Annals of Actuarial Science (2026), 1–27
doi:10.1017/S1748499526100281

ORIGINAL RESEARCH PAPER

Analyzing pension fund mortality with Gaussian processes in a subpopulation framework

Eduardo Fraga L. de Melo[1,2,3], Michael Ludkovski[4] and Rodrigo S. Targino[3]

[1]Universidade do Estado do Rio de Janeiro, Brazil; [2]Susep – Superintendência de Seguros Privados, Brazil; [3]School of Applied Mathematics, Fundacao Getulio Vargas, Brazil; and [4]Statistics and Applied Probability, UC Santa Barbara, USA
Corresponding author: Michael Ludkovski; Email: ludkovski@pstat.ucsb.edu

(Received 11 June 2025; revised 18 March 2026; accepted 23 March 2026)

Abstract

Pension fund populations often have mortality experiences that are substantially different from the national benchmark. In a motivating case study of Brazilian corporate pension funds, pensioners are observed to have mortality that is 40–55% below the national average, due to the underlying socioeconomic disparities. Direct analysis of a pension fund population is challenging due to very sparse data, with age-specific annual death counts often in low single digits. We design and study a collection of stochastic subpopulation frameworks that coherently capture and project pensioner mortality rates via deflator factors relative to a reference population. Superseding parametric approaches, we propose Gaussian process (GP)-based models that flexibly estimate age- and/or year-specific deflators. We demonstrate that the GP models achieve better goodness of fit and uncertainty quantification. Our models are illustrated on two Brazilian pension funds in the context of exogenous national mortality tables. The GP models are implemented in **R Stan** using a fully Bayesian approach and take into account over-dispersion relative to the Poisson likelihood.

Keywords: Gaussian processes; mortality deflators; pension fund longevity; subpopulation mortality

1. Introduction

The valuation of pension fund liabilities requires the use of consistent, realistic, and up-to-date mortality rates and also the forecast of these mortality rates into the future for cash flow discount or asset-liability management. Such requirements are embedded in modern solvency assessment principles and financial reporting standards; see Sandström (2016) and IASB (2017). The vast majority of studies in mortality rates forecasting make use of national population data when performing applications; see Cairns et al. (2009), Dowd et al. (2011), Lee and Carter (1992), Li and Lee (2005), Li (2013), and Renshaw and Haberman (2006), among others.

Since the national population is inevitably different from the pension fund population, there is an inherent basis risk in substituting the former mortality rates for the purposes of analyzing the latter. This basis risk may not present a meaningful problem when the difference between the national population and the underlying sub (selected) population is reasonable. For instance, in developed countries in Europe, the population mortality tends to be more homogeneous. However, in countries with high social inequalities, the income differences and healthcare access make the mortality rates significantly different among the selected and the overall national populations.

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Institute and Faculty of Actuaries. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

Brazil is a country with high social inequalities (OECD, 2018). According to supervisory authorities, less than 8% of Brazil's population has a pension plan,[1] and the fraction that does tends to be in the upper deciles of the national income distribution. As a result, national mortality rates are much higher than the ones for actuarial subpopulations, such as participants of pension funds or insurance customers. Our data indicates that, on average, the mortality rates of the pension fund subpopulation are *half* (49%) of the mortality of the national Brazilian population for ages 60+.

Mortality modeling for a pension fund must contend with two fundamental challenges. On the one hand, the population of interest is relatively small, so stand-alone models lack sufficient credibility: there are simply not enough death counts to reliably estimate mortality rates. On the other hand, the pensioners often have materially distinct mortality patterns, so that simple adjustments of the national trends are inadequate. In this paper, we aim to systematically explore and ultimately bridge this gap. To this end, we construct a spectrum of *deflator-based* stochastic models that relate the pension fund mortality to the reference population. As a starting point, we observe that the commonly used parametric deflators (such as deflators that are constant across ages) are too rigid and underfit. At the other extreme, allowing for independent age- (or year-) specific deflators leads to over-parametrization and overfitting. This motivates us to develop non-parametric deflators that are constrained to be smooth yet flexible enough. Specifically, we propose to employ Gaussian-process (GP)-based deflators, which naturally embed dependence while providing a fully probabilistic, data-driven, non-parametric framework.

Methodologically, we make three contributions. First, we provide a taxonomy of over half a dozen deflator models for small population mortality, systematically comparing them (including discussing multiple nesting relationships present) and offering a comprehensive "menu" for how to approach this problem. Second, we propose a new class of GP-driven deflators. Such models offer a valuable hybrid of leveraging the reference mortality table while building in sufficient flexibility. Third, with an emphasis on capturing model risk due to the small underlying population, we utilize Bayesian methods throughout our analysis, so that all our model hyperparameters come with priors. We employ the `Stan` package and underlying MCMC routines to provide uncertainty quantification, contributing to the literature on Bayesian mortality analysis.

Empirically, we are motivated by and illustrate our developments with two datasets of Brazilian pensioners. The pension fund in our primary case study (pension fund 1) consists of about 14,000 exposed lives with about 100 annual deaths. Compared to the national baseline, these pensioners have about half the mortality rate, with the respective longevity gap notably growing over time. We demonstrate that our proposed models outperform extant alternatives, including the two extremes of our modeling spectrum: a stand-alone GP model and a fully specified fixed-deflator model. Given that age-specific death counts in this fund are often 0 or 1, we moreover explore the observation likelihood, namely the overdispersion in death counts relative to the Poisson baseline. Our analysis provides insights into the mortality experience in Brazil and in developing countries more broadly, offering new takeaways on pension funds in highly heterogeneous settings. To confirm the above findings, we further explore a second Brazilian pension fund that is slightly smaller.

Our interest is on ages 60 and over (60+), ages that are most relevant for insurance and pension funds when paying annuities. This is another difference between our approach and the majority of papers in the mortality modeling literature. In this paper, Gaussian processes will be used in a subpopulation framework to forecast mortality at ages 60+ of small, specific pension funds' populations, using as reference the annual Brazilian national mortality tables.

The rest of the paper is organized as follows. Section 2 reviews existing methods for subpopulation models, as well as the Gaussian Process paradigm. Section 3 summarizes our Brazil pension

[1] https://www.investidorinstitucional.com.br/sessoes/investidores/seguradoras/40265-previdencia-aberta-atinge-5-3-da-populacao-brasileira-em-maio.html

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

fund dataset and the IBGE reference mortality tables. Section 4 presents the list of the considered deflator models, both from existing approaches as well as new ones proposed here for the first time. Section 5 illustrates and discusses the respective fits for our motivating dataset, while Section 6 provides a statistical assessment across all the fitted models. Appendix summarizes the models fitted in this paper, additional plots for the primary pension fund, and, finally, discussions for pension fund 2.

2. Subpopulation models

Given the small size of the modeled population, we directly work with death and exposed counts, disaggregated by age and calendar year. We employ the following notation:

- $d_{x,t}$ is the number of deaths at age x and calendar year t for the pension fund, where $x = \{60, \ldots, 89\}$ and $t = \{2013, \ldots, 2019\}$;
- $E_{x,t}$ is the risk exposure at age x and calendar year t for the pension fund;
- $m_{x,t}^{BRA}$ is the central mortality rate relative to age x and calendar year t of *BRA* – Brazilian national mortality rates.

To overcome the scarce data available about the pension fund experience, we adopt the solution of external reference tables. The idea of such relational models is to adjust the reference table to the experience of the pension fund. In our work, we consider as a benchmark the national Brazilian mortality tables (*BRA*), also known as the IBGE (Brazilian Institute of Geography and Statistics) tables. These tables, as well as the pension fund's data, are described in detail in Section 3.

Reflecting the limited number of deaths in the pension fund population, the common approach is to use a discrete (Poisson) likelihood where the reference mortality rate $m_{x,t}^{BRA}$ is part of the intensity parameter:

$$d_{x,t} \sim \text{Poi}\left(e^{\theta_{x,t}} m_{x,t}^{BRA} E_{x,t}\right), \tag{1}$$

and $e^{\theta_{x,t}}$ is the *deflator* applied to the reference population. Model (1) implies the fund mortality rate $m_{x,t} := \mathbb{E}[d_{x,t}]/E_{x,t}$ is given by $m_{x,t} = e^{\theta_{x,t}} m_{x,t}^{BRA}$. Hence,

$$\theta_{x,t} = \log m_{x,t} - \log m_{x,t}^{BRA}$$

is the difference on the log scale between the fund mortality and the reference mortality, or equivalently, the percent ratio between them. Since the funds' mortality is lower than the reference population's, $\theta_{x,t} < 0$.

An important modeling premise underlying the deflator framework is that the reference mortality surface $m_{x,t}^{BRA}$ is treated as fixed and exogenous. Any preprocessing, smoothing, or extrapolation embedded in the reference table therefore only affects the baseline level against which the pension fund mortality is measured and does not alter the stochastic structure of the deflators themselves. Consequently, the deflator process $\theta_{x,t}$ is interpreted as capturing relative deviations of the pension fund experience from a given reference trend, rather than as an attempt to recover an underlying "true" mortality surface.

Before describing our method in detail, we provide a review of existing deflator models. The deflator approach was introduced in Hardy and Panjer (1998), who started with a constant deflator $d_{x,t} \sim \text{Poi}(\Theta \cdot m_{x,t}^{BRA} \cdot E_{x,t})$. Beyond directly fitting Θ as a static parameter, they consider a sequential Bayesian setup, where Θ is endowed with a prior, which is updated over time based on observed $(d_{x,t}, E_{x,t})$. Specifically, they leverage Poisson-Gamma conjugacy to assign a Gamma prior $\Theta \sim \text{Gamma}(c, c)$ which has a mean of 1, and therefore, *a priori,* the portfolio is expected to have the same mortality as the reference population. Olivieri and Pitacco (2012) generalize to independent age-specific random effects, obtaining the posterior $\Theta_x|\mathcal{D} \sim \text{Gamma}(c + \sum_t d_{x,t}, c + \sum_t m_{x,t}^{BRA} E_{x,t})$.

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

Olivieri (2011) takes

$$d_{x,t} \sim \text{Poi}\left(\Theta_{x,t} \cdot m_{x,t}^{BRA} \cdot E_{x,t}\right) \qquad \text{where} \quad \Theta_{x,t} \sim \text{Gamma}(\alpha_{x,t}, \beta_{x,t})$$

and the hyperparameters α, β are sequentially updated over the years based on realized $d_{x,t}$. van Berkum et al. (2017) use age-specific Bayesian "random effects," $d_{x,t} \sim \text{Poi}(\Theta_x \cdot m_{x,t}^{BRA} \cdot E_{x,t})$, where m^{BRA} is estimated using a Lee-Carter model, and the age-specific deflators Θ_x are given either independent Gamma(c_x, c_x) priors (forcing $\mathbb{E}[\Theta_x] = 1$) or a log-AR(1) structure, $\log \Theta_x = \mu + \rho \log \Theta_{x-1} + \eta$, with a log-normal prior on $\log \Theta_{x_0}$, a uniform prior on the variance of η, and a logit-normal prior on ρ.

Making the deflators θ_x constant in time offers a coherent projection into the future. For instance, it is common to assume that the mortality trend of the small population (which is difficult to reliably estimate due to a paucity of data) matches that of the reference; see e.g. Olivieri and Pitacco (2012), Salhi et al. (2015), van Berkum et al. (2017), Li and Lu (2017), Hyndman et al. (2013), Salhi and Loisel (2017). Salhi et al. (2015) focuses on time-dependent "differential mortality law," i.e. θ_t depending only on year.

Closely related to our work is Tomas and Planchet (2015). They consider a sequence of models, starting with an endogenous nonparametric $d_{x,t} \sim \text{Poi}(E_{x,t} \cdot e^{f(x,t)})$ to the relational versions $d_{x,t} \sim \text{Poi}(E_{x,t} \cdot e^{f(\log m_{x,t}^{BRA})})$, $d_{x,t} \sim \text{Poi}(E_{x,t} \cdot m_{x,t}^{BRA} \cdot e^{f(x,t)})$, where the smooth functions $f(x,t)$ are estimated using local kernel weighted log-likelihood methods.

A related strand of literature directly proposes a parametric functional relationship between the mortality surfaces of the small and benchmark populations. In the original Brass model of Brouhns et al. (2002), the authors take $m_{x,t} = \theta_1 (m_{x,t}^{BRA})^{\theta_2}$ for parameters θ_1, θ_2 to be fitted by GLM. More generally, Brouhns et al. (2002) consider $m_{x,t} = \theta_0 f(m_{x,t}^{BRA}) + \theta_1$, for a given link function f (such as the logit). Plat (2009) directly models the ratio $P_{x,t} = (d_{x,t}/E_{x,t})/(d_{x,t}^{ref}/E_{x,t}^{ref})$ of the observed portfolio and reference mortality rate, imposing a linear relationship, $P_{t,x} = a_t + b_t x + \epsilon_{x,t}$. Bienvenüe and Rulliére (2012) use distortion of probability distributions to connect m^{BRA} and m.

A different way to frame the small population problem is via a multi-population approach. Such joint modeling of the larger and smaller subpopulation is especially appropriate when the smaller subpopulation is sufficiently large and moreover, when starting with raw data for both pools. For example, van Berkum et al. (2017) consider the CMI subpopulation in the United Kingdom, which consists of assured male lives and is roughly 20% of the total population, weighted towards pre-retirement ages. Hyndman et al. (2013) consider the joint modeling of female and male log-mortality by building functional time-series models for their sum and difference (the product-ratio method). This setup is, however, not applicable to our case, where the pension fund is much smaller, and moreover, we work with the fixed national mortality table. A related approach to sub-national populations is to first compute the principal age components of mortality and then estimate the respective time-dependent coefficients: $m_{x,t} = \sum_i a_i(t) Y_i^{ref}(x) + \epsilon_{x,t}$ (Alexander et al., 2017). Gravity models (Dowd et al., 2011) work in the Lee-Carter paradigm and introduce the dependence of the small population period effect on the larger population:

$$\kappa_t = \kappa_{t-1} + \phi(\kappa_{t-1}^{ref} - \kappa_{t-1}) + \mu + \rho \epsilon_t^{ref} + \sqrt{1-\rho^2}\epsilon_t,$$

similarly for the cohort term γ_{x-t}.

A nonparametric framework for capturing the co-dependence of mortality across multiple populations via multi-output Gaussian processes was proposed in Huynh and Ludkovski (2021). As an extension, Huynh and Ludkovski (2024) developed hierarchical multi-output GPs to capture multiple cause-of-death mortality rates across different countries and both genders. In these extensions, different populations (or causes of death) translate into distinct levels of a factor covariate, with the latter contributing a cross-population covariance term. In our context,

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

the national mortality table is already calibrated or smoothed out through statistical procedures applied by the National Institute of Statistics (IBGE), so there is no reason to smooth it again through a (multi-output) GP. Therefore, we employ GP regression within a single-output approach, i.e. solely for the pension fund mortality rates, whether directly or through deflators.

2.1 GP modeling

GP models (Oakley & O'Hagan, 2002) are one of the most popular regression methods thanks to the great flexibility they offer in the representation of complex non-linear input–output relationships. GPs define probabilistic models of functional behavior, offering high prediction power, interpretability, and the ability to provide both an interpolation of the data and an uncertainty quantification in the unexplored regions (Marrel et al., 2008). The GP model was first developed in spatial statistics under the name of kriging (Cressie, 1990) and rapidly gained attention in the machine learning community. The core idea is to assume the regression function is distributed according to a Gaussian process, which allows treating the regression function values as unknown quantities and estimating them from the training data. Over the years, the GP model has become popular in a wide range of applications, such as wind power generation (Mori & Kurata, 2008), vehicle design and navigation (Chen et al., 2014), modeling and prediction of natural hazards (Liu & Guillas, 2017; Rohmer & Idier, 2012), etc.

We call a stochastic process f a Gaussian process if for any finite collection of index points $\mathbf{y} = \{y_j, j = 1, \ldots, n\}$, the n-dimensional multivariate density function of the random vector $\mathbf{f} = [f(y_1), \ldots, f(y_n)]^\top$ is multivariate Gaussian (Rasmussen & Williams, 2005). A GP is completely specified by its mean and covariance functions. The mean function $m(y) := \mathbb{E}[f(y)]$ describes what the expected value of $f(y)$. The covariance or kernel function, denoted as $c(\cdot, \cdot)$, expresses the degree of dependence between two different function values as a function of two index points, $\mathrm{Cov}(f(y), f(y')) = c(y, y')$. In compact notation, this is written as

$$f \sim \mathrm{GP}\big(m(\cdot), c(\cdot, \cdot)\big).$$

In the GP regression framework, we usually work with a finite collection of index points, which is equivalent to working with a multivariate Gaussian distribution. The collection of index points y_j, for $j = 1, \ldots, n$, in the previous definition plays the role of covariates. In our case, these are age and/or calendar year. The vector of function values whose components are now associated with each of those covariates is then distributed according to a n-dimensional multivariate Gaussian distribution,

$$\begin{bmatrix} f(y_1) \\ \cdot \\ \cdot \\ f(y_n) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(y_1) \\ \cdot \\ \cdot \\ m(y_n) \end{bmatrix}, \begin{bmatrix} c(y_1, y_1) & \cdots & c(y_1, y_n) \\ \cdot & \cdots & \cdot \\ \cdot & \ddots & \cdot \\ c(y_n, y_1) & \cdots & c(y_n, y_n) \end{bmatrix} \right) = \mathcal{N}(m(\mathbf{y}), \mathbf{C}).$$

For mortality modeling, GPs were introduced in Ludkovski et al. (2018) as a data-driven approach for determining age-time dependence in mortality rates and jointly smoothing raw rates across these dimensions. The kernel $c(\cdot, \cdot)$ above then captures the correlation between different covariate coordinates. For example, it is expected that the mortality for age 70 in 2018, or $y_j = (70; 2018)$, is more correlated with $y_p = (69; 2017)$ than with $y_q = (60; 2015)$. The GP model provides uncertainty quantification associated with smoothed historical experience and generates full stochastic trajectories for out-of-sample forecasts. In this paper we analyze univariate GP models that work either with age y_{ag} or with year y_{yr} as the sole covariate, as well as a bivariate GP model that takes in both (y_{ag}, y_{yr}), see Sections 4.2-4.4.

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

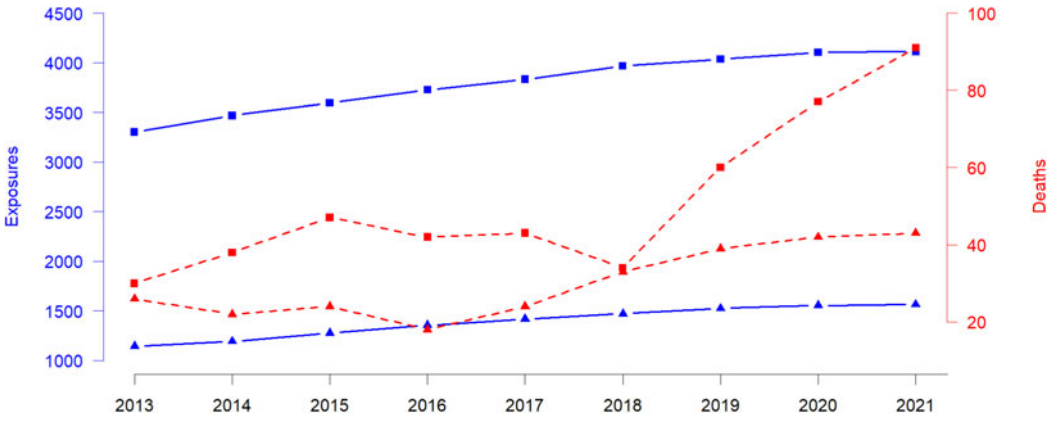


Figure 1. Time series of total exposure $E_t := \sum_x E_{x,t}$ (blue solid lines) and number of deaths $d_t := \sum_x d_{x,t}$ (red dashed lines) per year t for both pension funds. Data from 2013 to 2021, males, ages 60–89, pension fund 1 (squares) and pension fund 2 (triangles).

3. Pension fund population and reference mortality tables

Our analysis is motivated by a dataset summarizing mortality experience of two medium-size pension funds in Brazil that manage pension plans for employees of two firms, which in turn sponsor the pension plan, matching contributions for each one of their workers. By Brazilian law, pension funds are not allowed to write business for non-employees. Our sample is from firms with mainly office workers, belonging to the medium–high income segment. The pension fund population data comprise exposure to risk and number of deaths from 2013 to 2021 (9 years) for male and female employees and pensioners. Approximately two-thirds of the pensioners are males for both samples. Some of the pensioners, especially among the females, are surviving spouses of deceased employees.

Remark. In our dataset, the number of observed female lives and deaths was a fraction of males, leading to highly sparse exposure and mortality observations, such as the majority of ages with zero female deaths. The resulting level of sparsity rendered the female data unsuitable for robust model estimation and validation. Therefore, in the analysis below, we focus exclusively on male data. We emphasize that while our methodology is tailored to accommodate small populations (exposures on the order of $[10^3, 10^5]$), it nevertheless assumes a minimum data richness to enable smoothing and regularization techniques and capture meaningful patterns across ages and years.

Figure 1 shows the annual time series of total exposure $\sum_x E_{x,t}$ to risk (i.e. the number of male pensioners aggregated across ages 60–89) and the annual aggregate number of male deaths $\sum_x d_{x,t}$. There is a marked increase in the number of deaths in 2020 and 2021, probably linked to the COVID-19 pandemic. Although various extrapolation approaches could be used to smooth or impute post-COVID mortality and thus approximate a "normal" trajectory, our datasets end in 2021, and no post-pandemic observations are available. Implementing such corrections would therefore require strong assumptions about the unobserved recovery period, which could introduce more uncertainty than it resolves. For this reason, the analysis presented here relies on the observed data without attempting to reconstruct a counterfactual post-COVID path.

Figure 2 shows the age distribution of pensioners and deaths in the year 2018 for both pension funds when there were 4, 113 and 1, 564 male pensioners aged 60–89, respectively. For the primary sample, the exposure $E_{x,t}$ varies from 5 (for ages close to 90) to 250 (for ages close to 70). The number of deaths $d_{x,t}$ in 2018 ranges from none to 7.

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

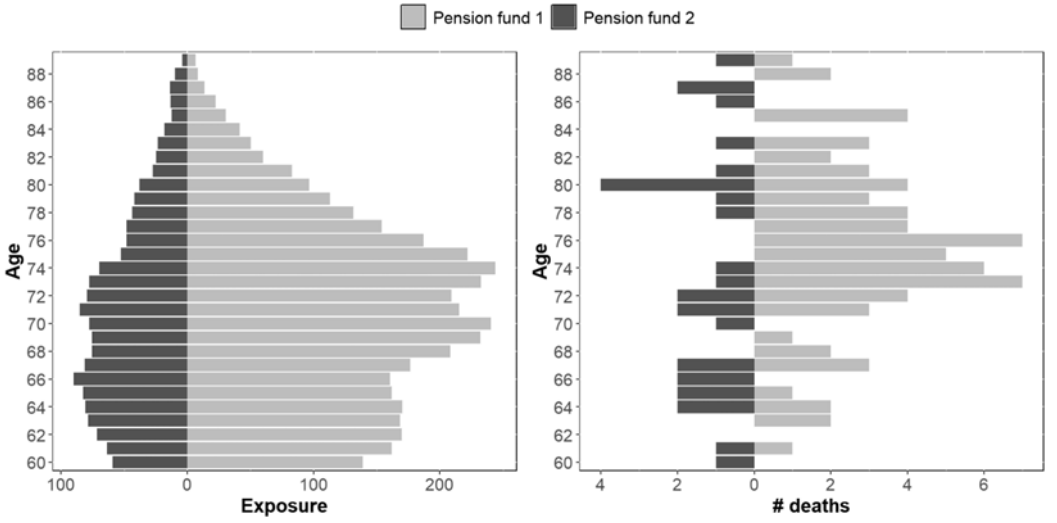


Figure 2. Age pyramids of exposures $E_{x,t}$ (*left*) and number of deaths (*right*) $d_{x,t}$ per age x for year $t = 2018$ for the pension funds 1 and 2: ages 60–89.

Figure 3 shows the reference Brazilian mortality tables (*BRA*) published by the Brazilian Institute of Geography and Statistics (IBGE, 2022).[2] Data is available from 1998 to 2021 for ages 0–80, male, female, and both genders. It is essential to keep in mind that these tables are heavily post-processed and do not represent raw data. According to IBGE (2016), the Gompertz law (mortality being exponential in age) is used to calibrate mortality rates for ages 60–80. As we are interested in older ages for the pension fund, we used the same IBGE estimated coefficients to extrapolate the mortality rates from age 81 to 89 for each year. In terms of time, the IBGE mortality tables are built based on the results of the Brazilian census, which is normally held every 10 years. The 2020 Brazilian Census was postponed to 2022 due to the COVID-19 pandemic. The BRA national mortality tables are available annually, according to an interpolation method described in IBGE (2024), which is based on a coherent Lee-Carter model; see United Nations (2015).

4. Mortality models

In this section, we introduce the approaches used to model the mortality of the pension funds' populations that were described in Section 3. First, we introduce the simplest frameworks: the use of a predetermined mortality table with no deflator and the use of a constant deflator relative to a reference population. Then we discuss how one may add age or calendar year dependence to the deflators. We conclude the presentation of the models by introducing GP-based models. All stochastic models in this paper follow the Bayesian paradigm, implying that all parameters should be assigned prior distributions and inference should be made based on posterior distributions. Prior distributions are generally vaguely informative, but as all the hyperparameters are easily interpretable, it is expected for specialists to have some degree of expert opinion on the range of the hyperparameters.

4.1 Practitioners' approach

The first two frameworks are common for practitioners in the insurance or pension fund industries: the use of known mortality tables with no deflator or deflated by a single fixed value for all

[2] https://www.ibge.gov.br/estatisticas/sociais/populacao/9126-tabuas-completas-de-mortalidade.html?=t=downloads

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

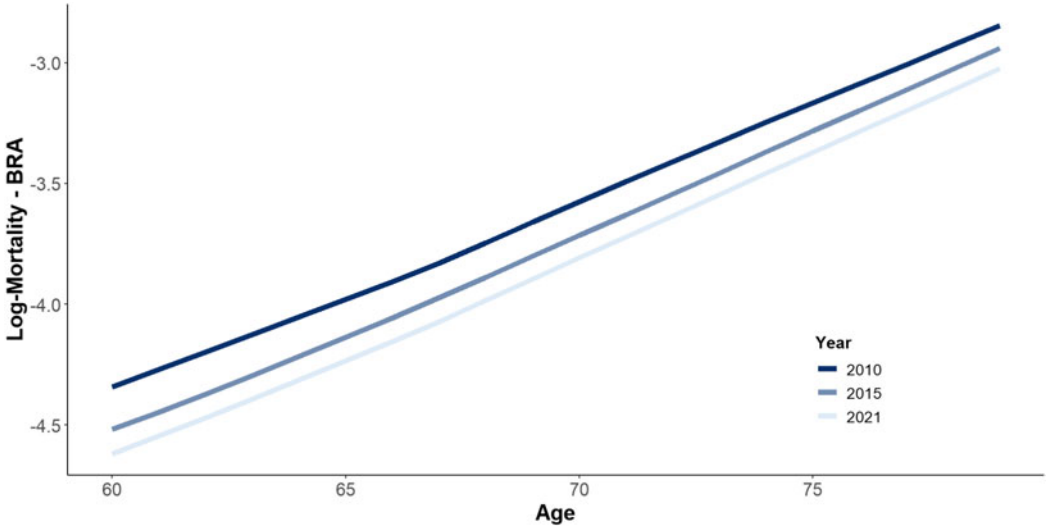


Figure 3. Brazilian national population (BRA) log-mortality rates for males for years 2010/2015/2021, ages 60–80.

ages. In the latter case, the fixed effect deflator model may be represented by a simple Generalized Linear Model (GLM) with a Poisson (or negative binomial) likelihood, where the offset variable is the product of the reference population mortality rates (IBGE) and the pension fund exposure. This approach is known in the industry as a *loading deduction* of a determined mortality/annuity table.

These two Fixed Deflator (**FD**) models are formally defined as:

Model FD-0 (no deflator). The number of deaths is deterministic given $E_{x,t}$:

$$d_{x,t} = m_{x,t}^{BRA} \cdot E_{x,t}. \tag{FD-0}$$

Model FD-1 (constant deflator):

$$\begin{aligned} d_{x,t} \mid \theta, \omega &\sim \text{NegBin}\left(e^{\theta} \cdot m_{x,t}^{BRA} \cdot E_{x,t},\ \omega\right) \\ \theta &\sim \mathcal{N}(-0.5, 0.5^2) \\ \omega &\sim \mathcal{N}(0, 1^2) I_{(0,\infty)}, \end{aligned} \tag{FD-1}$$

where

- NegBin$(\cdot, \omega)$ is a negative binomial distribution. The notation $d \sim \text{NegBin}(\mu, \omega)$ corresponds to $\mathbb{E}[d] = \mu$ and $\text{Var}(d) = \mu(1 + \omega)$. When $\omega = 0$, we recover the Poisson distribution $d \sim \text{Poi}(\mu)$; otherwise, ω captures the amount of variance overdispersion compared to the Poisson base case.
- θ is the constant deflator across all ages relative to the reference population. For example, if $\theta = -0.5$, then, on average, the mortality of the pension fund is $\exp(-0.5) = 61\%$ of the mortality of the reference population. Above we take -0.5 as the prior mean given a ballpark estimate of the deflators and the fact that the pensioners clearly have much lower mortality than the national population. See the end of Section 4.2 below for further discussion of this choice.

Note that the "model" FD-0 does not have any unknown parameters, and therefore, no prior distribution is needed. Moreover, it is not stochastic either, and it can be retrieved from the model FD-1 when $\theta \equiv 0$. For the bonified stochastic model FD-1, for a fixed reference population i, both

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

θ and ω are unknown with priors given as above. The need to generalize from the Poisson to the negative binomial observation likelihood is driven by the well-known overdispersion of death counts, cf. Section 5.1 below.

4.2 Age-dependent deflators

In this subsection, we present the age-dependent deflators' models (**AD**). All three models presented share the same likelihood function, their difference being in the structure (or lack thereof) imposed on the deflators.

The age-dependent deflator fixed effects model (AD-FE) does not impose any structure on the deflators, allowing, in principle, completely unrelated deflators across ages. The model is formally defined as follows:

Model AD-FE (age-dependent deflators fixed effects):

$$\begin{aligned} d_{x,t} &\sim \text{NegBin}\left(e^{\theta_x} \cdot m_{x,t}^{BRA} \cdot E_{x,t},\ \omega\right) \\ \theta_x &\sim \mathcal{N}(-0.5, 0.5^2) \\ \omega &\sim \mathcal{N}(0, 1^2) I_{(0,\infty)}, \end{aligned} \tag{AD-FE}$$

where θ_x is the deflator at age x relative to the reference population. The priors are assigned independently, see Table A.1, and are selected to maximally make the models comparable. For instance, we use the same prior mean -0.5 for θ_x as we did for FD-1. Note that FD-1 is a special case of AD-FE with constant θ_x across ages.

In order to impose structure on age deflators, an autoregressive age dependence among θ_x's has been suggested. In existing literature, van Berkum et al. (2017) proposed an autoregressive structure on θ_x; Li and Lu (2017) also use an autoregressive approach to the age-deflators, which is then applied to a reference population; Oliveira et al. (2021) use a random walk for θ_x in terms of x.

Model AD-AR (Age-dependent deflators with autoregressive predictor):

$$\begin{aligned} d_{x,t} \mid \theta_x, \omega &\sim \text{NegBin}\left(e^{\theta_x} \cdot m_{x,t}^{BRA} \cdot E_{x,t},\ \omega\right) \\ \theta_x \mid \theta_{x-1}, \rho &\sim \mathcal{N}\left(\mu + \rho\theta_{x-1}, 0.5^2(1-\rho^2)\right), \\ \theta_{60} &\sim \mathcal{N}(-0.5, 0.5^2) \\ \omega &\sim \mathcal{N}(0, 1^2) I_{(0,\infty)} \\ \rho &\sim \mathcal{N}(1, 1^2) I_{(0,1)} \\ \mu &= -0.5 \times (1-\rho). \end{aligned} \tag{AD-AR}$$

The structure of the AD-AR model is similar to the one used in van Berkum et al. (2017), where the prior choices ensure that the unconditional mean and variance of the deflators are the same for all ages. To see that, we first compute the mean and the variance of the deflators conditional on the autoregressive parameter ρ (which is itself random). As the prior for ρ restricts its values to the interval (0, 1), the autoregressive process for θ_x is stationary and therefore

$$\mathbb{E}[\theta_x \mid \rho] = \frac{\mu}{1-\rho} = -0.5 \text{ and } \text{Var}(\theta_x \mid \rho) = \frac{0.5^2(1-(\rho)^2)}{1-(\rho)^2} = 0.5^2.$$

Then, using the Law of Total Expectation/Variance, we observe that

$$\mathbb{E}[\theta_x] = \mathbb{E}[\mathbb{E}[\theta_x \mid \rho]] = -0.5$$

and

$$\text{Var}(\theta_x) = \mathbb{E}[\text{Var}(\theta_x \mid \rho)] + \text{Var}(\mathbb{E}[\theta_x \mid \rho]) = \mathbb{E}[0.5] + 0 = 0.5.$$

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

As $\mathbb{E}[\theta_{60}] = -0.5$ and $\text{Var}(\theta_{60}) = 0.5$, our prior assigned for the deflators in the AD-AR model is consistent with those from the AD-FE model.

In both models, for a reference population, the deflators for one age depend (stochastically) on the deflator of the previous age. In the AD-AR model, the parameter μ can be interpreted as the base level for all the deflators, regardless of age, while ρ provides the strength of the dependence between any age and its preceding one. As the deflator of each age depends on the deflator of the immediately younger group, we also set a probability distribution for the first age group being analyzed, i.e. the 60-year-olds.

As a different generalization of AD-FE, we consider GP models for the age-dependent deflators (AD-GP) relative to the reference population. Instead of forcing an autoregressive structure across age groups, the AD-GP model postulates that the dependence between deflators for different ages is given by a Gaussian process driven by y_{ag}. We use the squared-exponential kernel that takes the form

$$c_{ag}(y_j, y_p) := \sigma^2 \exp\left(-\frac{\left(y_{ag}^j - y_{ag}^p\right)^2}{2\phi_{ag}^2}\right). \tag{2}$$

This kernel has two hyperparameters: the process variance σ^2 and the length scale ϕ_{ag}.

For reference population i, let $x \mapsto \theta(x)$ denote now a function representing the deflator for age x. To be consistent with the previous models, when $x \in \{60, 61, \ldots, 89\}$, we write $\theta(x) = \theta_x$, but the reader should note that $\theta(\cdot)$ can be applied to any x, even if it is not an integer value. In the AD-GP model this function is assumed a priori to follow a GP.

Model AD-GP (Age-dependent deflators with Gaussian process):

$$\begin{aligned}
d_{x,t} \mid \theta(\cdot), \omega &\sim \text{NegBin}\left(e^{\theta(x)} \cdot m_{x,t}^{BRA} \cdot E_{x,t},\ \omega\right) \\
\theta(\cdot) \mid \sigma, \phi_{ag} &\sim \text{GP}(-0.5, c_{ag}(\cdot,\cdot)) \\
\omega &\sim \mathcal{N}(0, 1^2) I_{(0,\infty)} \\
(\sigma)^2 &\sim \mathcal{N}(0.5, 0.5^2) I_{(0,\infty)} \\
\phi_{ag} &\sim \mathcal{N}(4, 4^2) I_{(0,\infty)}.
\end{aligned} \tag{AD-GP}$$

As an abuse of notation, the mean function in the second row above is denoted as a single number, -0.5, meaning the deflator for all ages has the same expected value *a priori*.

To interpret the structure of (2), note that the covariance reaches its maximum value $c_{ag}(y_j, y_p) = (\sigma)^2$ when $y^j = y^p$, and when the ages y^j and y^p are far apart $c_{ag}(y^j; y^p) \approx 0$, so that $\theta(y^j)$ and $\theta(y^p)$ are statistically independent. The length scale parameter ϕ_{ag} controls how fast the dependence between two deflators decays in age. If ϕ_{ag} is large, the dependency between two different values of the function decays very slowly, and the function θ is expected to vary only a little, resembling almost a linear function. The process variance parameter σ^2 controls the magnitude of variation (i.e. the amplitude) of $\theta(\cdot)$.

For a fixed age x, we have that, a priori,

$$\theta(x) \mid \sigma, \phi_{ag} \sim \mathcal{N}\left(\beta_0, (\sigma)^2\right).$$

Therefore,

$$\mathbb{E}[\theta(x)] = \mathbb{E}[\mathbb{E}[\theta_x \mid \sigma, \phi_{ag}]] = \mathbb{E}[-0.5] = -0.5$$

and

$$\begin{aligned}
\text{Var}(\theta(x)) = \text{Var}(\theta_x) &= \mathbb{E}[\text{Var}(\theta_x \mid \sigma, \phi_{ag})] + \text{Var}(\mathbb{E}[\theta_x \mid \sigma, \phi_{ag}]) \\
&= \mathbb{E}[(\sigma)^2] + \text{Var}(-0.5) = 0.5 + 0.
\end{aligned}$$

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

Remark. In AD-GP, the deflator function $\theta(\cdot)$ is taken to have the SE kernel c_{ag} (2) in age. In fact, the AD-AR model can also be viewed as a GP model with the exponential Matérn-1/2 kernel $\theta(\cdot) \sim GP(-0.5, c_{M12,ag}(\cdot,\cdot))$, with $c_{M12,ag}(y^j, y^p) = \exp(-\phi|y^j_{ag} - y^p_{ag})|)$ for a (distinct from ϕ_{ag}) length scale hyperparameter $\phi \in \mathbb{R}_+$.

As is always the case in a Bayesian framework, it is important to revisit the prior assumptions about the model parameters. Throughout our analysis, our intent is for the priors to be mildly informative so that they are not strongly impacting the posteriors but still provide some guardrails on what is plausible. For the Brazilian pension funds in question, it is clear that the pensioners have vastly lower mortality than the IBGE reference. Consequently, we wish to start with a prior that leans towards $\theta < 0$, i.e. the deflators are negative. At the same time, we use a fairly wide prior standard deviation of 0.5. The upshot is that the precise prior mean (-0.5 everywhere above) is not so important. To illustrate this point, Figure 11 in the Online Appendix shows AD-GP results under the alternate choice of prior mean $\theta_i = -0.2$. As can be observed, the resulting posterior is almost identical to the one with $\theta_i = -0.5$, indicating that the data evidence strongly dominates the prior choice (as it should). Some minor upward shift in the posterior can be seen at the edges (ages less than 65 and over 85), where the influence of data is weaker. Moreover, the posterior variance is independent of the choice of the prior deflator mean, so the width of the band is invariant.

Further flexibility is provided by the opportunity to take non-constant prior means that depend on age. Indeed, expert knowledge suggests that deflators ought to be smaller for extreme old ages as the compensation law suggests convergence of respective mortality rates (see Gavrilov & Gavrilova, 2005; Richards, 2020). So far, we used a constant prior mean to make the priors of AD-GP, AD-AR, and FD-1 consistent with each other in order to facilitate their comparison. Figure 11 in the Online Appendix further demonstrates the results of AD-GP under the alternate age-dependent prior mean where $\mathbb{E}[\theta_i] = -0.50 + (i - 80) \times 0.04$ for $i > 80$. This represents a strong increase in the prior mean towards zero for higher ages. While the resulting posterior mean of θ in Figure 11 in the Online Appendix is pulled upwards, the deflators are still declining for $x > 80$, indicating that even with a strongly increasing prior trend, the evidence of the data pulls θ_i down. Also note that the width of the posterior band is almost identical to the original setting, showing that change in the prior mean has minimal influence on the posterior uncertainty.

4.3 Time-dependent deflators

In this subsection, we present the time-dependent deflators' models (TD), which follow the same structure as the age-dependent ones from Section 4.2. Instead of focusing on the role of age, our interest lies in introducing time-dependence for the deflators to capture the idea that the evolution of the pension funds' mortality and of the national populations follow different trends.

When deflator-based models are used for forecasting, two distinct sources of uncertainty arise. The first is associated with projecting the reference mortality table itself (the *reference trend*), which in practice may be obtained using standard mortality forecasting techniques, such as Lee–Carter or its coherent extensions. The second source stems from the stochastic evolution of the deflators $\theta(t)$ (the *deflator trend*), which captures the relative divergence between the pension fund and the reference population. In our framework, these two components are conceptually distinct and assumed independent, so that their respective forecast uncertainties can be combined when constructing predictive distributions for future pension fund mortality.

Similar to the notation in Section 4.2, we denote by $\theta(t)$ the function that returns the deflator for the year t when the reference population i is fixed. We first consider the autoregressive and the random walk specifications:

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

Model TD-AR (time-dependent deflator with autoregressive predictor):

$$\begin{aligned} d_{x,t} \mid \theta(\cdot), \omega &\sim \text{NegBin}(e^{\theta(t)} \cdot m_{x,t}^{BRA} \cdot E_{x,t},\ \omega) \\ \theta(t) \mid \theta(t-1), \rho &\sim \mathcal{N}\left(\mu + \rho\theta(t-1), 0.5^2(1-(\rho)^2)\right), \\ \theta(2013) &\sim \mathcal{N}(-0.5, 0.5^2) \\ \omega &\sim \mathcal{N}(0, 1^2) I_{(0,\infty)} \\ \rho &\sim \mathcal{N}(1, 1^2) I_{(0,1)} \\ \mu &= -0.5.(1-\rho). \end{aligned} \tag{TD-AR}$$

Similarly to the AD-AR setup, the TD-AR model defines a parametric stochastic relationship between the mortality deflators of different calendar years. Due to its similarities, the interpretation of the parameters is the same too.

As it may be hard, *a priori*, to pinpoint how the deflators in different years relate to each other, we propose the TD-GP model – the equivalent of AD-GP in the time domain. The analogous calendar-year kernel is

$$c_{yr}(y_j, y_p) := \sigma^2 \exp\left(-\frac{(y_{yr}^j - y_{yr}^p)^2}{2\phi_{yr}^2}\right). \tag{3}$$

Model TD-GP (time-dependent deflator with GP structure):

$$\begin{aligned} d_{x,t} \mid \theta(\cdot), \omega &\sim \text{NegBin}\left(e^{\theta(t)} \cdot m_{x,t}^{BRA} \cdot E_{x,t},\ \omega\right) \\ \theta(\cdot) \mid \sigma, \phi_{yr} &\sim \text{GP}(-0.5, c_{yr}(\cdot,\cdot)) \\ \omega &\sim \mathcal{N}(0, 1^2) I_{(0,\infty)} \\ (\sigma)^2 &\sim \mathcal{N}(0.5, 0.5^2) I_{(0,\infty)} \\ \phi_{yr} &\sim \mathcal{N}(4, 4^2) I_{(0,\infty)}. \end{aligned} \tag{TD-GP}$$

Once again, the GP kernel is chosen from the squared-exponential family (3), and the GP prior mean is constant across different years.

4.4 Direct modeling of the pension fund

The last set of models we consider employs Gaussian processes to model the pension fund population by itself, with no reference population. In lieu of having a baseline mortality, we de-trend the data using parametric prior means. Both a univariate GP that captures only age dependence and a bivariate age-year GP are considered. The respective prior means are (i) a Gompertz (1825) curve with no calendar year covariate (model GP-S1) and (ii) a Gompertz curve with a calendar year covariate (model GP-S2).

Model GP-S1 (S for single-population) directly captures the pension fund log-mortality $\psi(x)$, postulating it to be a function of age x and employing the GP to map from x to $\psi(x)$. GP-S1 assumes that the log-mortality is independent of time t and uses a prior mean $\mu_{age}(x) = \beta_0 + \beta_{ag} x$. This linear trend in x for $\psi(\cdot)$ matches the Gompertz assumption of mortality growing exponentially in age. As with all GP models in this paper, the kernel is chosen as the squared-exponential family specified in (2).

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

Model GP-S1 (single population GP with Gompertz prior):

$$\begin{aligned} d_{x,t} \mid \psi(\cdot), \omega &\sim \text{NegBin}(e^{\psi(x)} \cdot E_{x,t},\ \omega) \\ \psi(\cdot) \mid \beta_0, \beta_{ag}, \sigma^2, \phi &\sim \text{GP}\left(\mu_{ag}(\cdot), c_{ag}(\cdot,\cdot)\right) \\ \mu_{ag}(x) &= \beta_0 + \beta_{ag}(x-60) \\ \omega &\sim \mathcal{N}(0, 1^2) I_{(0,\infty)} \\ \beta_0 &\sim \mathcal{N}(-5, 1^2) \\ \beta_{ag} &\sim \mathcal{N}(0.1, 0.1^2) \\ \sigma^2 &\sim \mathcal{N}(0.5, 0.5^2) I_{(0,\infty)} \\ \phi_{ag} &\sim \mathcal{N}(4, 4^2) I_{(0,\infty)}. \end{aligned} \tag{GP-S1}$$

We emphasize that $\psi(x)$ is now the log-mortality rate, so it is on a different scale compared to the deflators θ_x considered in the previous section. To guarantee our prior views are consistent across all models, we compare model GP-S1 with AD-GP. Based on the likelihoods, both mortality rates should be the same, implying that

$$e^{\psi(x)} E_{x,t} = e^{\theta_x} m_{x,t}^{BRA} E_{x,t} \Rightarrow \psi(x) = \theta_x + \log\left(m_{x,t}^{BRA}\right).$$

Conditioning on the parameters of both models, on the one hand, we have that

$$\mathbb{E}[\psi(x) \mid \dots] = \beta_0 + \beta_{ag}(x-60) \tag{4}$$

and on the other, to have consistency with AD-GP,

$$\mathbb{E}[\psi(x) \mid \dots] = \mathbb{E}[\theta_x + \log\left(m_{x,t}^{BRA}\right) \mid \dots]. \tag{5}$$

Therefore, taking the expectation in (4) and (5), we would like to have

$$\mathbb{E}[\beta_0] + \mathbb{E}[\beta_{ag}](x-60) = -0.5 + \log\left(m_{x,t}^{BRA}\right), \tag{6}$$

where we used the fact that under AD-GP model assumptions the unconditional mean of the deflators is set to -0.5 for all ages. In order to define a reasonable prior mean for β_0 and β_{ag} we do a standard linear regression of $m_{x,t}$ against $(x-60)$ (note that g is fixed but t is not, so there are multiple observations for each age x) and then set the prior mean of β_0 to the resulting y-intercept minus 0.5, and the prior mean of β_{ag} to the resulting slope in $(x-60)$. For the population under analysis, the prior means for β_0 and β_{ag} are then chosen as, respectively, -5.0 and 0.1.

Remark. Placing a prior mean can be interpreted as parametrically de-trending the original data and then using the non-parametric GP to model the residuals. Because the globally imposed β's in μ_{ag} create a correlation between $\psi(x)$'s as x varies, the latter residual GP should have a weaker dependence structure, i.e. a shorter length scale ϕ_{ag} that decorrelates the above residuals relatively quickly across different ages.

Model GP-S2 generalizes GP-S1 by fitting a two-dimensional GP for the mortality surface in age-year. Accordingly, it also employs a bivariate prior mean that accounts for both the age x and the year t in the GP modeling the pension fund's log-mortality. The GP kernel in this case is chosen as the separable bivariate squared exponential kernel

$$c(y_j, y_p) = \sigma^2 \exp\left(-\frac{(y_{ag}^j - y_{ag}^p)^2}{2\phi_{ag}^2} - \frac{(y_{yr}^j - y_{yr}^p)^2}{2\phi_{yr}^2}\right) \tag{7}$$

with three hyperparameters $\sigma, \phi_{ag}, \phi_{yr}$.

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

Model GP-S2 (single population bivariate GP with linear prior mean):

$$\begin{aligned} d_{x,t} &\sim \text{NegBin}(e^{\psi(x,t)} \cdot E_{x,t},\ \omega) \\ \psi(\cdot,\cdot) \mid \beta_0, \beta_{ag}, \beta_{yr}, \sigma^2, \phi_{ag}, \phi_{yr} &\sim \text{GP}(\mu_{ag,yr}(\cdot,\cdot), c(\cdot,\cdot)) \\ \mu_{ag,yr}(x,t) &= \beta_0 + \beta_{ag}(x-60) + \beta_{yr}(t-2013) \\ \beta_{yr} &\sim \mathcal{N}(0, 0.1^2) \\ \phi_{yr} &\sim \mathcal{N}(4, 4^2) I_{(0,\infty)}. \end{aligned} \tag{GP-S2}$$

Revisiting (6), we wish to have

$$\mathbb{E}[\beta_0] + \mathbb{E}[\beta_{ag}]x + \mathbb{E}[\beta_{yr}]t = -0.5 + \log\left(m_{x,t}^{BRA}\right)$$

which is implemented by doing a multiple linear regression of $\log(m_{x,t}^{BRA})$ against age and calendar year (x, t) (keeping reference population i fixed) and then matching the respective coefficients in age and year. and the constant term (minus the -0.5 offset). After performing this exercise for $i = BRA$, we choose the priors for β_0, β_{ag}, and β_{yr} to be centered around -5.0, 0.1, and 0, respectively.

Prior distributions for the hyper-parameters of all models are summarized in Table A.1 in Appendix A.

5. Fitted models

In this section, we present the results obtained from training the 8 models (FD-1, AD-AR, AD-FE, AD-GP, TD-AR, TD-GP, GP-S1, GP-S2) on pension fund 1 data from 2013 to 2018. Supplementary analysis for an alternative pension fund is presented in Appendix B.

Our models were implemented in a Bayesian framework using the `Stan` library in `R` (Carpenter et al., 2017; Stan Development Team, 2018). We implemented the GP models with a method for scaling inference of GPs with stationary kernels in `Stan` using the Fast Fourier Transform, as described in Hoffmann and Onnela (2025). We run 3 chains, each with 10, 000 samples, a burn-in of 2, 000, and a thinning factor of 20. This provided 1, 200 samples for the posterior and predictive distributions for each model.[3]

5.1 Overdispersion parameter

Since all the eight models presented are based on a negative binomial likelihood, the estimated overdispersion parameter ω is comparable across models. The histograms of the posterior distributions of ω for several of the models are shown in Figure 4. The results are qualitatively similar: the mode of the posterior distribution of ω shifts to the right compared to the prior. For most models, the posterior mode is around 0.2, implying that the conditioned variance of the observed number of deaths is 20% higher than its conditioned mean, confirming the expected overdispersion that has been repeatedly observed in small mortality samples. The fact that the posterior of ω is consistent across all models provides internal validation that all the trainings converged and that the assumptions about the deflators are not impacting the assumed relationship between latent mortality and observed deaths. The 90% posterior intervals are far from zero, so we can reject the hypothesis of $\omega = 0$ at 95% confidence level and conclusively infer that the likelihood distribution is not Poisson.

[3] GP deflator models run in approximately 30 minutes, whereas GP-S1/2 models take about an hour on an Intel Core i5 CPU @ 1.70 GHz and 4 cores.

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

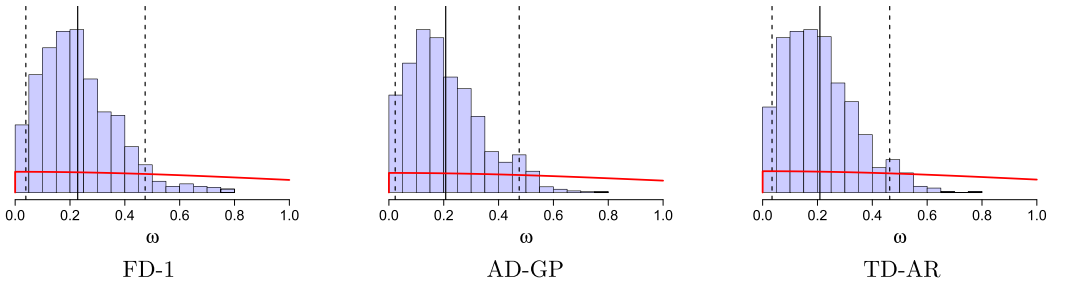


Figure 4. Inferred overdispersion parameters ω for Males. We show the prior density (red curve), the posterior histogram, the posterior mean (solid vertical line), and 90% quantile range (dashed lines).

Across all fitted models, we find that estimates of ω are significantly bigger than zero with high probability. For AD-GP in particular, the posterior probability of $\omega > 0.05$ exceeds 90%; considering that we start with a very weakly informative prior distribution (red density), $H_0 : \omega = 0$ may be rejected.

5.2 Deflators

Before analyzing specific model parameters, we compare the estimated fund deflators θ across the models. Recall that apart from model FD-1, we do not directly assign prior distributions to θ; rather, the prior for θ (and hence its posterior) is induced by priors of other model parameters.

Figure 5 presents the posterior distribution of the deflators in each of the models. Model FD-1 differs from all the others, as it only has a single age- and year-independent deflator, whose posterior distribution is presented as a histogram. For the three age-dependent (AD-) models the horizontal axis denotes the different ages, while in the time-dependent models (TD-) it denotes the year.

The single deflator estimated in FD-1 has a posterior density with mass concentrated between -1.0 and -0.5; a similar range is observed for the estimates in all other models. Hence the fund mortality is in the range of $\exp(-1) = 36.8\%$ to $\exp(-0.5) = 60.6\%$ of the national mortality.

The time trend is difficult to discern as we have access to just six training years (2013–2018); nevertheless, both TD-AR and TD-GP suggest a downward trend. The take-away is that there is a weakly identifiable decrease in θ's over the years, i.e. the relative gap between the fund mortality and the national population is getting wider. In Figure 5, we also present the predictive distribution for the deflators in 2019, which, as expected, yields a wider predictive range.

For pension fund 1, the age-dependent models highlight that the deflators are strongly non-constant across ages, indicating a major deficiency of FD-1 that assumes a constant θ. The disparity of deflators is more than $\max_{x,x'} |\theta_x - \theta_{x'}| > 0.5$ across ages, translating into a 30% relative ratio change. As we confirm below, a more flexible structure than FD-1 is necessary for acceptable goodness-of-fit. In the specific pension fund, we find a systematic pattern of increasing values of θ_x for younger ages ($x \in [60, 70]$), a U-shape between ages 70–80, and decreasing deflators for ages above $x > 80$. The models disagree on the strength of the pattern for ages 80–90.

Although the deflators are significantly age-dependent, the quantity of the available training data leads to quite wide uncertainty bands of the inferred deflators. As a result, there is insufficient data to identify the fine features of this age dependence, as highlighted by the differing patterns of AD-AR and AD-GP models.

It is also instructive to study the posterior uncertainty and variability of θ_x across ages. The haphazard pattern of the AD-FE estimates showcases the need to impose structure; otherwise, deflators for adjacent ages might end up drastically different (see ages 68 and 69 in Figure 5).

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

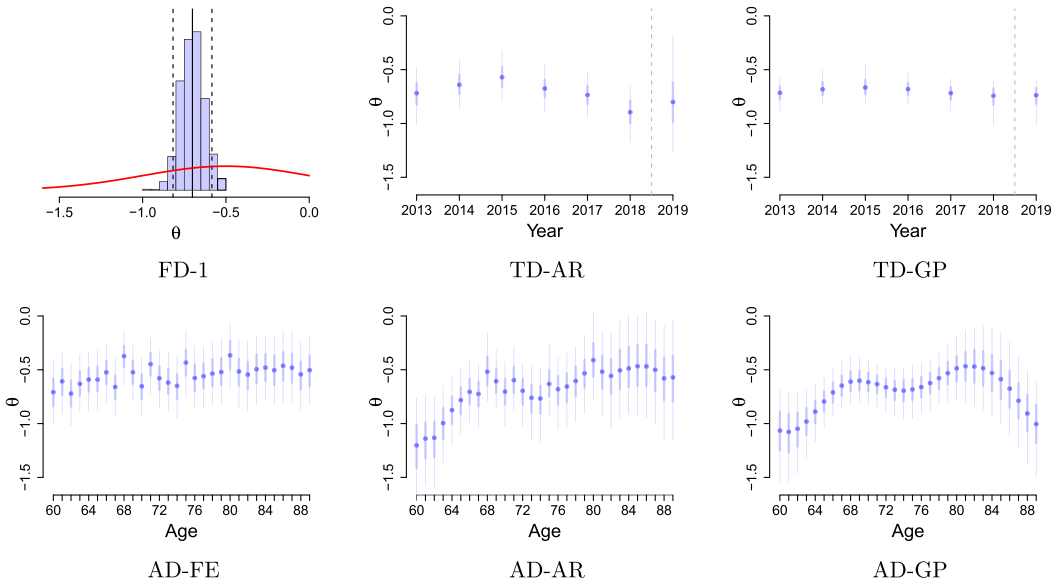


Figure 5. Inferred deflators $\theta(\cdot)$ across six models. For the FD-1 model we show the prior and posterior densities. For all other models, thicker error bars denote the 50% posterior credible interval, thinner bars the 90% interval, and the dots the posterior mean. For models AD-FE, AD-AR, and AD-GP; the horizontal axis denotes age, while for TD-AR and TD-GP, it denotes year. For the time-dependent models, 2019 is a forecast.

AD-AR achieves some of this smoothing; however, it also has very wide posterior uncertainty, indicating that the model is not able to learn much. Due to the smooth nature of GPs, their estimates are less volatile. This is a desirable feature, as we expect the gap between the funds' mortality rates and the country-wise mortality to be similar for nearby age groups. In turn, the assumed smoothness of the GP deflators allows for significant information fusion across ages, which leads to tighter posterior uncertainty bands and hence the most informative take on the age structure of θ's. Information fusion is most pronounced at the middle of the training range; note how the AD-GP error bars for ages 68-78 are significantly narrower compared to ages above or below and much narrower than those same ages in AD-FE. Similarly, the error bars of TD-GP are almost twice as narrow as those of TD-AR. Even more pronounced, the assumed smoothness in year makes the out-of-sample predictive band of TD-GP for 2019 much tighter than that of TD-AR.

Remark. We also tested a random walk of second order (AD-RW2) as an alternative smoothing approach; see Gascoigne et al. (2025). In the RW2 model, the distribution of the deflator at age x, θ_x, is recursively given in terms of θ_{x-1}, θ_{x-2}, the deflators at the "previous" two ages. However, there are two fundamental limitations of such AD-RW2 model. Firstly, it is non-stationary; in other words, the prior distribution of the deflators θ_x cannot be made invariant in x. This is in contrast to all other models we consider, which are stationary and where we specially pick priors such that θ_x has the same prior mean and prior variance across ages x and across all the models (FD-1, AD-AR, AD-GP, etc). This is essential to make the models comparable. Secondly, and linked to the above nonstationarity, the choice of the anchor ages inherently affects the overall prior of the deflator curve. As a result, this introduces yet another degree of freedom to be resolved. Even though it may provide smoothness in our application, θ_x variances greatly increase for ages away from 75, creating a fan-like effect on the posterior uncertainty. Given the limitations of using a nonstationary model, we did not include the AD-RW2 in the results.

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

5.3 Inferred mortality rates and death count distribution

We next shift our focus to the overall predictions of the pension fund's mortality in terms of respective mortality rates and predicted death counts, including uncertainty thereof. Depending on the model choice, mortality of the pensioners is either captured directly or implied by the corresponding deflator.

The top row of Figure 6 shows predicted log-rate (in blue) alongside two other quantities: the country-wide log-mortality (green squares) and the fund's raw mortality $\mu_{x,t}^{raw} = \log(d_{x,t}/E_{x,t})$ (red crosses) for the out-of-sample calendar year 2019 and three representative models: FD-1, AD-GP, and GP-S2. The bottom row of the figure displays the respective predictive distribution of age counts. For each age x, we show the observed number of deaths $d_{x,t}$ in 2019 as red crosses and a summary of the posterior predictive distribution in blue. The size of the circles represents the predictive probability assigned to a specific pair (age, number of deaths). The figure also includes the 50% and the 90% predictive intervals, represented, respectively, as thin and thick vertical lines. Since death counts are integers, the predictive distribution is discrete, taking on only 2–4 likely values due to the small size of the pension fund. Figure B.1 in the Appendix presents the complete results for all 8 models tested.

Since FD-1 assumes a single deflator for all ages, its resulting log-mortality curve is a parallel shift of the reference mortality. Because the latter is based on the Gompertz law, the resulting log-mortality is linear in x. In contrast, the age-dependence of AD-GP and GP-S2 allows for much more flexibility. The raw mortality rates reinforce the pattern of smaller deflators (i.e. larger gaps to national mortality) at ages below 70. Even though the pointwise forecasts (posterior means) present a decreasing pattern in the AD-GP model around ages 85–90, this can not be deemed statistically significant, as the Bayesian interval even for age 90 contains the point estimate for age 85.

While the log mortality rates provide some interpretation on the process driving the pension population's deaths, ultimately the only observable quantity is the actual number of deaths. Overall, the qualitative behavior of the models is largely consistent: for older (above 85) and younger (under 65) ages, all models assign high probability to zero or one deaths, while the highest expected number of deaths is at either 71 or 72 years. The only age whose number of deaths is consistently (i.e. across all models) underestimated is 84, where 4 deaths were observed and all models' maximum a posteriori (MAP) are at zero and the means are lower than one. Other than that, all observations fall at least inside the 90% predictive intervals, and all the models generate a "triangular" structure across ages, consistent with the total number of deaths in Figure 2. Note that while we aim to have the predictive intervals for $d_{x,t}$ contain the out-of-sample observations in the bottom row of Figure 6, there is no such interpretation for the log-mortality rates in the top row, where the presented uncertainty is about the *latent* mortality curve and not about raw observations (in other words, those error bars are omitting the extra uncertainty of the negative binomial).

Among the age-dependent models, the log-rates of AD-GP are the smoothest, while AD-FE presents the roughest, consistent with Figure 5. Once again, we observe the benefits of GP-based models, as smooth log-mortality rates are generally expected. On the other hand, the AD-GP results shown in Figures 6 and 8 exhibit behavior that may be inconsistent with the compensation law of mortality. Although the AD-GP approach provides smoothness, it may also permit implausible boundary behavior, such as decreasing mortality rates at ages above 85.

Since TD-AR and TD-GP have no age-dependency, their predicted log-mortality curves for 2019 are also parallel shifts of the BRA curves. As seen in Figure 5, the predicted point estimate for the 2019 deflators is very similar for TD-AR and TD-GP, making the resulting log-mortality curves almost indistinguishable in Figure 9 in the Online Appendix.

Predicted log-mortality rates for the single population models (GP-S1 and GP-S2) are found in the bottom row of Figure 9 in the Online Appendix. Of note, the rates inferred by GP-S2 look like

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

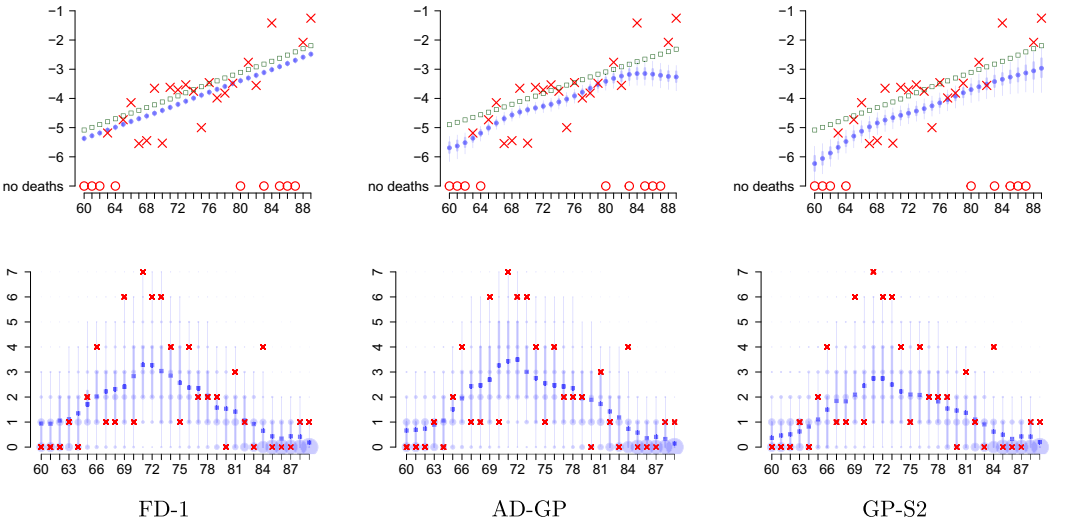


Figure 6. *Top:* predicted pension fund's male log-mortality rates as a function of age x for year 2019 (blue) induced by models FD-1, AD-GP, and GP-S2. We compare it to the reference population (BRA) mortality curve (green squares) and raw mortality (red crosses). *Bottom:* predicted number of deaths for each age $x \in \{60, \ldots, 89\}$ in $t = 2019$ (blue) relative to the observed number $d_{x,t}$ of deaths (red crosses). Circle sizes represent the probability assigned to the corresponding number of deaths, thin/thick lines represent 50%/90% posterior intervals, and the squares are the posterior means.

a smoothed version of those found by AD-AR, where the fund's rates are much lower than the national ones for younger ages (60–70) but show signs of convergence for older ages (80–90).

5.4 Specific model parameters

All model hyperparameters proposed are easily interpretable; Figure 12 in the Online Appendix shows the posteriors for AD-AR, AD-GP, TD-AR, and TD-GP models. Figure 13 in the Online Appendix presents the posterior distributions of key model parameters for the GP-based single population mortality models GP-S1 and GP-S2. Qualitatively comparing the posteriors (blue densities) with the assigned priors (red densities), we confirm that our priors are, indeed, weakly informative, as claimed in the previous section.

Presence of Age-Correlation/Persistence: In the two autoregressive models AD-AR and TD-AR, a prior distribution is only directly assigned to the slope ρ, whereas μ's are induced by the former. As seen in AD-AR's results, the posterior distributions for these two parameters end up being essentially equivalent (apart from a change in scale). The posterior mean of ρ for AD-AR is 0.78 with a 90% posterior quantile interval of [0.28; 0.99]. The slope parameter is related to deflator persistence. Recall that the random walk (RW) model is a special case of the autoregressive (AR) model, in which the slope parameter ρ is equal to 1. Thus, an AR model exhibits higher persistence when its slope parameter is closer to 1, but the process reverts to its mean fairly quickly. We observe that the ρ parameter of TD-AR is a bit lower than that of AD-AR, meaning that the age-dependent deflators have "longer memory" than time-dependent deflators.

Age and Year Length Scales: Figure 7 shows the posteriors of ϕ_{ag} and ϕ_{yr} across the various GP-based models. For the age length scale, the estimated 90% posterior ranges are [2.42;10.09] (posterior mean 5.52) for AD-GP, [2.22;12.11] (mean 6.73) for GP-S1, and [2.36;12.47] (mean 6.90) for GP-S2. This can be interpreted as fusing information across about $2\phi_{ag} \simeq 10$ age groups when estimating a given deflator. This estimate is consistent across the 3 models and matches other age length scales in the literature, which globally reflect the idea of a generation being about 15 years. We note that the posterior is very wide, so the models are not able to definitively infer

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

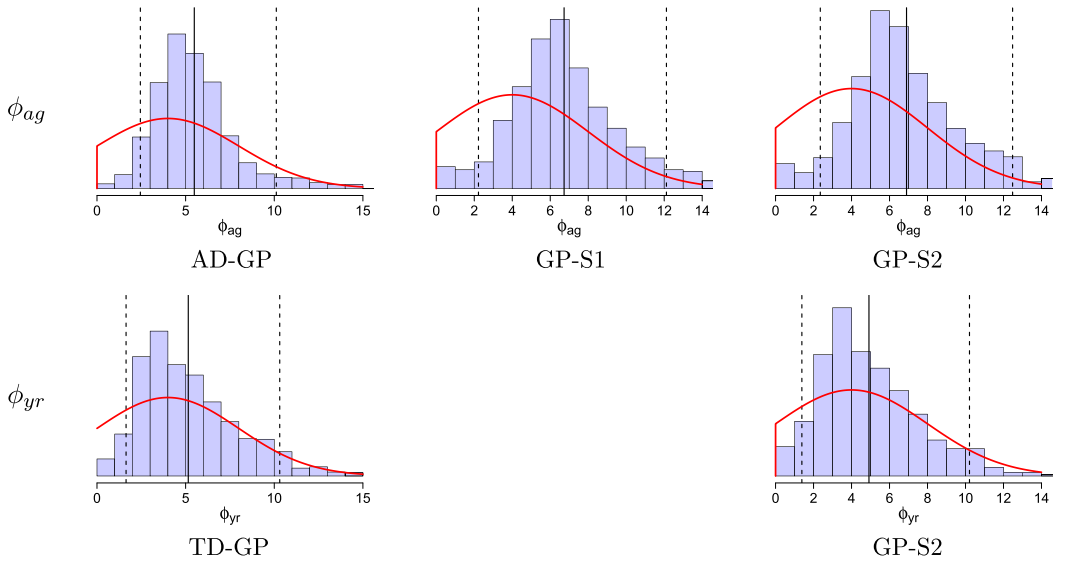


Figure 7. Posterior distribution of GP length scales: ϕ_{ag} in AD-GP, GP-S1, and GP-S2 (*top row*) and ϕ_{yr} in TD-GP, GP-S2 (*bottom row*) models. The red curve indicates the prior density as listed in Table A.1; the vertical solid (dashed) lines denote the posterior mean (90% posterior interval).

the age-persistence of deflators in the pension fund. While the dependence structure of a GP with a squared-exponential kernel is quite different from the autoregressive one, as a point of comparison, the GP structure implies that $cor(\theta_x, \theta_{x+1}) = \exp(-1/(2\phi_{ag}^2))$ while the AR gives $cor(\theta_x, \theta_{x+1}) = \rho$. Thus, $\rho = 0.9$ roughly matches $\phi_{ag} = 2.2$, and $\rho = 0.99$ translates to $\phi_{ag} = 7.1$.

The estimated temporal length scale ϕ_{yr} has a posterior mean of 5.1 in TD-GP (posterior 90% quantile interval of [1.64;10.30]) and 4.90 in GP-S2, again showcasing consistency. Furthermore, note that relative to the age-dependent deflators, $\phi_{yr} < \phi_{ag}$, i.e. the temporal correlation is weaker than across ages, which is intuitive and consistent with prior research.

Parametric Trends: the estimated age trend is $\beta_{ag} \simeq 0.12$ for GP-S1 and $\beta_{ag} \simeq 0.11$ in GP-S2. The estimated year trend in GP-S2 is $\beta_{yr} \simeq -0.078$ ($[-0.15; -0.01]$), which is large but unavoidable. This means that in 5 (9) years, the mortality rates will be approximately 30% (50%) lower, a rather unrealistic longevity improvement. The same coefficient estimated for the Brazilian national population is -0.016, considering the same ages and calendar years. This relative gap in mortality improvement cannot persist for long periods of time but could well hold over a decade or so, implying that the GP-based models are not suitable for long extrapolation in years. Mortality rate curves obtained with time deflators are increasing and smoother than the ones for the age deflator models, since the cross-section age structure of the reference population is kept for the pension fund population.

By construction, the estimates of β_0 should be consistent across models, corresponding to the estimated mortality at age 59. As shown in the figures, we get $\beta_0 \in [-7, -4]$ in all cases with very similar posterior means. One may notice a difference in posterior means of β_0 across the two GP models, with GP-S1 showing a lower β_0 compared to GP-S2. This is due to the use of a year covariate in GP-S2. Since the GP-S1 model does not take into account a linear decrease in log-mortality rates over time, β_0 is lower in this model to best fit lower mortality rates as time goes by.

When directly modeling mortality rates, one may get results for smaller subpopulations exhibiting unrealistic trends over the medium to long term. Within our sample, a notably high annual improvement rate was observed (see the posterior histograms of β_{yr} in Figure 13 in the

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

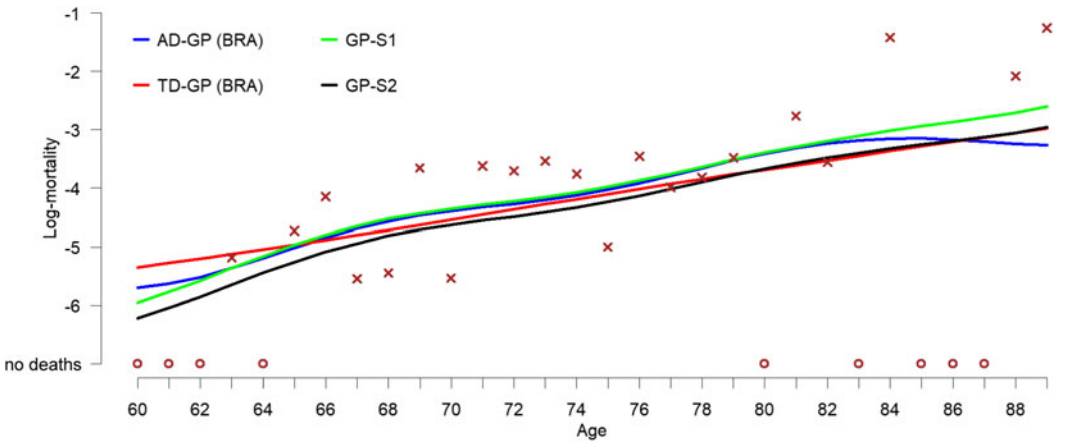


Figure 8. Predicted pension fund 2019 log-mortality rates for ages 60–89 based on the proposed GP-based models trained on 2013–2018. *x*'s indicate the observed raw mortality rates in 2019; note that for some ages there were zero recorded deaths, shown as circles at the bottom of the plot.

Online Appendix). Consequently, it appears advantageous to forecast mortality within (small) pension funds by employing a reference population that has been adjusted by deflators. Utilizing age deflators allows for some adaptability in addressing variations across different age groups in mortality patterns between the pension fund and the reference population, while still reflecting the improvements over time of the latter. Conversely, time deflators accommodate differences in temporal improvements between the two populations but offer limited flexibility in addressing idiosyncratic variations across different age groups. In this context, one may suggest to employ deflators on both coordinates: age and time. However, employing such an approach may offer minimal or no advantage compared to directly modeling the mortality rates of the pension fund.

5.5 Out-of-sample performance

Figure 8 presents predicted 2019 pension fund mortality across the GP-based models. We show out-of-sample projects of the AD-GP and TD-GP models as well as the forecasted log-mortalities for GP-S1 and GP-S2. Comparing models AD-GP and TD-GP to GP-S1/2 can shed light on the value of having a reference population and borrowing information from it, compared to a direct approach for the pension data.

The reader is reminded that, under the AD-GP, the calendar year dynamics of the pension fund population are determined by the calendar year dynamics of the reference population. On the other hand, for time-dependent deflator models, the age structure of the pension fund population is based on (deflated) reference populations, and there is effectively a single deflator across ages for a given year, but the time dynamics of the pension fund population are more flexible.

Under the TD-GP model, when forecasting the pension fund mortality, it is necessary to first forecast the reference population mortality rates. For the sample period, we used the available and published mortality rates, but this task may be done for non-observed future years by an adequate model for national populations such as, for example, the Lee and Carter (1992) model.

We observe that the 2019 projections are quite similar across the four models in Figure 8, with differences largely confined to the edges (ages less than 65 or over 85). GP-S1 and AD-GP almost exactly match each other in the middle of the range and give the best fit relative to the realized mortality. We also note that GP-S1 and GP-S2 are effectively parallel offsets of each other, i.e. yield the same projected age structure up to a constant that can be attributed to the year trend embedded in GP-S2.

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

6. Model assessment

In order to more rigorously compare the models, we use probabilistic performance metrics based on evaluating the model-based predictive distributions relative to the observed mortality data. We refer to Czado et al. (2009) for the evaluation of probabilistic forecasts for count data. We consider two proper scoring rules that are applicable for both Bayesian and classical, as well as parametric or non-parametric settings with discrete outcomes. Given a predictive probability mass function $p_{x,t}(\cdot)$ and the respective observed death count $d_{x,t}$, we first compute the log score

$$\text{score}_{\log} = -\frac{1}{N}\sum_{x,t} \log p_{x,t}(d_{x,t}) \tag{8}$$

where N is the number of observations. Since our predictive distributions $p_{x,t}(\cdot)$ are only available through simulation, so are the logarithmic scores. The best fit should present the lowest $\text{score}_{\log}$.

The other measure applied to assess the performance of the tested models is the rank probability score (RPS):

$$\text{score}_{RPS} = \frac{1}{N}\sum_{x,t}\left[\sum_{k=1}^{\bar{d}} \left(P_{x,t}(k) - \mathbb{I}(d_{x,t} \le k)\right)^2\right]. \tag{9}$$

Above, $P_{x,t}(k) = \mathbb{P}(d_{x,t} \le k) = \sum_{\ell=0}^{k} p_{x,t}(\ell)$ is the predictive CDF of $d_{x,t}$ for age x and year t. Empirically, we restrict to $k \le \bar{d} = 10$ given the single-digit distribution of observed age-specific deaths in our pension fund, see Figure 6 bottom row. Thus, $\mathbb{P}(d_{x,t} \le 10) \simeq 1$ for all x, t. This predictive CDF is compared to the realized observation $d_{x,t}$; lower $\text{score}_{\text{RPS}}$ again indicates better fit.

We also computed the mean absolute error (MAE) metric, though given that $d_{x,t}$ has a highly skewed asymmetric distribution, MAE is not very suitable and primarily focuses on ages 70–80, where predictive variance of $d_{x,t}$ is highest. Most models end up with very similar MAEs of about 1.01, except for FD-0, which has a MAE of 1.61.

Table 1 presents the results for each one of the models. The out- and in-sample results listed are the minimum, the mean, and the maximum of the cross-validated yearly log and RPS scores in the sample period 2013–2019 and ages 60–89. In each year, $N = 30$ in equations (8) and (9) for the out-of-sample, and $N = 210 - 30 = 180$ for the in-sample metrics. To compute the scoring rules, we performed a leave-one-year-out cross-validation strategy. As we have 7 years of data for the pension fund (2013–2019), we first remove 2013 and only use years 2014–2019 to fit the models. The predictive distributions for the number of deaths for each age group based on these adjusted models are then compared with the observed death counts and aggregated to give an RPS value and a log score. We then remove the year 2014 and fit the models based on years 2013 and 2015–2019, recomputing the predictive distributions and the performance scores (RPS and log score). By the end of the procedure, we are left with 7 measurements of RPS and log scores, for which we report the minimum, the mean, and the maximum. Note that because FD-0 does not provide a probabilistic forecast, it is omitted from Table 1.

The results confirm that FD-1 is vastly outperformed by all other models and that only capturing the year trend (TD-AR and TD-GP) is similarly inadequate. The best scores are achieved by AD-AR and AD-GP, as well as the GP-S1 model, which can be declared the overall winner. GP-S2 appears to strongly overfit, yielding the best in-sample performance but poor out-of-sample scores.

There is a regulatory requirement in Brazil for pension funds to perform a consistency test on the mortality table used for calculating technical provisions. This consistency test is a chi-square goodness-of-fit test that compares the observed number of deaths for a given set of ages with the expected number of deaths as defined by a mortality table. Accordingly, we conducted these tests using the AT-2000 M table, as well as the mortality rates estimated by our proposed AD-GP and GP-S1 models. The resulting p-values were: 25.67% for the AT-2000 M table, 60.37% for the AD-GP model, and 59.80% for the GP-S1 model.

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

Table 1. Minimum, mean, and maximum of the yearly performance indexes (8) and (9) for leave-one-out cross-validation across years 2013–2019, ages 60–89 for males. *BRA* is the reference population, except for models GP-S1 and GP-S2. Out-of-sample and in-sample results considering models described in Table A.1 and negative binomial likelihood. Bolded numbers indicate the best-performing model for the mean of each metric

	Out-of-sample			In-sample		
	Min	Mean	Max	Min	Mean	Max
Model	RPS					
FD-1	0.503	0.662	0.877	0.685	0.770	0.816
AD-FE	0.484	0.665	0.852	0.652	0.730	0.777
AD-AR	0.460	0.654	0.853	0.635	0.703	0.755
AD-GP	0.476	0.654	0.850	0.642	0.717	0.762
TD-AR	0.516	0.674	0.880	0.686	0.772	0.814
TD-GP	0.499	0.662	0.895	0.688	0.768	0.811
GP-S1	0.472	**0.649**	0.818	0.646	0.718	0.762
GP-S2	0.482	0.667	0.919	0.580	**0.620**	0.649
	Log score					
FD-1	1.156	1.428	1.671	1.492	1.593	1.644
AD-FE	1.146	1.433	1.660	1.446	1.543	1.603
AD-AR	1.115	1.422	1.645	1.421	1.506	1.582
AD-GP	1.136	1.425	1.663	1.541	1.559	1.593
TD-AR	1.173	1.452	1.703	1.501	1.583	1.658
TD-GP	1.153	1.440	1.742	1.576	1.590	1.602
GP-S1	1.135	**1.413**	1.596	1.538	1.548	1.553
GP-S2	1.127	1.428	1.677	1.333	**1.377**	1.426

Although none of the tables or models led to rejection of the null hypothesis – that is, all are statistically consistent with the observed data – the proposed models demonstrated a notably better fit. The higher p-values associated with the AD-GP and GP-S1 models suggest that they more accurately reflect the underlying mortality patterns of the sample, thereby reinforcing their potential applicability in actuarial practice.

7. Conclusion

Pension fund liability management requires the use of consistent, realistic, and up-to-date mortality rates for the fund, as well as their future projection for cash flow discounting. Practicing actuaries tend to rely on national population data for these purposes, in particular to use external benchmarks by official statistics offices that regularly update national table projections. To capture the basis risk between fund mortality and national data, we explored a variety of deflator-based models. Modeling the subpopulation as a deflated version of the national table leverages the stability and credibility of the national projections while still capturing the specific characteristics of the subpopulation.

This work demonstrates that GP-based models add a flexible and robust alternative to the menu of subpopulation mortality models available for actuarial applications. In the examples tested, GP-based models generated smooth mortality curves and were able to capture the variability in age deflators, outperforming more rigid parametric approaches. The ability of GPs to borrow strength across ages and years is particularly valuable in small populations, where data is sparse and traditional models may overfit or underfit.

The analyzed dataset is novel in the actuarial literature and highlights the high inequality in mortality rates across different groups in Brazil, where national mortality rates can be almost

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

double those of a selected subpopulation, such as participants of a pension fund. The subpopulation modeling approach, supported by GP-based deflators, proved effective in borrowing information from the reference population and producing credible mortality nowcasts even in the presence of significant heterogeneity and single-digit observed death counts. Indeed, the motivation for this study came from practitioners who questioned the adequacy of standard annuity tables, such as the AT-2000, for their specific populations. Our approach, based on the actual experience of Brazilian pension funds and validated through consistency tests required by the regulator, provides a better fit and a more transparent methodology for practitioners and regulators alike. Although GP-based deflators performed well in our empirical application, our aim is not to identify a universally superior model but to provide a flexible framework within which different specifications may be appropriate for different populations. In this sense, the relative performance observed here should be interpreted as an illustration rather than a general ranking of models. A related extension would be to consider (Bayesian) model averaging, namely blending the posteriors of several different models to come up with a composite deflator. In preliminary experiments for our main pension fund, averaging the models did not make a meaningful difference, in part because the statistical metrics reported in Table 1 are quite similar.

We also emphasize that the proposed deflator-based approach conditions on an externally supplied reference mortality surface. Historical estimation therefore focuses on modeling relative deviations rather than propagating uncertainty from the construction of national life tables. For forecasting, uncertainty naturally decomposes into a reference-trend component and a deflator-trend component, which we treat as independent and hence stackable. Future work could explore joint modeling strategies that explicitly propagate reference table uncertainty into subpopulation forecasts.

Another feature we encountered was that in the case study discussed, the observed yearly mortality improvements were unusually high, reflecting particularities of the population under analysis. In such situations, we recommend the use of informative priors, naturally afforded by our framework, that borrow information from the projections of the reference population, thus avoiding unrealistic extrapolations and ensuring more stable forecasts.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1748499526100281.

Acknowledgments. We are grateful to two anonymous referees whose careful reading and insightful comments improved the paper.

Data availability statement. The data underlying this study are proprietary and were provided by a Brazilian pension fund under a confidentiality agreement. Data access requests submitted to the authors will be forwarded to the data provider, and approval is subject to their internal policies and procedures. The `R` code that supports the findings of this study is available from the corresponding author upon reasonable request.

Competing interests. The authors declare none.

References

Alexander, **M.**, **Zagheni**, **E.**, & **Barbieri**, **M.** (2017). A flexible Bayesian model for estimating subnational mortality. *Demography*, **54**(6), 2025–2041.

Bienvenüe, **A.**, & **Rulliére**, **D.** (2012). Iterative adjustment of survival functions by composed probability distortions. *The Geneva Risk and Insurance Review*, **37**(2), 156–179.

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

Cairns, **A. J. G.**, **Blake**, **D.**, **Dowd**, **K.Coughlan**, **G. D.**, **Epstein**, **D.**, **Ong**, **A.**, & **Balevich**, **I.** (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, **13**(1), 1–35.

Carpenter, **B.**, **Gelman**, **A.**, **D.Hoffman**, **M.**, **Lee**, **D.**, **Goodrich**, **B.**, **Betancourt**, **M.**, **Brubaker**, **M.**, **Guo**, **J.**, **Li**, **P.**, & **Riddell**, **A.** (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, **76**(1), 1–32.

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

Chen, **T.**, **Dai**, **B.**, **Wang**, **R.**, & **Liu**, **D.** (2014). Gaussian-process-based real-time ground segmentation for autonomous land vehicles. *Journal of Intelligent Robotic Systems*, **76**, 563–582.

Cressie, **N.** (1990). The origins of kriging. *Mathematical Geology*, **22**, 239–252.

Czado, **C.**, **Gneiting**, **T.**, & **Held**, **L.** (2009). Predictive model assessment for count data. *Biometrics*, **65**(4), 1254–1261.

Dowd, **K.**, **Cairns**, **A. J.**, **Blake**, **D.**, **Coughlan**, **G. D.**, & **Khalaf-Allah**, **M.** (2011). A gravity model of mortality rates for two related populations. *North American Actuarial Journal*, **15**(2), 334–356.

Gascoigne, **C.**, **Riebler**, **A.**, & **Smith**, **T.** (2025). Smooth predictions for age-period-cohort models: A comparison between splines and random process. *BMC Medical Research Methodology*, **25**(1), 177.

Gavrilov, **L. A.**, & **Gavrilova**, **N. S.** (2005).The reliability theory of aging and longevity. In: E. J. Masoro & S. N. Austad (Eds.), *Handbook of the biology of aging* (pp. 3–42). Academic Press.

Gompertz, **B.** (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, **115**, 513–583. In a letter to Francis Baily, Esq.

Hardy, **M. R.**, & **Panjer**, **H. H.** (1998). A credibility approach to mortality risk. *ASTIN Bulletin: The Journal of the IAA*, **28**(2), 269–283.

Hoffmann, **T.**, & **Onnela**, **J.-P.** (2025). gptools: Scalable Gaussian process inference with Stan. *Journal of Statistical Software*, **112**, 1–31.

Huynh, **N.**, & **Ludkovski**, **M.** (2021). Multi-output Gaussian processes for multi-population longevity modelling. *Annals of Actuarial Science*, **15**(2), 318–345.

Huynh, **N.**, & **Ludkovski**, **M.** (2024). Joint models for cause-of-death mortality in multiple populations. *Annals of Actuarial Science*, **18**(1), 51–77.

Hyndman, **R. J.**, **Booth**, **H.**, & **Yasmeen**, **F.** (2013). Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography*, **50**(1), 261–283.

IASB (2017). IFRS. 17 insurance contracts. https://www.ifrs.org/issued-standards/list-of-standards/ifrs-17-insurance-contracts/.

IBGE (2016). *Procedimentos para obtenção de uma Tábua Completa de Mortalidade a partir de uma Tábua Abreviada – Brasil 2014*. https://ftp.ibge.gov.br/Tabuas_Completas_de_Mortalidade/Textos_metodologico_e_de_analise/Metodologia_para_transformar_uma_tabua_abreviada_em_completa_de_mortalidade.pdf.

IBGE. (2022). *Complete mortality tables*. https://www.ibge.gov.br/estatisticas/sociais/populacao/9126-tabuas-completas-de-mortalidade.html?=&t=resultados.

IBGE. (2024). *Projeções da População: Notas metodológicas 01/2024 Brasil e Unidades da Federação*. https://biblioteca.ibge.gov.br/visualizacao/livros/liv102111.pdf.

Lee, **R. D.**, & **Carter**, **L. R.** (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, **87**, 659–671.

Li, **H.**, & **Lu**, **Y.** (2017). Coherent forecasting of mortality rates: A sparse vector-autoregression approach. *ASTIN Bulletin: The Journal of the IAA*, **47**(2), 563–600.

Li, **J.** (2013). A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population Studies*, **67**(1), 111–126.

Li, **N.**, & **Lee**, **R.** (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, **42**(3), 575–594.

Liu, **X.**, & **Guillas**, **S.** (2017). Dimension reduction for Gaussian process emulation: An application to the influence of bathymetry on tsunami heights. *SIAM/ASA Journal of Uncertainty Quantification*, **5**, 787–812.

Ludkovski, **M.**, **Risk**, **J.**, & **Zail**, **H.** (2018). Gaussian process models for mortality rates and improvement factors. *ASTIN Bulletin: The Journal of the IAA*, **48**(3), 1307–1347.

Marrel, **A.**, **Iooss**, **B.**, **Van Dorpe**, **F.**, & **Volkova**, **E.** (2008). An efficient methodology for modeling complex computer codes with Gaussian processes. *Computational Statistics and Data Analysis*, **52**(10), 4731–4744.

Mori, **H.**, & **Kurata**, **E.** (2008). Application of Gaussian process to wind speed forecasting for wind power generation. In *2008 IEEE international conference on sustainable energy technologies* (pp. 956–959).

Oakley, **J.**, & **O'Hagan**, **A.** (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, **89**(4), 769–784.

OECD. (2018). OECD Economic Surveys: Brazil - Overview. https://www.oecd.org/economy/brazil-economic-snapshot/.

Oliveira, **G.**, **Loschi**, **R.**, & **Assunção**, **R.** (2021). Bayesian dynamic estimation of mortality schedules in small areas. arXiv: *2105.02203*.

Olivieri, **A.** (2011). Stochastic mortality: Experience-based modeling and application issues consistent with Solvency 2. *European Actuarial Journal*, **1**, 101–125.

Olivieri, **A.**, & **Pitacco**, **E.** (2012). Life tables in actuarial models: From the deterministic setting to a Bayesian approach. *AStA Advances in Statistical Analysis*, **96**(2), 127–153.

Plat, **R.** (2009). Stochastic portfolio specific mortality and the quantification of mortality basis risk. *Insurance: Mathematics and Economics*, **45**(1), 123–132.

Rasmussen, **C. E.**, & **Williams**, **C. K. I.** (2005). *Gaussian processes for machine learning*. Adaptive Computation and Machine Learning. The MIT Press.

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

Renshaw, **A. E.**, & **Haberman**, **S.** (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, **38**(3), 556–570.
Richards, **S. J.** (2020). A Hermite-spline model of post-retirement mortality. *Scandinavian Actuarial Journal*, **2020**(2), 110–127. doi: 10.1080/03461238.2019.1642239.
Rohmer, **J.**, & **Idier**, **D.** (2012). A meta-modelling strategy to identify the critical offshore conditions for coastal flooding. *Natural Hazards and Earth System Sciences*, **12**, 2943–2955.
Salhi, **Y.**, & **Loisel**, **S.** (2017). Basis risk modelling: A cointegration-based approach. *Statistics*, **51**(1), 205–221.
Salhi, **Y.**, **Thérond**, **P.**, & **Tomas**, **J.** (2015). A credibility approach of the Makeham mortality law. *European Actuarial Journal*, **6**(1), 61–96.
Sandström, **A.** (2016). *Handbook of solvency for actuaries and risk managers: Theory and practice*. CRC Press.
Stan Development Team (2018). *RStan: The R interface to Stan. R package version 2.18.2*.
Tomas, **J.**, & **Planchet**, **F.** (2015). Prospective mortality tables: Taking heterogeneity into account. *Insurance: Mathematics and Economics*, **63**, 169–190.
United Nations (2015). Estimating life tables for developing countries. *Population Division. Technical Paper No. 2014/4.*
van Berkum, **F.**, **Antonio**, **K.**, & **Vellekoop**, **M.** (2017). A Bayesian joint model for population and portfolio-specific mortality. *ASTIN Bulletin: The Journal of the IAA*, **47**(3), 681–713.

Appendix A. Model specification details

Table A.1. Model specifications for the pension fund number of deaths

Model	Likelihood	Predictor	Prior distributions
FD-0	$d_{x,t} = m^{BRA}_{x,t} E_{x,t}$	–	–
FD-1	$d_{x,t} \sim \text{NegBin}(e^{\theta} \cdot m^{BRA}_{x,t} E_{x,t}, \omega)$	–	$\theta \sim \mathcal{N}(-0.5, 0.5^2)$ $\omega \sim \mathcal{N}(0, 1^2) I_{(0,\infty)}$
Age-dependent deflators (AD)			
AD-FE	$d_{x,t} \sim \text{NegBin}(e^{\theta_x} \cdot m^{BRA}_{x,t} E_{x,t}, \omega)$	–	$\theta_x \sim \mathcal{N}(-0.5, 0.5^2)$ $\omega \sim \mathcal{N}(0, 1^2) I_{(0,\infty)}$
AD-AR	$d_{x,t} \sim \text{NegBin}(e^{\theta_x} \cdot m^{BRA}_{x,t} E_{x,t}, \omega)$	$\theta_x = \mu + \rho\theta_{x-1} + \epsilon_x$ $\epsilon_x \sim \mathcal{N}(0, 0.5^2(1-\rho)^2)$ $\mu = -0.5 \times (1-\rho)$	$\rho \sim \mathcal{N}(1, 1^2) I_{(0,1)}$ $\theta_{60} \sim \mathcal{N}(-0.5, 0.5^2)$ $\omega \sim \mathcal{N}(0, 1^2) I_{(0,\infty)}$
AD-GP	$d_{x,t} \sim \text{NegBin}(e^{\theta(x)} \cdot m^{BRA}_{x,t} E_{x,t}, \omega)$	$\theta(\cdot) \sim GP(-0.5, c_{ag}(\cdot,\cdot))$	$(\sigma)^2 \sim \mathcal{N}(0.5, 0.5^2) I_{(0,\infty)}$ $\phi_{ag} \sim \mathcal{N}(4, 4^2) I_{(0,\infty)}$ $\omega \sim \mathcal{N}(0, 1^2) I_{(0,\infty)}$
Time-dependent deflators (TD)			
TD-AR	$d_{x,t} \sim \text{NegBin}(e^{\theta_t} \cdot m^{BRA}_{x,t} E_{x,t}, \omega)$	$\theta_t = \mu + \rho\theta_{t-1} + \epsilon_t$ $\epsilon_t \sim \mathcal{N}(0, 0.5^2(1-(\rho)^2)$ $\mu = -0.5 \times (1-\rho)$	$\rho \sim \mathcal{N}(1, 1^2) I_{(0,1)}$ $\omega \sim \mathcal{N}(0, 1^2) I_{(0,\infty)}$ $\theta_{2013} \sim \mathcal{N}(-0.5, 0.5^2)$
TD-GP	$d_{x,t} \sim \text{NegBin}(e^{\theta(t)} \cdot m^{BRA}_{x,t} E_{x,t}, \omega)$	$\theta(\cdot) \sim GP(-0.5, c_{yr}(\cdot,\cdot))$	$(\sigma)^2 \sim \mathcal{N}(0.5, 0.5^2) I_{(0,\infty)}$ $\phi_{yr} \sim \mathcal{N}(4, 4^2) I_{(0,\infty)}$ $\omega \sim \mathcal{N}(0, 1^2) I_{(0,\infty)}$
Direct population mortality modeling			
GP-S1	$d_{x,t} \sim \text{NegBin}(e^{\psi(x)} E_{x,t}, \omega)$	$\psi(\cdot) \sim GP(\mu_{ag}(\cdot), c_{ag}(\cdot,\cdot))$ $\mu_{ag}(x) = \beta_0 + \beta_{ag}(x-60)$	$\beta_0 \sim \mathcal{N}(-5, 1^2)$ $\beta_{ag} \sim \mathcal{N}(0.1, 0.1^2)$ $\sigma^2 \sim \mathcal{N}(0.5, 0.5^2) I_{(0,\infty)}$ $\phi_{ag} \sim \mathcal{N}(4, 4^2) I_{(0,\infty)}$ $\omega \sim \mathcal{N}(0, 1^2) I_{(0,\infty)}$
GP-S2	$d_{x,t} \sim \text{NegBin}(e^{\psi(x,t)} E_{x,t}, \omega)$	$\psi(\cdot,\cdot) \sim GP(\mu_{ag,yr}(\cdot,\cdot), c_{ag,yr}(\cdot,\cdot))$ $\mu_{ag,yr}(x,t) = \beta_0 +$ $\beta_{ag}(x-60) + \beta_{yr}(t-2013)$	$\beta_0 \sim \mathcal{N}(-5, 1^2)$ $\beta_{ag} \sim \mathcal{N}(0.1, 0.1^2)$ $\beta_{yr} \sim \mathcal{N}(0, 0.1^2)$ $\sigma^2 \sim \mathcal{N}(0.5, 0.5^2) I_{(0,\infty)}$ $\phi_{ag} \sim \mathcal{N}(4, 4^2) I_{(0,\infty)}$ $\phi_{yr} \sim \mathcal{N}(4, 4^2) I_{(0,\infty)}$ $\omega \sim \mathcal{N}(0, 1^2) I_{(0,\infty)}$

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

Appendix B. Analysis for pension fund 2

The alternative pension fund 2 is about half as big as in the first case study. In 2019 for ages 60–89, it included 1,499 male retirees, with 39 deaths recorded. Compared to pension fund 1, this subpopulation has a higher mortality: the mean deflator (based on fitting the FD-1 model) is $\hat{\theta} = -0.29$ (whereas for pension fund 1 $\hat{\theta} = -0.50$, cf. Section 5.2).

In analogy to Figure 5, Figure B.1 presents the inferred deflators for the models fitted to this subpopulation. Apart from generally higher deflators compared to the fund analyzed in the main body, one striking difference is a much flatter age structure, with the deflators being almost constant (see the AD-GP results) across the entire age range. We also note that due to the smaller population size, the non-GP AD-FE and AD-GP models are extremely noisy, and their inferred age-specific deflators "bounce" around in an undesired manner. This is a telltale sign of overfitting and reaffirms the importance of smoothing across ages that is offered by the GP models. Moreover, the uncertainty around the AD-FE and AD-AR estimates is very large (error bar lengths exceed 1 in Figure B.1), so these models are unable to learn much due to the small population size; in contrast, the AD-GP uncertainty is significantly lower and on par with the uncertainty of FD-1. These features are confirmed in Table B.1, where AD-FE and AD-AR perform worse (out-of-sample) than even the constant-deflator FD-1 model, and AD-FE strongly overfits (large gap between in-sample and out-of-sample scores).

Pension fund 1, discussed in the main text, experienced substantial longevity gains in the analysis period (of more than 7% annually). Since such gains are not sustainable in the long run, our prior choices for the parameters governing mortality improvement were made to "tame" these observed gains. As a consequence, though, both the in- and out-of-sample performance of time-dependent models was negatively impacted (see Table 1). The population of pension fund 2 does not present similar longevity gains (we estimate a 1–2% annual gain, e.g. $\beta_{yr} = 0.016$), which provides an opportunity for time-dependent models to deliver improved performance. Indeed, in Table B.1, the TD-GP model achieves the best out-of-sample RPS and log-score metrics.

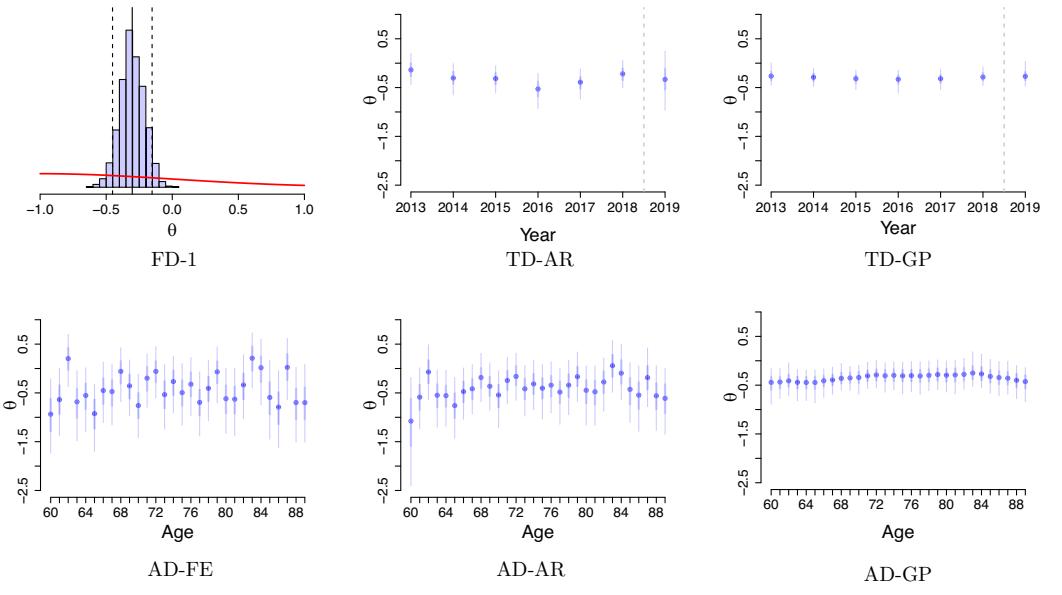


Figure B.1. Inferred deflators $\theta(\cdot)$ for males in pension fund 2. For the FD-1 model, we show the prior and posterior densities. For all other models, thicker error bars denote the 50% posterior credible interval, thinner bars the 90% interval, and the dots the posterior mean. For models AD-FE, AD-AR, and AD-GP, the horizontal axis denotes age, while for TD-AR and TD-GP, it denotes calendar year. For the time-dependent models TD-AR and TD-GP, 2019 is a forecast.

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press

Table B.1. Mean of the yearly performance indexes (8) and (9) for leave-one-out cross-validation across years 2013–2019 for pension fund 2. The shown metrics are for ages 60–89 for males. Out-of-sample and in-sample results considering models described in Table A.1 and negative binomial likelihood. Bolded numbers indicate the best-performing model for each metric

	Out-of-sample			In-sample		
	Min	Mean	Max	Min	Mean	Max
Model	RPS					
FD-1	0.375	0.517	0.653	0.449	0.539	0.625
AD-FE	0.386	0.528	0.681	0.377	0.499	0.602
AD-AR	0.356	0.525	0.670	0.401	0.506	0.604
AD-GP	0.390	0.520	0.669	0.426	0.529	0.632
TD-AR	0.367	0.517	0.656	0.432	0.536	0.641
TD-GP	0.357	**0.515**	0.642	0.442	0.536	0.643
GP-S1	0.371	0.519	0.663	0.421	0.529	0.639
GP-S2	0.373	0.520	0.663	0.406	**0.495**	0.602
	Log score					
FD-1	0.962	1.256	1.637	1.182	1.289	1.303
AD-FE	1.052	1.279	1.486	1.123	**1.221**	1.327
AD-AR	0.972	1.267	1.543	1.127	1.235	1.365
AD-GP	0.988	1.264	1.530	1.179	1.269	1.394
TD-AR	0.854	1.259	1.605	1.183	1.283	1.390
TD-GP	0.9890	**1.256**	1.570	1.178	1.288	1.366
GP-S1	0.960	1.261	1.640	1.172	1.277	1.406
GP-S2	0.937	1.264	1.564	1.110	1.225	1.318

As different subpopulations feature different mortality patterns, it should be understood that no single model will ever be the best for all datasets, and different specifications should always be tested and compared. Nevertheless, the aggregate results across the two pension funds showcase that GP-based models outperform other deflator approaches: AD-GP being the recommended choice for pension fund 1 and TD-GP for pension fund 2. We also see that the deflator-based methods tend to do better than the direct approach, as GP-S2 continues to overfit and GP-S1 offers a generally middling performance.

Cite this article: de Melo E F L, Ludkovski M and Targino R S (2026). Analyzing pension fund mortality with Gaussian processes in a subpopulation framework, *Annals of Actuarial Science*, 1–27. https://doi.org/10.1017/S1748499526100281

https://doi.org/10.1017/S1748499526100281 Published online by Cambridge University Press