Using Interest Rate Models to Improve Mortality Forecast

Giovanna Apicella*

Michel M. Dacorogna†

Emilia Di Lorenzo‡

Marilena Sibillo§

Abstract

Future evolution of mortality poses important challenges for life insurance, pension funds, public policy and fiscal planning. Indeed, when fair values, premium rates and risk reserves are computed, sound and accurate models to forecast stochastic longevity are needed. In this paper, we propose a methodological approach in order to improve the predictive accuracy of the existing survival models. The central idea is to model the ratio between the observed death rates and the corresponding fitted values obtained as outputs of a survival model we select, by means of the Cox-Ingersoll-Ross (CIR) model. For our numerical application, we choose to apply the CIR correction to the Cairns-Blake-Dowd (or M5) model. Using the Italian females mortality data and implementing the backtesting procedure, over both a static time horizon and fixed-length windows rolling one-year ahead through time, we empirically test the perfomance of the CBD model in forecasting death rates both for itself (CBD) and corrected by the CIR process (mCBD). On the basis of average measures of forecasting errors and information criteria we demonstrate that the mCBD model is a parsimonious model providing better results in terms of predictive accuracy than the CBD model.

Keywords: CBD model, Cox-Ingersoll-Ross process, ex-post forecasting performance, stochastic mortality models

*University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185 Rome, Italy; giovanna.apicella@uniroma1.it
†DEAR-Consulting, Scheuchzerstrasse 160, 8057 Zürich, Switzerland; michel@dacorogna.ch
‡University of Naples Federico II, complesso Monte S. Angelo, 80126 Naples, Italy; diloremi@unina.it
§University of Salerno, campus universitario, 84084 Fisciano (SA), Italy; msibillo@unisa.it

1 Introduction

The accuracy of mathematical models in predicting mortality is an important challenge given the strong impact of life expectancy forecasts for policy making in many significant social sectors. One of the aspects, which makes central the modeling of such a stochastic phenomenon, is the systematic incidence of longevity on life insurance and pension funds, particularly in the industrialized Countries. The stability and consistency of social welfare systems are put in danger worldwide due to the combined phenomenon of the progressive increase in life expectancy, along with the reduction of birth-rates in industrialized Countries (see for instance Dacorogna and Kratz (2015)).

In this paper we propose a methodological approach in order to improve the predictive accuracy of existing survival models; we aim at increasing the capacity of stochastic survival models to replicate the real data and accurately extrapolate it. Within this context we focus on one of the most popular survival models chosen as an example of application and, analyzing its ex-post forecasting performance, correct it by means of the Cox-Ingersoll-Ross (CIR) model. In particular, we perform our corrective model in the representative case of the Cairns-Blake-Dowd model, labeled as M5 in Cairns et al. (2009) where the Authors analyze eight different stochastic mortality models. Our study will focus on the reduction of the spread between the evidences of the real survival phenomenon and the corresponding forecasted one.

The paper is organized as follows: in Section 2 we explain the mathematical framework. In Section 3, we present and discuss the empirical methodology implemented in this study, namely the techniques applied for calibrating and backtesting the models, and describe the data used in the numerical application. In Section 4 we present the results and discuss the graphical and statistical assessment of the predictive performance of the two competing models. In Section 5 conclusions are given.

2 Mathematical Framework

In this Section we set the mathematical framework and the actuarial assumptions useful for the following development of the paper.

2.1 Notation

Let us consider the instantaneous death rate at time t known as the *force of mortality*. Posing $q_{x,\Delta x,t}$ the probability that an individual aged x at time t dies before age $x + \Delta x$, the force of mortality $\mu_{x,t}$ related to this individual can be expressed as follows:

$$\mu_{x,t} = \lim_{\Delta x \to 0} \frac{q_{x,\Delta x,t}}{\Delta x} \tag{1}$$

If Δx is sufficiently small, we can write:

$$q_{x,\Delta x,t} \simeq \mu_{x,t} \Delta x \tag{2}$$

so that the product on the right can be interpreted as a probability. It is commonly assumed that, within each year of age, $\mu_{x,t}$ remains constant, as in (3) where k and h are numbers such that, respectively, $x + k$ is a non-integer age and $t + h$ is a non-integer duration:

$$\mu_{x+k,t+h} = \mu_{x,t}, \quad \text{if} \quad k, h \in [0, 1) \tag{3}$$

Let $S_t(x)$ be the probability that T_0, the random future lifetime for a newborn, is longer than x attained at time t:

$$S_t(x) = I\!P[T_0 > x, t], \quad t \geq 0$$

If we assume that $S_t(x)$ is differentiable with respect to x, it is straightforward to show that the force of mortality is given by:

$$\mu_{x,t} = \frac{-\frac{\partial S_t(x)}{\partial x}}{S_t(x)} = -\frac{\partial}{\partial x} \ln S_t(x) \tag{4}$$

When $\mu_{x,t}$ is known, equation (4) becomes a differential one; integrating between 0 and x, and posing the initial position $S_t(0) = 1$, we have:

$$S_t(x) = \exp\left[-\left(\int_0^x \mu_{u,t} du \right) \right] \tag{5}$$

Being:

$$_h q_{x,t} = \frac{S_t(x) - S_t(x + h)}{S_t(x)} \tag{6}$$

it results in:

$$ {}_hq_{x,t} = 1 - \exp\left[- \left(\int_x^{x+h} \mu_{u,t} du \right) \right] \tag{7} $$

Recalling formula (3), we have immediately:

$$ p_{x,t} = \exp\left[-\mu_{x,t}\right]; \quad q_{x,t} = 1 - \exp\left[-\mu_{x,t}\right] $$

where $q_{x,t}$ is the probability for a live aged x at time t to die within 1 year, that is before age $x+1$, and $p_{x,t}$ the probability for the same person to be alive at age $x+1$.

Data concerning deaths and living people (*exposed to risk*) are organized in matrices: the first contains $d_{x,t}$, the number of deaths in one year among the people aged x in calendar year t and the second containing $E^c_{x,t}$, the exposed to risk, with x varying from 0 to the ultimate age and t varying with the calendar year (cf. Currie (2014)). The exposed to risk are assumed to be *central exposed to risk*, an average of the living people aged x in t.

The Mortality Coefficient is given by (cf. Olivieri and Pitacco (2010) and Pitacco (2007)):

$$ m_{(x,x+h),t} = \frac{\int_x^{x+h} \mu_{u,t} S_t(u) du}{\int_x^{x+h} S_t(u) du} \tag{8} $$

and the Central Death Rate is defined as follows, taking into account equation (6) and posing $h = 1$ in (8):

$$ m_{x,t} = m_{(x,x+1),t} = \frac{S_t(x) - S_t(x+1)}{\int_x^{x+1} S_t(u) du} \tag{9} $$

2.2 The CIR Process and its Meaning in the Context of Mortality Modelling

As in the previous Section, we denote by x the age and by t the calendar year when the age is measured. We consider then the random variable (rv) $Y_{x,t}$ expressed as the ratio between the observed central death rates and the corresponding fitted values, outputs of the mortality model chosen for describing the mortality behavior (see Di Lorenzo et al. (2006)). We have:

$$ Y_{x,t} = \frac{B_{x,t}}{\mu_{x,t}} \tag{10} $$

where $B_{x,t}$ denotes the observed central death rates and $\mu_{x,t}$ the baseline provided by the chosen mortality model for describing the same rates. As to the meaning of the baseline, we remind the reader that under condition (3):

- the forces of mortality and the central death rates coincide, that is $\mu_{x,t} = m_{x,t}$;
- the maximum likelihood estimate $\hat{\mu}_{x,t}$ of the force of mortality $\mu_{x,t}$ is given by the crude death rate for age x in calendar year t, $\hat{m}_{x,t}$ (see Pitacco et al. (2009));
- the mortality rate can be computed as follows: $q_{x,t} = [1 - \exp(-\mu_{x,t})] = [1 - \exp(-m_{x,t})]$.

For a fixed age $\ddot{x}$, the idea is to model the dynamics over time of $Y_{\ddot{x},t}$ as a Cox-Ingersoll-Ross (CIR) stochastic process as in the following stochastic differential equation:

$$dY_{\ddot{x},t} = \alpha(\beta - Y_{\ddot{x},t})dt + \sigma\sqrt{Y_{\ddot{x},t}}dW_t \tag{11}$$

which, for $\alpha, \beta > 0$, corresponds to a continuous time first-order autoregressive process where the rv $Y_{\ddot{x},t}$ is elastically pulled toward a long-term mean, β, at a speed α.

If the starting point of $Y_{\ddot{x},t}$ is non-negative and the coefficients satisfy the condition:

$$2\alpha\beta \geq \sigma^2 \tag{12}$$

$Y_{\ddot{x},t}$ can never become negative.

Within this structure, provided that condition (12) is fulfilled, in our opinion $Y_{\ddot{x},t}$ has a similar behavior as the interest rate process proposed in Cox et al. (1985). We think they are both characterized by the same empirically relevant properties:

- negative values are precluded;
- if the process reaches zero, it can subsequently become only positive;
- the absolute variance of the process increases when the process itself increases;
- there exists a steady state distribution for $Y_{\ddot{x},t}$.

The process $Y_{x,t}$ can be applied giving to it two different meanings:

1. As a correction factor:

   The procedure leads to a method for constructing projected values starting from the chosen mortality model: its outputs can be corrected simply multiplying them by the CIR process in (11) with properly estimated parameters, as in the form:

   $$B_{x,t} = Y_{x,t}\mu_{x,t} \tag{13}$$

2. As a mortality model benchmark:

   Briefly speaking, when the calibration of the CIR process in (11) provides an optimal value of the long-term mean β close to 1, we can immediately infer, with the confidence coming from the empirical calibration, that the chosen mortality model is able to fit well the real data themselves.

2.3 The CBD (or M5) Model

The model we choose to test our methodology is the CBD model of Cairns et al. (2006). They fitted the following model to the mortality rates $q_{x,t}$ (see also Cairns et al. (2009)):

$$\log \frac{q_{x,t}}{(1-q_{x,t})} = \beta_x^{(1)} k_t^{(1)} + \beta_x^{(2)} k_t^{(2)} \tag{14}$$

where the left term of the equation is the *logit* function of $q_{x,t}$ and the terms $\beta_x^{(i)}, i = 1, 2$, modulating the effect of the time indices $k_t^{(i)}, i = 1, 2$, across ages, are assumed to be parametric, as follows:

$$\beta_x^{(1)} = 1 \quad and \quad \beta_x^{(2)} = (x - \bar{x})$$

where $\bar{x}$ is the mean age in the sample range. Basing on equation (14), the predictor structure of the CBD model can be formulated as:

$$logit(q_{x,t}) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) \tag{15}$$

Such a model does not require parameter constraints to ensure unique parameter estimates, namely the model has no identification problem, as it is explained in Cairns et al. (2009).

3 Empirical Methodology

In this Section, we describe the empirical methodology for testing our model and discuss the data used in the numerical applications. The first step of the method consists in the choice of the survival model to correct by the CIR process. As written above, we choose the CBD model, which is one of the most popular survival models within the actuarial literature. Nevertheless, the procedure we propose is suited to any choice of survival model.

3.1 Backtesting Procedure

We empirically test the performance of the CBD model in forecasting the central death rates first by itself and then corrected by the CIR process $Y_{x,t}$ as in (13).

In order to be able to compare the forecast of survival against the realized one, we use the following backtesting procedure: we split the sample of available reliable mortality data into two time intervals and use the corresponding data sets as follows:

- the first one (within the "lookback" window (cf. Dowd et al. (2013)) for fitting $\mu_{x,t}$ from the CBD model and calibrating the CIR process $Y_{x,t}$
- the second one (within the "lookforward" window) for estimating how accurately the forecasts produced by the CBD model perform out-of-sample, with and without our correction. It is an "out-of-sample" test because the data used here have not been used to calibrate the model.

Using this methodology allows us to test the model with data that were never seen before as it would be the case in the real use of such models to predict future development of mortality.

We choose the historical central death rates (as defined in (9)) as the target variables to fit and test in the predictions. In subsections 3.4 and 3.5, we separately deal with the operations we carry out over the two different time intervals, namely calibration and backtest of the forecast.

3.2 Overall Data Set

Our study focuses on the Italian females mortality. We use the mortality data provided by (Human Mortality Database (2016). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)) (data downloaded on October 20, 2016). However, we limit ourselves to the time interval [1906,2012], since it is recommended to use with extra caution the data prior to 1906 due to problems of data quality (see Glei (2015)). We thus collect the following data from HMD: the number of deaths $d_{x,t}$ and the central exposures to the risk of death $E^c_{x,t}$, with $x = 0, ..., 90$ and $t = 1906, ..., 2012$. Within the implementation of the methodology following a static approach, as described in subsection 3.1, we use the data from 1906 up to 1977 as the lookback window and the rest [1978,2012] as the backtesting window. Furthermore, in order to test the stability of our methodological approach, we implement the backtesting procedure also in a dynamic setting, making use of the rolling windows of data described in the next subsection.

3.3 Rolling Windows of Data

We perform the backtesting procedure explained in subsection 3.1 over both the whole data set, as described in subsection 3.2, and by using a fixed-length rolling window one-year-ahead through time, in order to verify if the re-optimization on the rolling windows provides updated parameters able to catch better the dynamics of the system than the static optimization.

We arrange our whole data set relating to the time interval $[1906, 2012]$ into two different ways obtaining in both cases 36 time horizons to be used for implementing our overall methodology. The lookback and the lookforward windows are sized according to the following two arrangements:

1. 67 years in the lookback window and 5 years in the lookforward one;

2. 62 years in lookback window and 10 years in the lookforward one.

In order to provide a few examples of how the rolling windows move forward through time, in what follows we briefly describe the first two time horizons within the first and the second arrangement of the overall data

set. Concerning the first arrangement, the first time horizon is set such that the corresponding lookback and lookforward windows are the following time intervals: [1906,1972] and [1973,1977] respectively. Rolling one-year ahead with respect to the previous time horizon, the second one is characterized by the lookback and lookforward windows: [1907,1973] and [1974,1978], and so on. The same mechanism rules the second arrangement of the data set, whose time horizons, as previously mentioned, are made up of 62 points lookback windows and 10 points lookforward windows. Thus, the lookback and lookforward windows corresponding to the first time horizon are, respectively, the time intervals [1906,1967] and [1968,1977], while those ones relating to the second time horizon are, respectively, [1907,1968] and [1969,1978].

3.4 Calibration of the Models

In this Section, we describe the calibration procedures used for the two models: the CBD and the CIR model. We remind the reader that, when we use a fixed time horizon, the lookback window includes years from 1906 to 1977.

3.4.1 Calibration of the CBD Model Over a Fixed Look-back Window

We perform the fitting of the CBD model using the package "StMoMo", provided in **R**. It is an implementation of the generalized Age-Period-Cohort family of stochastic mortality models (see Villegas et al. (2016a), (2016b)). We concentrate on the one-year mortality rate $q_{x,t}$. For the parameters estimation, we maximize the log-likelihood function based on the assumption that the random numbers of deaths $D_{x,t}$ are independent and follow a Binomial distribution (B) conditionally on $(q_{x,t})$ (Haberman and Renshaw (2011)):

$$D_{x,t} \sim \mathcal{B}(E^0_{x,t}, q_{x,t}). \tag{16}$$

In order to obtain the estimation function with the required initial exposures, say $E^0_{x,t}$, we compute them by adding half the matching reported number of deaths to the central exposures (cf. Pitacco (2007)). Following what we stated above, we start from the inputs: $d_{x,t}$ and $E^0_{x,t}$, with $x = 18, ..., 90$ and $t = 1906, ..., 1977$. Basing on the following choices about the features of the model: a logit link function; a predictor structure

as in (15); an empty set of parameter constraints (see Section 2.3), we get as outputs the fitted mortality rates $\hat{q}_{x,t}, x = 18, ..., 90, t = 1906, ..., 1977$, from which we compute the corresponding death rates $\hat{m}_{x,t}$.

3.4.2 Calibration of the CIR Process Over a Fixed Lookback Window

As illustrated in Section 2.2, the CIR process in (11) describes the dynamics of $Y_{x,t}$ over time and for fixed ages. Our application is focused on the following representative ages: 18, 35, 40, 45, 65. After fixing the age $\ddot{x}$, we obtain the $Y_{\ddot{x},t}$ time series as in (10), by calculating the ratios between the observed crude death rates $B_{\ddot{x},t}$ and the corresponding fitted ones $\hat{m}_{\ddot{x},t}$, for $t = 1906, ..., 1977$. For instance, the ratio relating to Italian females aged 18 in the starting year of the lookback window, that is 1906, is computed as follows:

$$Y_{18,1906} = \frac{B_{18,1906}}{\hat{m}_{18,1906}} \tag{17}$$

and, considering the same individuals, the overall time series to which we calibrate the CIR process in (11) is: $(Y_{18,1906}, Y_{18,1907}, ..., Y_{18,1977})$.

We estimate the vector of parameters of the CIR process $\theta = (\alpha, \beta, \sigma)$ by maximization of the log-likelihood function $\ln L(\theta)$ based on the known probability density for the process. We borrow and adapt to $Y_{\ddot{x},t}$ the model for the term structure of interest rates described in Cox et al. (1985). Within such structure, the probability density of the state variable at time t, conditional on its value at the current time s, for some $s < t$, has indeed a closed form formula. We perform the optimization by using the **MATLAB** function *fminsearch*, looking for the minimum of $-\ln L(\theta)$. We highlight that we do not set any constraint on the parameters among the inputs of the minimizing function.

In Table 1, we show the optimal parameters provided by the calibration of the CIR process $Y_{\ddot{x},t}$ over the lookback window and the corresponding value of the log-likelihood function. We see that all the optimal parameters $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}$, are positive and that the condition (12) is always fulfilled, verifying *a posteriori* that we did not need to put any constraint on them.

Table 1: Estimated parameters and log-likelihood of the CIR process $Y_{\ddot{x},t}$

Age	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\ln L(\hat{\theta})$
18	0.1538	3.3172	0.2016	-0.2978
35	0.1286	1.2782	0.0638	1.2756
40	0.0892	1.0869	0.0708	1.2524
45	0.0174	1.2788	0.0498	1.6627
65	0.1687	0.8172	0.0272	2.3366

The CIR process describing the dynamics over time of $Y_{\ddot{x},t}$, when $\ddot{x}$ is 18, is characterized by the highest long term mean and the biggest volatility. The estimated long term mean is bigger than 1 for all the ages but 65, denoting the propensity of the CBD model to underestimate the real crude death rates observed over the lookback window. For a deeper understanding of these considerations, see Section 4.2.3.

3.4.3 Calibration of the CBD and CIR Models Over Rolling Windows

Making use of the techniques explained in subsections 3.4.1 and 3.4.2, we calibrate the CBD model and the CIR model to the observed death rates and to the factors $Y_{x,t}$, respectively, over the fixed-length windows rolling one-year-ahead through time described in subsection 3.3; we deepen, at a later step (see Sections 3.5.3 and 4.3), if the optimal parameters provided over time by the calibration adapt and catch better the dynamics of the system over the corresponding forecast periods.

3.5 Lookforward Window: Forecasting

In this Section, we move on the forecasting procedure. When a fixed time horizon is used, the lookforward window includes years from 1978 to 2012.

3.5.1 Forecasting of the CBD Model Over a Fixed Lookforward Window

We perform the forecasting of the CBD Model 35 years-ahead from 1977, using the package "StMoMo", which provides, exploiting the outcomes of the fitting procedure described in Section 3.4.1, among other information, a matrix with the central projection of the mortality rates, $\dot{q}_{x,t}$, with $t = 1978, ..., 2012$, from which we compute the corresponding forecasts of the death rates $\dot{m}_{x,t}$ with $t = 1978, ..., 2012$.

3.5.2 Simulation of the CIR process Over a Fixed Lookforward Window

Examining further the features of the CIR model as it was designed in Cox et al. (1985), Cox et al. pointed out that the distribution of the state variable at time t given its current value at time s, for some $s < t$, up to a scale factor, is a non central chi-square distribution. As discussed in Glasserman (2003), such a property can be used to simulate the process (11) exactly on a discrete time grid.

In light of these remarks, exploiting the optimal parameters provided by the calibration, we simulate 1'000'000 paths of the square-root diffusion (11) over the time interval [1978,...,2012], by sampling from the transition density. The time step is one year and the starting point is $Y_{\ddot{x},1977}$.

Each average across paths represents our "best estimate" of the realization of the ratio $Y_{\ddot{x},t}$ in a given year within the lookforward window. The outcome of the simulation procedure is thus the following "expected" projection: $(\tilde{Y}_{\ddot{x},1978}, \tilde{Y}_{\ddot{x},1979}, ..., \tilde{Y}_{\ddot{x},2012})$. The same results can be obtained by computing the conditional expected value of the CIR process $Y_{\ddot{x},t}$ (cf. Cox et al. (1985)).

3.5.3 Forecasting of the CBD and CIR Models Over Rolling Windows

Exploiting the techniques described in subsections 3.5.1 and 3.5.2, we perform, respectively, the forecasting of the CBD model and the simulation of the CIR process 5 and 10 years-ahead, according to the chosen arrangement of our data set (see subsection 3.3). Such a rolling-window analysis lets us backtest for checking the predictive performance of the two models on 36 different time horizons for each arrangement of the data set.

3.6 The Predictive Ability of the CBD Model

In this Section, we examine the performance of the CBD model over both the lookback and the lookforward window. The aim of this Section is to deepen the predictive capacity of the CBD model in order to test if it gives satisfactory results. In particular, at this stage, we mainly focus on the graphical outcomes provided by the fitting and the forecasting of the CBD model (see sections 3.4.1 and 3.5.1, respectively) and displayed in Figure 1. We report in Table 2 the Root Mean Square Error, ψ_{lb}, measuring the goodness of fit of the CBD model over the lookback period.

We select two ages among the five under study, namely 18 and 65 as representative examples. In Figure 1, we display the realized crude death rates (black lines) plotted against the fitted crude death rates (grey lines in the left panel) and the forecasts of crude death rates (grey lines in the right panel) through the CBD model. The graphs on the right and left panel are displayed with the same scale on the y-axis, for consistency reasons. The CBD model depicts a downward trend in the evolution of mortality affecting all ages we consider. This is in line with the longevity increase we experience since a century. Both graphs in the left panel display a peak in the trend of the observed and fitted mortality rates in 1918; such a jump, which is much more evident for age 18, was due to the outbreak of the 1918 pandemic flu known as Spanish flu. For age 18, modest reductions in the mortality trend have occurred since the 1950s, while we see a more pronounced trend for age 65, which is in line with the fact that life expectancy increases because of better health of the aging people. Concerning the graphs on the left side of Figure 1, we can see that the calibration of the CBD model provides a good fit to the data extrapolated from the mortality tables and related to the time interval [1906,1977]. We get similar graphs for the other ages considered for this study.

In Table 2 we report the RMSE of the CBD model (ψ_{lb}, "lb" for lookback) for measuring its accuracy in fitting the crude death rates. We refer to Section 4.2.2 for a deeper examination of the statistics, including the RMSE we use for assessing the forecasting performance of both the CBD model and our innovative one. The low value taken by the RMSE confirms the graphical evidences in Figure 1. In particular, the most satisfactory fit throughout the lookback period concerns age 35, characterized by the smallest value of the RMSE. We can derive further evidences about the behavior of the CBD model over the lookback period

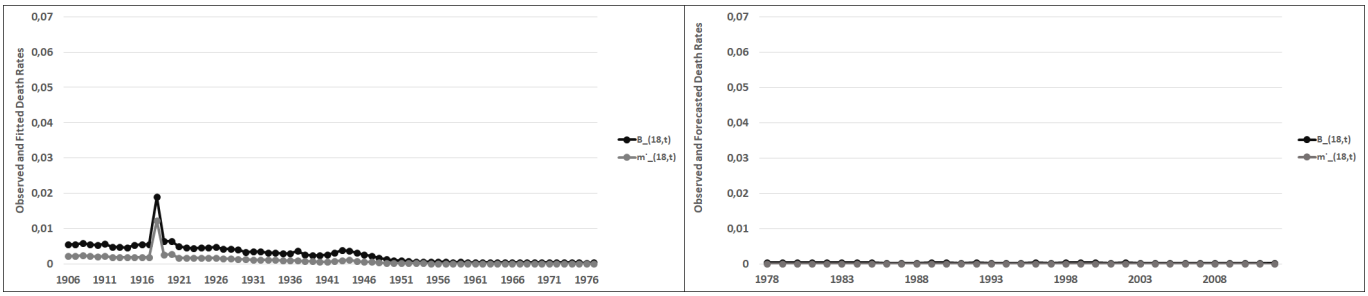
from the value taken at each calendar year by the ratio $Y_{\ddot{x},t}$ as defined in (10), or, more generally, from the optimal long term mean of the CIR process (11) as provided by the calibration described in Section 3.4.2. If $Y_{\ddot{x},t}$ is smaller than 1, the fitted crude death rate coming from the fitting of the CBD model for calendar year t is larger than the corresponding realized crude death rate; this means that the CBD model overestimates the real data in a given year t. The farther $Y_{\ddot{x},t}$ is from 1, the larger is the distance between the observed central death rate and the fitted one. Within our application, the estimated long term mean is bigger than 1 for all the ages but 65 (see Table 1), denoting, as above already observed, the propensity of the CBD model to underestimate the real central death rates observed over the lookback window for all ages except for 65. Such a propensity appears also in the graphs on the right side of Figure 1, showing the observed crude death rates on the time interval [1978,2012] and the forecasts, for both ages 18 and 65, obtained implementing the CBD model as in Section 3.5.1. Even though the CBD model is not designed for modelling mortality at lower ages, its least satisfactory fitting and forecasting performance concerns the highest age under study, namely age 65.

The overall outcomes of the CBD model forecasting turn out to be adequate, with values of the RMSE that, as we point out in Section 4.2.2, are even smaller than those we get from the calibration procedure. In particular, it is worth noting that, in the graphs in the right panel of Figure 1, the grey lines appear to be very close to the black ones because of the scale on the y-axis. Graphs with broader scales are provided in Section 4.2.1. In Section 4.2.2 we use quantitative tools for showing that the CBD model forecasting performance can be further improved through our methodology. For this purpose, we firstly explain, in the following Section 3.7, how to compute the "adjusted" projections starting from the outcomes of the CBD model.

Table 2: RMSE (unit 10^{-4}) of the CBD model fitting performance

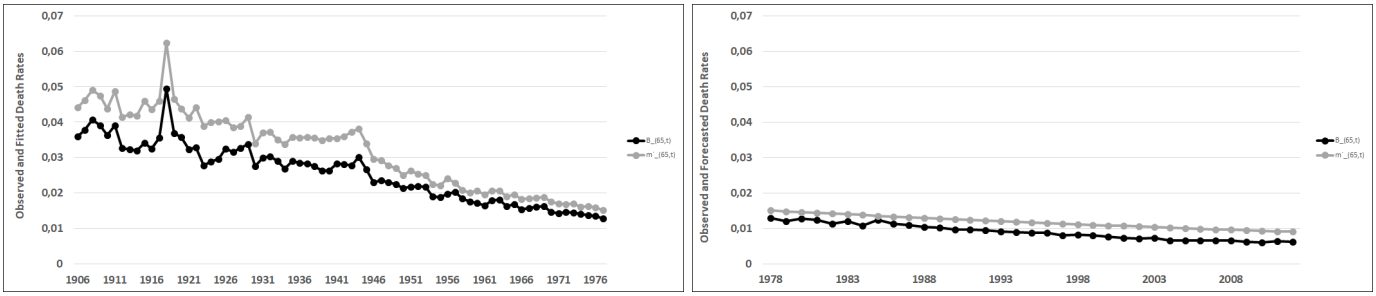
Age	18	35	40	45	65
ψ_{lb}	22.6	5.6	10.9	29.8	69.0

NOTE: ψ_{lb}, RMSE of the CBD model fitting performance



(a) Age 18. Fitting on the lookback window

(b) Age 18. Forecasting over the lookforward window



(c) Age 65. Fitting on the lookback window

(d) Age 65. Forecasting over the lookforward window

Figure 1: Lookback and lookforward window: the CBD model fitting and forecasting performance. Black lines: crude death rates observed on the lookback window; grey lines: death rates fitted by the CBD model on the same period.

3.7 The "Adjusted" Projection: the "New" Model mCBD

Based on (10), we use the following formula to correct the CBD forecast:

$$\breve{m}_{\ddot{x},t} = \tilde{Y}_{\ddot{x},t}\dot{m}_{\ddot{x},t} \tag{18}$$

to construct "adjusted" projections of the death rates $m_{\ddot{x},t}, t = 1978, ..., 2012$. We denote such projections by $\breve{m}_{\ddot{x},t}, t = 1978, ..., 2012$. We remind the reader that $\tilde{Y}_{\ddot{x},t}$ is the "best estimate" of the realization of the ratio $Y_{\ddot{x},t}$, for a given age $\ddot{x}$ and in a given year t within the lookforward window, obtained as in Section 3.5.2, while $\dot{m}_{\ddot{x},t}$ represents the forecasted death rate for the same age and the same year coming from the CBD model (see Section 3.5.1).

In line with what we have explained in the previous Sections, in the follow-up we will handle, over the lookforward window, the forecasted death rates provided by:

1. the stand-alone CBD (hereinafter CBD) model, from which we get $\dot{m}_{\ddot{x},t}$;
2. the forecasts $\dot{m}_{\ddot{x},t}$ of the CBD model, matched with the Monte Carlo sample paths generated from the CIR process (11), globally providing us with $\tilde{m}_{\ddot{x},t}$ as in (18). We denote by mCBD the model exploiting the synergy between the two just mentioned models.

4 Assessing the Forecasting Performance of the CBD and the mCBD Models

In this section, we study the forecasting performance of the CBD and the mCBD models through the procedures described in Section 3. The section starts introducing the statistical measures we use for this assessment.

4.1 Forecasting Errors and Information Criteria: a General Overview

Let us suppose n pairs of estimates or predictions $(P_i, i = 1, ...n)$ provided by a given model and the corresponding observations $(O_i, i = 1, ..., n)$. The average model-estimation(or prediction) error can be written as (see Willmott and Matsuura (2005)):

$$\bar{e}_\gamma = \left[\sum_{i=1}^{n} w_i |e_i|^\gamma / \sum_{i=1}^{n} w_i \right]^{\frac{1}{\gamma}} \tag{19}$$

where $|e_i| = |O_i - P_i|$ are the absolute values (or magnitudes) of individual errors, $\gamma \geq 1$ and w_i is a scaling assigned to each $|e_i|^\gamma$ for reflecting its hypothesized influence on the total error.

Let us define the two following estimators of the average error expressed in (19): the Root Mean Square Error (RMSE), which is the most commonly used statistical measure of the quality of an estimator or predictor, and the Mean Absolute Error (MAE). For illustration purposes, here we recall some notations, which we are going to specify within our application in the following subsection 4.2.3.

For the RMSE we use equation (19), putting $w_i = 1$ and $\gamma = 2$; for the MAE, we use equation (19), setting $w_i = 1$ and $\gamma = 1$.

Compared to the MAE, in the computation of the RMSE each individual error e_i affects the total error (namely the sum of all the individual errors) in proportion to its square, rather than just its magnitude, so that large errors have a stronger influence on the total square error than small ones. Furthermore, the RMSE increases with the variance of the frequency distribution of the error magnitudes; in other words, the total square error grows as the total error is concentrated within a decreasing number of increasingly large individual errors (for a more complete discussion, we refer the reader to Willmott and Matsuura (2005)).

The mCBD model implies the increase in the number of parameters used to predict the outcome. Concerning this topic, in what follows we assess if the improvements made by adjusting the forecast by means of the CIR process justifies the increase in the number of parameters. As stressed in Burnham and Anderson (2004), on the basis of the famous Box's statement "All models are wrong but some are useful" (see Box (1976)), models, by definition, are only approximations to unknown reality or truth. Neither the CBD

nor the mCBD model can perfectly reflect the real dynamics of the survival phenomenon, but, within the model-based inference framework, we can assess which of the two methods performs better on the basis of three principles: *simplicity and parsimony*, *multiple working hypotheses* and *strength of evidence*. Between our two competing hypotheses, the CBD model is the one characterized by the lowest number of parameters (two against the overall five parameters characterizing the mCBD model). In order to deepen the impact of the mCBD larger number of parameters, we look at information criteria. Here, we assess if, within our application, the simplest model for the number of parameters, namely the CBD model, should be preferred, computing the Bayesian Information Criterion (BIC) or Schwarz Criterion:

$$BIC = n \ln(\tau/n) + k \ln(n) \tag{20}$$

where: n, the number of points in the lookforward window, is equal to 35 under both the settings; k, the number of parameters, is 2 or 5, according to the concerned model (the CBD model and the mCBD model, respectively); τ represents the residual sum of squares. We see that the BIC penalizes models with a large number of parameters. Thus, it is well suited to compare between the CBD and the mCBD models.

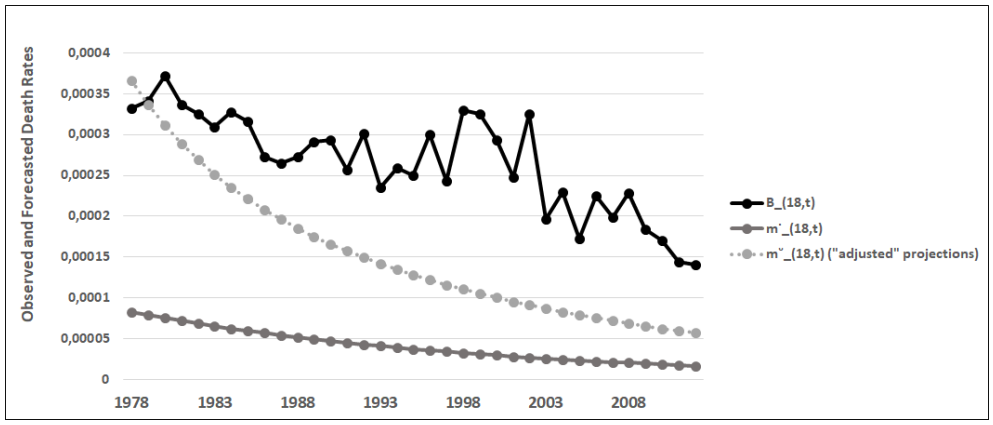
4.2 Over the Fixed Look-forward Window [1978-2012]

In this section, we assess the quality of the forecast both graphically and statistically. The graphical analysis provides us with an insight of the dynamics of the forecasts of death rates and of the errors associated with them, while the statistical measures give us an estimate of their magnitude.

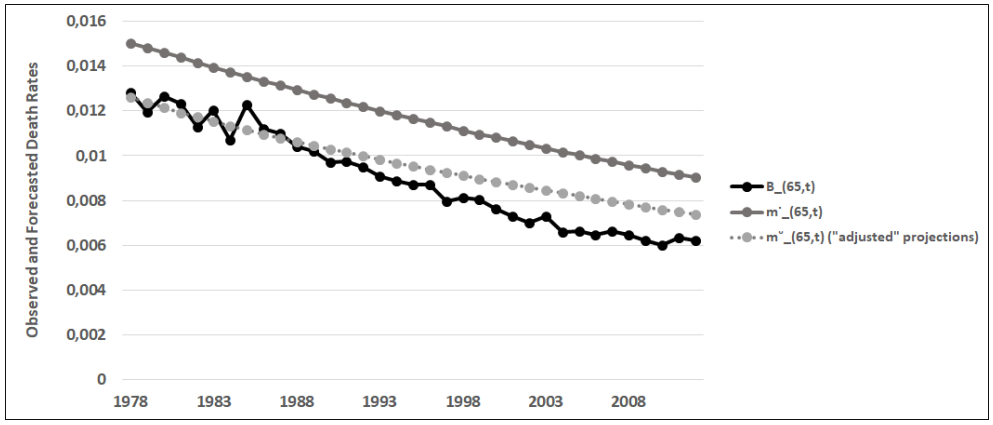
4.2.1 Graphical Results

For each of the five fixed ages under study and for $t = 1978, \ldots, 2012$, we collect the graphical evidences of:

- the realized death rates, $B_{\ddot{x},t}$ (the black line);
- the projections of death rates, $\dot{m}_{\ddot{x},t}$ (the dark grey line);
- the "adjusted" projections of death rates, $\breve{m}_{\ddot{x},t}$ (the dotted light grey line).



(a) Age 18



(b) Age 65

Figure 2: Observed and forecasted crude death rates for Italian females aged differently. Observed crude death rates: black line, CBD forecasted death rates: dark grey line, mCBD forecasted death rates: dotted light grey line.

In Figure 2, we show the results for the two illustrative ages $\ddot{x} = 18, 65$. In both graphs, the dotted grey line is closer to the black one than the dark grey line, meaning that, within this empirical application, just at a first sight it is evident that, thanks to the correction factor $\tilde{Y}_{\ddot{x},t}$, the mCBD model is able to improve the forecasting performance of the CBD model for both ages and in each calendar year. Indeed, on the basis of the graphs displaying the observed and forecasts for the death rates for all ages under study over the same forecasting horizon, we can infer that the mCBD model predicts better than the CBD model in 100% of the ages and the calendar years we examine.

4.2.2 Statistical Assessment of the Forecasting Performance

This section is devoted to assess and compare the forecasting performance of the CBD and the mCBD models looking at both the forecasting errors and the information criteria.

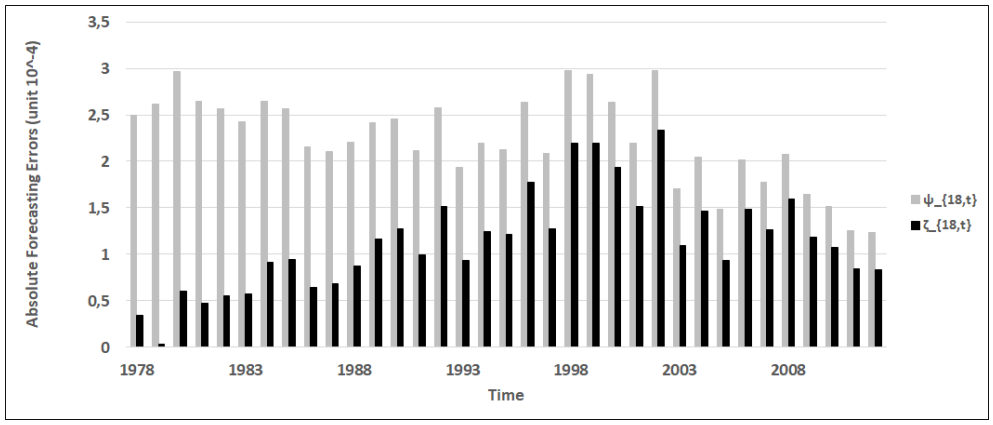
4.2.3 Measuring Forecasting Errors

We define the magnitudes of forecasting errors of the CBD and the mCBD models, respectively, as $\psi_{\ddot{x},t}$ and $\zeta_{\ddot{x},t}$, where the subscript $\ddot{x}$ denotes that the age is fixed; ψ and ζ are thus only functions of time $t \in [1978, 2012]$. Both $\psi_{\ddot{x},t}$ and $\zeta_{\ddot{x},t}$ are designed to measure the absolute forecasting error between the real crude death rate, observed in a given year of the lookforward window, and the corresponding forecast; in particular:

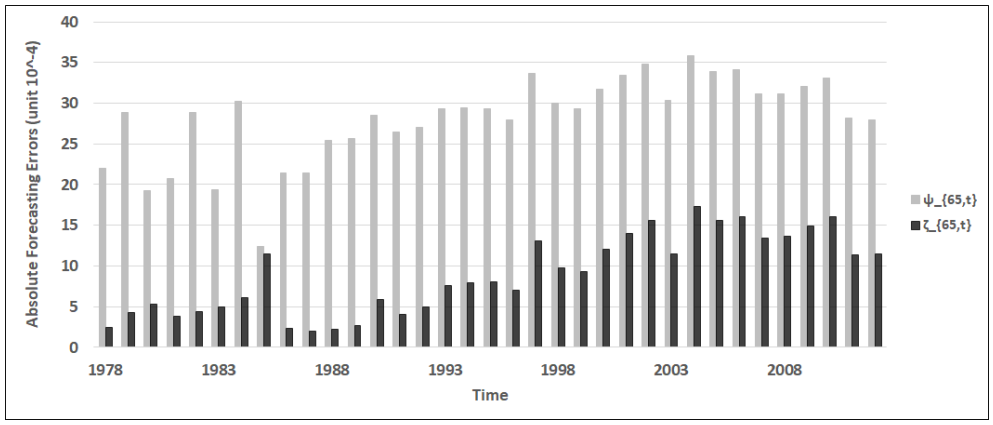
- $\psi_{\ddot{x},t}$ quantifies the absolute forecasting error between the realized crude death rates, $B_{\ddot{x},t}$, and their projections, $\dot{m}_{\ddot{x},t}$, obtained as output of the CBD model;
- $\zeta_{\ddot{x},t}$ quantifies the absolute forecasting error between the realized crude death rates, $B_{\ddot{x},t}$, and their "adjusted" projections, $\breve{m}_{\ddot{x},t}$, computed through the mCBD model as in (18).

In Figure 3, we show, again for the two illustrative ages 18 and 65, the absolute forecasting errors of the mCBD model $\zeta_{\ddot{x},t}$ (the black bars) and of the CBD model $\psi_{\ddot{x},t}$ (the grey bars), over the lookforward window $[1978, ..., 2012]$ (unit 10^-4). Confirming the evidences provided in Figure (2), $\zeta_{\ddot{x},t}$ turns out to be smaller than $\psi_{\ddot{x},t}$ in 100% of both ages and calendar years. This happens for all the ages under study.

Starting from the absolute forecasting errors, $\psi_{\ddot{x},t}$ and $\zeta_{\ddot{x},t}$, for each age under study, we compute the RMSE and the MAE of the CBD model (ψ^* and $\bar{\psi}$, respectively) and the RMSE and the MAE of the mCBD model (ζ^* and $\bar{\zeta}$, respectively). In Table 3, we report such measures of the average forecasting error, multiplied by 10^{-4}.



(a) Age 18



(b) Age 65

Figure 3: Absolute forecasting errors of the CBD model and the mCBD model (unit 10^{-4}).

Black bars: absolute forecasting errors of the mCBD model; grey bars: absolute forecasting errors of the CBD model.

Table 3: Measures of the average forecasting error (unit 10^{-4})

Age	CBD		mCBD	
	ψ^*	$\bar{\psi}$	ζ^*	$\bar{\zeta}$
18	2.29	2.24	1.26	1.14
35	2.68	2.56	1.79	1.56
40	3.00	2.95	2.18	2.01
45	3.54	3.48	1.69	1.48
65	28.60	28.13	10.13	8.94

NOTE: ψ^* and $\bar{\psi}$ are, respectively the RMSE and the MAE of the CBD model; ζ^* and $\bar{\zeta}$ are, respectively, the RMSE and the MAE of the mCBD model

If, for each age under study, we look at ψ^* compared to $\bar{\psi}$ and at ζ^* compared to $\bar{\zeta}$, we notice that for all ages and for both the CBD model and the mCBD model, the RMSEs, ψ^* and ζ^*, are slightly bigger than the corresponding MAEs, $\bar{\psi}$ and $\bar{\zeta}$ respectively. For both the CBD model and the mCBD model, the largest difference between the RMSE and the MAE corresponds to age 65, for which, in both cases, the distribution of the error magnitudes is characterized by the highest standard deviation. In Table 4, we report the standard deviation of the absolute forecasting errors distribution of the CBD (σ_ψ) and the mCBD (σ_ζ) models, for the two illustrative ages 18 and 65; both these distributions are indeed characterized by the highest standard deviation when age 65 is concerned.

Table 4: Standard deviation of the distribution of the absolute forecasting errors (unit 10^{-5})

Age	σ_ψ	σ_ζ
18	4.7	5.3
65	52.2	48.3

NOTE: σ_ψ and σ_ζ are the standard deviations of the distribution of the absolute forecasting errors of the CBD model and the mCBD model, respectively

It is important to highlight, in Table 3, that the mCBD model is characterized by smaller values of both the RMSE and the MAE than the CBD model, this meaning that it turns out to be more accurate in predicting future death rates for each age under study.

A synthetic measure of the ability of the factor $\tilde{Y}_{\ddot{x},t}$ in correcting, on average and throughout the lookforward window, the outcomes of the CBD model, is what we call "Improvement Rate" (IR). In particular, we consider the two following measures: IR^* $((\zeta^* - \psi^*)/\psi^*)$, quantifying how much, in percent, the mCBD model is able to reduce the RMSE of the CBD model; and $\bar{IR}$ $((\bar{\zeta} - \bar{\psi})/\bar{\psi})$, measuring the percent gain provided by the mCBD model in the average forecasting performance, when MAE is concerned.

Since, as reported in Table 3, both ψ^* and $\bar{\psi}$ are always bigger than, respectively, ζ^* and $\bar{\zeta}$ for all ages, IR^* and $\bar{IR}$ are always negative; the higher such rates are, in absolute terms, the larger the magnitude of the correction provided, on average, by $\tilde{Y}_{\ddot{x},t}$ is. In Table 5, IR^* and $\bar{IR}$ take the highest absolute value for age 65; therefore, among the five ages under study, the mCBD model provides the largest improvement of the predictive accuracy of the CBD model for age 65, both when the RMSE and the MAE are concerned. We consider this observation important for its implications in the pension demographic assessment.

Table 5: Improvement Rate

Age	18	35	40	45	65
IR^*	-45%	-33%	-27%	-52%	-65%
$\bar{IR}$	-49%	-39%	-32%	-58%	-68%

NOTE: IR^*: Improvement Rate when the RMSE is concerned; $\bar{IR}$: Improvement Rate when the MAE is concerned

By looking at Tables 1 and 5, we can infer an intuitive and general empirical rule about the features of $\tilde{Y}_{\ddot{x},t}$. As stated in Section 2.2, when the optimal long-term mean β is very close to 1, the chosen mortality model fits well to the data within the lookback window used for calibrating $Y_{\ddot{x},t}$. Pulled toward 1, $\tilde{Y}_{\ddot{x},t}$ has a weaker and weaker ability to drive the mortality outcomes of the model toward the real data. Within our application, we find evidence of that with age 40, for which $\hat{\beta}$ is equal to 1.08 and both IR^* and $\bar{IR}$ take the smallest (absolute) value.

4.2.4 Bayesian Information Criterion

As the next step, we look at the BIC as defined in (20) to assess the quality of the improvement in our forecast. We denote by "BICsa" ("sa" for stand-alone) the Bayesian information criterion related to the the CBD model and by "BICm" the one related to the mCBD. Table 6 reports "BICsa", "BICm", ΔBIC, that is the first difference between "BICsa" and "BICm", the percent *Change rate* (computed as $(\Delta BIC/BICsa)$). We can see that "BICm" is smaller than "BICsa" for all the ages under study; the *Change rate* is indeed always negative. The highest spread between "BICm" and "BICsa", 15.4%, is found for age 65. ΔBIC suggests that, for all the ages under study, we have a decisive evidence against the *null hyphotesis*, which is the one with the highest BIC value, namely the CBD model.

As illustrated in Kass and Raftery (1995), the Bayes factor is a summary of the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another. The Bayesian Information Criterion provides a rough approximation to the logarithm of the Bayes factor, which is easy to use and does not require evaluation of prior distributions (we refer the reader to Kass and Raftery (1995), Weakliem (1999) and Raftery (1999) for a detailed discussion about Bayes Factor and BIC).

Table 6: Bayesian Information Criterion

Age	$BICsa$	$BICm$	ΔBIC	*Change Rate*
18	-579.68	-610.94	31.26	*-5.4%*
35	-568.59	-586.11	17.52	*-3.1%*
40	-560.67	-572.24	11.57	*-2.1%*
45	-549.05	-590.04	40.99	*-7.5%*
65	-402.89	-464.83	61.94	*-15.4%*

NOTE: "BICsa" is the Bayesian Information Criterion of the CBD model; $BICm$ is the Bayesian Informatin Criterion of the mCBD model; ΔBIC is the first difference between "BICsa" and "BICm"; *Change rate* represents the percent difference between "BICsa" and "BICm"

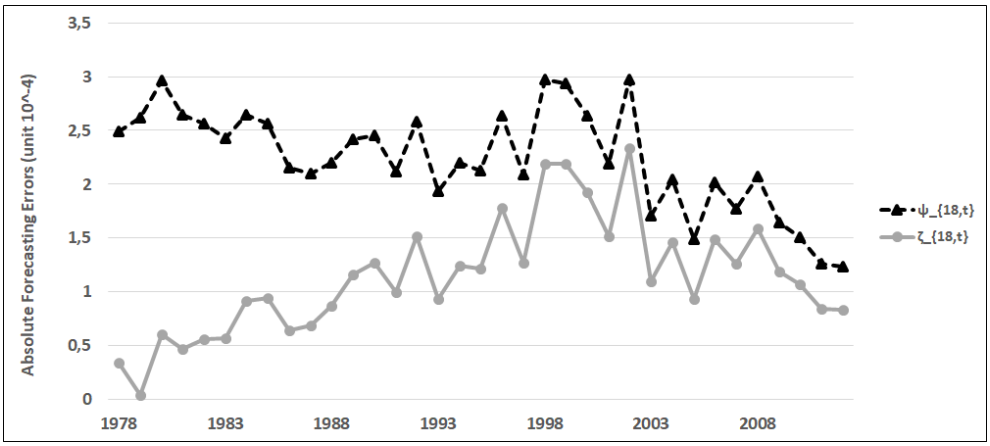


Figure 4: Absolute forecasting errors trend over the lookforward window [1978,2012] for age 18. Black line: absolute forecasting errors of the CBD model; grey line: absolute forecasting errors of the mCBD model"

4.3 Assessing the Forecasting Performance Over Rolling Windows

In this section, we describe the results of the forecasting procedure performed over rolling windows as presented in subsection 3.5.3.

Before illustrating such results, let us briefly focus on Figure 4, displaying how the absolute forecasting errors of the CBD model (the black line) and those of the mCBD model (the light grey line) behave over the time interval $[1978, 2012]$ for age 18, chosen as an example. $\zeta_{\ddot{x},t}$, plotted against time, presents a trend that can be observed for all the ages under study. As previously mentioned, performing our methodology over the fixed-length windows rolling one-year-ahead through time described in subsection 3.3, allows us to verify if the optimal parameters provided over time by the calibration can adapt and catch better the dynamics of the system over the corresponding forecast periods.

Let us now examine in detail the resulting predictive performance of the CBD and the mCBD model: $\tilde{Y}_{\ddot{x},t}$ turns out to improve the predictive performance of the CBD model in all cases. As an example, we can look at: the age 18, the 62-years lookback window [1941,2002] and the 10-years lookforward window [2003,2012]. As to this settings, Figure 5 presents over the lookforward window:

- the realized death rates, $B_{18,t}$ (the black line);
- the projections of death rates, $\dot{m}_{18,t}$ (the dark grey line), by the CBD model;

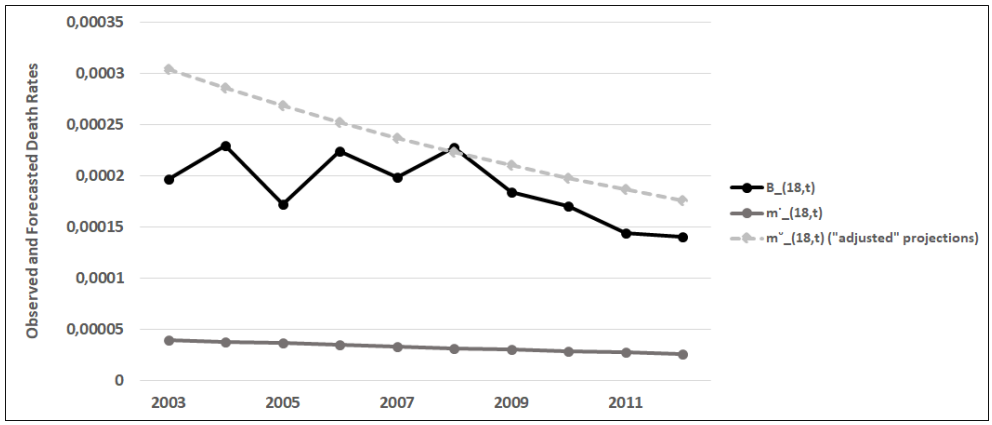


Figure 5: 36th 10-years lookforward window: realized and forecast death rates for age 18. Black line: realized crude death rates; dark grey line: projections of crude death rates obtained as output of the CBD model; dotted light grey line: "adjusted" projections of crude death rates coming from the mCBD model.

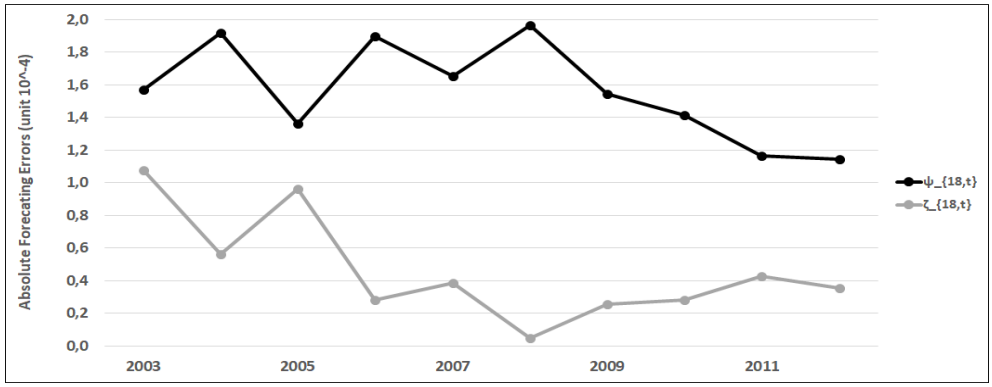


Figure 6: Age 18 and 36th 10-years lookforward window: absolute forecasting errors trend over the lookforward window [2003,2012]. Black line: absolute forecasting errors of the CBD model; light grey line: absolute forecasting errors of the mCBD model.

- the "adjusted" projections of death rates, $\breve{m}_{18,t}$ (the dotted light grey line), by the mCBD model.

In Figure 5, we see that $\breve{m}_{18,t}$ is, for each $t \in [2003, 2012]$, much closer to the corresponding realized crude death rate than $\dot{m}_{18,t}$. In Figure 6 we also show, for age 18 and setting as unit 10^{-4}, $\psi_{18,t}$ (the absolute forecasting error between the realized crude death rates, $B_{18,t}$, and their projections, $\dot{m}_{18,t}$, obtained as output of the CBD model) and $\zeta_{18,t}$ (which quantifies the absolute forecasting error between the realized crude death rates, $B_{18,t}$, and their "adjusted" projections, $\breve{m}_{18,t}$, computed through the mCBD model), $t = 2003, ..., 2012$. In this Figure, we see how $\psi_{18,t}$ (black line) and $\zeta_{18,t}$ (light grey line) behave over the

36th 10-years lookforward window, namely the period [1978,2012], for age 18. We note that the absolute forecasting errors of the mCBD have lost their trend. Moreover, in this specific example, the mCBD model is characterized by an initially high absolute forecasting error, that decreases after 2005 and becomes much more stable after 2006.

For all the other time horizons (whether the lookforward windows are made up by 5 years or 10 years) we find that, after implementing our methodology over rolling windows, the absolute forecasting errors of the mCBD model have no trend any more. We can thus infer that, when making use of rolling windows, $\zeta_{\ddot{x},t}$ looses its trend and the dynamics of the system is better captured by the re-optimization on the rolling windows (dynamic case) than it is in the static case.

4.4 Predictive Accuracy under the Static and the Dynamic Approach

As we have explained in Section 3, within our application, we implement the backtesting procedure over both an unchanging calibration sample (static approach) and fixed-length windows for the calibration sample that is rolling one year ahead through time (dynamic approach). In Section 4 we have assessed the forecasting performance of the CBD and the mCBD models, finding out that, under both the static and the dynamic approach, the mCBD model provides better results, in terms of predictive accuracy, than the CBD model.

The last step of our study consists in comparing the forecasting performance of the CBD and the mCBD models between the static and the dynamic case, in order to understand if the predictive ability of our models deteriorates going further and further in time and, in such a case, how much accuracy we are able to gain using the dynamic approach. For this purpose, we compute and compare the RMSE of our models in the static and in the dynamic case, by exploiting two different methods.

With the first method, the computation of the RMSEs of the two models is performed over the forecasting time horizon [1978-2012], that is the lookforward window we use for backtesting the CBD and the mCBD models in Sections 3.5.1 and 3.5.2, under the static approach. As illustrated in Section 4.2.3, using a fixed lookforward window gives rise, for the CBD and the mCBD models, to the RMSEs reported in Table

Table 7: First method: RMSEs of the CBD and the mCBD models under the static and the dynamic approach (unit 10^{-4})

Model	Static Approach	Dynamic Approach	*Accuracy Gain*
CBD	2.29	2.22	*-3.2%*
mCBD	1.26	0.73	*-42.3%*

3. Such outcomes are compared to the RMSEs we get under the dynamic approach described in Section 3.5.3, when 36 10-years lookforward windows are concerned. Under the dynamic approach, the lookback windows roll one-year ahead through time, so each of them, compared to the previous one, includes the same years except for the first and the last ones. It results that all the years within the time interval (1978, 2012), apart from those at the beginning and at the end, are involved in the forecasting procedure more than once; we thus have available, for them, at least one and at most 36 corresponding realizations of the absolute forecasting error, except that the forecasting horizon changes: the second year in the first case becomes the first year when we move the calibration sample from one year. In order to associate only one forecasting error to the corresponding year, for each year we compute the RMSE across the 36 lookforward windows of the dynamic approach for both the CBD and the mCBD models. In Table 7 we report the RMSEs of the CBD model and the mCBD one, in both the static and the dynamic case, and the rate we call "Accuracy Gain". For each model, the "Accuracy Gain" is the percentage change of its RMSE when moving from the static to the dynamic approach. Using the dynamic approach allows us to enhance the predictive accuracy of both the CBD and mCBD models; in particular, the forecasting performance of the mCBD model has been remarkably improved under the dynamic setting. The problem with this procedure is that we compute an average error over different forecasting horizons. In order to compare the errors over similar forecasting horizons, we design a second method.

With the second method, we implement the backtesting procedure under the following settings: under the static approach, we use the time interval [1906-1967] as a fixed lookback window (the calibration sample) and the time interval [1968-2012] as a fixed lookforward window; under the dynamic approach, we arrange the overall data set described in Section 3.2 in such a way to exploit 41 different time horizons: each of

Table 8: Second method: RMSEs of the CBD and the mCBD models under the static and the dynamic approach (unit 10^{-4})

Span	CBD			mCBD		
	Static	Dynamic	*Accuracy Gain*	Static	Dynamic	*Accuracy Gain*
1968-1977	3.44	3.44	*0.0%*	1.29	1.29	*0.0%*
1978-1987	2.55	2.55	*0.0%*	1.21	0.74	*-39.5%*
1988-1997	2.30	2.19	*-4.9%*	1.47	0.82	*-44.3%*
1998-2007	2.35	2.21	*-5.6%*	1.82	0.96	*-47.4%*
2008-2012	1.58	1.49	*-5.9%*	1.22	0.24	*-80.4%*

them is characterized by 62-years lookback windows and 10-years lookforward windows. The lookback and the lookforward windows of the first time horizon coincide with the unchanging lookback and lookforward windows we employ under the static approach. The last lookback window is made up by the years from 1946 to 2007 and the last lookforward window includes years from 2008 to 2017.

We compare, between the static and the dynamic case, the RMSEs of the CBD and the mCBD models over four consecutive 10-years forecasting spans, starting from 1968. Since the mortality data concerning the Italian population are available until 2012, the last span includes only five years. In Table 8 we report, for the four forecasting spans under study, the RMSEs of the CBD model and the mCBD one under the static and the dynamic approach and the "Accuracy Gain". Concerning the first span, [1968-1977], the CBD and the mCBD model equally perform in the static and the dynamic case since, under both the approaches, the calibration procedures involve mortality data relating to the same lookback window, so that the same resulting parameters are exploited for forecasting 10-years ahead. The different reasonings underlying the static and the dynamic approaches and the resulting different outcomes become evident from the second forecasting span. Indeed, the forecasts corresponding to the second decade are obtained: in the static case, still using the parameters of the CBD and the mCBD models provided by their calibration to the lookback window [1906-1967], while, in the dynamic case, exploiting the parameters of the CBD and mCBD models resulting from their calibration to the lookback window [1916-1977]. Similar remarks can be done for the

subsequent forecasting spans: while in the static case the forecasting procedure resorts always to the same historical information, regardless of the prediction horizon we are interested to, in the dynamic case it is based on the latest 62 data points preceding the first year of the prediction span. Under the dynamic approach, we are thus able to update the historical data sample to be used for the calibration, filtering out the old information and incorporating the most recent one, time by time. It results that, as the "Accuracy Gain" in Table 8 suggests, the farther in time we forecast, the more the precision of both the CBD and the mCBD model is enhanced if we select the dynamic approach as our "modus operandi". This behavior is to be expected, since, in the static case, the forecasting horizon becomes longer and longer. The improvement of the predictive accuracy takes a very remarkable magnitude for the mCBD model, meaning that the mCBD model not only outperforms the CBD model in forecasting central death rates, but also shows a stronger potential to gain accuracy in the long-run under the dynamic approach.

5 Conclusion

The aim of the paper is to improve the performance of the mortality models making their behavior closer to the real trend of the mortality phenomenon. The central idea is to model the ratio between the observed central death rates and the corresponding fitted values obtained as outputs of the chosen mortality model, by means of the Cox-Ingersoll-Ross (CIR) stochastic model. This approach leads to a method for constructing projected death rates starting from the chosen mortality model and also to a benchmark for valuing the capacity of the mortality model in fitting the real data.

For our numerical application, we choose to apply the CIR correction to the $M5$ version of the Cairns-Blake-Dowd model Cairns et al. (2006). Exploiting the Italian females mortality data and making use of the backtesting procedure, we empirically test the performance of the CBD model in forecasting central death rates both for itself (CBD) and in connection with the CIR process $Y_{x,t}$ Cox et al. (1985) (mCBD). By means of average measures of the forecasting errors and information criteria, we demonstrate that using the mCBD model provides better results in terms of predictive accuracy than CBD. Such a methodology exploits the synergy between the CBD model and a CIR process describing the dynamics of the ratio between the observed central death rates and the corresponding fitted values obtained as output of the

CBD model on the lookback window; on the lookforward window the forecasts coming from the CBD model are matched with the Monte Carlo sample paths generated from the CIR process. Following Box (1976), we believe that all models are wrong and thus that scientists cannot obtain a "correct" one by excessive elaboration. Therefore, as William of Occam suggests, we seek a parsimonious description of the natural phenomenon, that is, in our case, the survival phenomenon. Combining the CBD model with a CIR process lets us obtain good results exploiting only three additional parameters compared to the CBD model. Even when correcting the results for the additional parameters, the BIC is improved by 5% on average by mCBD over CBD, showing that the gained accuracy compensates the additional complexity of the model. Moreover, using a dynamic setting for the optimization improves the quality of the mCBD model remarkably making it suitable for pricing of longevity bonds.

To conclude, we want to point out that our procedure is not restricted to the choice of the CBD survival process. The overall applicability and validity of our numerical procedure can be tested on any survival model.

References

Alho, J. M. and Spencer, B. D. (1990). Error model for Official Mortality Forecasts. *Journal of the American Statistical Association*, 85, Issue 411:609–616. Available at `http://www.jstor.org/stable/2289992`.

Booth, H. and Tickle, L. (2008). Mortality Modelling and Forecasting: a Review of Methods. *Annals of Actuarial Science*, 3, Issue 1–2:3–43. Available at `https://doi.org/10.1017/S1748499500000440`.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, 71:791–799. Available at `http://www.jstor.org/stable/2286841`.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods and Research*, 33(2):261–304. available at `http://dx.doi.org/10.1177/0049124104268644`.

Cairns, A. J. G., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73(4):355–367.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and 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, Issue 1. DOI:10.2139/ssrn.1340389.

Cox, J. C., Ingersoll, E., J., and Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2):385–407. Available at `http://www.jstor.org/stable/1911242`.

Currie, I. D. (2014). On fitting generalized linear and non-linear models of mortality. *Scandinavian A ctuarial Journal*, pages 356–383. 10.1080/03461238.2014.928230.

Dacorogna, M. and Kratz, M. (2015). Living in a stochastic world and managing complex risks. *Available at SSRN: http://ssrn.com/abstract=2668468*.

Di Lorenzo, E., Sibillo, M., and Tessitore, G. (2006). A Stochastic Proportional Hazard Model for the Force of Mortality. *Journal of Forecasting*, 25:529–536.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. (2013). Backtesting Stochastic Mortality Models: an ex post evaluation of multiperiod-ahead density forecasts. *North American Actuarial Journal*, 14(3):281–298. DOI:10.1080/10920277.2010.10597592.

Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Springer, ISBN: 978-0-387-21617-1.

Glei, D. A. (updated by Gabriel Borges on September 27, 2015). About mortality data for Italy. *Available at http://www.mortality.org/hmd/ITA/InputDB/ITAcom.pdf*.

Haberman, S. and Renshaw, A. (2011). A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, 48:35–55.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430):773–795. Available at `http://www.jstor.org/stable/2291091`.

Olivieri, A. and Pitacco, E. (2010). *Introduction to insurance mathematics. Technical and financial features of risk transfers*. Springer, ISBN 978-3-642-16028-8.

Park, Y., Choi, J. W., and Kim, H. (2006). Forecasting Cause-Age Specific Mortality Using Two Random Processes. *Journal of the American Statistical Association*, 101(474):472–483. Available at `http://www.jstor.org/stable/27590710`.

Pitacco, E. (2007). *Matematica e tecnica attuariale delle assicurazioni sulla durata di vita*. Lint, ISBN 978-88-8190-182-1.

Pitacco, E., Denuit, M., Haberman, S., and Olivieri, A. (2009). *Modelling Longevity Dynamics for Pensions and Annuity Business*. Oxford University Press, ISBN:9780199547272.

Raftery, A. E. (1999). Bayes Factors and BIC - Comment on "A critique of the Bayesian Information Criterion for Model Selection. *Sociological Methods and Research*, 27(3):411–427.

Basel Committee on Banking Supervision (August 2013). Longevity risk transfer markets: market structure, growt drivers,and impediments, and potential risks. *Available at* *http://www.bis.org/publ/joint31.pdf*.

European Commission (2012). Green Paper: "Confronting demographic change: a new solidarity between the generations". *Available at* *http://eur-lex.europa.eu/legal-content/IT/TXT/?uri=URISERV%3Ac10128*.

Human Mortality Database (2016). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). *Available at* *http://www.mortality.org* *or* *http://www.humanmortality.de*. (data downloaded on October 20, 2016).

United Nations (2002). Report of the Second World Assembly on Ageing, Madrid, 8-12 April 2002. *Available at* *https://documents-dds-ny.un.org/doc/UNDOC/GEN/N02/397/51/PDF/N0239751.pdf?OpenElement*.

Villegas, A., Millossovich, P., and Kaishev, V. (2016a). Package 'StMoMo'. *Available at* *https://cran.r-project.org/web/packages/StMoMo/StMoMo.pdf*.

Villegas, A., Millossovich, P., and Kaishev, V. (2016b). StMoMo: An R Package for Stochastic Mortal-

ity Modelling. *Available at* *https://cran.r-project.org/web/packages/StMoMo/vignettes/StMoMoVignette.pdf*.

Weakliem, D. L. (1999). A Critique of the Bayesian Information Criterion for Model Selection. *Sociological Methods and Research*, 27:359–397.

Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1):79–82. DOI:10.3354/cr030079.