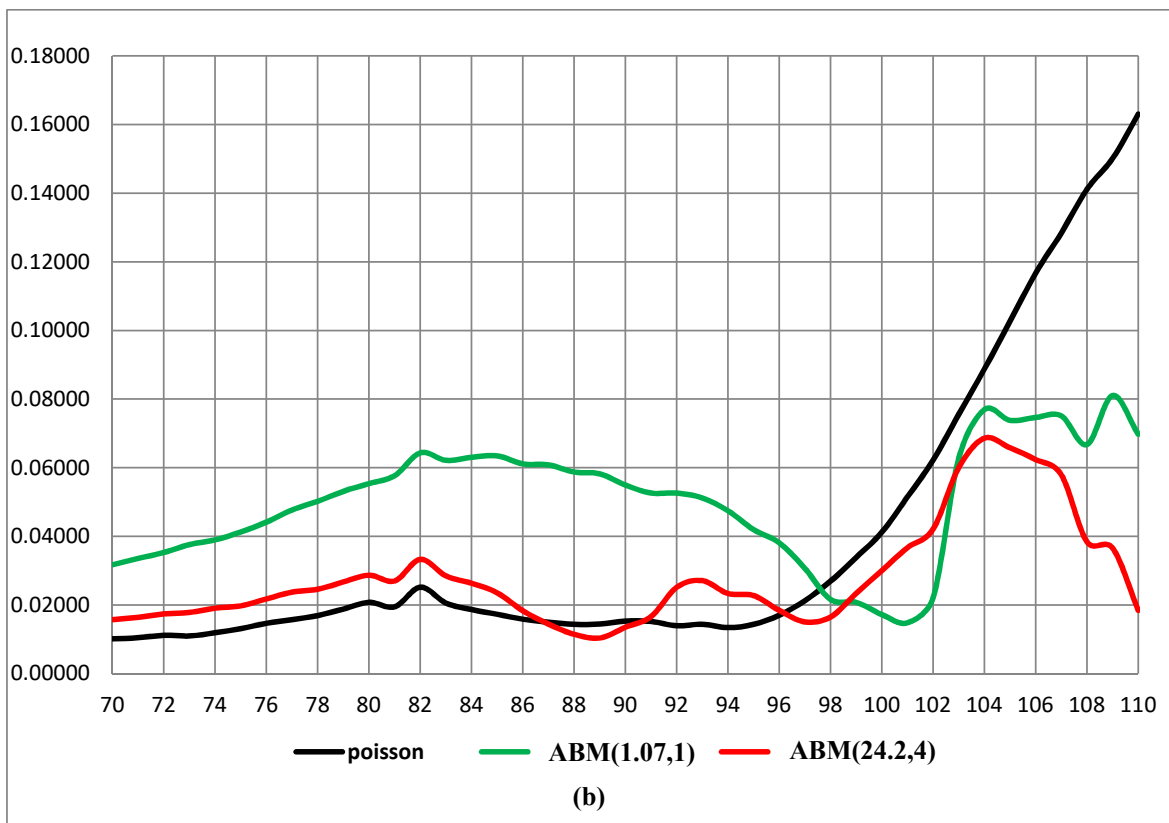
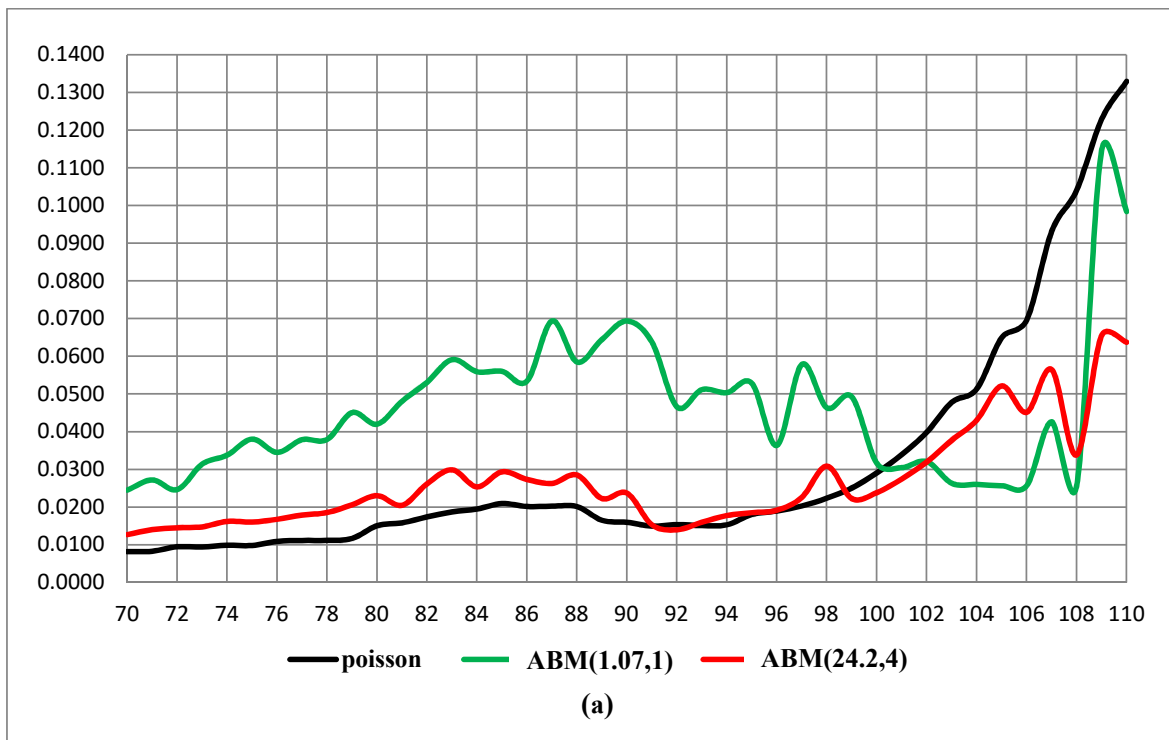


Figure 2. RMSE for projecting mortality rate, USA, 2001–2007. (a) by age; (b) by cohort.

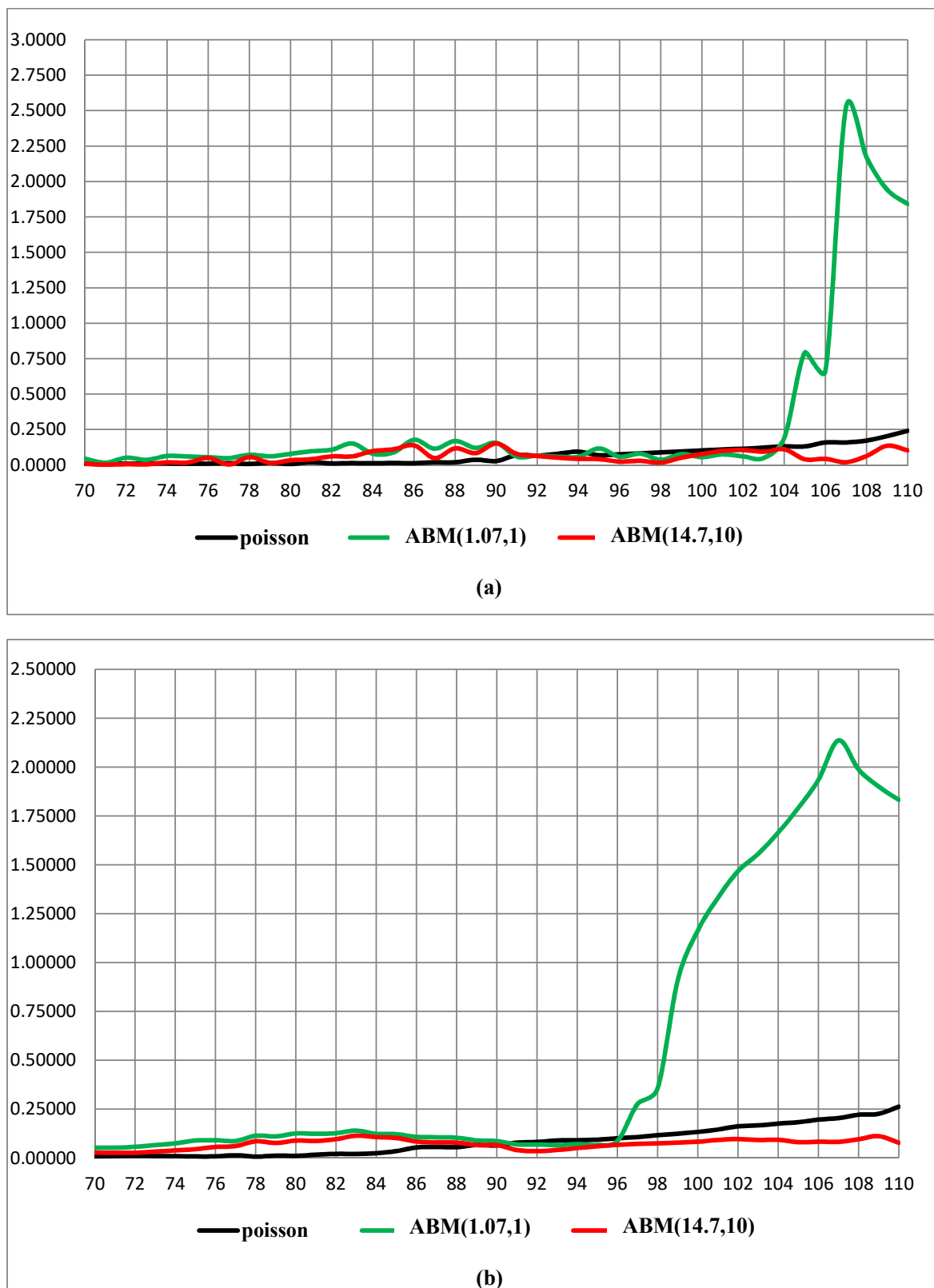


Figure 3. RMSE for projecting mortality rate, Ukraine, 2001–2009. (a) by age; (b) by cohort.

5. Discussion

Several extensions to the LC model assume that the number of deaths is distributed Poisson or negative binomial. These distributions have offered adequate mortality projections in several populations reported in the literature. It is not implausible that cases where

these two failed were not reported. Rather than deciding a priori to choose a particular distribution, we aimed to enrich the LC model by allowing a richer pool of candidate distributions. The chosen distribution would be the one providing the best projection for the population and age range under study. To achieve this goal, we proposed a new class of counting distributions on the non-negative integers, the ABM class, which belongs to a wider class of natural exponential families characterized by their variance functions. This class includes the Poisson and negative binomial distributions which are included in an infinitely countable set of additional members. A data mining approach was adopted whereby the model is fitted using a training set and tested using a test set with the RMSE used to pick the winning model. As an alternative to parameter estimation via the singular value decomposition (SVD) used in the classical LC model, we adopted Bayesian estimation, harnessing the Markov Chain Monte Carlo (MCMC) methodology. While we do not suggest that MCMC is superior to SVD (for a comparison of the two, see [Ichikawa et al. 2021](#)), we still promote the former since the Bayesian framework frees us from the burden of calculating the normalizing constant of the ABM. This is indeed a great plus, even though running the MCMC requires more computer time for the mid-size databases of the kind used for national mortality projections. We note, however, that the use of MCMC is rather costly and might fail if the dataset is huge. So, perhaps other Bayesian techniques such as Variational Bayes can be more helpful. However, employing such a suggestion is beyond the scope of this paper. We examined ABM models for three countries and established that, for the countries examined, the commonly adopted Poisson distribution is justified except for a very old age for which an alternative member of the ABM class offers better projection. We do not claim that to suggest a superior model. When deciding on an underlying model, one can adopt as an example the Poisson model or the negative binomial model. Rather than adopting a model, we suggest adopting a class of models (the ABM) comprising the Poisson, negative binomial, and numerous other counting distributions. The superiority of this approach lies in the ability to choose a model amongst candidate models. Since no one single model necessarily fits every population and every age group well, the ABM class could allow picking, as an example, the Poisson for members of the population aged under 50, the negative binomial for those aged 50 to 80, and another member of the class to the oldest-old. The suggested criteria for preferring one member of the class over another is the mean squared projection errors (RMSE). This advantage is gained at the cost of additional complexity, which is justified given the financial benefits associated with more accurate modeling. We conclude that it is no longer appropriate to assume a single distribution for the whole process of mortality projection. Instead, for every country and every relevant range of ages, a desirable approach is to pick a member of the ABM class that provides the best mortality projection. In the numerical study reported here, neither the Poisson nor the negative binomial distributions adequately serve the very oldest-old and superior alternatives are within reach in the suggested novel ABM class of distributions.

Author Contributions: Conceptualization, Y.A., S.K.B.-L. and U.M.; methodology, Y.A., S.K.B.-L. and U.M.; software, Y.A.; validation, Y.A., S.K.B.-L. and U.M.; formal analysis, Y.A., S.K.B.-L. and U.M.; investigation, Y.A., S.K.B.-L. and U.M.; resources, Y.A., S.K.B.-L. and U.M.; data curation, Y.A. and U.M.; writing—original draft preparation, Y.A., S.K.B.-L. and U.M.; writing—review and editing, Y.A., S.K.B.-L. and U.M.; visualization, Y.A., S.K.B.-L. and U.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: the database of Human Mortality Database (<https://www.mortality.org> access on 8 March 2013).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1

The marginal posterior probability function of the time index k_t is

$$f(k_t | D, \alpha, \beta, k_{-t}, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1) \propto f(D_{t_{\min}} | k_{t_{\min}}, \alpha, \beta, p_1) \times f(k_{t_{\min}} | \theta, \sigma_w^2) \times \prod_{j=t_{\min}+1}^{t_{\max}} f(D_j | k_j, \alpha, \beta, p_1) \times f(k_j | k_{j-1}, \theta, \sigma_w^2),$$

where

$$f(D_t | \alpha, \beta, k_t, p_1) = \prod_x \exp \left[d_{xt} \left(\ln \frac{E_{xr} \mu_{xt}}{p_1 + E_{xt} \mu_{xt}} + \sum_{i=1}^{p_2-1} \frac{1}{i} \frac{p_1^i}{(p_1 + E_{xt} \mu_{xt})} \right) + \frac{p_1^{p_2}}{(p_2-1)(p_1 + E_{xt} \mu_{xt})^{p_2-1}} \right].$$

For the remaining expressions we distinguish between three cases:

1. For $t = t_{\min}$, the marginal posterior probability function of the time index k_t is:

$$f(k_t | D, \alpha, \beta, k_{-t}, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1) \propto f(D_t | \alpha, \beta, k_t, p_1) \times f(k_t | \theta, \sigma_w^2) \times f(k_{t+1} | k_t, \theta, \sigma_w^2),$$

where

$$f(k_t | \theta, \sigma_w^2) = \exp \left(-\frac{1}{2\sigma_w^2} (k_t - \theta)^2 \right) \text{ and}$$

$$f(k_{t+1} | k_t, \theta, \sigma_w^2) = \exp \left(-\frac{1}{2\sigma_w^2} (k_t - k_{t-1} - \theta)^2 \right).$$

2. For $t_{\min} < t < t_{\max}$, the marginal posterior probability function of the time index k_t is:

$$f(k_t | D, \alpha, \beta, k_{-t}, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1) \propto f(D_t | \alpha, \beta, k_t, p_1) \times f(k_t | k_{t-1}, \theta, \sigma_w^2) \times f(k_{t+1} | k_t, \theta, \sigma_w^2),$$

where

$$f(k_t | k_{t-1}, \theta, \sigma_w^2) = \exp \left(-\frac{1}{2\sigma_w^2} (k_t - k_{t-1} - \theta)^2 \right) \text{ and}$$

$$f(k_{t+1} | k_t, \theta, \sigma_w^2) = \exp \left(-\frac{1}{2\sigma_w^2} (k_{t+1} - k_t - \theta)^2 \right).$$

3. For $t = t_{\max}$, the marginal posterior probability function of the time index k_t is:

$$f(k_t | D, \alpha, \beta, k_{-t}, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1) \propto f(D_t | \alpha, \beta, k_t, p_1) \times f(k_t | k_{t-1}, \theta, \sigma_w^2),$$

where

$$f(k_t | k_{t-1}, \theta, \sigma_w^2) = \exp \left(-\frac{1}{2\sigma_w^2} (k_t - k_{t-1} - \theta)^2 \right).$$

Appendix A.2

The marginal posterior probability function of β_x is

$$f(\beta_x | D, \alpha, \beta_{-x}, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1) \propto \prod_{j=x_{\min}}^{x_{\max}} f(D_j | \alpha, \beta_j, k, p_1) \times f(\beta_j) \propto f(D_x | \alpha, \beta_x, k, p_1) \times f(\beta_x),$$

where

$$f(D_x | \alpha, \beta_x, k, p_1) = \prod_t \exp \left[d_{xt} \left(\ln \frac{E_{xr} \mu_{xt}}{p_1 + E_{xt} \mu_{xt}} + \sum_{i=1}^{p_2-1} \frac{1}{i} \frac{p_1^i}{(p_1 + E_{xt} \mu_{xt})} \right) + \frac{p_1^{p_2}}{(p_2-1)(p_1 + E_{xt} \mu_{xt})^{p_2-1}} \right],$$

and

$$f(\beta_x) = \exp\left(-\frac{1}{2\sigma_\beta^2}(\beta_x)^2\right).$$

Appendix A.3

The marginal posterior probability function of α_x is

$$f(\alpha_x | D, \alpha_{-x}, \beta, k, \theta, \sigma_\alpha^2, \sigma_\beta^2, \sigma_w^2, p_1) \\ \propto \prod_{j=x_{\min}}^{x_{\max}} f(D_j | \alpha_j, \beta, k, p_1) \times f(\alpha_j) \propto f(D_x | \alpha_x, \beta, k, p_1) \times f(\alpha_x),$$

where

$$f(D_x | \alpha_x, \beta, k, p_1) = \\ \prod_t \exp\left[d_{xt} \left(\ln \frac{E_{xt}\mu_{xt}}{p_1 + E_{xt}\mu_{xt}} + \sum_{i=1}^{p_2-1} \frac{1}{i} \frac{p_1^i}{(p_1 + E_{xt}\mu_{xt})} \right) + \frac{p_1^{p_2}}{(p_2-1)(p_1 + E_{xt}\mu_{xt})^{p_2-1}} \right],$$

and

$$f(\beta_x) = \exp\left(-\frac{1}{2\sigma_\alpha^2}(\alpha_x - \alpha_{0x})^2\right).$$

References

- Antonio, Katrien, Anastasios Bardoutsos, and Wilbert Ouburg. 2015. Bayesian Poisson log-bilinear models for mortality projections with multiple populations. *Eur. Actuar. J.* 5: 245–281. [CrossRef]
- Awad, Yaser, Shaul K. Bar-Lev, and Udi Makov. 2016. *A New Class Counting Distributions Embedded in the Lee–Carter Model for Mortality Projections: A Bayesian Approach*. Technical Report No. 146. Israel: Actuarial Research Center, University of Haifa.
- Azman, Shafiqah, and Dharini Pathmanathan. 2020. The GLM framework of the Lee–Carter model: A multi-country study. *Journal of Applied Statistics* 49: 752–63. [CrossRef]
- Baltes, B. Paul, and Jacqui Smith. 2003. New frontiers in the future of aging: From successful aging of the young old to the dilemmas of the fourth age. *Gerontology* 49: 123–35. [CrossRef] [PubMed]
- Bar-Lev, K. Shaul, and Ad Ridder. 2019. Monte Carlo methods for insurance risk computation. *International Journal of Statistics and Probability* 8: 54–74. [CrossRef]
- Bar-Lev, K. Shaul, and Ad Ridder. 2021a. Exponential dispersion models for overdispersed zero-inflated count data. *Communications in Statistics-Simulation and Computation*. Available online: <https://doi.org/10.1080/03610918.2021.1934020> (accessed on 10 August 2021).
- Bar-Lev, K. Shaul, and Ad Ridder. 2021b. New exponential dispersion models for count data—The ABM and LM classes. *ESAIM: Probability and Statistics* 25: 31–52. [CrossRef]
- Bar-Lev, K. Shaul, and Clestin C. Kokonendji. 2017. On the mean value parametrization of natural exponential families—A revisited review. *Mathematical Methods of Statistics* 26: 159–75. [CrossRef]
- Brouhns, Natacha, Michel Denuit, and Jeroen K. Vermunt. 2002. A Poisson log-bilinear regression approach to the construction of project life-table. *Insurance: Mathematics and Economics* 31: 373–93.
- Buettner, Thomas. 2002. Approaches and experiences in projecting mortality patterns for the oldest-old. *North American Actuarial Journal* 6: 14–29. [CrossRef]
- Czado, Claudia, Antoine Delwarde, and Michel Denuit. 2005. Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics* 36: 260–84. [CrossRef]
- Danesi, L. Ivan, Steven Haberman, and Pietro Millosovich. 2015. Forecasting mortality in subpopulations using Lee–Carter type models: A comparison. *Insurance: Mathematics and Economics* 62: 151–61. [CrossRef]
- Delwarde, Antoine, Michel Denuit, and Christian Partrat. 2007. Negative binomial version of the Lee–Carter model for mortality forecasting. *Applied Stochastic Models in Business and Industry* 23: 385–401. [CrossRef]
- Ellison, Joanne, Erengul Dodd, and Jonathan J. Forster. 2020. Forecasting of cohort fertility under a hierarchical Bayesian approach. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 183: 829–56. [CrossRef]
- Graziani, Rebecca. 2020. Stochastic Population Forecasting: A Bayesian Approach Based on Evaluation by Experts. *Developments in Demographic Forecasting* 49: 21–42.
- Hilton, Jason, Erengul Dodd, Jonathan J. Forster, and Peter W. F. Smith. 2019. Projecting UK mortality by using Bayesian generalized additive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 68: 29–49. [CrossRef]
- Hunt, Andrew, and David Blake. 2020. A Bayesian Approach to Modeling and Projecting Cohort Effects. *North American Actuarial Journal* 25: S235–S254. [CrossRef]

- Ichikawa, Shota, Hiroyuki Yamamoto, and Takumi Morita. 2021. Comparison of a Bayesian estimation algorithm and singular value decomposition algorithms for 80-detector row CT perfusion in patients with acute ischemic stroke. *La Radiologia Medica* 126: 795–803. [CrossRef]
- Idrizi, Olgerta. 2018. The Heteroscedasticity Impact on Actuarial Science: Lee Carter Error Simulation. *European Journal of Engineering and Formal Sciences* 1: 1–12. [CrossRef]
- Jorgensen, Bent. 1987. Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society, Series B* 49: 127–62.
- Jorgensen, Bent. 1997. *The Theory of Exponential Dispersion Models*. Monographs on Statistics and Probability 76. London: Chapman and Hall.
- Kogure, Atsuyuki, Takahiro Fushimi, and Shinichi Kamiya. 2019. Mortality Forecasts for Long-Term Care Subpopulations with Longevity Risk: A Bayesian Approach. *North American Actuarial Journal* 25: S534–44. [CrossRef]
- Lee, D. Ronald, and Lawrence R. Carter. 1992. Modelling and forecasting the time series of US mortality. *Journal of the American Statistical Association* 87: 659–71.
- Légaré, Jacques, Yann Décarie, Kim Deslandes, and Yves Carrière. 2015. Canada's oldest old: A population group which is fast growing, poorly apprehended and at risk from lack of appropriate services. *Population Change and Lifecourse Strategic Knowledge Cluster Discussion Paper Series/Un Réseau Stratégique de Connaissances Changements de Population et Parcours de vie Document de Travail* 3: 1. Available online: <https://ir.lib.uwo.ca/pclc/vol3/iss2/1> (accessed on 1 April 2021).
- Letac, Gerard, and Marianne Mora. 1990. Natural real exponential families with cubic variance functions. *Annals of Statistics* 18: 1–37. [CrossRef]
- Liu, Zhen, Xiaoqian Sun, Leping Liu, and Yu-Bo Wang. 2020. Bayesian Poisson Log-normal Model with Regularized Time Structure for Mortality Projection of Multi-population. *arXiv* arXiv:2010.04775.
- Njenga, N. Carolyn, and Michael Sherris. 2020. Modeling mortality with a Bayesian vector autoregression. *Insurance: Mathematics and Economics* 94: 40–57. [CrossRef]
- Pedroza, Claudia. 2006. A Bayesian forecasting model: predicting U.S male mortality. *Biostatistics* 7: 530–50. [CrossRef]
- Renshaw, Arthur, and Steven Haberman. 2008. On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee–Carter modelling. *Insurance: Mathematics and Economics* 42: 797–816. [CrossRef]
- Shair, N. Syazreen, Muhammad A. S. Rosmizan, Mohammad J. M. S. Ting, and Mohd A. A. Zaini. 2018. Projected Malaysian Life Table: Evaluations of The Lee–Carter and Poisson Log-Bilinear Models. *International Journal of Modern Trends in Social Sciences* 1: 60–72.
- Wang, Duolao, and Pengjun Lu. 2005. Modelling and forecasting mortality distributions in England and Wales using the Lee–Carter model. *Journal of Applied Statistics* 32: 873–85. [CrossRef]
- Wong, S. T. Jackie, Jonathan J. Forster, and Peter W. F. Smith. 2018. Bayesian mortality forecasting with overdispersion. *Insurance: Mathematics and Economics* 83: 206–21. [CrossRef]