Sapienza Università di Roma

Dipartimento di Scienze Statistiche
Sezione di Statistica Economica ed Econometria

Francesca Di Iorio Stefano Fachin

Forecasting mortality rates and life expectancy in the year of Covid-19

DSS Empirical Economics and Econometrics Working Papers Series

DSS-E3 WP 2020/1

Dipartimento di Scienze Statistiche
Sezione di Statistica Economica ed Econometria
“Sapienza” Università di Roma
P.le A. Moro 5 – 00185 Roma - Italia

http://www.dss.uniroma1.it

Forecasting mortality rates and life expectancy in the year of Covid-19

Francesca Di Iorio
University of Naples Federico II

Stefano Fachin*
"Sapienza" University of Rome

JEL: C12, C33, C55
Keywords: Mortality forecasting, life expectancy forecasting, Lee-Carter, factor model, Covid-19.

Abstract

Forecasting mortality rates and life expectancy is an issue of critical importance made arguably more difficult by the effects of current Covid-19 pandemic. In this paper we compare the performances of a simple random walk model (benchmark), three variants of the standard Lee-Carter model (Lee-Carter, Lee-Miller, Booth-Maindonald-Smith), the Hyndman-Ullah functional data analysys model, and a general factor model. We use both symmetric and asymmetric loss functions, as the latter are arguably more suitable to capture preferences of forecast users such as insurance companies and pension and health system planners. In a counterfactual study, designed exploiting the particular features of Italian data, we reproduce the likely impact of Covid-19 on forecasts using 2020 as a jump-off year. To put the results in perspective, we also carry out out a general assessment on 1950-2016 data for three countries with very diverse demographic profiles, France, Italy and USA. While the results with these latter datasets suggest that in normal conditions the Lee-Miller and Hyndman-Ullah models are somehow superior, from the counterfactual study the best option appears to be the Booth-Maindonald-Smith model.

1 Introduction

The last century witnessed an extraordinary decline in mortality rates at all ages, and consequent increase in life expectancy. At birth, for both sexes combined, the latter grew on a world wide basis from less than 50 years in 1955-60 to over 72 years in 2015-2020, and in high-income countries from less than 65 to more than 80 years[1]. Planning of adequate health and pension systems for ageing populations has thus become a critical policy issue in developed countries, spurring interest in models for mortality rates forecasting. A keystone of this literature was provided about thirty years ago by Lee and Carter (1992), who proposed a model with a single latent factor which rapidly became standard[2]. In the last few decades various variants of this model been proposed, as well as radically different models such as that based on a functional data approach by Hyndman and Ullah (2007). However, the issue is far from settled, as no model has yet proved to be consistently superior. For instance, the conclusion of the extensive review by Booth, Hyndman, Tickle and de Jong (2006) is that performances seem to depend on period, variable of interest (mortality rates or life expectancy) and country.

This uncertainty raises the further question of the impact the current Covid-19 emergency can be expected to have on the comparative performences of various popular models (introduced in section 2) for next vintage of forecasts. To answer this question we designed a counterfactual exercise, described in detail in section 4.2. To put these results in perspective, the findings of the counterfactual exercise are preceded, in section 4.1, by a wide-ranging comparison of forecasts obtained from these models for three countries with quite different demographic profiles and recent history, France, Italy and USA. A distinctive feature of our evaluation is the use, along with traditional symmetric loss functions, of asymmetric functions. The latter are likely to be particularly relevant for insurance companies and health and pension system planners. The set-up of the forecast comparison (data, loss functions and computational details) is described in section 3.

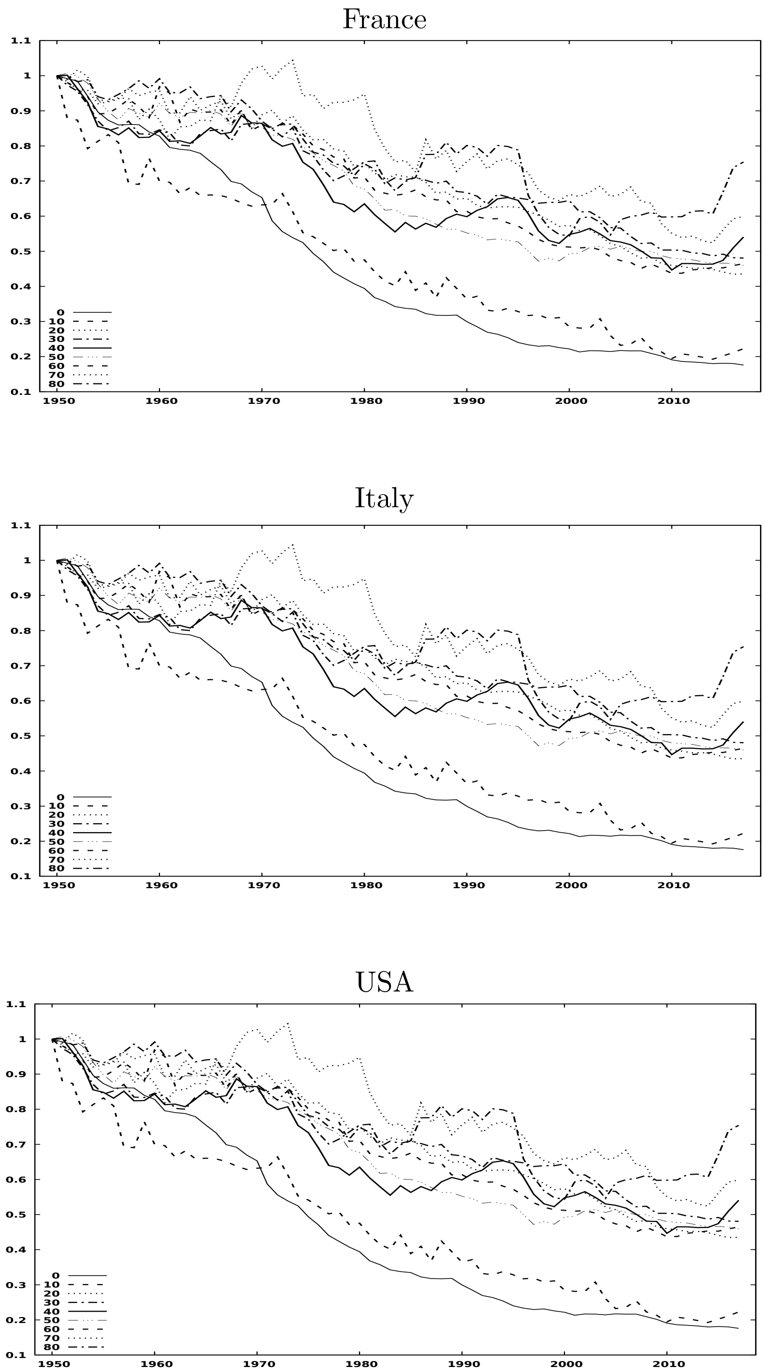
2 Forecasting mortality rates

To fix ideas, consider Fig. 1 which reports the time series 1950-2016 of total mortality rates for selected ages in the range 0-80 in France, Italy and USA. Although the time paths of the various rates do differ across ages as well as across the three countries, visual inspection of these plots suggests two main stylised facts. First, all mortality rates seem to be non-stationary with a negative drift; this is confirmed by formal tests, reported in Table 2 in Section 3.1. Second, some stochastic trends common to different age groups seem to be present. More precisely, in all countries one of these trends clearly drives the rapid decline of the rates for ages 0 and 10, and different one(s) the slowest fall of the rates of the adult and old ages.

[1]Source: UN World Population Prospects 2019, https://population.un.org/wpp/.

[2]Its standing is well described by the two following quotations: "The most well-known method of mortality forecasting" (Booth, 2006, p. 555); "[...] the most widely cited and used method in mortality prediction and applications" (Tsai and Yang, 2015, p. 1).

Figure 1: Total mortality rates for selected ages, 1950-2016 (1950=1)



Defining m_x the log of the central death rate for age x, the first point suggests as the simplest possible forecasting model a random walk with drift (briefly, RW). Assuming t is the last year of the estimation period (often called in the literature "jump-off year"), this model yields the h-periods ahead forecast

$$\widehat{m}_{xt+h} = h\widehat{\theta}_x + m_{xt} \tag{1}$$

where $\widehat{\theta}_x$ is the estimated drift. Note that the rate for the jump-off year is set at the observed value.

The problem with this model is that allowing for a different drift for each age x may easily lead to overfitting, which typically affect adversely the forecasting performance. Some robustness may be obtained following the suggestion of the second stylised fact, that all rates may depend upon some latent common trend. In this vein, in a path-breaking contribution Lee and Carter (1992), henceforth LC, studied US rates using a model based on the assumption that a single latent common factor, k_t, may represent the improvement in living conditions causing the general decline in death rates at all ages. Defining b_x the coefficient ("loading") projecting this latent trend k onto the death rate for age x,

$$m_{xt} = a_x + b_x k_t + \varepsilon_t$$

where a_x is the average mortality rate for age x. Estimates of latent trend and loadings may obtained by Singular Value Decompostion of the matrix of the centred death rates, with constraints imposed to obtain a unique solution[3]. The estimate of k is then further iteratively adjusted to match estimated and observed total number of deaths for each year. Finally, forecasts of the death rates are obtained from ARIMA forecasts of the common factor. In practice (see *e.g.* Booth, Hyndman, Tickle and de Jong, 2006) these are typically obtained from a random walk model, so that

$$\widehat{k}_{t+h} = h\widehat{\delta} + \widehat{k}_t.$$

where $\widehat{\delta}$ is the estimated drift. Then,

$$\begin{aligned} \widehat{m}_{xt+h} &= \widehat{a}_x + \widehat{b}_x \widehat{k}_{t+h} \\ &= \widehat{a}_x + \widehat{b}_x \left(h\widehat{\delta} + \widehat{b}_x \widehat{k}_t \right). \end{aligned} \tag{2}$$

Lee and Miller (2001), henceforth LM, proposed some marginal adjustments to the LC model. The most important are the introduction of constant adjustment to ensure a perfect fit in the last observation and the adjustment of k to match in each year estimated and observed life expectancy at birth, $e(0)$. This amended version of the LC model is now widely used in place of the original version (Booth, 2006). A second LC variant is due to Booth, Maindonald and Smith (2002), henceforth BMS, who adjust k to fit the age *distribution* of deaths rather than their total number.

A more radical departure from the LC framework was proposed by Hyndman and Hullah (2007), henceforth HU. The central idea of HU is to consider the matrix of the death rates as a surface in the three-dimensional space (age, time, rates), where age is continuous, and model it using Functional Data Analysis, writing

$$m_{xt} = a(x) + \sum\nolimits_{j=1}^{J} k_{jt} b_j(x) + \varepsilon_t(x) \tag{3}$$

where $a(x)$ is a smooth function of age, the $b(x)$'s are smooth basis functions, and the $k's$ are latent components evolving over time. The $a(x)$ are estimated taking the average over time of splines fitted separately to the data of each year, while the $k's$ and the $b(x)'s$ by principal components. The number of components J is unconstrained, and it can be chosen for instance minimising the forecast error on a subsample not used for estimation. In practice, HU in their empirical examples use 3 or 4. Finally, the forecasts of the rates are based upon forecasts of the latent $k's$ obtained from a state space formulation of damped exponential smoothing.

The HU model differs from the LC model in two respects. First, the restrictive and unwarranted assumption of a single factor is abandoned. Second, there is a massive use of non-parametric techniques. The first point can be also pursued applying a general factor

[3] More precisely, $\sum_{x=1}^{\Omega} b_x = 1$ and $\sum_{t=1}^{T} k_t = 0$, so that $a_x = T^{-1} \sum_{t=1}^{T} m_{xt}$.

model (henceforth GFM) with an empirically determined number J of non-stationary factors

$$m_{xt} = a_x + \sum\nolimits_{j=1}^{J} k_{jt} b_{xj} + \varepsilon_t. \quad (4)$$

A GFM for mortality rates with optimal, empirically determined number of factors is easy to estimate using the consistent information criteria derived by Bai (2004) for non-stationary settings. This models, which allows for different latent trends driving the decline in death rates of different age groups, is at the same time more general and flexible than LC's, simpler than HU's, and easy to estimate by Principal Components. Since the idea is to assess the explanatory and forecasting power of a general model with unrestricted number of factors, we will deliberately avoid to apply to the estimated factors a calibration procedure such as those embedded in all the LC variants. In this respect this model is similar to HU's.

Forecsts of mortality rates are obtained from forecasts of the factors as

$$\widehat{m}_{xt+h} \;=\; \widehat{a}_x + \sum\nolimits_{j=1}^{J} \widehat{k}_{jt+h} \widehat{b}_{xj}.$$

where $\widehat{k}_{jt+h}$ is the optimal ARIMA forecast for factor j, easily computed for instance using automatic ARIMA modelling.

The idea of using a model with more than one factor is not new. French and O'Hare (2013) proposed to apply a Dynamic Factor Model for stationary variables to the differenced mortality rates, and then cumulate the forecasts to obtain the mortality rates. Clearly, using a factor model especially designed for non-stationary data is more satisfactory. Haldrup and Rosenskjold (2019) proposed instead a factor model with the number of factors J constrained *a priori* to four, assumed to respectively capture a general trend and those specifically associated to early ages, mid-20's-early 30's and adult ages in general. This identification is ensured by parametric estimation of the loadings (for instance, those of the second factor converge to zero after about age 10). While this approach is consistent with the demographic tradition of parametric forecasting models, it appears unecessarily rigid, it is computationally demanding and difficult to automate. We thus did not include it in our comparison. A summary of the different competing models we did consider is in Box 1.

Finally, once a set of forecasts of mortality rates is available we may compute by means of standard formulas (see, *e.g.*, Arias, 2012) forecasts of life expectancy, either at birth or at a different age of special interest, such as retirement age. Since life expectancy is a non-linear function of the mortality rates, the quality of these forecasts will be assessed on its own. We will examine two different measures: life expectancy at birth, a measure of general interest, and at 65 years, a typical retirement age. This latter measure is of special interest for health and pension system planning.

Box 1 Summary of the forecasting models considered

1. Random Walk (RW): $\widehat{m}_{xt+h} = h\widehat{\theta}_x + m_{xt}$.

2. Lee-Carter (1992; LC), Lee-Miller (2001; LM), Booth, *et al.* (2002; BMS): $\widehat{m}_{xt+h} = \widehat{a}_x + \widehat{b}_x \left(h\widehat{\delta} + \widehat{b}_x \widehat{k}_t \right)$, where $\widehat{k}_{jt+h}$ is the ARIMA(0,1,0) forecast of $\widehat{k}$, calibrated in slightly different ways in the three models. Other minor computational differences also present.

3. Hyndman and Hullah (2007; HU): $\widehat{m}_{xt+h} = \widehat{a}(x) + \sum_{j=1}^{J} \widehat{k}_{jt+h} \widehat{b}_j(x)$, where $\widehat{k}_{jt+h}$ is a forecast of $\widehat{k}$ obtained by damped exponential smoothing.

4. General Factor Model (GFM): $\widehat{m}_{xt+h} = \widehat{a}_x + \sum_{j=1}^{J} \widehat{k}_{jt+h} \widehat{b}_{xj}$, where $\widehat{k}_{jt+h}$ is the optimal ARIMA(p,1,q) forecast. No calibration applied to $\widehat{k}$.

3 Forecast evaluation: set-up

3.1 Data

To assess the forecasting ability of the methods outlined in section 2 in normal conditions we use data for total mortality rates, at single years of age, for three advanced countries, France, Italy and USA, for the period 1950-2016[4]. Discarding observations before 1950 we shall exclude the outliers associated with the two world wars and the Spanish influenza pandemic.

As it can be appreciated from Table 1, over the 1950-2016 period these three countries followed quite different demographic trends[5]. Italy clearly underwent the most dramatic ageing process, with the share of population 0-14 dropping from nearly 25% in the 1960's to less than 14% in the 2010's and that of population over 65 more than doubling, from about 10% to over 21%. The corresponding share is smaller but still sizeable in France (about 18%, from a 1960's initial value of 12%) and much smaller in the USA (about 14% in the 2010's, starting from less than 10% in the 1960's). Interestingly, the final 0-14 shares in France and USA are not very different, about 18%-19%. Given the pretty different initial values (about 26% in France and 30% in the USA), this implies a sharper decline in the USA. In fact, in this country the 10% decline is approximately as large as in Italy. Finally, life expectancy at birth, approximately 70 years in the 1960's in all the three countries, increased definitely more in Italy and France than in the USA, reaching respectively about 82 years in the first two countries and less than 79 years in the latter. Summing up, these three countries definitely provide a good example of three quite different "situations" (Booth, 2006) in which to test the forecasting ability of the competing methods.

Table 1: France, Italy and USA: 0-14 and 65+ population shares ($\times 100$) and life expectancy at birth

		1960-69	1970-79	1980-89	1990-99	2000-09	2010-16	Δ
0-14	*France*	25.7	24.0	21.3	19.6	18.6	18.4	*-7.4*
	Italy	24.7	24.0	19.4	15.2	14.2	13.8	*-10.9*
	USA	30.0	25.5	21.9	21.9	21.0	19.5	*-10.5*
65+	*France*	12.1	13.5	13.4	15.0	16.4	18.2	*6.1*
	Italy	10.2	12.1	13.5	16.4	19.4	21.5	*11.3*
	USA	9.5	10.7	12.1	12.6	12.4	14.1	*4.7*
$e(0)$	*France*	70.7	72.7	75.2	77.7	80.1	82.3	*11.5*
	Italy	70.1	72.7	75.4	78.1	80.8	82.7	*12.5*
	USA	70.2	72.2	74.5	75.8	77.5	78.6	*8.5*

$e(0)$: life expectancy at birth;

Δ: difference between 2010-17 and 1960-69 averages.

Source: elaborations on data from "World Development Indicators", World Bank.

Our forecasting exercise will be carried out for rates at single year of age up to 94, and for the aggregate of all ages over 95. As anticipated above, log rates can safely taken to be $I(1)$. In all countries standard ADF-GLS tests allowing for a deterministic trend

[4]Source: Human Mortality Database, https://www.mortality.org/.

[5]The following comparisons are based on data from the "World Development Indicators" database of the World Bank, which start in 1960. Given that the aim is drawing a broad picture, the lack of data for the 1950's, included in the estimation sample, is not a particular issue.

rejected the null hypothesis in small fractions of cases, always fully compatible with the significance levels of the test (cf. Table 2).

Table 2: ADF-GLS tests: fraction of rejections ×100

α	*1.0*	*5.0*	*10.0*
France	2.1	2.1	3.1
Italy	6.7	8.5	9.5
USA	0.0	2.1	4.2

$H_0 : m_x$ is $I(1)$,
$H_1 : m_x$ is $I(0)$+trend.
$x = 0, \ldots, 95+$.

3.2 Loss functions

An often overlooked point of forecast assessment is the choice of the loss function, usually confined to symmetric functions, such as the average of quadratic or absolute errors. However, these functions may not reflect accurately the risk function of forecasts users. Given the variables of interest, mortality rates and life expectancy, we can consider a stylised picture with forecast users divided in two groups.

The first group includes agents offering only term life insurance contracts covering the risk of death. In a stylised set-up this group may be denominated "insurance companies". These forecast users need to plan the flow of future payments to beneficiaries of term life insurance contracts, which are a positive function of future mortality.

The second group includes agents providing healthcare services and paying old age pensions, thus essentially covering the opposite risk of survival, or longevity risk[6]. These agents, whom will be denominated "healthcare systems and pension funds", need to forecast a demand increasing with life expectancy.

Under- and over-estimation of future mortality rates (implying errors of opposite direction in the estimation of life expectancy) will clearly have different implications for the two groups of forecast users. For insurance companies an underestimation of future mortality rates will cause an underestimation of future payment flows, with possible, serious, liquidity problems. On the other hand, an overestimation will imply an over-accumulation of reserves, a suboptimal but not particularly critical condition. Defining the estimation error for mortality rate at age x and time t as $\epsilon_{xt+h} = m_{xt+h} - \widehat{m}_{xt+h}$, the loss function of insurance companies will thus arguably attach a greater weight to positive forecast errors of mortality rates than to negative ones.

The same holds for healthcare systems and pension funds with respect to forecasts of life expectancy. Overestimation of future life expectancy (defining the estimation error for age x as $\epsilon_{e(x)t+h}{=}e\left(x\right)_{t+h} - \widehat{e\left(x\right)}_{t+h}$, $\epsilon_{e(x)t+h} < 0$) will cause overestimation of health and pension expenditure, also a suboptimal but not critical condition. On the other hand, underestimation ($e\left(x\right)_{t+h} > \widehat{e\left(x\right)}_{t+h}$, so that $\epsilon_{e(x)t+h} > 0$) will lead to plan an inadequate flow of health services, pension payments and required revenues, an

[6]Health and long term care insurance policies will fall in this category as well. For the sake of simplicity, and with no much loss of generality, they are ignored in the following discussion. To take them into account we would simply need to enlarge the second group to include specialised insurance companies offering those products only.

increasingly serious problem for the advanced economies (see, e.g., the lecture by the chairman of the Bank of Italy, Visco, 2015).

Our evaluation exercise should thus be extended to include some asymmetric loss functions as well as standard symmetric loss functions. For simplicity, in the main evaluation exercise we shall consider only the case of asymmetric functions attaching greater weight to positive errors than negative ones (that is, penalising underestimation more than overestimation; in short, positive asymmetry). In the light of what argued above for mortality rates this may be interpreted as capturing the loss function of insurance companies offering term life insurance contracts, while in the case of life expectancy those of healthcare systems and pension funds[7].

The specific form of the functions can be derived following Elliott, Komunjer and Timmerman (2005), who define a general loss function governed by two shape parameters: p, the power of the error, and $\alpha \in (0, 1)$, the asymmetry coefficient. For a generic forecast error ϵ_{t+h} and a vector θ of parameters of the model used to produce the forecast,

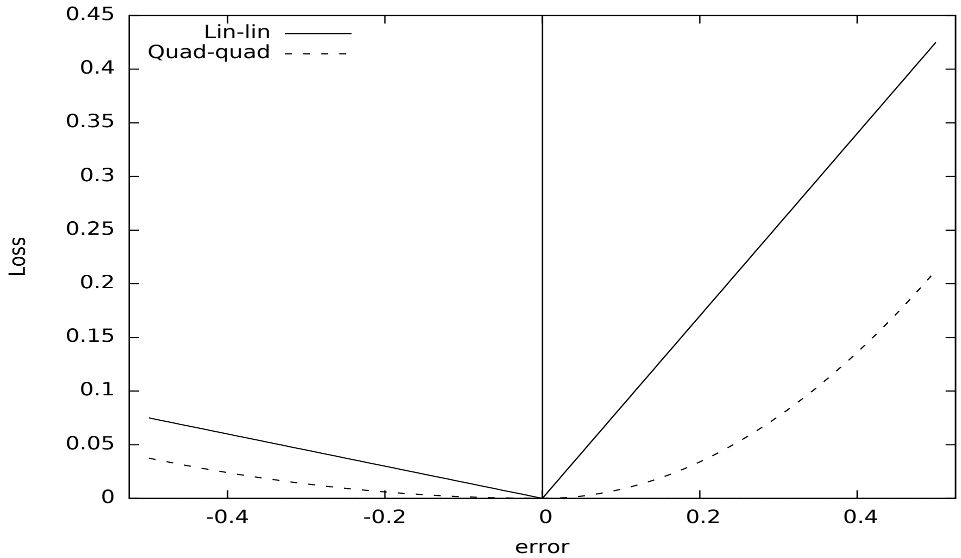
$$L(p, \alpha, \theta) \equiv [\alpha + (1 - 2\alpha) \cdot 1(\epsilon_{t+h} < 0)] |\epsilon_{t+h}|^p \tag{5}$$

Setting $\alpha = 0.5$ we have symmetric loss, respectively quadratic for $p = 2$ and in absolute value for $p = 1$. Choosing any $\alpha \neq 0.5$ generates right ($\alpha > 0.5$) and left ($\alpha < 0.5$) asymmetry. Since we need a rather strong right asymmetry, without much loss of generality we set $\alpha = 0.85$, which generates the "Quad-quad" ($p = 2$) and "Lin-lin" ($p = 1$) functions plotted in Fig. 2.

Summing up, we consider four different loss functions:

1. *Quadratic*: $L(p = 2, \alpha = 0.5, \theta) \equiv \epsilon_{t+h}^2$
2. *Absolute*: $L(p = 1, \alpha = 0.5, \theta) \equiv |\epsilon_{t+h}|$
3. *Quad-quad*: $L(p = 2, \alpha = 0.85, \theta) = [0.85 - 0.70 \cdot 1(\epsilon_{t+h} < 0)] \epsilon_{t+h}^2$
4. *Lin-lin*: $L(p = 1, \alpha = 0.85, \theta) = [0.85 - 0.70 \cdot 1(\epsilon_{t+h} < 0)] |\epsilon_{t+h}|$

Figure 2: Quadratic and linear loss functions with positive asymmetry ($\alpha = 0.85$)



[7] Conversely, loss functions with negative asymmetry (greater weight to negative errors, penalising overestimation more than underestimation) may be interpreted as representing those of healthcare systems and pension funds for forecasts of mortality rates, and those of insurance companies offering term life contracts for forecasts of life expectancy.

3.3 Computational details

For LC, LM, BMS and HU we used the R package "demography" by Hyndman (2006). This implementation is standard in the literature (it is for instance recommended in the review by BMS). The general factor model has been estimated with number of factors selected on the basis of the information criterion IPC_2 by Bai (2004), and automatic ARIMA forecasting of the factors with model selection based on the BIC criterion. For both tasks we used own code written in Hansl, the programming language of the free econometric program Gretl. For the HU model following Hyndman and Hullah (2007) we set $J = 4$. Some robustness check showed the model to be quite robust to the choice of this parameter.

In view of the available data we decided to assess forecasts for horizon h up to 10 years. We examined separately the short-, medium- and long-term performances looking at forecasts one, five and ten years ahead ($h = 1, 5, 10$), and the overall performance looking at the average over the entire path. We start with a first set of models estimated on the sample 1950-1999 ($T = 50$, reasonably large), used to forecast from 2000 ($h = 1$) up to 2004 ($h = 5$) and finally 2009 ($h = 10$). We then roll on until we reach the estimation sample 1950-2007 ($T = 58$), yielding forecasts from 2008 ($h = 1$) up to 2012 ($h = 5$) and finally 2016 ($h = 10$). We thus have a set of eight different forecasts: set 1, data 1950-1999, forecasts 2000-2009; set 2, data 1950-2000, forecasts 2001-2010; set 3, data 1950-2001, forecasts 2002-2011; etc., until set 8, data 1950-2007, forecasts 2008-2016. For each loss function we then compare the forecasting performance of the six different models for mortality rates, life expectancy at birth and life expectancy at a typical retirement age, 65 years. The comparison is based on a Model Confidence Set (Hansen, Lunde and Nason, 2011) at the 5% significance level.

4 Forecast evaluation

4.1 Basic assessment

We discuss first the results of an assessment of the performances which can be expected in normal conditions. This will provide some perspective to the results of the counterfactual study reported in the next section.

The amount of the results to be discussed here is proportional to the many dimensions of the exercise: six models, three variables, four loss functions, four forecasting horizons ($h = 1, 5, 10$, and the entire path from $h = 1$ to $h = 10$). The full details are reported in Tables A.1 (mortality rates), A.2 (life expectancy at birth) and A.3 (life expectancy at 65). Here we shall discuss the synthesis across the three countries reported in Table 3, where the performances for each variable are classified by type of Loss function for the aggregate of all forecasting horizons, and in Table 4, for each variable and forecast horizon for the aggregate of all Loss functions. In these tables we report two summay measures: the number of times a model is in the 5% MCS, denoted by $\#(MCS)$, and the number of times its loss is the smallest of the MCS, $\#(\mathit{Min\ Loss})$. For each class of loss function the models delivering the best performance are those with the highest values of these statistics, marked in bold face.

The first obvious message of Table 3 (aggregate of all forecasting horizons) is that the MCS is always very wide. This is evident from the inflated $\#(MCS)$ statistics, which are of very little use for comparison purposes. For instance, taking the first row of Table 3 (symmetric loss functions, forecasting mortality rates) we find for two models $\#(MCS)$ = 20 and 21 and for three more only slightly smaller values, 15 and 16. Only one model performs cleary worse, with a score of 11. Of course, in view of the small number of forecasting periods the fact that the performances tend to be approximately equivalent is not really surprising.

To reach some mode definite conclusion we then look at #(*Min Loss*), the number of times a model has the smallest loss of the MCS. In other words, we look at the point estimates of the losses, even if for all models in the MCS the differences among them are not significant. This produces a complex picture, with findings highly variable across models and variables. Broadly speaking, the performances of LM and HU appear to be rather consistently good with both types of loss functions, while on the opposite those of BMS and LC rather consistently poor. GFM and RW stand in between, with performances which may be good for one variable/loss type combination but poor for a different one. Going into some more detail, we can say that LM seems to dominate for Mortality rates forecasting, while HU for Life expectancy, both at birth and at 65 years.

The superiority of the LM and HU models is essentially confirmed by the breakdown by forecasting horizon, see Table 4. LM is essentially the best model for Mortality rate forecasting at all horizons, and HU for Life expectancy at birth; things are less clear for Life expectancy at 65 years, with LC dominating at 1 year, LM at 5 years, HU at 10 years and for the entire 1-10 years path.

Table 3: Comparative performances by variable and type of Loss function (aggregate of all forecasting horizons)

		RW	GFM	LC	LM	BMS	HU
Loss function				*Mortality rates*			
Symmetric	#(*MCS*)	16	15	16	20	11	**21**
	#(*Min Loss*)	1	6	1	**10**	0	6
Asymmetric	#(*MCS*)	17	16	6	**20**	8	10
	#(*Min Loss*)	6	3	1	**9**	0	5
				Life expectancy at birth			
Symmetric	#(*MCS*)	19	16	21	22	22	**24**
	#(*Min Loss*)	0	0	1	5	2	**16**
Asymmetric	#(*MCS*)	22	**23**	15	14	12	**23**
	#(*Min Loss*)	4	**7**	2	4	0	**7**
				Life expectancy at 65			
Symmetric	#(*MCS*)	8	5	15	18	10	**21**
	#(*Min Loss*)	0	0	4	11	0	**9**
Asymmetric	#(*MCS*)	17	16	16	20	9	**24**
	#(*Min Loss*)	1	3	**9**	4	0	7

#(*MCS*): number of times model is in the 5% MCS for each h;
#(*Min Loss*): number of times model has the minimum loss in the MCS;
Bold = *best*.

Table 4: Comparative performances by variable and forecast horizon (aggregate of all loss functions)

		RW	GFM	LC	LM	BMS	HU
horizon		*Mortality rates*					
1	#(*MCS*)	10	4	4	**12**	8	10
	#(*Min Loss*)	0	0	0	**11**	0	1
5	#(*MCS*)	8	11	8	**10**	6	**10**
	#(*Min Loss*)	2	**5**	0	4	0	1
10	#(*MCS*)	8	9	7	**9**	1	8
	#(*Min Loss*)	3	1	2	1	0	**5**
1-10	#(*MCS*)	7	7	3	**9**	4	9
	#(*Min Loss*)	2	3	0	3	0	**4**
		Life expectancy at birth					
1	#(*MCS*)	**12**	10	11	5	5	11
	#(*Min Loss*)	2	1	0	0	0	**9**
5	#(*MCS*)	9	9	11	11	11	**12**
	#(*Min Loss*)	0	0	2	**4**	2	**4**
10	#(*MCS*)	**12**	10	6	11	10	**12**
	#(*Min Loss*)	1	3	0	3	0	**5**
1-10	#(*MCS*)	8	10	8	9	8	**12**
	#(*Min Loss*)	1	3	0	3	0	**6**
		Life expectancy at 65					
1	#(*MCS*)	10	4	**12**	8	8	10
	#(*Min Loss*)	0	0	**8**	1	0	3
5	#(*MCS*)	4	6	8	**12**	5	11
	#(*Min Loss*)	0	1	2	**6**	0	3
10	#(*MCS*)	6	6	4	9	3	**12**
	#(*Min Loss*)	1	2	1	2	0	**6**
1-10	#(*MCS*)	5	5	7	9	3	**12**
	#(*Min Loss*)	0	0	2	**6**	0	4

#(*MCS*): number of times model is in the 5% MCS for horizon h, all loss functions;
#(*Min Loss*): number of times model has the minimum loss in the MCS;
Bold = *best*.

4.2 Counterfactual robustness assessment

Forecasts computed in 2021 will use 2020 as a jump-off year. Therefore, assuming that the Covid-19 pandemic will be under control by 2021, that vintage of forecasts will be based on a jump-off year characterised by anomalous mortality. As of early November 2020, this has been essentially concentrated in the older age groups: 98.7% of Covid-19 deaths have been of people over 60, and 95.6% of people over 70 (see Table 6 in Task force Covid-19, 2020). This implies that some of the conclusions reached above may continue to hold for the forecasts for the younger age groups, whose data will show no anomalies, but they may not for the older ones, which instead will.

For instance, in the LM model the coefficients are constrained to have a perfect fit for each age in the jump-off year. If the mortality rates for older groups go back to lower levels from 2021 onwards, its forecasts for these groups will be upward biased. For some loss functions this may not be a problem, while for others it will. In those cases models not adjusting the estimates to achieve a perfect fit in the last year of the sample may perform better than the LM model.

On the other hand, the HU model, estimated using robust techniques, is a priori unlikely to suffer from the use of an anomalous jump-off year.

Clearly, the picture is so complex that predicting the overall impact of Covid-19 on the comparative performances of the various models on the basis of the characteristics of the currenty anomaly and of the models is practically impossible. However, as anticipated in the Introduction, we can reach a conclusion carrying out a conceptually simple counterfactual study. This counterfactual study should answer the following question: "Given that we have data for 1950-2016, and that we use 1950-2006 for estimation while keeping those for 2007-2016 to assess forecasting performances at horizons from 1 (2007) to 10 (2016), *which would have been the conclusion of our comparison if the jump-off year 2006 had been contaminated by extramortality as 2020?*".

The natural objection to this proposal is that constructing data mimicking conditions of the *current* year appears to be in practice an hardly possible task. Fortunately, one of the countries of our dataset, Italy, has particular features making it possible. In Italy in 2015 "the action of flu viruses during the wintertime, associated to the lethal effects of a particularly hot summer" (Blangiardo, 2020, p. 2) caused about 50,000 excess deaths. This is still significantly more than the about 40,000 deaths registred by the official Covid-19 Italian integrated surveillance database as of early November 2020 (Task force Covid-19, 2020)[8]. Further, the 2015 extra mortality was essentially concentrated in the older age groups, exactly as it happened so far in 2020. The simple average of the mortality rates ($\times 1000$) over 65 jumped from 99.5 in 2014 to 109.3 in 2015, falling to 99.4 in 2016. On the other hand, the average under 65 was 1.4 in 2014 and 2015, and declined marginally to 1.3 in 2016[9].

Defining the "normal" mortality rate in 2015 for age x, $\overline{m}_{x2015}$, as the linear interpolation of those for 2014 and 2016,

$$\overline{m}_{x2015} = 0.5\,(m_{x2014} + m_{x2016})$$

a simple estimate of the impact of the 2015 extraordinary conditions is given by the ratio $m_{x2015}/\overline{m}_{x2015}$. Using this estimate we can compute *counterfactual mortality rates* m^*_{xt}, that is, the mortality rates which would have been registered in year t under (extraordinary) conditions similar to those of 2015, as

$$m^*_{xt} = \frac{m_{x2015}}{\overline{m}_{x2015}} m_{xt}. \tag{6}$$

Using (6) with $t = 2006$ we can compute counterfactual rates for the jump-off year used for forecasts running from 2007 (horizon $h = 1$) up to 2016 (horizon $h = 10$). Note that

[8]Given the dramatic impact of Covid-19, the fact that in 2015 extra mortality may have been possibly even higher may appear surprising. However, it should not be forgotten that in 2015 excess mortality was spread over the entire nation and most of the year, while as of early November 2020 the effects of the Covid-19 epidemic have been mostly concentrated in a relatively small area over quite a short period (respectively, some of the Northern regions and about two months, March and April; see Istat-ISS, 2020a,b). This concentration in time and space caused the dramatic overcrowding of the intensive care units in the hospitals of the worse hit areas and the decision to lockdown the entire nation in March 2020, which limited overall mortality. The effects of the so-called "second wave" of the epidemic have been so far quantitatevely smaller.

[9]Averages computed from rates for 5-years age groups. Data downloaded on November, 12, 2020 from www.istat.it.

since both the 2015 and 2020 anomalies had an impact essentialy only on the older age groups, we modify the rates for ages $x \geq 65$ only.

As some of the models we are comparing require population data for calibration purposes, we also need to construct counterfactual population series (Pop^*_{xt}) consistent with these counterfactual mortality rates. Recalling that population is measured at Janury, 1^{st} and the mortality rates are annual averages, these series can obtained[10] as

$$Pop^*_{xt+1} = \begin{cases} \left(1 - m^*_{x-1,t}\right) Pop_{x-1,t} & t = 2006 \\ \left(1 - m_{x-1,t}\right) Pop^*_{x-1,t} & t > 2006 \end{cases}$$

for $x \geq 65$.

We can now run our forecasting competition with this counterfactual dataset. Comparing these results with those obtained with the actual datasets (in which the year 2006 had no anomalies) we can establish if the models performing better in the base competition are robust to the use of an outlier as a jump-off year, or, on the contrary, if in these circumstances they are outperformed by other models. Since in the counterfactual dataset only the mortality rates above 65 years have been modified, the comparison is carried out only for mortality rates and life expectancy at 65.

An important point which needs to be discussed is the shape of the loss function. With models using actual jump-off year data, extramortality in that year will introduce a positive bias in mortality rates forecasts, and a negative bias in those of life expectancy. Thus, to have a complete picture we need to consider both functions with a positive asymmetry, which penalise more strongly underestimation, and functions with a negative asymmetry, which instead penalise more strongly over-estimation. More precisely, we shall use functions with $\alpha = 0.15$. From equation (5) this value is easily seen to yield loss functions with negative asymmetry which are the mirror image of those with a positive asymmetry ($\alpha = 0.85$) used in the main assessment exercise.

The detailed results are reported in Tables A.4-A.6 in the Appendix. Here we discuss a summary measure, the number of times each model delivers the minimum loss over all four horizons for all loss functions.

First of all, consider in Table 6 the results for the two symmetric loss functions. In this case this summary measure ranges from zero to 8.

With actual data for mortality rates forecasting the LM model, delivering the minimum loss at all horizons and with all loss functions, is a clear winner. This is consistent with the results of the previous section, which essentially singled out LM and HU as the models with the best performances. Confirming the need of assessing separately the forecasts for the two variables, results for life expectancy are different, and rather surprising. The best model is the basic LC model, which delivers the minimum loss four times. We then have the HU and the naive RW model, both with a score of two.

With counterfactual data the outcome is completely different, with BMS essentially dominating for both variables. Since in this LC variant the coefficient estimates are *not* adjusted to have a perfect fit in the jump-off year we must conclude that the gain of having robust forecasts for the older groups outweights the loss of not having well-calibrated forecasts for younger groups.

[10]Since we are constructing an artificial population, ignoring the approximations caused by the use of central mortality rates is not an issue.

Table 5: Counterfactual robustness assessment: forecasting with symmetric loss functions, jump-off year 2006. Number of times each model delivers the minimum loss (aggregate over forecasts at 1, 5 and 10 years and on the entire 1-10 path).

	RW	GFM	LC	LM	BMS	HU
		#(Min Loss) with actual data				
mortality rates	0	0	0	**8**	0	0
life expectancy at 65	2	0	4	0	0	2
		#(Min Loss) with counterfactual data				
mortality rates	0	0	1	0	7	0
life expectancy at 65	0	0	0	0	**8**	0

Let us now move to the results with asymmetric loss functions. These have been introduced in section 3.2 to represent the different requirements of two stylised groups of users, insurance companies offering term life insurance contracts and healthcare systems and pension funds. The loss functions can be expected to be as follows

(i) *Insurance companies*: positive asymmetric loss for mortality rates; negative asymmetric loss for life expectancy.

(ii) *Healthcare systems and pension funds*: negative asymmetric loss for mortality rates; positive asymmetric loss for life expectancy.

In case (i), Table 6, the results are partially different from those under symmetric loss. First, with actual data LM is the best model for both variables. Of course, the explanation of this different outcome is simple: the loss function of insurance companies does not penalise heavily a positive bias in forecasts. Although with the countefactual data this bias is likley to be significant for the older age groups, the LM is still the best forecasting model for mortality rates, while for life expectancy BMS (with no constant adjustment, thus robust to the Covid-191 anomaly) performs better.

Table 6: Counterfactual robustness assessment: forecasting for insurance companies offering term life insurance contracts, jump-off year 2006. Number of times each model delivers the minimum loss (aggregate over forecasts at 1, 5 and 10 years and on the entire 1-10 path).

	RW	GFM	LC	LM	BMS	HU
		#(Min Loss) with actual data				
mortality rates	0	0	0	**8**	0	0
life expectancy at 65	2	0	2	**4**	0	0
		#(Min Loss) with counterfactual data				
mortality rates	2	0	0	**4**	2	0
life expectancy at 65	0	0	0	0	**8**	0

mortality rates: positive asymmetric loss functions ($\alpha = 0.85$);
life expectancy at 65: negative asymmetric loss functions ($\alpha = 0.15$).

Finally,in case (*ii*), Table 7, we find that with actual data the choice is between LM (for mortality rates) and, rather surprisingly, the naive RW (for life expectancy). With counterfactual data BMS is instead by far the best for both variables.

Table 7: Counterfactual robustness assessment: forecasting for healthcare systems and pension funds, jump-off year 2006. Number of times each model delivers the minimum loss (aggregate over forecasts at 1, 5 and 10 years and on the entire 1-10 path).

	RW	GFM	LC	LM	BMS	HU
			#(Min Loss) with actual data			
mortality rates	0	0	0	**8**	0	0
life expectancy at 65	**4**	1	2	0	0	1
			#(Min Loss) with counterfactual data			
mortality rates	0	0	0	1	**5**	1
life expectancy at 65	0	0	0	1	**7**	1

mortality rates: negative asymmetric loss functions ($\alpha = 0.15$);
life expectancy at 65: positive asymmetric loss functions ($\alpha = 0.85$).

Summing up, while with actual data considering both symmetric and asymmetric loss functions over all horizons LM delivers the best forecasts, if we had extramortality in the jump-off year these would have been generally given by BMS. We should however remark that in a few cases LM works best in these case as well.

5 Conclusions

Forecasting mortality rates and life expectancy is becoming an increasingly important issue for health and pension systems and the insurance industry at large, and its challenges are particulary severe in the year of the Covid-19 emergency. In this paper we carried out a counterfactual exercise based on Italian data, aimed at assessing the forecasting performances which can be expected in the current anomalous conditions by the classical single factor model by Lee and Carter (1992), some of its variants, the functional data model by Hyndman and Ullah (2007), and a general factor model with number of factors empirically determined. A novel feature of our exercise is the use of both symmetric and asymmetric loss functions. The latter, although rarely used, may be argued to represent more realistically than the symmetric ones the cost functions of the users of forecasts of mortality rates and life expectancy. To put the counterfactual exercise in perspective we introduced it with an extensive comparison of the performances for France, Italy and USA with estimation period 1950-2006 and rolling forecasts for 2007-2016 is that no model can be singled out as clearly superior to the other. Formally, the Model Confidence Sets computed following Hansen, Lunde and Nason (2011) often include several models. This said, broadly speaking the performances of the Lee-Miller (2001) variant of the Lee-Carter model and the Hyndman and Ullah (2007) model are often better than those of the other models. These conclusions are however not robust to the presence of extra mortality in the older age groups in the final year of the sample (or jump-off year), of a scale comparable to that of the current Covid-19 pandemic. Our counterfactual exercise suggests that in these circumstances the best option is the Booth, Maindonald and Smith (2002) variant of the Lee-Carter model.

References

[1] Arias, E. (2012) “United States Life Tables, 2008”, *National Vital Statistics Reports*, 61.

[2] Bai, J. (2004) “Estimating cross-section common stochastic trends in nonstationary panel data”, *Journal of Econometrics,* 122, 137–183.

[3] Blangiardo, G. C. (2020) “Demographic effects of COVID-19: scenarios for mortality” downloaded on 20/7/2020 from https://www.istat.it/it/files//2020/03/Presidents-Report.pdf

[4] Booth, H. (2006) “Demographic forecasting: 1980 to 2005 in review”, *International Journal of Forecasting,* 22, 547– 581.

[5] Booth H., J. Maindonald, L. Smith (2002). “Applying Lee-Carter under conditions of variable mortality decline”, *Population Studies*, 56, 325–336.

[6] Booth, H., R.J. Hyndman, L. Tickle, P. de Jong (2006) “Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions”, *Demographic Research*, 15, 289-310.

[7] Elliot, G., I. Komunjer, A. Timmerman (2005) “Estimation and Testing of Forecast Rationality under Flexible Loss”, *Review of Economic Studies*, 72, 1107–1125.

[8] French, D., and C. O’Hare (2013) “A Dynamic Factor Approach to Mortality Modelling” *Journal of Forecasting*, 32, 587-599.

[9] Maravall, A., V. Gómez, G. Caporello (1996) "Statistical and Econometrics Software: TRAMO and SEATS," Statistical and Econometrics Software tramoseats, Banco de España; Statistical and Econometrics Software Homepage, revised 2015.

[10] Haldrup, N. and C.P. T. Rosenskjold (2019) “A Parametric Factor Model of the Term Structure of Mortality”, *Econometrics*, 7, 9; doi:10.3390/econometrics7010009

[11] Hansen, P. , A. Lunde, J.M. Nason (2011) “The Model Confidence Set”, *Econometrica*, 79, 453-497.

[12] Hyndman R.J. (2006) “demography: Forecasting Mortality, Fertility, Migration and Population Data” https://cran.r-project.org/web/packages/demography/demography.pdf

[13] Hyndman R.J., M.S. Ullah (2007) “Robust forecasting of mortality and fertility rates: a functional data approach” *Computational Statistics and Data Analysis*, 51, 4942–56.

[14] Istituto Nazionale di Statistica-Istituto Superiore di Sanità (2020a) “Impact of the Covid-19 epidemic on the total mortality of the resident population in the first quarter of 2020” downloaded on 20/7/2020 from https://www.istat.it/en/archivio/242152

[15] Istituto Nazionale di Statistica-Istituto Superiore di Sanità (2020b) “Impatto dell’epidemia Covid-19 sulla mortalità totale della popolazione residente periodo Gennaio-Maggio 2020” downloaded on 16/9/2020 from https://www.epicentro.iss.it/coronavirus/pdf/Rapp_Istat_Iss_9luglio.pdf

[16] Lee, R. D., L. R. Carter (1992) “Modeling and forecasting U.S. mortality”, *Journal of the American Statistical Association*, 87, 659–71.

[17] Lee, R. D.,T. Miller (2001) “Evaluating the performance of the Lee–Carter method for forecasting mortality”, *Demography*, 38, 537–49.

[18] Task force Covid-19 (2020) “Epidemia COVID-19, Aggiornamento nazionale: 7 novembre 2020” Dip. Malattie Infettive e Servizio di Informatica, Istituto Superiore di Sanità, Roma, Italy.

[19] Tsai, C. C.-L., and S. Yang (2015) “A Linear Regression Approach to Modeling Mortality Rates of Different Forms”, *North American Actuarial Journal*, 19, 1–23.

[20] Visco, I. (2015) “Il rischio di longevità e i cambiamenti dell’economia” *Lectio magistralis*, Università degli Studi di Trieste.

Corresponding author:
Stefano Fachin
Dip. di Scienze Statistiche, "Sapienza" University
P.le A. Moro 5
00185 Roma Italy

E-mail: stefano.fachin@uniroma1.it.
Tel: +-39-6-49910834.

Table A.1: Forecasting Mortality rates, ages 0-95+

	h=1						*h=5*						*h=10*						*h=1-10*					
													France											
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quadratic	×	×	×	⊗	×	×		⊗	×	×	×	×	×	×	×	×		⊗	×	⊗	×	×	×	×
Absolute	×		×	⊗	×	×		⊗	×	×	×	×	×	×	×	×		⊗		⊗	×			×
Quad-quad	×	×	×	⊗	×	×		⊗	×			×	×	×	×	×	×	⊗	×	×		×	×	⊗
Lin-lin	×	×		×	×	⊗		⊗					×	×		×		⊗		×				⊗
													Italy											
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quadratic	×	×	×	⊗	×	×	×	×	×	×	×	⊗	⊗	×		×			×	⊗			×	×
Absolute	×			⊗	×	×	×	×	×	⊗	×	×	×			⊗			×			⊗	×	×
Quad-quad	×			⊗	×	×	⊗	×		×		×	⊗	×		×			⊗	×				
Lin-lin	×			⊗		×	⊗	×		×			⊗	×		×			⊗					
													USA											
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quadratic	×			⊗		×	×	⊗	×	×		×			⊗			×		×	×	×		⊗
Absolute				⊗			×	×	×	⊗		×			×			⊗						⊗
Quad-quad				⊗			×		×	⊗	×	×		×	⊗			×	×			⊗		×
Lin-lin	×			⊗	×	×	×	×		⊗	×	×		⊗	×	×		×				⊗		

×: *model in the MCS at the 5% confidence level*

⊗: *model with the lowest loss in the MCS*

Asymmetry parameter: 0.85.

Table A.2: Forecasting Life expentancy at birth

	h=1						*h=5*						*h=10*						*h=1-10*					
France																								
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quadratic	×	×	×	×	×	⊗		×	×	×	×	⊗	×	×	×	×	×	⊗		×	×	×	×	⊗
Absolute	×	×	×	×	×	⊗	×	×	×	×	×	⊗	×	×		×	×	⊗	×	×	×	×	×	⊗
Quad-quad	×	×	×			⊗	×	×	×	×	×	⊗	×	×	×	×	×	⊗		×				⊗
Lin-lin	⊗	×	×			×	×	×				⊗	×	⊗				×		⊗				×
Italy																								
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quadratic	×		×	×	×	⊗			×	×	⊗	×	×		×	⊗	×	×			×	⊗	×	×
Absolute	×		×	×	×	⊗			⊗	×	×	×	×		×	⊗	×	×	×		×	⊗	×	×
Quad-quad	⊗	×	×	×	×	×	×		×	⊗	×	×	×	×	⊗	×	×	×	×	×	×	⊗	×	×
Lin-lin	×	⊗					×	×	⊗	×	×	×	⊗	×	×	×	×	×	⊗	×	×	×		×
USA																								
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quadratic	×	×	×			⊗	×	×	×	×	⊗	×	×	×	×	×	⊗	×	×	×	×	×	×	×
Absolute	×	×	×			⊗	×	×	×	⊗	×	×	×	×	×	⊗	×	×	×	×	×	×	×	×
Quad-quad	×	×	×			⊗	×	×	×	⊗	×	×	×	×	×	⊗	×	×	×	⊗		×	×	×
Lin-lin	×	×	×			⊗	×	×	×	⊗	×	×	×	×	×	⊗	×	×	×	⊗				

×: *model in the MCS at the 5% confidence level*

⊗: *model with the lowest loss in the MCS.*

Asymmetry parameter: 0.85.

Table A.3: Forecasting Life expectancy at 65

	h=1						*h=5*						*h=10*						*h=1-10*					
												France												
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quadratic	×		⊗	×	×	×		×	×	×		⊗		×				⊗		×	×			⊗
Absolute	×		×	×	×	⊗		×	×	×		⊗		×				⊗						⊗
Quad-quad	×	×	×	×	×	⊗		×	×	×		⊗	×	⊗		×		×	×	×	×	×		⊗
Lin-lin	×	×	×		×	⊗	×	⊗	×	×		×	×	⊗				×	×	×				⊗
												Italy												
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quadratic			⊗						×	⊗	×	×	×		×	⊗	×	×			×	⊗	×	×
Absolute			⊗				×		×	⊗	×	×	×		×	⊗	×	×	×		×	⊗	×	×
Quad-quad	×		⊗	×		×	×	×	⊗	×	×	×	×	×		×	×	×	×	×	⊗	×	×	×
Lin-lin	×	×	⊗			×	×	×	⊗	×		×	⊗	×	×	×		×	×	×	⊗	×		×
												USA												
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quadratic	×			⊗	×	×				⊗						×		⊗				⊗		×
Absolute	×		⊗	×	×	×				⊗		×				×		⊗				⊗		×
Quad-quad	×		⊗	×	×	×				⊗	×	×				×		⊗				⊗		×
Lin-lin	×	×	⊗	×	×	×				⊗	×	×			×	×		⊗			×	⊗		×

×: *model in the MCS at the 5% confidence level*

⊗: *model with the lowest loss in the MCS.*

Asymmetry parameter: 0.85.

Table A.4: Forecasting mortality rates in Italy with actual and counterfactual data: loss values for each model (row min = 1)

	h=1						*h=5*						*h=10*						*h=1-10*					
	actual data																							
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quad	8.3	11.8	17.1	**1**	8.9	2.4	1.8	2.4	3.1	**1**	2.7	1.4	1.4	2.1	2.8	**1**	3.2	1.5	1.5	1.8	2.3	**1**	2.1	1.3
Abs	2.0	3.0	3.4	**1**	2.5	1.2	2.0	3.0	3.4	**1**	2.5	1.2	1.2	1.6	1.8	**1**	2.1	1.3	1.2	1.6	1.8	**1**	1.8	1.2
Q-q$^+$	8.3	11.5	17.0	**1**	8.9	2.4	1.8	2.3	3.1	**1**	2.7	1.4	1.4	2.0	2.8	**1**	3.3	1.5	1.5	1.8	2.3	**1**	2.1	1.3
L-l$^+$	1.9	2.5	3.1	**1**	2.4	1.1	1.1	1.4	1.7	**1**	2.2	1.2	1.1	1.4	1.7	**1**	2.2	1.2	1.2	1.4	1.7	**1**	1.8	1.1
Q-q$^-$	8.4	13.7	18.0	**1**	8.9	2.6	1.8	2.7	3.2	**1**	2.7	1.5	1.5	2.0	2.3	**1**	2.1	1.3	1.5	2.0	2.3	**1**	2.1	1.3
L-l$^-$	2.4	5.9	5.2	**1**	2.9	1.9	1.5	2.6	2.3	**1**	1.6	1.7	1.4	2.4	2.1	**1**	1.6	1.5	1.4	2.4	2.1	**1**	1.6	1.5
	counterfactual data																							
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quad	9.5	1.4	**1**	8.7	1.05	1.5	7.0	5.1	3.7	4.2	**1**	5.8	6.1	2.9	2.2	2.6	**1**	3.7	2.1	2.6	2.2	1.5	**1**	2.9
Abs	2.7	1.3	1.2	2.5	**1**	1.3	2.1	2.1	1.9	1.6	**1**	2.1	2.2	1.8	1.7	1.5	**1**	1.9	1.6	1.8	1.6	1.4	**1**	1.8
Q-q$^+$	8.7	6.3	3.2	8.0	**1**	6.6	1.6	6.3	4.0	**1**	1.02	7.3	2.2	5.1	3.5	**1**	1.5	6.8	**1**	5.9	4.6	1.1	2.1	6.5
L-l$^+$	2.3	4.0	2.3	2.2	1	4.1	1.3	4.9	3.9	**1**	2.1	5.3	1.3	3.7	3.0	**1**	2.2	4.2	**1**	3.6	2.9	1.1	1.9	3.8
Q-q$^-$	21.6	4.1	3.7	17.4	**1**	4.1	1.8	8.5	7.5	**1**	2.8	9.0	14.6	2.0	2.1	5.9	**1**	2.1	7.2	2.1	2.0	4.7	**1**	2.1
L-l$^-$	3.7	1.02	1.3	3.6	1.4	**1**	3.9	1.8	2.0	2.9	**1**	1.7	4.8	2.0	2.2	3.0	**1**	1.8	3.4	1.6	1.8	2.6	**1**	1.5

Counterfactual data: modified to have in 2006 excess mortality in age groups 65+ on a scale comparable to that caused in 2020 by Covid-19.

Abbreviations Quad: Quadratic; Abs: absolute value; Q-q: Quad-quad; L-l: Lin-lin; "+" asymmetry parameter 0.85, "−" asymmetry parameter 0.15.

Table A.5: Forecasting life expectancy at 65 in Italy with actual and counterfactual data: loss values for each model (row min = 1).

	h=1						h=5					
	actual data											
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quad	2.9E+02	1.0E+03	**1**	7.5E+02	1.09E+02	40.9	**1**	2.8E+03	1.8E+02	4.7E+02	5.5E+03	1.9E+02
Abs	17.0	31.7	**1**	27.4	33.0	6.4	**1**	53.2	13.3	21.6	74.1	13.8
Q-q$^+$	1.6E+03	1.0E+03	**1**	4.3E+03	6.2E+03	2.3E+02	**1**	2.8E+03	1.8E+02	2.6E+03	3.1E+03	1.9E+02
L-l$^+$	96.3	31.7	**1**	1.6E+02	1.9E+02	36.2	**1**	53.2	13.3	1.2E+02	4.2E+02	13.8
Q-q$^-$	51.0	1.0E+03	**1**	1.3E+02	1.9E+02	7.2	**1**	2.8E+03	1.8E+02	82.4	9.7E+02	1.9E+02
L-l$^-$	3.0	31.7	**1**	4.8	5.8	1.1	**1**	53.2	13.3	3.8	13.1	13.8
	counterfactual data											
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quad	1.7	1.7	3.7	1.5	**1**	1.6	12.0	10.5	18.6	8.3	**1**	7.7
Abs	1.3	1.3	1.9	1.2	**1**	1.3	3.5	3.2	4.3	2.9	**1**	2.8
Q-q$^+$	1.7	1.7	3.7	1.5	**1**	1.6	12.0	10.5	18.6	8.3	**1**	7.7
L-l$^+$	1.3	1.3	1.9	1.2	**1**	1.3	3.5	3.2	4.3	2.9	**1**	2.8
Q-q$^-$	3.1	3.0	6.4	2.5	**1**	2.5	90.4	69.0	1.6E+02	43.7	**1**	37.5
L-l$^-$	1.3	1.3	1.9	1.2	**1**	1.3	3.5	3.2	4.3	2.9	**1**	2.8

Counterfactual data: modified to have in 2006 excess mortality in age groups 65+ on a scale comparable to that caused in 2020 by Covid-19.

Abbreviations Quad: Quadratic; Abs: absolute value; Q-q: Quad-quad; L-l: Lin-lin; "+" asymmetry parameter 0.85, "−" asymmetry parameter 0.15.

Table A.6: Forecasting life expectancy at 65 in Italy with actual and counterfactual data: loss values for each model (row min = 1).

	h=10						*h=1-10*					
	actual data											
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quad	1.5	17.3	1.8	**1**	57.2	3.0	1.1	3.1	**1**	2.2	11.2	10.5
Abs	1.2	4.2	1.3	**1**	7.6	1.7	1.1	2.0	**1**	1.4	3.8	1.1
$Q\text{-}q^+$	**1**	11.7	1.2	3.8	2.2E+02	2.0	1.5	1.02	1.1	3.8	19.6	**1**
$L\text{-}l^+$	**1**	3.4	1.1	4.7	35.3	1.4	1.9	**1**	1.3	3.4	9.2	1.4
$Q\text{-}q^-$	8.3	98.0	10.3	**1**	57.2	16.8	1.1	8.0	1.5	**1**	5.1	1.7
$L\text{-}l^-$	6.9	23.6	7.6	**1**	7.6	9.8	1.8	6.6	2.2	**1**	2.3	2.4
	counterfactual data											
Loss	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*	*RW*	*GFM*	*LC*	*LM*	*BMS*	*HU*
Quad	2.1E+02	1.6E+02	2.6E+02	1.1E+02	**1**	98.6	5.5	4.6	8.2	3.6	**1**	3.3
Abs	14.7	12.8	16.0	10.7	**1**	9.9	2.5	2.3	3.2	2.0	**1**	1.9
$Q\text{-}q^+$	37.9	28.9	45.1	20.1	**1**	17.4	1.9	1.6	2.9	1.3	**1**	1.2
$L\text{-}l^+$	2.6	2.3	2.8	1.9	**1**	1.8	1.3	1.2	1.6	1.04	1.2	**1**
$Q\text{-}q^-$	1.2E+03	9.3E+02	1.5E+03	6.5E+02	**1**	5.6E+03	8.2	6.9	12.2	5.4	**1**	5.0
$L\text{-}l^-$	83.1	72.5	90.6	60.5	**1**	56.3	3.3	3.1	4.2	2.7	**1**	2.6

Counterfactual data: modified to have in 2006 excess mortality in age groups 65+ on a scale comparable to that caused in 2020 by Covid-19.

Abbreviations Quad: Quadratic; Abs: absolute value; Q-q: Quad-quad; L-l: Lin-lin; "+" asymmetry parameter 0.85, "−" asymmetry parameter 0.15.