Pessimistic predictions in stochastic mortality modelling: Implications in the life annuity framework

Giovanna Apicella

Department of Economics and Statistics, University of Udine,
via Tomadini 30/A, 33100 Udine, Italy
giovanna.apicella@uniud.it

Abstract. In this paper, we propose a case study based on the Italian unisex mortality experience. By means of a backtesting analysis, we show that five selected stochastic mortality models are overly pessimistic about mortality across post-retirement ages, and over different forecasting horizons. Furthermore, we estimate some financial consequences of mortality overprediction, in relation to actuarial and economic measures that are intrinsically based on longevity forecasts.

Keywords. life annuities; mortality forecasts, anti-selection effects, mortality improvements.

M.S.C. classification. 65P05
J.E.L. classification. C63, G22, G52

1 Introduction

Over the last decades, several countries have performed progressive and substantive pension reforms, involving modification of pension system rules and parameters; e.g., see [18] in relation to the European Union Member States. Population ageing and the financial consequences of economic crises have contributed to the need for such changes. One of the key goals of the evolving pension paradigms concern the "adequacy" of retirement benefits for pensioners and the "financial sustainability" of such benefits for pension providers and governments. In this context, the definition of a new balance between private and public pension schemes is prominent. Furthermore, individuals are becoming more responsible for retirement planning and thus more active players with respect to saving and

Mathematical Methods in Economics and Finance – m^{2}ef
Vol. 17/18, No. 1, 2022/23 – Published online on July 29, 2024.

investment choices in their working years and wealth decumulation afterwards ([47]).

Longevity insurance products, such as annuities, can provide retirement income, while protecting individuals against longevity risk and old-age poverty. The related actuarial valuations, such as, for instance, pricing and reserving, require to forecast the mortality profile of annuitants. One of the most deployed approaches to forecast mortality rates relies on extrapolative models, that deduce information about age patterns and underlying mortality trend from the past data and convey this information to the future through stochastic time series methods ([8]). Examples, in this respect, are represented by the Lee-Carter model ([36]) and the Cairns-Blake-Dowd (CBD) model ([15]), that are relatively simple in their predictor structures, and by their extensions. Extrapolative models implicitly assume regularity between the past and the future. This aspect makes the user's judgement relevant, in that a proper choice of the mathematical formulation and of the length of the fitting horizons can lead to some important targets of longevity modelling, such as adequate smoothness and shape of future mortality trends ([60, 38]), along with plausible predicted levels of uncertainty of mortality forecasts across ages ([14]). Nevertheless, predicting the changing pace of the downward mortality trend over time has been challenging. Indeed, it is empirically verified that longevity improvements have been consistently underestimated, and thus mortality rates overestimated. Furthermore, a recent slowdown in mortality improvement rates has been observed in many developed countries since 2010, particularly at ages over 50 ([21]).

In this paper, we aim to examine the financial implications of overly pessimistic mortality forecasts on actuarial measures related to life annuities.

A first aim of the paper is to address the systematicity of mortality overprediction across post-retirement ages, and over different forecasting horizons. Such an attitude is investigated through a backtesting analysis of five stochastic mortality models and of their forecasting performance, relative to the Italian population. A second research objective relates to the consequences of mortality overprediction on the valuation of life annuity contracts. In respect, we measure the financial impact of biases in mortality forecasts on annuity expected present discounted values in the backtesting framework set out for models' fitting and forecasting. Furthermore, we give sense of the extent by which the mortality forecasts from different mortality models should be "optimistic" to mimic the mortality profile expected for annuitants, in terms of quantiles of the distribution of forecasted mortality rates. In the remainder of this Section, we outline the research steps presented in the paper and highlight the related outcomes.

We design a case study based on the Italian unisex mortality experience, aiming to estimate to which extent selected extrapolative mortality models overestimate mortality and to study the persistence of such an attitude over time, by means of a backtesting strategy with rolling windows of data. The stochastic mortality models under study are: the Lee-Carter model, the Renshaw-Haberman model ([55, 27]), the age-period-cohort (APC) model ([20]), the Cairns-Blake-Dowd model and the M6 model ([16]). Furthermore, we estimate some financial consequences of mortality overprediction, in relation to actuarial and economic measures that are intrinsically based on longevity forecasts. Our examples relate to the actuarial value of annuity contracts and cover also their actuarial fairness, namely the value that the contract provides to the annuitant. The annuity contract in our example case is set in different time frames, according to the examined forecasting horizon, and provides payouts between ages 66 and 100. The availability of observed mortality rates allows to portray the realized mortality profile of the annuitant and thus to capture how the overprediction of mortality rates across the ages involved in the contract impacts on the assessment of the insurer's liabilities. Depending on the mortality model, overpredicting mortality translates into an underestimation of the expected present value of the annuity payouts from around 5% (APC) to around 16% (M6), on average over time. This evidence sheds light on some relevant consequences: besides the underestimation of the longevity risk associated with the annuity from the insurer's perspective, the introduction of biases in those measures designed to measure the actuarial fairness of an insurance contract. For instance, the money's worth ratio (MWR) is the ratio between the expected present discounted value of annuity payments (EPDV) and the premium cost of the annuity ([45, 43]). The two values that are compared against each other may rely on different approaches to forecast mortality; for instance, they may result from the adoption of different mathematical formulations and different underlying data from statistical offices, on one side, and insurance companies, on the other one. Such different choices can highly affect forecasting outputs. Accordingly, measures similar to the MWR may incorporate, to some extent, models' (in)abilities to capture mortality improvements over time, and thus may provide biased estimations of the magnitude of actuarial fairness.

In this paper, we adopt a similar rationale to the one underpinning the MWR and we propose a measure we denominate "mortality scenario gap (MSG)". Such indicator is a ratio between two actuarial values for the same annuity contract. MSG expresses in the numerator the expected present discounted value of the annuity based on the mortality forecasts for the general population obtained from stochastic models, according to a given quantile of the distribution of the simulated mortality rates. In the denominator, it expresses the expected present discounted value of the same annuity contract, but based on the mortality rates from annuitants life tables. MSG thus serves as a rule of thumb to measure to which extent the mortality forecasts for the unisex general population, as obtained from different mortality models, should be "optimistic" to reflect the mortality outcomes expected for annuitants. In this respect, we adopt the Lee-Carter and the APC models to obtain mortality forecasts for the Italian unisex population and we use the mortality rates from Italian annuitants life tables known as A62I ([1, 2]), that derive forecasts from methods based on the Lee-Carter model, as adopted by the Italian National Institute for Statistics. This allows us to filter out partially the effects of prediction biases from different modelling approaches. We find that accounting for anti-selection effects, according to the A62I life tables, is almost equivalent to assume that future mortality rates will realize according to a low mortality scenario (2.5% quantile of the distribution of simulated mortality rates) for a population in which the share of male is dominant (80%), based on the projections provided by model APC. Indeed, for this specific case, the mortality scenario gap attains a value of 0.99.

The paper is organized as follows. Sections 2 and 3 describe the mathematical framework and the empirical methodologies, including the backtesting strategy and the methods for mortality models' fitting and forecasting. Section 4 describes the assessment of models' forecasting performance. Section 5 addresses the implications of mortality overprediction on annuity valuation and discusses the mortality scenario gap. Section 6 concludes.

2 Mathematical framework

2.1 Notation and notions

We use notation and basic notions of actuarial mathematics in life insurance and pensions as in [49]. Throughout the paper, we define x as the age of the individual and we denote by t the calendar year spanning from time t to time $t+1$. $c = t - x$ denotes the year of birth of the individual.

Deaths and exposed to risk

$d_{x,t}$ denotes the number of deaths observed between time t and $t+1$ among the people aged x in year t, and $E^c_{x,t}$ expresses the central exposed to risk at age x in calendar year t, that is generally approximated by the average number of individuals with age in the interval $(x, x+1)$ between times t and $t+1$. When only data on the central exposed to risk are available, an approximation of the initial exposed to risk is given by $E_{x,t} \approx E^c_{x,t} + \frac{1}{2} d_{x,t}$.

Fundamental definitions based on the random residual lifetime

Let $\{T_x; x \geq 0\}$ be random variables defining the random lifetime of individuals aged x. For instance, let T_0 denote the random lifetime, measured in years, of a newborn. The probability distribution of T_0 can be used to determine the probability distribution of any T_x, by conditioning on the event of survival to age x:

$$T_x = (T_0 - x) | T_0 > x.$$

Based on the cumulative distribution function of T_x, we can define $q_{x,t}$, namely the one-year mortality rate. It expresses the probability that an individual aged x at time t dies between time t and time $t+1$, and thus before reaching age $x+1$. Given the definition of the random lifetime T_x, $q_{x,t}$ can be expressed as a conditional probability:

$$q_{x,t} = P(T_x \leq 1) = P(T_0 \leq x + 1 | T_0 > x). \tag{1}$$

$p_{x,t}$, the one-year survival rate, defines the probability that an individual aged x at time t survives age $x+1$. Based on the definition of the random lifetime, $p_{x,t}$ can be expressed as the following conditional probability:

$$p_{x,t} = P(T_x > 1) = P(T_0 > x + 1 | T_0 > x) = 1 - q_{x,t}. \tag{2}$$

As shown in equation 2, $p_{x,t}$ and $q_{x,t}$ are probabilities of complementary events, and thus add up to 1.

For any integer k, based on relations between actuarial measures, the probability for an individual aged x at time t to survive further k years is obtained as the product of the one-year survival rates characterizing the individual from age x in year t to age $x+k-1$ in year $t+k-1$:

$${}_kp_{x,t} = P(T_x > k) = P(T_0 > x + k | T_0 > x) = p_{x,t} \cdot p_{x+1,t+1} \cdots p_{x+k-1,t+k-1}. \tag{3}$$

Fundamental definitions based on the force of mortality

A further measure of mortality is the force of mortality $\mu_{x,t}$, that is the instantaneous death rate for an individual aged x at time t, as follows:

$$\mu_{x,t} = \lim_{\Delta x \to 0} \frac{{}_{\Delta x}q_{x,t}}{\Delta x}, \tag{4}$$

where ${}_{\Delta x}q_{x,t}$ denotes the probability for an individual aged x to die in the interval $[x, x + \Delta x]$. For Δx small, this probability is proportional to the amplitude Δx of the interval, according to the coefficient $\mu_{x,t}$.

$${}_{\Delta x}q_{x,t} \approx \mu_{x,t} \Delta x. \tag{5}$$

Throughout the paper, consistently with [16] and further contributions in the actuarial literature, we assume that the force of mortality remains constant within each year of integer age and each calendar year. As stressed in [19], this assumption has two main implications. In first place, it allows to state the relationship between the force of mortality $\mu_{x,t}$ and the one-year mortality rate $q_{x,t}$, as follows:

$$q_{x,t} \approx 1 - exp(-\mu_{x,t}), \tag{6}$$

In second place, if we regard the number of deaths as a random variable $(D_{x,t})$, then $D_{x,t}$ can follow a Poisson ($\mathcal{P}$) distribution conditionally on $\mu_{x,t}$:

$$D_{x,t} \sim \mathcal{P}(E^c_{x,t} \mu_{x,t}),$$

with $\mathbb{E}(D_{x,t}/E^c_{x,t}) = \mu_{x,t}$.

Following this relation, $D_{x,t}$ can follow alternatively a Binomial ($\mathcal{B}$) distribution conditionally on $q_{x,t}$:

$$D_{x,t} \sim \mathcal{B}(E_{x,t} q_{x,t}), \tag{7}$$

with $\mathbb{E}(D_{x,t}/E_{x,t}) = q_{x,t}$.

Accordingly, the maximum likelihood estimates (MLEs) of $q_{x,t}$ and $\mu_{x,t}$ are, respectively:

$$\hat{\mu}_{x,t} = \frac{d_{x,t}}{E^c_{x,t}}, \tag{8}$$

$$\hat{q}_{x,t} = \frac{d_{x,t}}{E_{x,t}}. \tag{9}$$

$\hat{\mu}_{x,t}$ is also known as central mortality rate and denoted by $m_{x,t}$. Following the relation in Equation (6),

$$q_{x,t} \approx 1 - exp(-m_{x,t}). \tag{10}$$

Term life annuities: actuarial measures
In this paper, we consider life annuities that are paid in arrears, namely life annuities that offer a stream of unitary amounts that would be paid at the end of each year during the policy duration provided that the annuitant is alive. We focus on term life annuities and denote by n the maximum number of years for which annuity payouts can be received by the insured. x is the age of the insured at the time of the annuity issue t. The expected present discounted value (EPDV) of such annuity is given by:

$$a_{x,t:\overline{n}|} = \sum_{h=1}^{n} (1+i)^{-h} \ {}_h p_{x,t}, \tag{11}$$

where i denotes the interest rate being the technical basis.

2.2 Mortality models

[16] perform a quantitative comparison of eight stochastic mortality models to describe and explain the evolution of mortality rates. The models they consider are represented according to one of the following structures:

$$log\ m_{x,t} = \sum_i \beta^{(i)}_x k^{(i)}_t \gamma^{(i)}_{(t-x)}, \tag{12}$$

or

$$logit\ q_{x,t} = \sum_i \beta^{(i)}_x k^{(i)}_t \gamma^{(i)}_{(t-x)}, \tag{13}$$

where $logit\ q_{x,t} = log \frac{q_{x,t}}{1-q_{x,t}}$.

In Equations (12) and (13), functions $\beta^{(i)}_x$ describe age-related effects, $k^{(i)}_t$ express period-related effects, and $\gamma^{(i)}_{(t-x)}$ describe cohort-related effects.

Within our study, we use the following models within the family of generalized Lee-Carter models: Lee-Carter(LC) ([36]), Renshaw and Haberman (RH) ([27]), and age-period-cohort (APC) ([20]). Furthermore, we use Cairns-Blake-Dowd

(CBD) model ([15]) and one of its cohort-based extension known as M6 model ([16]). The respective model structures are reported in Table 1.

We thus select the most basic model structures, LC and CBD, and consider simple extensions that add only a function to reflect cohort-related effects, without a corresponding age modulating parameter to be estimated.

As to LC, $\beta_x^{(1)}$ describes the age-specific general shape of mortality. The age/period term, $\beta_x^{(2)} k_t^{(2)}$, consists of a time index $k_t^{(2)}$ that captures the overall mortality evolution over time and an age function $\beta_x^{(2)}$ that modulates the temporal change in the level of mortality across ages.

We consider a substructure of the Renshaw and Haberman (RH) model which was proposed by [27], in which the age modulating term for the cohort effect is set equal to 1. The considered RH model differs from LC only by the presence of one cohort function, $\gamma_{(t-x)}^{(3)}$, that determines the lifelong effects for each generation, identified by the year of birth $t - x$. A further substructure of RH model we use is the so-called age-period-cohort (APC) model, where both age modulating parameters (for the time index and for the cohort effect) are set equal to 1.

CBD model incorporates two time-varying functions, $k_t^{(1)}$ and $k_t^{(2)}$, allowing to capture the imperfect correlation in mortality improvements at diverse ages. The intercept $k_t^{(1)}$ expresses the general level of mortality in year t and affects mortality dynamics in the same way at all ages. The influence on mortality rate dynamics exerted by the slope, $k_t^{(2)}$, is modulated by a pre-specified parametric function of age, $x-\bar{x}$, where $\bar{x}$ is the average age in the considered sample range. Compared with CBD model, M6 incorporates one cohort function, $\gamma_{(t-x)}^{(3)}$.

All models, except for CBD, require a given number of constraints for identifiability, see [16, 19].

One of the key features of the mortality models under study is that they can be projected stochastically so that to obtain probabilistic forecasts of future mortality rates. This can be done by using time series methods to forecast period and cohort functions, as explained in more detail in Section 3.2.

Table 1. Structures of the mortality models implemented in the study

Model	Formula
Lee-Carter (LC)	logit $q_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} k_t^{(2)}$
Renshaw-Haberman (RH)	logit $q_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} k_t^{(2)} + \gamma_{t-x}^{(3)}$
Age-period-cohort (APC)	logit $q_{x,t} = \beta_x^{(1)} + k_t^{(2)} + \gamma_{t-x}^{(3)}$
Cairns-Blake-Dowd (CBD)	logit $q_{x,t} = k_t^{(1)} + k_t^{(2)}(x - \bar{x})$
M6: CBD + cohort effect	logit $q_{x,t} = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + \gamma_{t-x}^{(3)}$

3 Empirical methodology

3.1 Data and backtesting strategy

We use the data from the Human Mortality Database ([31]), relative to the exposed to risk $E^c_{x,t}$ and the deaths $d_{x,t}$ among the total Italian population from 1906 to 2020, so that to select the most reliable time span of data. Data are available for ages 0-110+.

Within our study, we aim to shed light on some stylized features of the forecasting behaviour of the mortality models we consider, namely an attitude that is consistently verified, independently on the specific time horizon over which forecasts are made. This allows to understand commonalities in the forecasting behaviour that turn out to be general, since systematically verified beyond the choice of a given time horizon for calibrating the models and forecasting them. The analysis provides also evidence on the mortality model in our sample that provides the most accurate forecast of the mortality pattern from age 65 to age 99 for the Italian total population. Our case study focuses on annuity contracts providing benefits from age 66 and with duration 35 years. As stressed by [50], the life annuities offered in the markets may be characterized by different policy conditions that relate also to different time frames, with respect to both the accumulation and payment phase. The results shown in this paper are not strictly depending on the chosen time frame for the annuity payouts. Indeed, the policy duration defines only the length of the time horizon over which we obtain forecasts.

The methodology we use relies on backtesting strategies.

We make use of a backtesting strategy with fixed-length windows rolling one-year-ahead through time; see for instance, [22, 41, 4]). Such a strategy employs lookback windows to calibrate the mortality models on the observed data, and lookforward windows to forecast the models and test their performance against realized mortality.

As already mentioned, our case study focuses on annuity contracts with duration 35 years. Accordingly, we allocate 35 years of data to the lookforward windows. Coherently, we use lookback windows made up by 70 years of data, in such a way to reserve two thirds of the data to the fitting and one third of the data to the assessment of the models' out-of-sample performance, according to a rationale that has already been employed in the literature, for instance in [4]. Overall, by exploiting the whole data sample we have available, we can implement the backtesting procedure over eleven time horizons. The first lookback window spans calendar years from 1906 to 1975 and the respective lookforward window extends from 1976 to 2010. The last lookback window is made up by calendar years 1916 up to 1985, and the respective lookforward window spans the time period 1986-2020. In what follows, in line with the state-of-the-art, we define the first year of the lookforward window "stepping off year" ([22]).

In the second stage of the analysis, we perform model forecasting over a random future time horizon, 35 five years-ahead from 2020, based on the fitting to the 70 more recent years of data. In this context, we design a practical case study aimed to further discuss the possible implications of mortality models' forecasting behaviour on measures based on the annuity values.

3.2 Models' calibration and forecasting

We calibrate and forecast the stochastic mortality models under study in the **R** software environment (`https://www.r-project.org/`) through the package *StMoMo* ([58]), which provides specific tools for stochastic mortality modelling, in the framework of generalized age-period-cohort models. The models can be defined by specifiyng the respective features, relative to the link function, the predictor structure, and the set of parameter constraints, or by using the predefined commands. The parameter estimation is performed through maximization of the log-likelihood function based on the assumption that the distribution for the random number of deaths $D_{x,t}$ is Binomial conditionally on $q_{x,t}$, as specified in Equation (7). We thus target the mortality rate $q_{x,t}$ with a logit link function. The central mortality rates $m_{x,t}$ are then obtained according to formula (10). The fitting function requires as input matrices of deaths $d_{x,t}$ and initial exposed to risk $E_{x,t}$. We select data relative to the integer ages x in the interval $[50-99]$, while the calendar years vary according to the specific lookback window. As explained in the previous section, lookback windows roll one-year ahead. Accordingly, each lookback window contains the same data as the previous one except for the first and the last data point. Such procedure allows to update the calibration sample and base the parameter estimation on progressively new data.

Both the forecasting and simulation of mortality rates relies on the modelling of the period-related functions $k_t^{(i)}, i = 1, 2$ and of the cohort-related function $\gamma^3_{(t-x)}$, by means of time series techniques. A common approach is to adopt a multivariate random walk with drift to describe the dynamics of the period-related effects ([27]) and to use a ARIMA(p, d, q) specification for the cohort effects, that is identified based on the BIC and on the biological reasonableness of the forecasts ([14]). In relation to Finnish and Swedish mortality data, [40] selects the optimal ARIMA(p, d, q) specification based on Bayesian Information Criterion (BIC), over the combinations $d \in \{0, 1, 2\}$ and $(p, q) \in \{0, 1, 2\}$, finding that two times differencing is never an optimal choice. As in [57], we select the optimal ARIMA(p, d, q) specification for the cohort-related function of the models RH, APC and M6, by restricting the domain of values for d to $\{0, 1\}$. For these models, we apply the *auto.arima* function from the *forecast* package in **R** ([33]) to the time series of the estimated parameters $\gamma^3_{(t-x)}$, over each lookback window. Such a function performs a search over possible model specifications, under given settings (e.g., maximum p, d and q order and drift inclusion), according to the values of information criteria, such as the Akaike Information Criterion (AIC) and the BIC. For the data set we consider, for all the lookback windows, the optimal ARIMA specifications turn out to be always integrated and, in the vast majority of cases, with both the AR and the MA component included. In this respect, ARIMA$(1, 1, 2)$ is the most prevalent specification over the most recent lookback windows and across models RH, APC and M6.

We assess the uncertainty in mortality projections that arise from the estimation of the parameters of the mortality models by means of semiparametric bootstrap ([11]), through the functionalities of the *StMoMo* package.

4 Assessment of mortality overprediction

We intend to test the forecasting accuracy, across both ages and time, of the stochastic mortality models under study, to assess the overall forecasting performance in relation to the mortality profile of a potential annuitant at post-retirement ages until 99. Tracking the ageing of such a person along the policy duration means targeting the following sequence of point forecasts, $\tilde{m}_{x,t}$, provided by the mortality models, where x and t move simultaneously one year forward:

$$\tilde{m}_{65,t}, \tilde{m}_{66,t+1}, \ldots, \tilde{m}_{99,t+34}.$$

We compare these point forecasts of mortality rates against the realized values that can be observed in the data, $\dot{m}_{x,t}$, over all the lookforward windows where we temporally set the duration of the annuity contract. For instance, in relation to the last lookforward window, we target the following 35 couples of one forecasted death rate and its respective observed counterpart, associated with the same age and calendar year:

$$(\dot{m}_{65,1986}, \tilde{m}_{65,1986}), (\dot{m}_{66,1987}, \tilde{m}_{66,1987}), \ldots, (\dot{m}_{99,2020}, \tilde{m}_{99,2020}).$$

A similar procedure is used in [3].

As the lookforward windows move one year forward in time, we generally denote by lf_1 the stepping off year and by lf_{35} the last calendar year included in a lookforward window.

We use the root mean square error (RMSE) to measure the average errors made by the models under study in relation to the targeted mortality rates. The RMSE is commonly used also to assess the forecasting performance of mortality models, for instance in [57, 37]. We obtain the RMSE with respect to our data as follows:

$$\bar{e} = \left[\sum_{i=1}^{35} w_i \, |e_i|^2 / \sum_{i=1}^{35} w_i \right]^{\frac{1}{2}}, \tag{14}$$

where $|e_i| = |(\dot{m}_{65+i-1,lf_i} - \tilde{m}_{65+i-1,lf_i})/\dot{m}_{65+i-1,lf_i}|, i = 1, \ldots, 35$, is the relative individual error for i-th couple of forecasted and observed mortality rate in the lookforward window and w_i is a scaling associated with each $|e_i|^2$ to express its assumed relative contribution to the total error. We give each $|e_i|^2$ the same weight, this implying $w_i = 1, \forall i$.

In Figure 1, we display a comparative analysis of the magnitude of the RMSE $\bar{e}$ associated with models LC, RH, APC, CBD and M6. Such analysis allows to gain visual insights into the discrepancies in the forecasting performance across models and prediction time spans, as we display RMSEs on all the lookforward windows considered in our case study. The different lookforward windows are displayed on the horizontal axis and are labeled with integer numbers between 1 and 11, with $1 = [1976, 2010], 2 = [1977, 2011], \ldots, 11 = [1986, 2020]$. Models LC, RH, APC, CBD and M6 are associated with colours blue, orange, dark green, cyan and purple, respectively. On any given lookforward window, the higher the bar the greater the RMSE score. Model APC is characterized by the lowest RMSE systematically, on 9 lookforward windows out of 11. Model M6 shows the poorest predictive accuracy among the considered models. Model LC ranks first, in terms of RMSE, on lookforward windows 10 and 11. We split the age interval in three bands, $[65, 74]$, $[75, 84]$ and 85+, to examine in more detail the relative merits of the five models for making short-term, medium-term and long-term forecasts, respectively. It is worth highlighting that model LC persistently shows the highest predictive accuracy throughout the lookforward windows in relation to the integer ages in $[65, 74]$, among the models under study, but to a small extent (15% better RMSE than the other models for this age range on average over all lookforward windows).

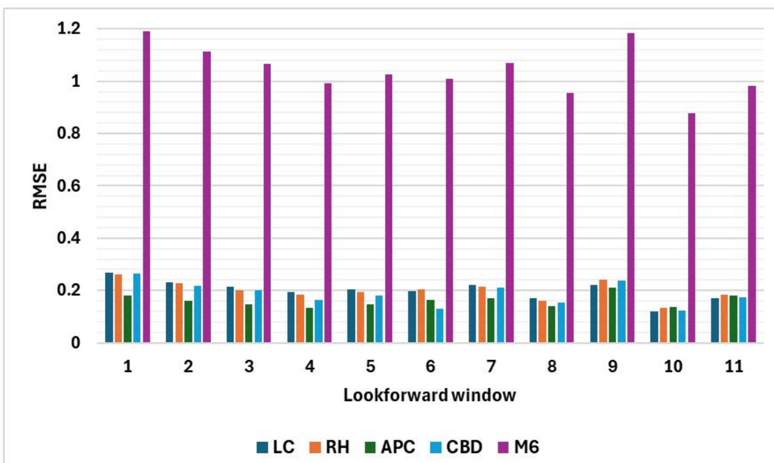

Fig. 1. Comparison, in terms of RMSE, between the forecasting performances of models LC, RH, APC, CBD and M6 over 11 different lookforward windows. Lookforward windows roll one-year ahead through time and are labeled with integer numbers between 1 and 11, with $1 = [1976, 2010], 2 = [1977, 2011], \ldots, 11 = [1986, 2020]$, on the horizontal axis.

To obtain synthetic information about whether the mortality prediction error consists in a misestimation from above (overprediction error) or from below (underprediction error) with respect to the true data, we extrapolate from the individual errors the following ratio between the observed and the model-generated mortality rates, that is of immediate interpretation, as the perfect matching corresponds to the value of 1:

$$Y_i = \frac{\dot{m}_{65+i-1,lf_i}}{\tilde{m}_{65+i-1,lf_i}}, \tag{15}$$

with $i = 1, \ldots, 35$.

[4] uses this ratio as a benchmark for the fitting performance of stochastic mortality models. Indeed, the closer the ratio Y to 1, the lower the fitting error. Futhermore, they provide a stochastic representation of this error to obtain a correction factor for models' predictions. In the context of this study, Y_i is tasked with measuring how much pessimistic the mortality forecasts arising from the chosen mortality model are, compared with the observed data, for each combination of age and calendar year under study, the lookforward window being fixed. For instance, a higher prediction than the mortality rate that is actually observed results in a Y ratio lower than 1. Since we are interested in assessing how well the various models manage to represent the mortality pattern of the annuitant, we measure the average pessimism or overprediction attitude of the models across the various ages attained by the annuitant as we move forward in time along the policy duration. Such average pessimism is expressed by the arithmetic mean of the individual Y_is over each lookforward window, $\bar{Y}$.

In Figure 2, we display how $\bar{Y}$ varies over the considered lookforward window, for models LC (blue), RH (orange), APC (dark green), CBD (cyan) and M6 (purple). We thus provide graphical insights, to evaluate, comparatively, the average pessimism attitude of the models, dynamically as we move from a lookforward window to the next one. All the mortality models under study generate pessimistic mortality predictions compared to the realized data. Nevertheless, APC model turns out to be systematically the least biased model. On average, all the lookforward windows being considered, mortality overprediction bias for model APC is around 8%, that is considerably below the bias assessed for model M6 (around 43%) and mildly below the bias characterizing models LC, RH, and CBD (around 14%). For all models, the pattern of the average pessimism bias shows more variability on the most recent lookforward windows.

The model chosen for predicting mortality contributes to define the demographic assumption that underlies the computation of actuarial measures related to life annuities. For instance, the effect of the user's modelling choice is embedded in the survival probabilities that explicitly appear in equation (11), which defines the expected present discounted value of term life annuity contracts. Since they result from the product of one-year survival rates, h-years survival probabilities, with $h > 1$, compound the effect of biases in mortality forecasting across ages. We thus perform a comparative analysis of the survival curves that are obtained from the outcomes of the five stochastic mortality models under study. In particular, it is relevant to examine survival curves in relation to the post-retirement ages that are involved in our case study. In this respect, we focus on the probabilities that an invidual aged 65 survives h years ahead, with

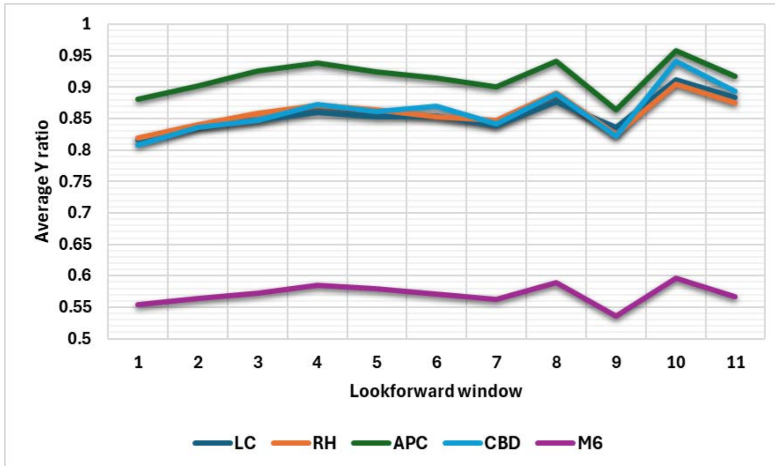

Fig. 2. Pattern of average mortality overprediction bias (average Y ratio) of models LC, RH, APC, CBD and M6, over 11 lookforward windows. Lookforward windows roll one-year ahead through time and are labeled with integer numbers between 1 and 11, with $1 = [1976, 2010], 2 = [1977, 2011], \ldots, 11 = [1986, 2020]$, on the horizontal axis.

$h = 1, \ldots, 35$. Our numerical evidence shows that the pessimistic predictions of the mortality pattern of the annuitants imply expecting lower survival prospects than actually realized. This occurs systematically over the examined lookforward windows.

As in the previous analysis, we show these results through a graphical analysis, that makes more immediate the comparison among models with respect to a common benchmark, being represented by the realized survival curve; the latter curve has been built based on the mortality rates actually observed for the considered annuitants' ages, over a given prediction time span. By way of an example, we focus on the seventh lookforward window within our backtesting scheme, spanning the calendar years 1982-2016. Accordingly, our analysis targets the probabilities ${}_hp_{65,1982}, h = 1, \ldots, 35$, where 65 and 1982 denote the age of the individual and the respective calendar year when such age is attained. We display this graphical evidence in Figure 3, where survival probabilities are plotted against the years ahead h with respect to which they are computed. Six curves are displayed, depending on the mortality rates underlying the survival functions: observed mortality rates (black) or forecasts from models LC (blue), RH (orange), dark green (APC), light blue (CBD) or M6 (purple).

As it is shown in the Figure, the discrepancies between models appear more and more marked as the number of years ahead h increases. The gap becomes narrower, instead, when approaching the maximum number of years ahead considered in this study, namely when considering the probability that an individual aged 65 survives ages close to 100. Model APC provides the closest approximation to the realized survival curve, while M6 turns out the least accurate model in our sample. Furthermore, the gap between the survival curves obtained from models RH, LC and CBD is not enouugh consistent and sharp to determine the dominance of one these models over the others. As shown in the previous analysis, this result holds until lookforward window 8; afterwards, model LC turns out to be the most accurate among all the considered mortality models.

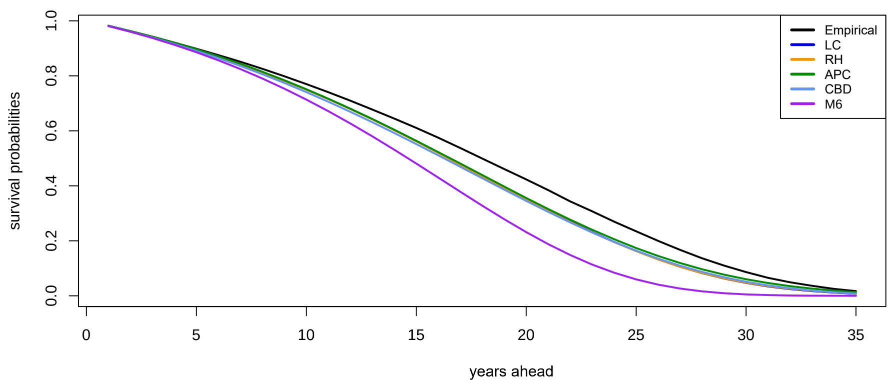

Fig. 3. Curves displaying the probabilities for a person aged 65 in calendar year 1982 to survive further h years, with $h = 1, \ldots, 35$. The curves are based on forecasts from models LC, RH, APC, CBD and M6 and are compared against the empirical survival curve, that is based on the mortality rates actually realized for the same ages and calendar years.

5 The impact of mortality overprediction on annuity values

In this Section, we provide a simple case study that quantifies, in monetary terms, the impact of models' forecasting behaviour on annuity values.

[35] study the impact of mortality improvements for post-retirement ages on annuity values. In this respect, they show how allowing for future mortality improvements is associated with an increase in the present value of annuity payments and analyse the age ranges giving the greatest contribution to such an additional cost, in relation to UK data.

Annuity values are sensitive to the choice of the model used to project future mortality ([27]) and, given the choice of a specified model, rely on the user's judgement about the data that are informative to capture the underlying dynamics of mortality. For instance, [3] show that implementing dynamic forecasting procedures leads to more accurate estimations of the fair value of annuity contracts.

In the context of our study, we show that the considered stochastic mortality models capture only partially the magnitude of the mortality improvements occurred in the past and thus predict higher mortality levels than actually observed in the data. We measure the financial impact of such a systematic overprediction attitude on annuity values in the backtesting framework set out for models' fitting and forecasting. In this way, we can measure the extent of the misestimation of annuity values as we move forward in time.

We focus on an immediate term life annuity, providing a stream of unitary cash flows during the policy duration, for at most 35 years, conditional on the annuitant's survival. For this contract, we compute the expected present discounted value of the annuity payments (EPDV). We assume that the policy is issued on a person aged 65 in the observation year, when the actuarial valuation has to be performed, and that provides benefits from the age 66. We discount the stream of payments with a constant technical rate of return equal to 2%. The backtesting framework we use allows to compute the EDPV that relies on forecasts from stochastic mortality models and to compare it against the EDPV that we could deem true on the demographic side, in that it reflects the realized longevity outcomes over the time horizon where the life annuity is temporally framed and for those post-retirement ages that are involved in the actuarial valuation. Furthermore, the backtesting framework with rolling windows allows to assess over time the misestimation of annuity values that results from the persistent forecasting inaccuracy observed in the previous analysis.

Consistent with equation (11), in the remainder of the analysis, we denote by $a^{E,j}_{65,lf_1:\overline{35}|}$ the EPDV whose computation relies on the mortality rates that are actually observed in the $j-$th lookforward window for the post-retirement ages from 65 to 99. While the superscript identifies the empirical nature of the underlying mortality trajectory and the lookforward window this trajectory relates to, the subscript denotes that the annuitant is aged 65 in the first year of the lookforward window, and that the annuity has duration 35 years. In Figure 4, we display how the empirical EPDV evolves over the eleven lookforward windows we consider in our study. Data updating drives the empirical EPDV toward progressively higher values, as displayed by the blue curve in Figure 4 and by the dotted trend line linking the minimum to the maximum value reached by the EPDV over time. On average, we observe an increase of 0.10 monetary units when moving from one lookforward window to the next one, that reflects the longevity improvements revealed by progressively new data.

We use the empirical EPDV as a benchmark to assess the estimation bias that arises from relying on the mortality models' forecasts. We denote by $a^{M,j}_{65,lf_1:\overline{35}|}$ the theoretical EPDV, relative to the j-th lookforward window ($j = 1, \dots, 11$), for every choice of the mortality model $M = LC, RH, APC, CBD, M6$.

On each lookforward window and for each mortality model, we measure the relative difference between the the theoretical and the empirical EPDV. We denote such measure by $Spread^{TE}_{j,M}$, to emphasize that it expresses a spread

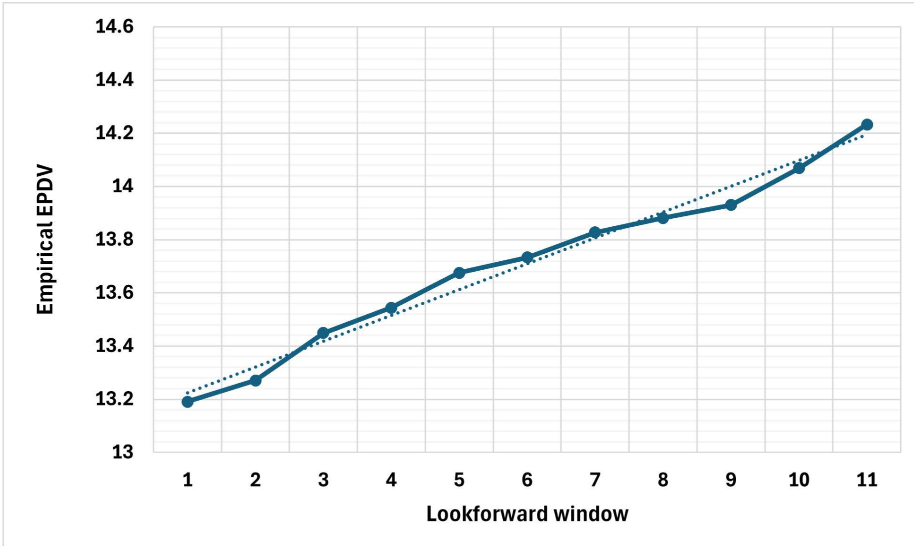

Fig. 4. Temporal evolution, over eleven lookforward windows, of the expected present discounted value (in monetary units) of an immediate term life annuity on a person aged 65. This annuity provides payouts for 35 years, conditional on the annuitant's survival. The underlying survival probabilities are obtained from the observed mortality rates in each lookforward window.

between two EPDV's that differ only by the nature of the underlying survival probabilities, being either model-generated or based on realized mortality rates. The subscript denotes, instead, dependence on the choice of the lookforward window j and of the mortality model M. We compute $Spread^{TE}_{j,M}$ as follows:

$$Spread^{TE}_{j,M} = \frac{a^{M,j}_{65,lf_1:\overline{35}|} - a^{E,j}_{65,lf_1:\overline{35}|}}{a^{E,j}_{65,lf_1:\overline{35}|}} \tag{16}$$

In Figure 5, we propose a bar chart to show the extent by which the over-prediction attitude of different mortality models leads to biased estimates of the EPDV of the annuity contract under study. Indeed, in Figure 5 we display the magnitude (percent values) of $Spread^{TE}_{j,M}$ in relation to each lookforward window (x-axis) and for all the considered mortality model, each identified by a different coloured bar. Since all models underestimate longevity, in all cases $Spread^{TE}_{j,M}$ turns out to be negative, this denoting that the theoretical EPDV consistently approximates from below the empirical one. For instance, on lookforward window 11, choosing model APC as a predictor of mortality evolution leads to an estimate of the fair value of the annuity contract that is 5.56% lower than it can be retrospectively observed (13.37 monetary units against 14.23). In relation to the same lookforward window, $Spread^{TE}_{11,LC}$ is -5.32%. Generally speaking, it is systematically verified that model APC provides the closest estimate to the empirical EPDV over lookforward windows 1-8, with model LC being ranked the second in the sample of considered models. The opposite holds for lookforward windows 9-11.

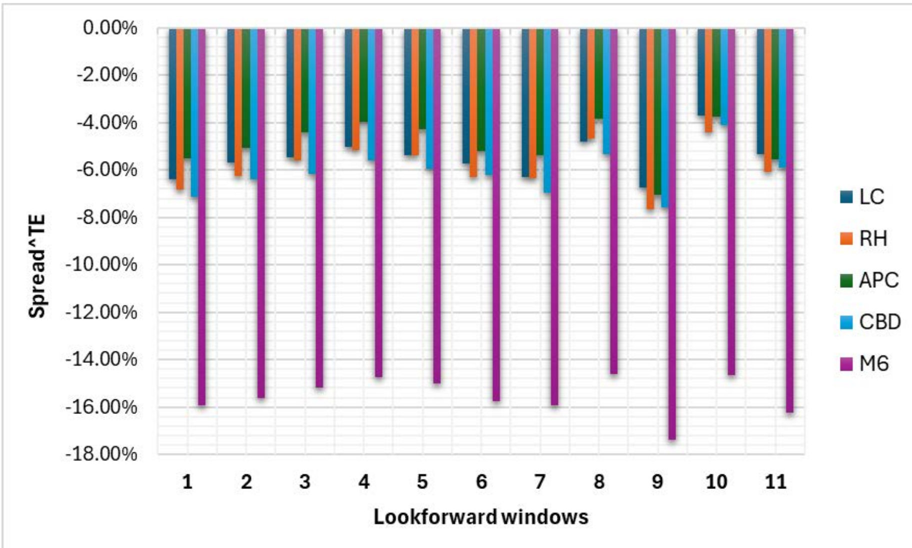

Fig. 5. Magnitude (in percent values) of the spread between the theoretical (model-based) and the empirical expected present discounted values of an example annuity contract, on different lookforward windows and for different mortality models.

5.1 Optimism in mortality prediction and anti-selection effects

In the previous section, we have shown how the systematic errors made by some extrapolative models in the family of generalized age-period-cohort stochastic mortality models impact on the computation of the expected present discounted value of an annuity contract. The consequence of mortality overprediction is the underestimation of the longevity risk associated with the annuity contract and thus of its actuarial value. Biases in mortality models' behaviour have implications on a variety of economic and actuarial measures that are based on longevity forecasts, for instance those designed to express the value that life annuities provide to annuitants, namely to the demand-side of the contract.

Empirical evidence shows that voluntary demand for annuities at retirement is remarkably lower than what is predicted by standard economic theory. This long-standing phenomenon, known as "annuity puzzle" ([46]), has been observed and studied globally ([13, 34, 39, 17]), with a few exceptions, for instance, Switzerland ([5]). Vast literature shows that behavioural and institutional factors contribute to the observed mismatch between theory and observed behaviour ([6, 51, 54]).

Extensive literature evaluates demand for annuities in the framework of portfolio choice and life-cycle models of consumption, for instance [61], [25], [42], [30], [29], and [59]. Besides utility based-measures, as in [12, 10, 44, 7], a commonly adopted measure for the value of an annuity is the so-called "money's worth ratio" (MWR), that is the expected present discounted value of annuity payments (EPDV) relative to the premium cost of the annuity ([45, 43]).

According to the mathematical notation illustrated in section 2, the money's worth ratio can be expressed by the following ratio:

$$\frac{a_{x,t:\overline{n}|}}{P},$$

where $a_{x,t:\overline{n}|}$ is the expected present discounted value of the stream of annuity payments and P denotes the premium paid when the contract is underwritten by the annuitant.

A MWR equal to 1 indicates that, on average, the annuitant can expect to get back one monetary unit, in today's terms, for every monetary unit she paid in premiums. A MWR lower than 1 denotes that, on average, the individual will get back less in annuity payouts than what she invested in the annuity. A MWR different from 1 is tipically identified in the literature as actuarially unfair pricing or non-neutral pricing in actuarial sense ([9]). [53] report money's worth values for annuities of about 92 cents per premium dollar, in relation to the US retail insurance market in 2020.

A MWR different from 1 reflects the cost of adverse selection, along with administrative costs and profit margins ([26]). Indeed, mortality tables for annuitants, that are tipically used for pricing, assume lower mortality than the one expected for the general population. In this respect, according to [23], MWR calculations that relate population mortality tables to the ones underlying annuity prices represent quantitative measures of the degree of adverse selection in the annuity markets. Such calculations indeed reveal the return that the typical individual in the general population receives from the annuity, compared to the typical annuity purchaser.

Adverse selection is extensively studied in relation to insurance markets. [24] find evidence that individuals self-select into annuities having more valuable characteristics for them, in terms of the time profile of annuity payouts. Indeed, annuities providing more of their payments in later periods (i.e., back-loaded annuities) are selected more, on average, by long-lived individuals, while annuities paying out more in case of an early death are mainly chosen by short-lived annuitants. Also other contributions point to the fact that individuals have information about their mortality risk type, that goes beyond what can be inferred from their observable characteristics ([28, 32, 52]). [9] stress that if an individual perceives the own mortality risk as greater than the one underlying the pricing of the annuities, then the MWR could be significantly lower than unity, even though the premium can be thought as neutral according to actuarial standards.

[56] study the mortality assumptions underlying the valuation of pension liabilities in several countries, compared with the population mortality rates.

In the Italian context, [26] perform a study on MWR's, based on different mortality tables and on a wide dataset of annuity prices. They show that using life tables relative to the general Italian population delivers lower MWRs than adopting annuitants mortality tables; the related decline, that quantifies the costs of adverse selection, is about 10%.

Within our case study, we adopt a similar rationale to the one underpinning the MWR and we propose a measure we denominate "Mortality scenario gap (MSG)". Such indicator is a ratio between two actuarial values for an annuity contract. MSG expresses in the numerator the expected present discounted value of the annuity based on the mortality forecasts for the general population generated by stochastic models, according to a given quantile of the distribution of the simulated mortality rates. In the denominator, it expresses the expected present discounted value of the same annuity contract, but based on the mortality rates from annuitants life tables. MSG thus serves as a rule of thumb to measure to which extent the mortality forecasts for the unisex general population, as obtained from different mortality models, should be "optimistic" to reflect the mortality outcomes expected for annuitants. In this respect, as explained in the remainder of the section, we adopt models LC and APC to obtain mortality forecasts for the Italian unisex population, while we use the mortality rates from Italian annuitants life tables known as A62I ([1]). These life tables derive forecasts from methods based on model LC, as adopted by the Italian National Institute for Statistics. We can thus partially filter out the effects of selecting very heterogeneous approaches to mortality forecasts when constructing the MSG.

We focus on an immediate term life annuity, that offers a stream of unitary cash flows at the end of each year during the policy duration, conditional on the annuitant's survival, starting from age 66.

We use the methods and tools illustrated in section 3.2 to fit models LC and APC on the mortality data relative to calendar years 1951-2020 and to forecast them 35 years ahead. We obtain central projections of the mortality rates and focus on the mortality profile of an individual aged 65 in 2021 until age 99: $\{\tilde{m}^{M,C}_{65,2021}, \tilde{m}^{M,C}_{66,2022}, \ldots, \tilde{m}^{M,C}_{99,2055}\}$, with $M = LC, APC$.

We assess the uncertainty that originates from the estimation of the parameters of the two models, by means of semiparametric bootstrap, that is performed through the *bootstrap* function of the **R** package *StMoMo*, as explained in section 3.2. From the boostrapped models LC and APC we obtain 5000 simulated trajectories for the target mortality rates, for the next 35 years. We select the 2.5% quantile of the distribution of simulated mortality rates as a cut point to reflect low mortality scenario, and 97.5% quantile to represent high mortality scenario. The spread between the two selected quantiles is the amplitude of the 95% confidence interval. In relation to both quantiles, we derive the mortality profile of a person aged 65 in 2021, up to her age 99, according to the high mortality scenario, $\{\tilde{m}^{M,H}_{65,2021}, \tilde{m}^{M,H}_{66,2022}, \ldots, \tilde{m}^{M,H}_{99,2055}\}$ and the low mortality scenario, $\{\tilde{m}^{M,L}_{65,2021}, \tilde{m}^{M,L}_{66,2022}, \ldots, \tilde{m}^{M,L}_{99,2055}\}$. For each model, the three sets of mortality rates represent the baseline for the computation of the survival probabilities ${}_hp^C_{65,2021}$, ${}_hp^H_{65,2021}$ and ${}_hp^L_{65,2021}$, $h = 1, \ldots, 35$, that are involved in the computation of the expected present discounted value for the annuity contract under study. We denote by $a^{M,C}_{65,2021:\overline{35}|}$, $a^{M,H}_{65,2021:\overline{35}|}$ and $a^{M,L}_{65,2021:\overline{35}|}$ the EPDV under the three considered mortality scenarios: central, high and low, respectively, with the superscript M denoting the fact that we use projections from mortality models LC and APC. In Table 2, we report, for each model, the EPDV's of the annuity contract under the low, central and high mortality scenarios for the Italian unisex population.

Table 2. Expected present discounted values of an example annuity contract based on central, low and high mortality scenarios, models LC and APC.

	Model LC	Model APC
$a^{M,C}_{65,2021:\overline{35}\|}$	16.36	16.68
$a^{M,L}_{65,2021:\overline{35}\|}$	17.66	18.24
$a^{M,H}_{65,2021:\overline{35}\|}$	15.02	15.09

As expected, extrapolating mortality trends from the 70 most recent years of data leads to higher EPDV's for the annuity contract we consider with respect to the theoretical values obtained in the backtesting framework of the previous section. The uncertainty in the estimation of the EPDV, arising from the models' forecast error along with parameter uncertainty, is higher for model APC than for model LC, as measurable by the spread $a^{M,L}_{65,2021:\overline{35}|} - a^{H,L}_{65,2021:\overline{35}|}$.

[48] provides an overview of the institutions that are involved in the development of standard mortality tables that serve as a reference for pension and insurance providers, in several countries. With respect to the Italian population, ANIA (Italian Association of Insurance Companies) developed two-dimensional mortality tables for life annuitants according to ages and cohorts, labeled A1900-2020 ([1, 2]). These mortality tables were used to construct a one-dimensional table (A62) according to an age-shifting method that references the 1962 cohort. Separate tables relate to different types of annuity contracts (immediate-A62I, deferred-A62D and group life annuities-A62C). We use A62I mortality tables for immediate life annuity contracts. A62I tables are based on the population mortality projections carried out by Istat (Italian National Institute of Statistics) and account also for the differentials, in terms of life expectancy, that are expected to characterize the subpopulation of annuitants compared with the general population (through specific self-selection factors). Relevant information is provided also by the Italian National Order of Actuar-

ies, for instance at `https://www.ordineattuari.it/articoli/news/2014/2/studio-ania-su-nuove-basi-demografiche/`, where data can be retrieved.

Although referencing the unisex population of annuitants, separate tables are available for different proportions of male and female annuitants in the sample. By way of an example, we target the following proportions of male and female annuitants, respectively: 80%-20%, 50%-50%, 20%-80%. For each gender composition, we use the mortality rates q_x in the tables to obtain survival probabilities and to compute the EPDV for the immediate term life annuity contract under study. We denote such EPDV by $a^A_{65,2021:\overline{35}|}$ where the superscript A refers to the fact that the computation relies on annuitants mortality tables. In Table 3, we report the EPDV values of the annuity contract under study, based on annuitants mortality tables (A62I), for different proportions of male and female annuitants.

Table 3. EPDV values based on annuitants mortality tables (A62I), for different proportions of male and female annuitants.

	80%-20%	50%-50%	20%-80%
$a^A_{65,2021:\overline{35}|}$	18.45	19.06	19.71

Due to the higher longevity prospects for females, $a^A_{65,2021:\overline{35}|}$ increases with the share of female population.

The mortality scenario gap is computed as follows:

$$MSG = \frac{a^{M,k}_{65,2021:\overline{35}|}}{a^{A,g}_{65,2021:\overline{35}|}}, \tag{17}$$

where $M = LC, APC$. $k = C, L, H$ denotes the central, low or high mortality scenario, respectively, and $g = 0.80, 0.50, 0.20$ denotes the male to female proportion considered.

We report the values of MSG under each considered case in Table 4. The closer the ratio to 1 the more aligned the mortality trajectories underlying the EPDVs that are compared against each other.

The MSG is lower than 1, for any considered choice of: the mortality projection model, the mortality scenario, and the gender composition. This result is intuitive, since annuitants are generally expected to experience higher longevity outcomes with respect to the general population. Nevertheless, from Table 4 we can infer information about how much mortality forecasts for the general population should be optimistic to mimic the mortality profile that A62I tables expect for the annuitants. In this respect, we find that accounting for anti-selection effects, according to the A62I life tables, is almost equivalent to assume that future mortality rates will realize according to a low mortality scenario (2.5% quantile

Table 4. Mortality scenario gap that arises from comparing EPDVs based on population life tables (according to models LC and APC) in the numerator with the EPDVs based on annuitants life tables in the denominator (according to different gender proportions).

		$a^{A}_{65,2021:\overline{35}\rvert}$ 80%-20%	50%-50%	20%-80%
Model LC	$a^{M,C}_{65,2021:\overline{35}\rvert}$	0.89	0.86	0.83
	$a^{M,L}_{65,2021:\overline{35}\rvert}$	0.96	0.93	0.90
	$a^{M,H}_{65,2021:\overline{35}\rvert}$	0.81	0.79	0.76
Model APC	$a^{M,C}_{65,2021:\overline{35}\rvert}$	0.90	0.88	0.85
	$a^{M,L}_{65,2021:\overline{35}\rvert}$	0.99	0.96	0.93
	$a^{M,H}_{65,2021:\overline{35}\rvert}$	0.82	0.79	0.77

of the distribution of simulated mortality rates) for a population in which the share of male is dominant (80%), based on the projections provided by model APC. Indeed, for this specific case, the mortality scenario gap attains a value of 0.99.

Our empirical findings provide a kind of rule of thumb to understand to which extent moving donward through the distribution of simulated mortality rates for the general population may be useful to predict the potential outcomes of higher longevity within the valuation of annuity contracts. Furthermore our results show empirically how actuarial valuation is impacted by the following two factors. On the one hand, as it is shown in the backtesting framework of section 4, central projections from some extrapolative mortality models tend to be overly pessimistic about future mortality. On the other hand, as expected, accounting for anti-selection effects implies considering optimistic mortality patterns, according to the lowest quantiles of the distribution of mortality rates. Accordingly, it is important to disentangle the systematic predictions biases of some stochastic mortality models from the actual impact of heterogeneity in the mortality risk profiles of the annuitants and of the general population.

As a concluding remark, we stress that the MSG is impacted by the proportion of male and females in the population. This reveals that the assumptions about the female share of the population along with the assumptions about how the gender gap in longevity may develop in the future may be important drivers of annuity valuation.

6 Conclusions

In this paper, we provide a case study based on the Italian unisex mortality experience, to assess mortality overprediction as as systematic attitude of some extrapolative mortality models. Furthermore, we study the persistence of such an attitude over time, by means of a backtesting strategy with rolling windows of data. Furthermore, we estimate some financial consequences of mortality overprediction, in relation to actuarial and economic measures that are intrinsically based on longevity forecasts. Our examples relate to the actuarial value of annuity contracts and cover also their actuarial fairness, namely the value that the contract provides to the annuitant. Depending on the selected mortality model, overpredicting mortality translates into an underestimation of the expected present value of the payouts of an example annuity contract from around 5% (APC) to around 16% (M6), on average over time. This evidence sheds light on some relevant consequences: besides the underestimation of the longevity risk associated with the annuity from the insurer's perspective, the introduction of biases in those measures designed to measure the actuarial fairness of an insurance contract. For instance, measures similar to the MWR may incorporate, to some extent, models' (in)abilities to capture mortality improvements over time, and thus may provide biased estimations of the magnitude of actuarial fairness.

Within our case study, we adopt a similar rationale to the one underpinning the MWR and we propose a measure we denominate "mortality scenario gap (MSG)". Such indicator is indeed a ratio between two actuarial values for an annuity contract, but it serves as a rule of thumb to measure to which extent the mortality forecasts from different mortality models should be "optimistic" to reflect the mortality outcomes expected for annuitants. For instance, we find that accounting for anti-selection effects, according to the A62I life tables from ANIA, is almost equivalent to assume that future mortality rates will realize according to a low mortality scenario (2.5% quantile of the distribution of simulated mortality rates) for a population in which the share of male is dominant (80%), based on the projections provided by model APC.

Our empirical findings provide a kind of rule of thumb to understand to which extent moving donward through the distribution of simulated mortality rates for the general population may be useful to predict the potential outcomes of higher longevity within the valuation of annuity contracts. Furthermore we stress that the assumptions about how the gender gap in longevity may develop in the future and prediction biases from mortality models can be important drivers of annuity valuation.

Future research may deepen the role of mortality models' prediction biases and of the assumptions about how the gender longevity gap could evolve in the future on the size of the degree of adverse selection measured in annuity markets, in more complex settings, when larger information about annuity market prices and take-up rates are available.

References

1. ANIA. Le basi demografiche per rendite vitalizie A1900-2020 e A62. Relazione tecnico-metodologica, 2014. `https://www.ordineattuari.it/media/169536/relazione_tecnica_a62_final.pdf`
2. ANIA. Le basi demografiche per assicurazioni di rendita A1900-2020 e A62, 2015 `https://www.ania.it/studi-e-rapporti-demografici`.

3. Apicella, G., and Sibillo, M.: Corrective factors for longevity projections in a dynamic context. European Actuarial Journal, **8** (2018), 53–68. doi: 10.1007/ s13385-018-0166-6.
4. Apicella, G., Dacorogna, M., Di Lorenzo, E., and Sibillo, M.: Improving the forecast of longevity by combining models. North American Actuarial Journal, **23-2** (2019), 298–319, doi: 10.1080/10920277.2018.1556701.
5. Avanzi, B.: What is it that makes the Swiss annuitise? A description of the Swiss retirement system. Australian Actuarial Journal, **16-2** (2010), 135–162
6. Benartzi, S., Previtero, A., and Thaler, R. H.: Annuitization Puzzles. Journal of Economic Perspectives, **25-4** (2011), 143–64, doi: 10.1257/jep.25.4.143
7. Bernard, C., De Gennaro Aquino, L., and Levante, L.: Optimal annuity demand for general expected utility agents. Insurance: Mathematics and Economics, **101** (2021), 70–79, doi: 10.1016/j.insmatheco.2020.07.004
8. Booth, H., and Tickle, L.: Mortality modelling and forecasting: a review of methods. Annals of Actuarial Science, **2(1–2)** (2008), 3–43, doi: 10.1017/S1748499500000440
9. Boyer, M. M., Box-Couillard, S., and Michaud, P.: Demand for annuities: Price sensitivity, risk perceptions, and knowledge. Journal of Economic Behavior and Organization, **180** (2020), 883–902, doi: 10.1016/j.jebo.2019.03.022
10. Bravo, J. M., and El Mekkaoui de Freitas, N.: Valuation of longevity-linked annuities. Insurance: Mathematics and Economics, **78** (2018), 212–229, doi: 10.1016/j.insmatheco. 2017.09.009
11. Brouhns, N., Denuit, M., and Van Keilegom, I.: Bootstrapping the Poisson Log-Bilinear model for mortality forecasting. Scandinavian Actuarial Journal, **2005-3** (2005), 212–224, doi: 10.1080/03461230510009754
12. Brown, J., and Poterba, J.: Joint life annuities and the demand for annuities by married couples. Journal of Risk and Insurance, **67-4** (2000), 527–553, doi: 10.2307/253849
13. Brown. J. R.: Private pensions, mortality risk, and the decision to annuitize. Journal of Public Economics, **82-1** (2001), 29–62, doi: 10. 1016/S0047-2727(00)00152-3
14. Cairns, A. G. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M.: Mortality density forecasts: An analysis of six stochastic mortality models. Insurance: Mathematics and Economics, **48-3** (2011), 355–367, doi: 10.1016/j.insmatheco. 2010.12.005
15. Cairns, A. G. J., Blake, D., and Dowd, K.: A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. Journal of Risk and Insurance, **73-4** (2006), 687–718, doi: 10.1111/j.1539-6975.2006.00195.x
16. Cairns, A. G. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I.: A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. North American Actuarial Journal, **13-1** (2009), 1–35, doi: 10.1080/10920277.2009.10597538
17. Cannon, E., and Tonks, I.: Cohort mortality risk or adverse selection in annuity markets? Journal of Public Economics, **141** (2016), 68–81, doi: 10.1016/j.jpubeco.2016.07.002
18. Carone, G., Eckefeldt, P., Giamboni, L., Lained, V., and Sumner, S.P.: Pension reforms in the eu since the early 2000's: Achievements and challenges ahead, European Commission. Directorate-General for Economic and Financial Affairs. European Economy. Discussion Paper 042 (2016). ISSN 2443-8022 (online). `https://economy-finance.ec.europa.eu/system/files/2020-06/dp042_en.pdf`

19. Currie, I.D.: On fitting generalized linear and non-linear models of mortality. Scandinavian Actuarial Journal, **2016-4** (2016), 356–383, doi: 10.1080/03461238.2014.928230
20. Currie, I.D., Durban, M., and Eilers, P. H. C.: Generalized linear array models with applications to multidimensional smoothing. Journal of the Royal Statistical Society. Series B (Statistical Methodology), **68-2** (2006), 259–280, doi: 10.1111/j.1467-9868.2006.00543.x
21. Djeundje, V.B., Haberman, S., Bajekal, M., and Lu, J.: The slowdown in mortality improvement rates 2011-2017: a multi-country analysis. European Actuarial Journal, **12** (2022). doi: 10.1007/s13385-022-00318-0
22. Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M.: Backtesting stochastic mortality models. North American Actuarial Journal, **14-3** (2010), 281–298, doi: 10.1080/10920277.2010.10597592
23. Finkelstein, A., and Poterba, J.: Selection effects in the United Kingdom individual annuities market. The Economic Journal, **112-476** (2002), 28–50, doi: 10.1111/1468-0297. 0j672
24. Finkelstein, A., and Poterba, J.: Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Markets. Journal of Political Economy, **112** (2004), 183–208, doi:/10.1086/379936
25. Friedman, B. M., and Warshawsky, M. J.: The cost of annuities: Implications for saving behavior and bequests. The Quarterly Journal of Economics, **105-1** (1990), 135–154, doi: https://doi.org/10.2307/2937822
26. Guazzarotti, G., and Tommasino, P.: The annuity market in an evolving pension system: lessons from Italy (2008). Working Paper 77/08. Center for Research on Pensions and Welfare Policies, `https://www.cerp.carloalberto.org/wp-content/uploads/2008/12/wp_77.pdf?f6fa34`
27. Haberman, S., and Renshaw, A.: A comparative study of parametric mortality projection models. Insurance: Mathematics and Economics, **48-1** (2011), 35–55, doi: 10.1016/j.insmatheco.2010.09.003
28. Hamermesh, D. S.: Expectations, Life Expectancy, and Economic Behavior. The Quarterly Journal of Economics, **100-2** (1985), 389–408, doi: 10.2307/ 1885388
29. Han, N.-W., and Hung, M.-W.: The annuity puzzle and consumption hump under ambiguous life expectancy. Insurance: Mathematics and Economics, **100** (2021), 76–88, doi: 10.1016/j.insmatheco.2021.04.011
30. Horneff, V., Maurer, R., Mitchell, O. S., and Rogalla, R.: Optimal life cycle portfolio choice with variable annuities offering liquidity and investment downside protection. Insurance: Mathematics and Economics, **63** (2015), 91–107, doi: 10.1016/j.insmatheco.2015.03.031
31. Human Mortality Database. Max Planck Institute for Demographic Research (Germany), university of California, Berkeley (USA), and French Institute for Demographic Studies (France). http://www.mortality.org (2023) (data downloaded on December 31, 2023)
32. Hurd, M. D., and McGarry, K.: The Predictive Validity of Subjective Probabilities of Survival. The Economic Journal, **112-482** (2002), 966–985, doi: 10. 1111/1468-0297.00065
33. Hyndman, R. H., and Khandakar, Y.: Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software, **27-3** (2008), 1–22, doi: 10.18637/jss.v027.i03
34. Inkmann, J., Lopes, P., and Michaelides, A.: How Deep Is the Annuity Market Participation Puzzle? The Review of Financial Studies, **24-1** (2011), 279–319, doi: 10.1093/rfs/hhq080.

35. Khalaf-Allah, M., Haberman, S., and Verrall, R.: Measuring the effect of mortality improvements on the cost of annuities. Insurance: Mathematics and Economics, **39** (2006), 231–249, doi: 10.1016/j.insmatheco.2006.02.005
36. Lee, R. D., and Carter, L. R.: Modeling and forecasting U.S. mortality. Journal of the American Statistical Association, **87-419** (1992), 659–671, doi: 10.1080/01621459. 1992.10475265
37. Levantesi, S., Lizzi, M., and Nigri, A.: Enhancing diagnostic of stochastic mortality models leveraging constrast trees: an application on Italian data. Quality & Quantity, **58** (2024), 1565-1581, doi: 10.1007/s11135-023-01711-x
38. Li, H., and Li, J.S.: Optimizing the Lee-Carter approach in the presence of structural changes in time and age patterns of mortality improvements. Demography, **54** (2017), 1073-1095, doi: 10.1007/s13524-017-0579-x
39. Lockwood, L. M.: Bequest motives and the annuity puzzle. Review of Economic Dynamics, **15-2** (2012), 226–243, doi: 10.1016/j.red.2011.03.001
40. Lovász, E.: Analysis of Finnish and Swedish Data with Stochastic Mortality Models. European Actuarial Journal, **1-2** (2011), 259–289, doi: 10.1007/s13385-011-0039-8.
41. Maccheroni, C., and Nocito, S.: Backtesting the Lee-Carter and the Cairns-Blake-Dowd Stochastic Mortality Models on Italian Death Rates. Risks, **5-3** (2017), 34, doi: 10. 3390/risks5030034.
42. Maurer, R., Mitchell, O. S., Rogalla, R., and Kartashov, V.: Lifecycle portfolio choice with systematic longevity risk and variable investment-linked deferred annuities. The Journal of Risk and Insurance, **80-3** (2013), 649–676, doi: 10.1111/j.1539-6975. 2012.01502.x
43. Maurer, R., Rogalla, R., and Siegelin, I.: Participating payout life annuities: Lessons from Germany. Astin Bulletin **43-2** (2013), 159–187, doi: 10.1017/asb.2013.10
44. Milevsky, M. A., and Huang, H.: The utility value of longevity risk pooling: Analytic insights. North American Actuarial Journal, **22-4** (2018), 574–590, 2018. doi: 10.1080/10920277.2018.1467271
45. Mitchell, O. S., Poterba, J. M., Warshawsky, M. J., and Brown, J. R.: New Evidence on the Money's Worth of Individual Annuities. American Economic Review, **89-5** (1999), 1299–1318, doi: DOI:10.1257/aer.89.5.1299
46. Modigliani, F.: Life Cycle, Individual Thrift, and the Wealth of Nations. The American Economic Review, **76-3** (1986), 297–313.
47. OECD. Improving retirement incomes considering behavioural biases and limited financial knowledge. OECD Pensions Outlook 2018, OECD Publishing Paris (2018). doi: `10.1787/pens_outlook-2018-8-en`
48. OECD. Mortality and the Provision of Retirement Income. OECD Publishing, Paris, (2023). doi: https://doi.org/10.1787/a10a6c09-en
49. Olivieri, A., and Pitacco, E.: Introduction to Insurance Mathematics. Technical and Financial Features of Risk Transfers. Springer Berlin, Heidelberg (2011). doi: 10.1007/ 978-3-642-16029-5
50. Olivieri, A., and Pitacco, E.; Time Restrictions on Life Annuity Benefits: Portfolio Risk Profiles. Risks, **10-8** (2022), 164, doi: 10.3390/risks10080164
51. Peijnenburg, K., Nijman, T., and Werker, B. J.: The annuity puzzle remains a puzzle. Journal of Economic Dynamics and Control, **70** (2016), 18–35, doi: 10.1016/j.jedc.2016.05.023
52. Perozek, M.: Using Subjective Expectations to Forecast Longevity: Do Survey Respondents Know Something We Don't Know? Demography, **45-1** (2008), 95–113, doi:10.1353/dem.2008.0010

53. Poterba, J. M., and Solomon, A.: Discount rates, mortality projections, and money's worth calculations for US individual annuities. Working Paper 28557, National Bureau of Economic Research (2021). `http://www.nber.org/papers/w28557`
54. Ramsay , C. M., and Oguledo, V. I.: The Annuity Puzzle and an Outline of Its Solution. North American Actuarial Journal, **22-4** (2018), 623–645, doi: 10.1080/10920277.2018.1470936
55. Renshaw, A., and Haberman, S.: A cohort-based extension to the Lee-Carter model for mortality reduction factors. Insurance: Mathematics and Economics, **38-3** (2006), 556–570, doi:10.1016/j.insmatheco.2005.12.001
56. Sithole, T. Z., Haberman, S., and Verrall R.: Second International comparative study of mortality tables for pension fund retirees. British Actuarial Journal, **17-3** (2012), 650–687, doi: 10.1017/51357321712000207
57. Van Berkum, F., Antonio, K., and Vellekoop, M.: The impact of multiple structural changes on mortality predictions. Scandinavian Actuarial Journal, **2016-7** (2016), 581–603, doi: 10.1080/03461238.2014.987807
58. Villegas, A. M., Millossovich, P., and Kaishev, V. K.: StMoMo: An R package for stochastic mortality modelling. Journal of Statistical Software, **84-3** (2018). doi: 10.18637/jss.v084.i03
59. Wei, J., Cheng, X., Jin, Z., and Wang, H.: Optimal consumption–investment and life-insurance purchase strategy for couples with correlated lifetimes. Insurance: Mathematics and Economics, **91** (2020),244–256, doi: 10.1016/j.insmatheco.2020.02.006
60. Wong-Fupuy, C., and Haberman, S.: Projecting mortality trends: Recent Developments in the United Kingdom and the United States. North American Actuarial Journal, **8 -2** (2004), 56–83, doi: 10.1080/10920277.2004.10596137
61. Yaari, M. E.: Uncertain lifetime, life insurance, and the theory of the consumer. Rev. Econom. Stud., **32-2** (1965), 137–150, 1965. doi: 10.2307/2296058

Mathematical Methods in Economics and Finance – m^{2}ef
Vol. 17/18, No. 1, 2022/23
ISSN print edition: 1971-6419 – ISSN online edition: 1971-3878
Web page: `http://www.unive.it/m2ef/` – E-mail: `m2ef@unive.it`