



Forecasting life expectancy in an international context

Tiziana Torri*, James W. Vaupel

Max Planck Institute for Demographic Research, Konrad-Zuse-Str. 1 18057 Rostock, Germany

ARTICLE INFO

Keywords:

Best-practice levels
Geometric Brownian motion
Geometric mean-reverting process
ARIMA models
Monte Carlo simulation

ABSTRACT

Over the past two centuries, the life expectancy has more than doubled in many countries, for both males and females. The levels of the countries with the highest life expectancies have risen almost linearly. We exploit this regularity by using the classic univariate ARIMA model to forecast future levels of best-practice life expectancy. We then compare two alternative stochastic models for forecasting the gap between the best-practice level and life expectancy in a particular population. One of our approaches is based on the concept of discrete geometric Brownian motion; our other approach relies on a discrete model of geometric mean-reverting processes. A key advantage of our strategy is that the life expectancies forecast for different countries are positively correlated because of their tie to the forecast best-practice line. We provide illustrations based on Italian and US data.

© 2011 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

The life expectancy in the countries doing best has been rising at a remarkably steady pace – about three months per year – since 1840 (Oeppen & Vaupel, 2002). The analysis of a somewhat different, longer time series since 1750 suggests four segments of linear increase, but in this alternative analysis the rise since 1900 is also close to linear (Vallin & Meslé, 2009). We exploit these findings to apply the classic univariate ARIMA model to data since 1900 in order to forecast best-practice life expectancy.

The life expectancy levels in different countries tend to be positively correlated. The life expectancy in countries close to the record level tends to increase by about three months per year, though it is sometimes a bit slower. The life expectancy in countries below the best-practice level tends to increase faster, as the countries catch up. Forecasts of a population's life expectancy based solely on historical data for that country fail to make use of these cross-national correlations. Here we forecast both the best-practice level of life expectancy and the gap between a population's level and this record level. We thus put forecasting in an international context and take advantage of empirical regularities across countries.

Because there is no compelling theoretical reason to think that life expectancies are approaching a limit and because the empirical evidence does not show any deceleration of the increase in best-practice life expectancy (Oeppen & Vaupel, 2002; Vaupel, 2010; Vaupel et al., 1998), we simply extrapolate from the past to predict the future (Wilmoth, 1998). We use ARIMA models (Box & Jenkins, 1976, p. 93) to forecast the best-practice life expectancy, and two alternative models to forecast the gap between a country's life expectancy and the record level.

Important research on forecasting life expectancies includes that of Alho and Spencer (2005), Booth (2006), Keilman (2001), Lee and Carter (1992) and Li and Lee (2005). Assuming the convergence of future mortality, Li and Lee (2005) modified the original Lee–Carter model to generate coherent mortality forecasts for a group of populations. They take advantage of common features to produce converging mortality forecasts. Although in a completely different setting, we also make the assumption of converging mortality, with the difference being the limit imposed on mortality via the best-practice level.

Two approaches to forecasting life expectancy that are more closely related to ours were proposed by Andreev and Vaupel (2006) and Lee (2006). Andreev and Vaupel (2006) assume that best-practice life expectancy will rise linearly and non-stochastically and that the current gap between a population's life expectancy and the record

* Corresponding author. Tel.: +33 146987116.
E-mail address: ttorri@scor.com (T. Torri).

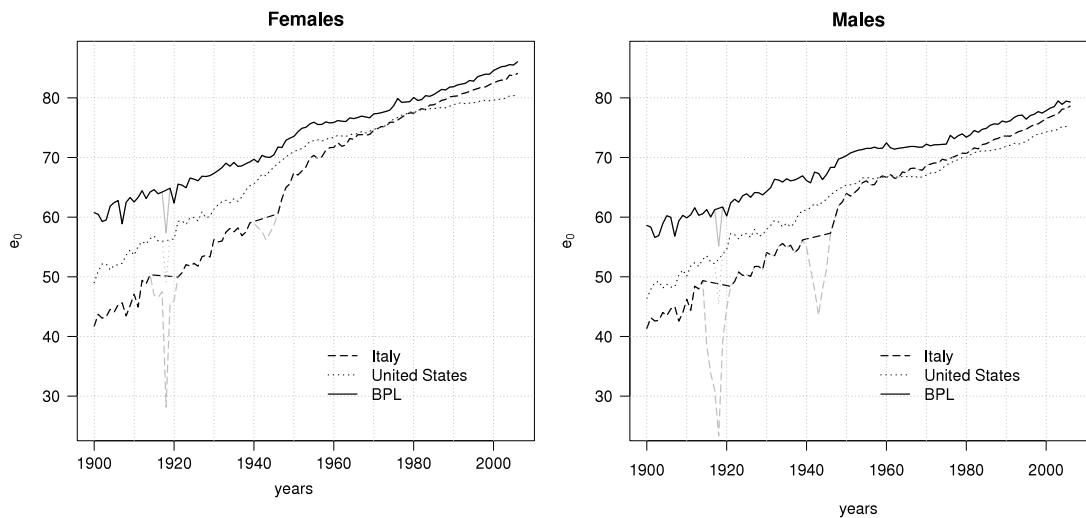


Fig. 1. Period life expectancies at birth, e_0 , for Italy and the United States, together with the best-practice level (BPL), 1900–2006. Data for females and males. The transient shocks (in grey) were excluded from our analysis because our focus was on long-term trends.

level will remain constant. This is straightforward, but too simplistic. Lee (2006) models the changes in a population's life expectancy in a stochastic setting, assuming that the mean changes are a linear function of the gap. In this case, the model suffers from a conceptual drawback: stochastic future life expectancies can exceed the best-practice levels. This assertion is further analyzed in Section 3.2. In this article, we take advantage of models from time series analysis to overcome these limitations.

A new range of applications of life expectancy forecasts is represented by mortality-linked securities, new financial products designed to transfer the longevity risk on financial markets. Such products are built on mortality indexes, such as future life expectancies.

The next section presents the data for Italy and the United States which we used in our analysis. Section 3 describes the models fitted to the best-practice levels and the gaps. An application of the models to the data is presented in Section 4, followed by a critical discussion of the methods and conclusions.

2. Data description

Our illustrative forecasts were based on Italian and US data, taken from the Human Mortality Database (HMD, 2010) and the Berkeley Mortality Database (BMD, 1995). Italian mortality data in the HMD cover the period from 1872 to 2006. With regard to the United States, the HMD only includes data between 1933 and 2006. However, death rates by single years of age, prepared by the Office of the Chief Actuary in the US Social Security Administration, cover the period from 1900 to 1995. By combining these two sources, we are able to reconstruct mortality data from 1900 to 2006, for both countries. Life tables and the corresponding life expectancies at birth have been calculated by the authors to guarantee consistency in methodologies.

Best-practice life expectancies from 1900 to 2006 were derived from data on the 37 countries in the HMD. A plot of life expectancies at birth for males and females in

Italy and the United States, together with the best-practice values, is provided in Fig. 1. The steady increase in life expectancies at birth that is observed in the first half of the 20th century is an effect of the reduction in infant and child mortality. This steady increase is interrupted by obvious troughs during the First and Second World Wars and the Spanish Flu pandemic, which increased adult general mortality. However, because we are interested in forecasting long-term trends, not transient shocks, these exceptional events are excluded from our analysis. Their values are replaced with those obtained from the interpolation between adjacent values. In the second half of the 20th century, improvements in survival after the age of 65 largely affected the improvements in life expectancies.

The increase in the female best-practice level, measured by the slope of the regression line, is 0.25, about 3 months per year (Oeppen & Vaupel, 2002). This progress is outperformed by the Italian and US female life expectancies, with slopes equal to 0.43 and 0.30, respectively. The same results are observed for male life expectancies, with Italy and the United States having slopes of 0.38 and 0.25 respectively, versus the 0.20 observed for the best-practice level.

As is shown in Fig. 1, the female life expectancy in Italy tends to converge towards the best-practice level, maintaining a constant distance from it in recent decades. In contrast, the female life expectancy in the United States first moves closer to, and then lags behind, the best-practice level. In 1900, female life expectancy in the United States was 49 years, and lagged about 10 years behind the world leader, New Zealand (non-Maori) (Oeppen & Vaupel, 2002). Later on, the United States managed to catch up, recording a gap of only 2 years in 1950. Since 1980, the gains in life expectancy have slowed, and the lag behind the record holder has grown. In the year 2000, the life expectancy for US females was 79.7 years, about 5 years lower than the highest level of life expectancy, in Japan (Oeppen & Vaupel, 2002). The plot of life expectancies for

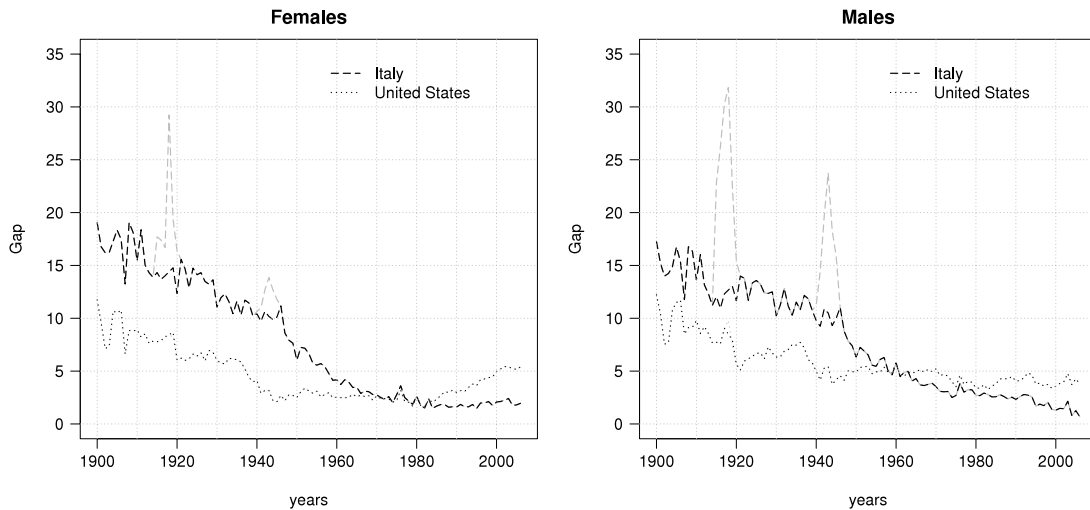


Fig. 2. Gap between the best-practice level and country-specific life expectancies at birth, for Italy and the United States, 1900–2006. Data for females and males. The transient shocks (in grey) were excluded from our analysis because our focus was on long-term trends.

Italian males shows a converging trend towards the best-practice level. The corresponding values for US males show a converging trend towards the best-practice levels until 1950 and a gap that is maintained constantly thereafter.

To gain a better understanding of the relationship between the best-practice level and country-specific life expectancies, we calculated the gap between the two curves. Fig. 2 plots the gaps for both countries and both sexes. We observe big reductions in the gap in periods characterized by high levels of the gap, whereas small gains were observed when the gap was small. This behavior is less clear for the United States, which is characterized by a leveling-off of the male gap and a widening of the female gap (Fig. 2).

3. Methods

We forecast the life expectancies in individual countries by forecasting the best-practice level and the gap between the national performance and the best-practice level. Separate calculations were performed for the two sexes. By taking advantage of the pronounced linear trend observed in best-practice levels, and assuming the persistence of the trend observed in the past, we extrapolated the stochastic process using the classic ARIMA model in time series analysis.

The situation becomes more complicated when we wish to model the gap. If we consider best-practice levels to impose an upper bound on the country-specific life expectancy, then zero is a lower bound for the gap. We have to model the gap in such a way that it never falls below the value of zero. To solve this problem, we work with the logarithmic transformation of the variables of interest.

Here, we describe the behavior of the gaps with two alternative discrete models: the geometric Brownian motion and the geometric mean-reverting process. Empirical estimates of the uncertainty associated with the future values of the life expectancy are provided by a Monte Carlo simulation of the future paths of the gap and the best-practice levels.

3.1. The ARIMA model

Classic time series analysis is based on the theory of stationary stochastic processes.

Processes exhibiting non-stationary behaviors are included in the class of autoregressive integrated moving average (ARIMA) models (Box & Jenkins, 1976, p. 93, and Brockwell & Davis, 1996, p. 177). In general notation, we have an ARIMA(p, d, q) model, where p is the order of the autoregressive process, d indicates the order of integration, namely the number of times that the series must be differenced in order to make it stationary, and q is the order of the moving average process. The ARIMA(p, d, q) model for a stochastic process y_t is given by:

$$\Delta^d y_t = \delta + \sum_{i=1}^p \phi_i \Delta^d y_{t-i} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}, \tag{1}$$

where the constant parameter δ is the drift term, indicating the average change in the series over time; ϕ_i are the parameters of the autoregressive part, and θ_j are the parameters of the moving average part. The symbol Δ^d is the difference operator of order d , and ϵ_t is a sequence of independent and identically distributed random variables with mean zero and variance σ^2 . The model selection strategy developed by Box and Jenkins (1976, p. 333) is characterized by three steps: model identification, model estimation, and diagnostic checking.

3.2. The discrete geometric Brownian motion

The discrete geometric Brownian motion (GBM) is a discrete-time stochastic process, such that the mean and variability of the changes in the process at time t are proportional to the value of the process in t .

In general, given a discrete stochastic process G_t following a GBM, the stochastic difference equation is equal to:

$$G_{t+1} - G_t = \mu G_t + \sigma G_t \epsilon_t, \tag{2}$$

with ϵ_t being white noise, and where the constants μ and σ are the drift and volatility of G_t , respectively. In more detail,

the first term on the right-hand-side of the equation refers to the expected variation of G_t in the unitary time interval given the drift term, and the second term represents the variation of G_t .

Working with the logarithmic transformation of the random variable G_t makes calculations easier, guaranteeing that the stochastic variable G_t will be positive when we perform forecasts. We work with the new function $F_t = \ln(G_t)$; we substitute it in Eq. (2), and use the approximation $\exp(x) \simeq 1 + x$ to obtain the following solution for the variable F_{t+1} :

$$F_{t+1} = F_t + \mu + \sigma \varepsilon_t, \quad (3)$$

which is the formula of a random walk with drift, corresponding to an ARIMA(0, 1, 0), with a non-zero mean.

Note that F_t follows a normal distribution and G_t follows a log-normal distribution.

The main differences between the model just described and those proposed by Lee (2006) are as follows. For the sake of simplicity, we only consider one of Lee's six models here, all of which describe the relationship between changes in the country-specific life expectancy and the best-practice level. The discrete version of Lee's model is:

$$\begin{aligned} e_{t+1}^0 - e_t^0 &= \phi + \alpha(e_t^{\text{Best}} - e_t^0) + \varepsilon_t \\ &= \phi + \alpha G_t + \varepsilon_t. \end{aligned} \quad (4)$$

The model assumes that a country-specific life expectancy, e_t^0 , increases at some constant rate, ϕ , and moves a proportion α of the gap towards the best-practice level in the unitary time interval. The disturbances, ε_t , are assumed to be white noise or autocorrelated of order one.

To compare the GBM with the model proposed by Lee, in this context we describe the best-practice level using a straight line and show that changes in the life expectancy over time are given by the following equation:

$$\begin{aligned} e_{t+1}^0 - e_t^0 &= (e_{t+1}^{\text{Best}} - e_t^{\text{Best}}) - (G_{t+1} - G_t) \\ &= \beta - (\mu G_t + \sigma G_t \varepsilon_t) \\ &= (\beta - \mu G_t) - \sigma G_t \varepsilon_t, \end{aligned} \quad (5)$$

where β is the slope of the line interpolating the best-practice level: its constant increase over time. The first set of parentheses on the right-hand-side of the last equation corresponds to the first two terms of Eq. (4), while there is no correspondence between the two models' random components. The GBM assumes variability of the changes in life expectancies which is proportional to the gap, G_t ; while the Lee model assumes independent and identically distributed random disturbances. Moreover, the GBM implies a smaller variation of the gap when it is close to zero, preventing future values of the life expectancy from rising above the best-practice level. This is not the case for Lee's model, which can generate paths of future life expectancy which are above the forecast best-practice level.

3.3. The discrete geometric mean-reverting process

The discrete geometric mean-reverting process (MRP) is a discrete-time stochastic process which assumes that

some forces exist that act on the process when it is far from an equilibrium level, leading it back to the equilibrium. The process fluctuates around the reversion level, and hence is characterized by non-explosive behavior, with a limited long-term variance.

The discrete-time stochastic process G_t following a MRP can be represented by the following stochastic difference equation:

$$G_{t+1} - G_t = \eta G_t [m - \ln(G_t)] + \sigma G_t \varepsilon_t, \quad (6)$$

where ε_t is white noise, m is the long-run equilibrium level of the logarithm of the gap G_t , and η indicates its speed of reversion. The constant σ represents the volatility of G_t .

Here again, it is preferable to work with the logarithmic transformation of the random variable G_t , in order to make the calculations easier and to guarantee that the stochastic variable G_t will be positive. We work with the new function $F_t = \ln(G_t)$; we substitute it in Eq. (6) and use the approximation $\exp(x) \simeq 1 + x$ to obtain the following solution for the variable F_{t+1} :

$$F_{t+1} = (1 - \eta)F_t + \eta m + \sigma \varepsilon_t, \quad (7)$$

which is an AR(1) with a non-zero mean for $|1 - \eta| < 1$.

Note that F_t follows a normal distribution and G_t follows a log-normal distribution.

4. Results

We apply the methods presented in Section 3 to the data described in Section 2, namely the period life expectancy at birth for the United States and Italy, and the best-practice levels from 1900 to 2006. To identify the most suitable model for our data, we begin testing for the stationarity of the series, checking for the presence of a unit-root when there is a trend in the series. We apply the Dickey-Fuller test (Dickey & Fuller, 1979).

We first analyze the best-practice levels, and conclude that we cannot reject the null hypothesis of a linear trend in the data and the non-stationarity of the series. Similar results are obtained from the analysis of the logarithm of the different life expectancies at birth. The Italian data, both male and female, are characterized by the presence of a linear trend and non-stationarity. The US female data are non-stationary as well, but are not characterized by any linear trend. US male life expectancies at birth are characterized by a stationary behavior around a linear trend. The results of the test provide information as to which model best represents the data: a random walk with drift in the presence of a unit-root and a linear trend, and an AR(1) when we reject the null hypothesis of a unit-root. Although the two models presented in Sections 3.2 and 3.3 do not always correspond to the characteristics of the data, we apply them to our data. We strongly believe in the theoretical framework of the models, and moreover, it is well known that the best model for fitting the data is not always also the best model for forecasting them. We judge the forecasting performances of the models later, in Section 4.4.

4.1. Forecasting best-practice levels

The pronounced linear trend observed in best-practice levels helped us in forecasting. We assumed the persistence of the past observed trends and extrapolated the

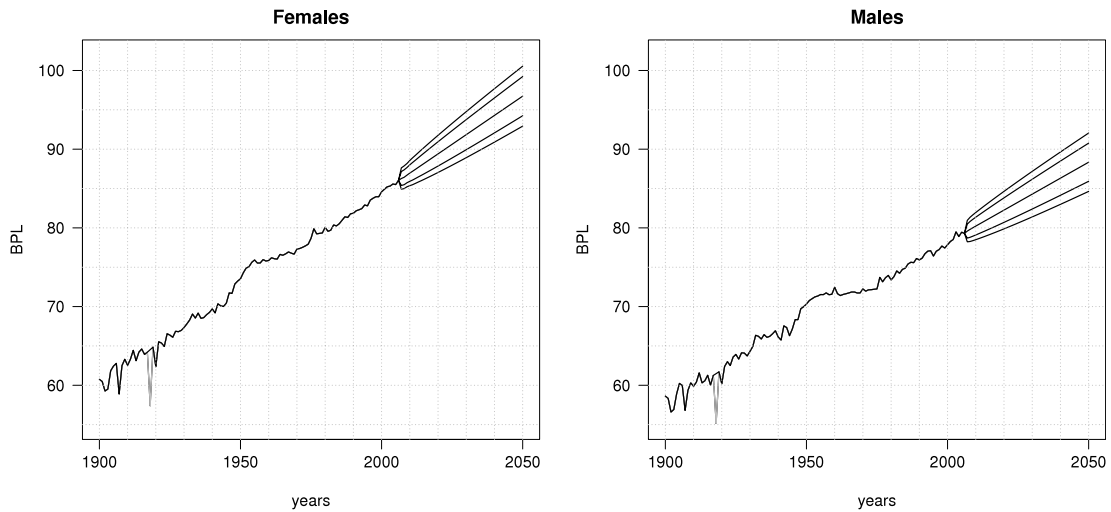


Fig. 3. Actual and forecast best-practice level (BPL) using an ARIMA model, together with the associated 80% and 95% prediction intervals, 1900–2050. Data for females and males. The transient shocks (in grey) were excluded from our analysis because our focus is on long-term trends.

Table 1
Estimated parameters of the ARIMA model selected for the best-practice levels on the data from the period 1900–2006, by sex.

	Females	Males
	ARIMA(2, 1, 1)	ARIMA(1, 1, 1)
δ	0.244	0.203
ϕ_1	-0.246	0.235
ϕ_2	-0.294	-
θ_1	-0.377	-0.710

Table 2
Forecasts of the best-practice level (BPL) in 2050 produced by the ARIMA model, with 80% and 95% prediction intervals (PI). The model was evaluated on data from the period 1900–2006, by sex.

	Females	Males
BPL	96.59	88.38
80% PI	(94.10–99.09)	(85.96–90.81)
95% PI	(92.78–100.41)	(84.67–92.09)

stochastic process using the classic ARIMA models (see Section 3.1). A further analysis shows that the ARIMA(2, 1, 1) for females and the ARIMA(1, 1, 1) for males are the most suitable models for describing the data. We can write the models as follows:

$$\Delta e_t^{\text{Best}} = \delta + \phi_1 \Delta e_{t-1}^{\text{Best}} + \phi_2 \Delta e_{t-2}^{\text{Best}} + \epsilon_t + \theta_1 \epsilon_{t-1},$$

$$\Delta e_t^{\text{Best}} = \delta + \phi_1 \Delta e_{t-1}^{\text{Best}} + \epsilon_t + \theta_1 \epsilon_{t-1},$$

where the first equation refers to females and the second to males. The values of the estimated parameters are given in Table 1.

For the estimation of the future values of the best-practice levels, we considered the same models and the parameters obtained by fitting the ARIMA model to the data.

The estimated future values of the record life expectancy in the year 2050, together with 80% and 95% prediction intervals, are provided in Table 2 and plotted in Fig. 3. We forecast that, in the year 2050, there will be a country enjoying a life expectancy at birth of 97 years for

females and 88 years for males. If we look at the results of the 95% prediction intervals, we find that a value of the life expectancy higher than 100 years is possible for females. For males, life expectancies could reach an approximate value of 92 years at best.

4.2. Forecasting the gap using the discrete geometric Brownian motion

We now aim to model and forecast the second component of our approach: the gap (G_t) between the national life expectancy and the best-practice level. Specifically, we use the GBM model presented in Section 3.2. It assumes that the changes in the variable G_t have a mean and variability that are proportional to the value of the variable G_t itself. Such behavior can be seen in the actual trends of the gaps: a country shows a slow increase when it is close to the best-practice level and a faster increase when it is relatively distant from it. On the one hand, it makes sense to assume that a country in an unfavorable position has a great potential to improve its mortality, or that its mortality may improve relatively quickly by profiting from the improvements which other countries have already achieved. On the other hand, for those countries with a small gap – namely those which are performing almost as well as the best-practice country – the model assumes small changes occurring on average. It is very difficult to achieve further improvements in mortality when the life expectancy is already very close to the best-practice level.

By including stochasticity in the process, we allow the gap to diverge from its median value. The model assumes that the variability of the process is proportional to the gap, meaning that the bigger the gap, the wider the range in which the process can move.

As has already been noted, we work with the logarithm of the gap and estimate Eq. (3) via maximum-likelihood, obtaining the parameters μ and σ presented in Table 3. The estimated average change, μ , of the process F_t is negative for all of the data, representing a decreasing trend of the

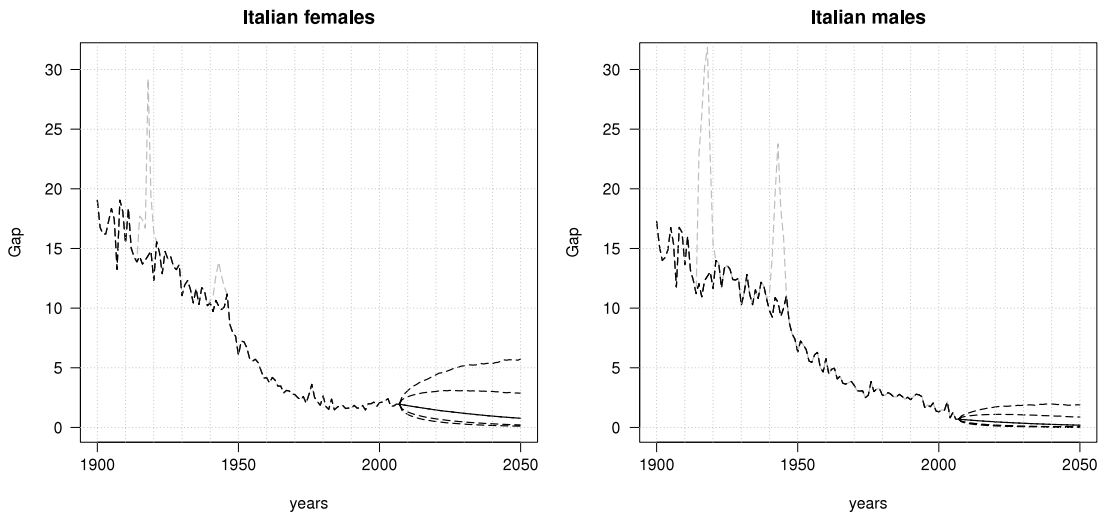


Fig. 4. Actual and forecast gaps using the GBM, with the associated 80% and 95% prediction intervals. Italy, 1900–2050; data for females and males. The transient shocks (in grey) were excluded from our analysis because our focus was on long-term trends.



Fig. 5. Actual and forecast gaps using the GBM, with the associated 80% and 95% prediction intervals. United States, 1900–2050; data for females and males. The transient shocks (in grey) were excluded from our analysis because our focus was on long-term trends.

Table 3

Estimated parameters for the gap from the GBM. The data are for Italy and the United States, 1900–2006, for both sexes.

	Females		Males	
	Italy	US	Italy	US
μ	−0.023	−0.008	−0.029	−0.010
σ	0.154	0.156	0.195	0.129

series of the logarithm of the gap. A stronger reduction is estimated for the Italian data.

Having estimated the parameters of the model, we can easily simulate the future values of the logarithm of the gap via Eq. (3). We then apply the exponential function to the results and obtain the future value G_t .

The distributions of the future values of the gaps are plotted in Figs. 4 and 5. We predict an overall decreasing trend of the median value of the gaps. The prediction intervals are asymmetric and wider in the upper intervals.

This is not surprising, as we assumed that the gap is log-normally distributed. The model also produced implausible prediction intervals for US females. This seemingly erroneous result is due mainly to the high value of the gap observed in the recent years. Indeed, the model assumes that the variability of the gap is proportional to the value of the gap. This, together with the hypothesis of log-normality of the gap, yields overly wide 95% prediction intervals. This result calls into question the ability of the model to predict the uncertainty for data with high current values of the gap.

The future median values of the gap in the year 2050, together with the corresponding 80% and 95% prediction intervals, are given in the top panel of Table 5. Very low values are expected for the Italian gap, whereas those for the US gap are higher. Nevertheless, we forecast a reduction of the US gap with respect to current values.

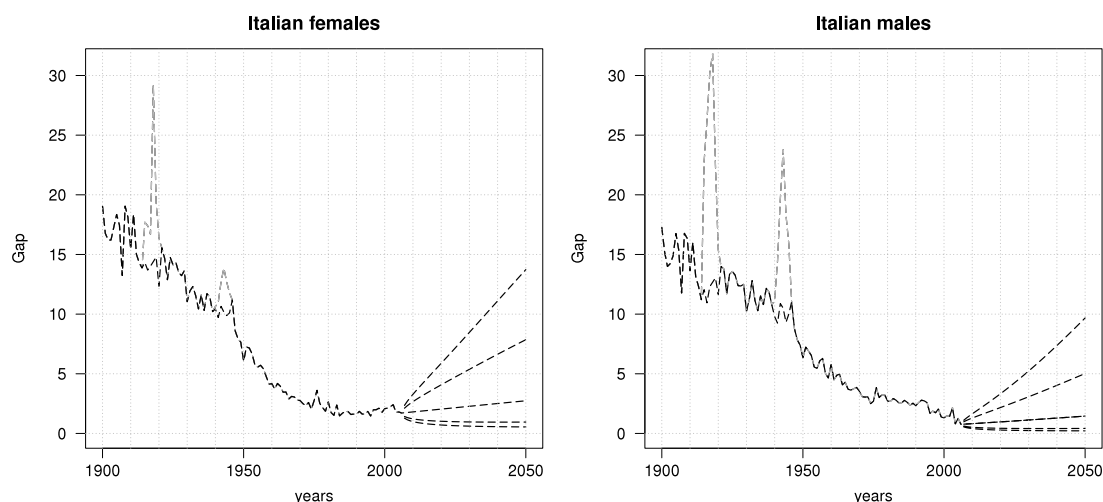


Fig. 6. Actual and forecast gaps using the MRP, together with the associated 80% and 95% prediction intervals. Italy, 1900–2050; data for females and males. The transient shocks (in grey) were excluded from our analysis because our focus was on long-term trends.

4.3. Forecasting the gap using the discrete geometric mean-reverting process

The MRP described in Eq. (6) differs from the GBM in the drift term and the variability. In the MRP, the drift term is positive if the current level of the variable G_t is lower than the equilibrium level m , and vice versa. In other words, the process is attracted towards the equilibrium level m , with an intensity that is proportional to the distance between the process and the equilibrium level.

We can justify the fluctuation of the gap around the equilibrium level in the following way. It can be assumed that the government of a country is in a state of ‘alert’ when the gap is large, and above the equilibrium level. In contrast, the government of a country with a small gap, below the equilibrium level, can be more ‘relaxed’ and less concerned about its performance with regard to mortality. The alternation of periods with increased and reduced levels of attention to mortality justifies the assumed fluctuation of the gap around an equilibrium level. The equilibrium level differs between countries, reflecting the different potentials of countries with respect to improvements in mortality.

In addition, the MRP has a limited long-term variance. The assumption of a limited long-term variance also seems reasonable in a globalized world where knowledge spreads fast, and the mortality rates of different countries tend to converge and are less likely to diverge. We hypothesized that by assuming a limited long-term variance of the process, the prediction intervals would be narrower than for the GBM model.

We work with the logarithm of the gap and estimate Eq. (7) via maximum-likelihood in order to obtain the parameter estimation presented in Table 4. Using the estimated parameters, we simulate the future values of the logarithm of the gap, F_t , using Eq. (7), and then exponentiate them to obtain future values of G_t .

The distributions of future gaps are plotted in Figs. 6 and 7. The forecast long-run equilibrium levels seem to be close to the last observed value of the gap. The presence

Table 4

Estimated parameters from the MRP model of the gap. Data for Italy and the United States, 1900–2006, for both sexes.

	Females		Males	
	Italy	US	Italy	US
m	1.736	1.616	1.459	1.744
η	0.011	0.037	0.010	0.055
σ	0.155	0.157	0.179	0.123

Table 5

Forecast gap in 2050 using the GBM and MRP models, together with 80% and 95% prediction intervals (PI). The models are evaluated on the data period 1900–2006, for both sexes.

GBM	Italy	US
	Females	
Gap	0.63	3.67
80% PI	(0.17–2.33)	(0.98–13.74)
95% PI	(0.08–4.66)	(0.49–27.65)
Males		
Gap	0.22	2.44
80% PI	(0.04–0.95)	(0.84–6.89)
95% PI	(0.02–2.10)	(0.47–12.04)
MRP	Italy	US
	Females	
Gap	2.74	5.05
80% PI	(0.95–7.85)	(2.49–10.24)
95% PI	(0.54–13.73)	(1.72–14.89)
Males		
Gap	1.45	5.54
80% PI	(0.42–5.02)	(3.34–8.95)
95% PI	(0.21–9.67)	(2.67–11.53)

of a reversion level characterizing the process is especially visible in the US plots, an effect of the higher values of the estimated parameters η . It is clear that by assuming a

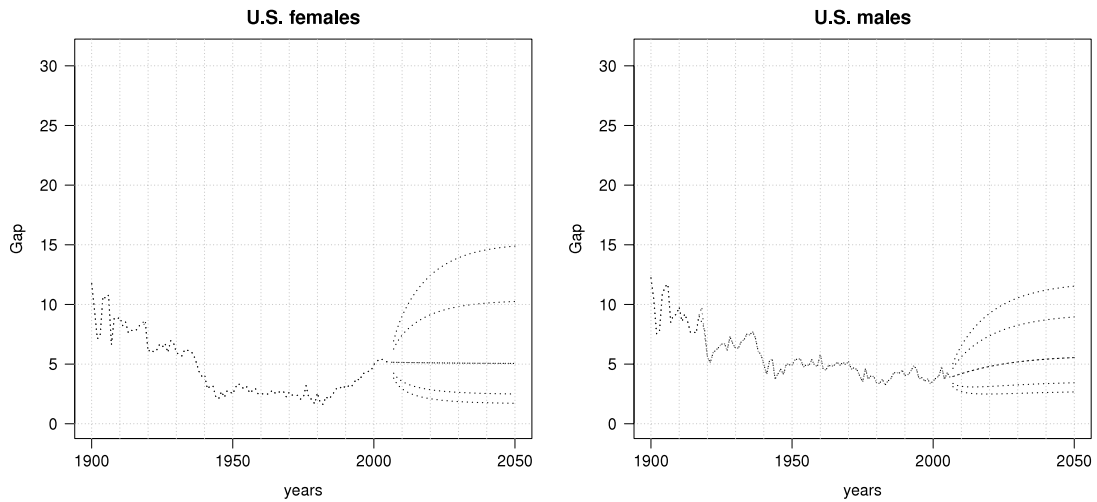


Fig. 7. Actual and forecast gaps using the MRP, together with the associated 80% and 95% prediction intervals. United States, 1900–2050; data for females and males. The transient shocks (in grey) were excluded from our analysis because our focus was on long-term trends.

limited long-term variance of the process, we prevented the gap from assuming overly extreme values.

The forecast values of the gap in the year 2050, together with the corresponding 80% and 95% prediction intervals, are given in Table 5. The MRP always returns higher values of the gap than the GBM, reaching a difference of 3.1 years in US males. The more pronounced reduction in the gap returned by the GBM corresponds to the prediction of faster improvements in the country-specific life expectancy than in the best-practice level. The MRP's assumption of the convergence of the gap to a long-term trend, produces future values of the life expectancy which improve at a speed equal to that of the best-practice level.

More plausible values of the prediction intervals are returned by the MRP, solving the problem of the GBM for US females.

4.4. Forecasting life expectancy at birth

We estimated the distributions of future values of country-specific life expectancies at birth by combining simulated future paths of the gap and the best-practice level obtained via Monte Carlo simulation. The innovations were generated from a multi-normal distribution with mean zero and covariance matrix Σ , evaluated on the residuals of the fitted best-practice levels and gaps. We utilize the covariance matrix to provide information on the tendency of the two processes to move together, with implications for the uncertainty of the forecast life expectancy. We would obtain different prediction intervals if we were ignoring this information, implicitly assuming a null correlation between best-practice levels and the gaps. The covariances between the gaps of different countries are not influential for the analysis. Using a large number of simulations, we obtain a good approximation of the analytical probability distribution of the variables of interest G_t and the best-practice line.

The estimated future life expectancies for the year 2050, along with the corresponding 80% and 95% prediction intervals, are presented in Tables 6 and 7 and shown in

Table 6

Observed life expectancies at birth in 2006, $e_0(2006)$, and forecast life expectancies at birth in 2050, $e_0(2050)$, using the GBM and MRP models, with 80% and 95% prediction intervals (PI), produced by the United Nations, the US Census Bureau and within the UPE project. Data for Italy and the United States. Females.

Females	Italy	US
$e_0(2006)$	84.09	80.65
GBM		
$e_0(2050)$	96.14	92.88
80% PI	(93.53–98.65)	(84.83–95.81)
95% PI	(91.81–99.98)	(72.12–97.13)
MRP		
$e_0(2050)$	93.99	92.01
80% PI	(89.58–96.85)	(88.25–94.70)
95% PI	(84.59–98.27)	(84.86–95.96)
UN		
$e_0(2045-50)$	88.4	85.8
US CB		
$e_0(2050)$	86.6	85.3
UPE		
$e_0(2050)$	89.8	–

Figs. 8 and 9. Both forecasting approaches for the gap are presented. The results show similar median values, reflecting the predominance of the effect of the best-practice line on the overall forecast of life expectancy more than that of the gap. In 2050, Italian females are expected to have a life expectancy of approximately 96 years according to the GBM and 94 years according to the MRP. Starting from a life expectancy of 84 in 2006, we projected a 10- or 12-year increase over the next forty-four years. The corresponding value for males is approximately 87 years according to both models, starting from a value of 79 years in 2006. The United States is expected to have a life expectancy of approximately 92 years for females, with both models. Males reach a life expectancy of 85 years with the GBM and 83 years with the MRP. We estimated an increase of 12 years for females and 10 or 8 years for males.

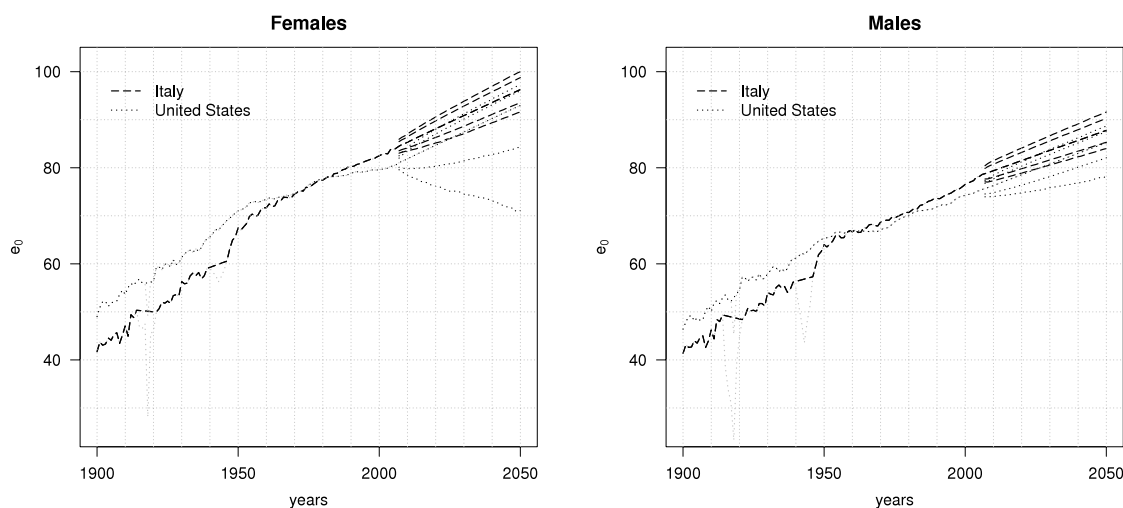


Fig. 8. Actual and forecast life expectancies at birth, e_0 , using a GBM, together with the associated 80% and 95% prediction intervals for Italy and the United States, 1900–2050. Data for females and males. The transient shocks (in grey) were excluded from our analysis because our focus was on long-term trends.

Table 7

Observed life expectancies at birth in 2006, $e_0(2006)$, and forecast life expectancies at birth in 2050, $e_0(2050)$, using the GBM and MRP model, with 80% and 95% prediction intervals (PI), produced by the United Nations, the US Census Bureau and within the UPE project. Data for Italy and the United States. Males.

Males	Italy	US
$e_0(2006)$	78.62	75.50
GBM		
$e_0(2050)$	87.85	85.22
80% PI	(85.35–90.31)	(82.19–87.59)
95% PI	(84.01–91.61)	(77.94–88.63)
MRP		
$e_0(2050)$	86.28	82.93
80% PI	(83.04–88.92)	(80.27–85.36)
95% PI	(79.56–90.28)	(78.64–86.62)
UN		
$e_0(2045–50)$	82.5	80.8
US CB		
$e_0(2050)$	80.6	80.9
UPE		
$e_0(2050)$	85.7	–

Overall, using the GBM we predict an average increase in life expectancy over the coming forty-four years of 0.27 years per year for Italian females, 0.21 years per year for Italian males, 0.28 for US females and 0.22 for US males. The corresponding values obtained with the MRP are slightly lower: 0.25 for Italian females, 0.20 for Italian males, 0.28 for US females and 0.17 for males. Comparing these results with the corresponding ones for the best-practice levels, namely 0.24 for females and 0.20 for males, we deduce the predicted convergence or divergence of a country-specific life expectancy relative to the best-practice level. For those data returning an average increase in life expectancies which is higher than the best-practice level, we predict a long-term convergence to it. For those with the same average increase, we predict a persistent

constant gap; while for those with a lower average increase we predict a divergent trend between life expectancy and the best-practice level.

Both models yield future life expectancy results which are characterized by a diverging male-female gap. Modeling the male and female mortalities independently, without introducing any constraint, we are implicitly maintaining the better female performances over time. The diverging male-female gap observed in the best-practice line is translated into a diverging male-female gap in the country-specific life expectancy.

Additional information is provided by the prediction intervals of the life expectancies: the 95% prediction intervals tend to be between 8 and 9 years wide in the year 2050. Implausible values of the prediction intervals for the United States are obtained by the GBM, a problem which does not exist when we use the MRP model, which is characterized by a limited long-term variability, and hence more stable future values.

To gain a better understanding of the results of our models, we took two additional steps. First, we compared the median values of our life expectancies in 2050 with the ones published by the United Nations (UN), Eurostat, the US Census Bureau, and [Alho, Jensen, and Lassila \(2008\)](#) within the UPE project. Second, we undertook an out-of-sample evaluation of the performances of the two methods, computed on three different time periods.

We present the results of our forecasts, together with those from other sources, in [Tables 6 and 7](#). Note that there are systematic differences between our forecasts and the other ones: in general, we have a more optimistic view of the future. This is especially true if we compare our results with those from the US Census Bureau and the UN. The US Census Bureau and the UN rely on the idea of a maximum life expectancy, and assume that as countries approach this maximum value, the annual improvement will become smaller. The UN chose (in 2004) values of maximum life expectancies (92.5 years for women and 87.5 years for men) which are lower than our predicted best-practice lines. The UN median future values of the Italian life

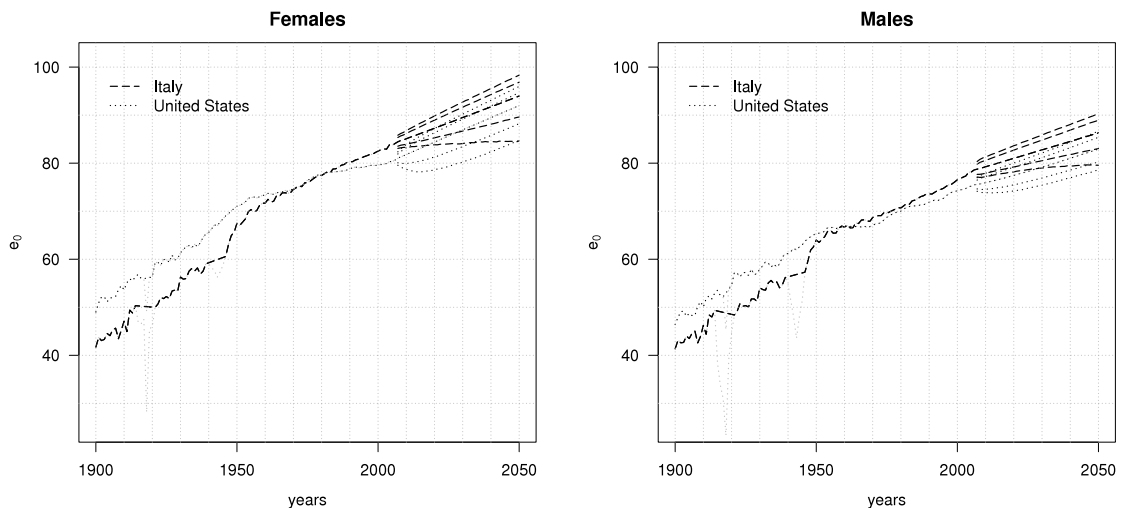


Fig. 9. Actual and forecast life expectancies at birth, e_0 , using a MRP model, together with the associated 80% and 95% prediction intervals for Italy and the United States, 1900–2050. Data for females and males. The transient shocks (in grey) were excluded from our analysis because our focus was on long-term trends.

expectancies therefore fall outside our 95% prediction intervals. Although our forecasts differ remarkably from those of the UN, note that past official forecasts have been systematically too low. The forecasts made within the UPE project are notably higher than the UN's forecasts and the US Census Bureau forecasts, but lower than our forecasts.

In the following section we provide an out-of-sample evaluation of the performances of our two models, for both countries and both sexes. We analyze three different samples of data. First, we fit the model on the data from 1900 to 1950 and predict until 2006, the last observed value in the data. Then we augment the period of data used to fit the models by ten more years, until 1960, and finally by twenty-five years until 1975. We expand the observation period and make predictions on a progressively less distant future. In the first case we are predicting life expectancies for the next 56 years, in the second case for the next 46 years, and in the last case only for the next 31 years.

In Table 8 we present the mean errors calculated between the observed life expectancy at birth and our forecast of it. The GBM almost always returns positive values, meaning that the model, on average, overestimates the future life expectancy. On the other hand, the MRP shows less biased results, with a tendency to underestimate the life expectancy. If we look at the magnitude of the mean errors, we note that the GBM performs better on Italian data, while the MRP model is better for the US data.

We provide a more complete view of the results in Figs. 10–13, where the forecasts of life expectancies at birth, computed using three different data periods, are plotted against the corresponding observed value.

5. Discussion

Modeling a country-specific life expectancy as a linear function plus a smooth gap function, which by construction is always positive, we guarantee that the projected country-specific life expectancies converge, but never rise

Table 8

Mean errors of the forecast life expectancy at birth, produced considering three different forecasting periods: 1951–2006, 1961–2006, and 1976–2006. Results are presented for both the GBM and the MRP, by sex and by country.

GBM		
	Italy	US
Females		
1951–2006	0.414	2.812
1961–2006	1.419	3.224
1976–2006	0.397	2.219
Males		
1951–2006	2.125	3.906
1961–2006	2.192	3.121
1976–2006	−0.399	0.424
MRP		
	Italy	US
Females		
1951–2006	−6.312	−0.171
1961–2006	−1.868	0.952
1976–2006	−0.975	0.975
Males		
1951–2006	−4.843	0.225
1961–2006	−2.445	0.803
1976–2006	−1.813	−0.787

above the forecast best-practice line. This approach, however, has one limitation: we cannot apply the logarithm transformation to the record-holding countries, which exhibit a gap of zero. The logarithmic transformation, considered in both the GBM and the MRP, fails in this specific case.

However, this problem is limited to 6 of the 33 countries present in the Human Mortality Database (Iceland, Japan, the Netherlands, Norway, New Zealand (non-Maori) and Sweden), if we consider data from 1900; and the problem can be fixed, albeit in an ad hoc manner, by adding a small positive value to the argument of the logarithm.

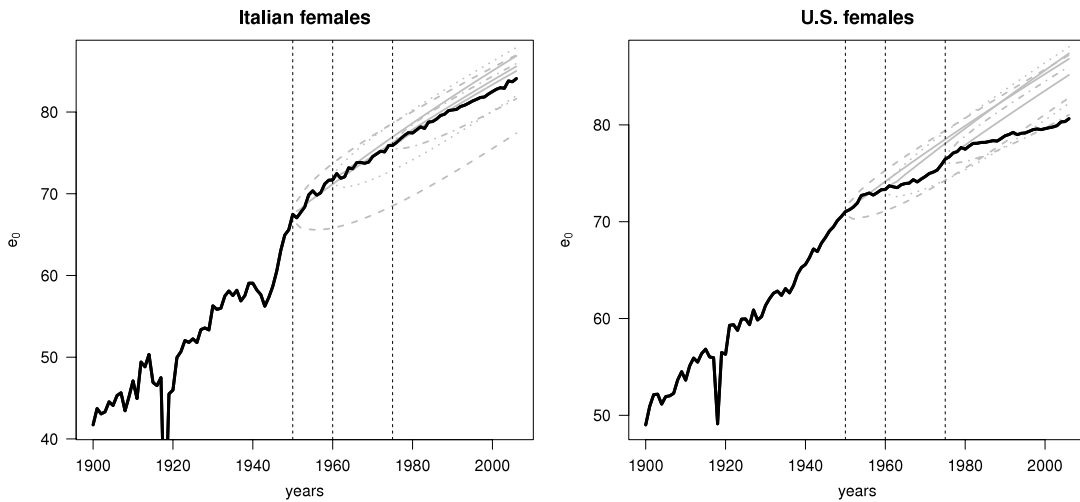


Fig. 10. Actual and forecast life expectancies at birth, e_0 , using a GBM model, with associated 95% prediction intervals, for Italy and the United States. Forecast periods: 1951–2006, 1961–2006, 1976–2006. Females.

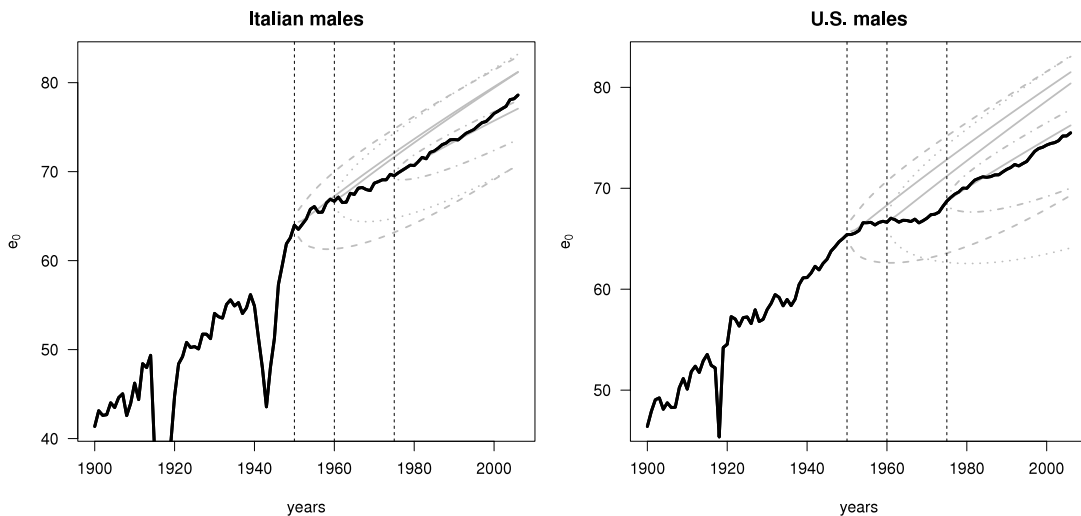


Fig. 11. Actual and forecast life expectancies at birth, e_0 , using a GBM model, with associated 95% prediction intervals, for Italy and the United States. Forecast periods: 1951–2006, 1961–2006, 1976–2006. Males.

The fact that the gap cannot fall to zero implies that the record-holding country cannot be determined exactly. This likewise turns out not to be a major problem, as the definition of the record-holding country is a relative concept, defined with respect to other countries' life expectancies.

It is worth considering again the extremely wide prediction intervals of the future gap yielded by the GBM applied to US female data. These wide intervals are a result of the assumption made in the model about the uncertainty of the process. It was assumed that the variability of the gap is proportional to the gap itself. This assumption may be problematic. If, on the one hand, it makes sense to assume big improvements in life expectancies for a country which is lagging behind the best-practice level, it makes less sense, on the other hand, to assume possible big changes in the opposite direction, with a sharp deterioration in life expectancies. If this is the case, the model may be deficient

in capturing the uncertainty for countries which are far from the best-practice level.

When the GBM seems to fail, we can use the results of the MRP instead.

6. Conclusions

The innovative approach that we take in this study combines separate forecasts of best-practice levels and gaps.

We forecast the pronounced stochastic linear trend of the best-practice levels and the logarithmic behavior of the gap using ARIMA models. More specifically, we consider a discrete geometric Brownian motion and mean-reverting process model. The two alternative models produced similar results in terms of the median values of life expectancies in the year 2050, due to the predominance of the best-practice line on the overall forecast of

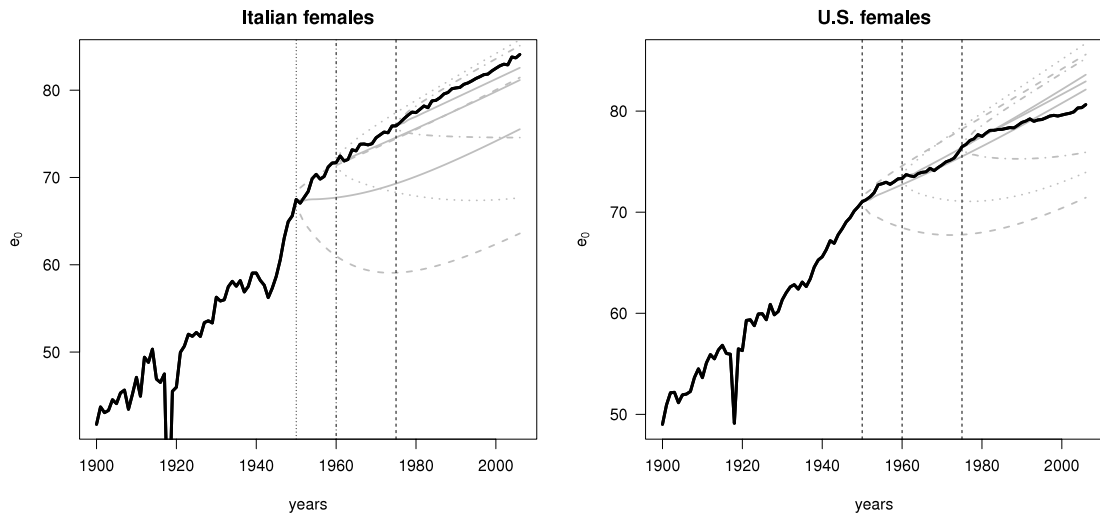


Fig. 12. Actual and forecast life expectancies at birth, e_0 , using a MRP model, with associated 95% prediction intervals, for Italy and the United States. Forecast periods: 1951–2006, 1961–2006, 1976–2006. Females.

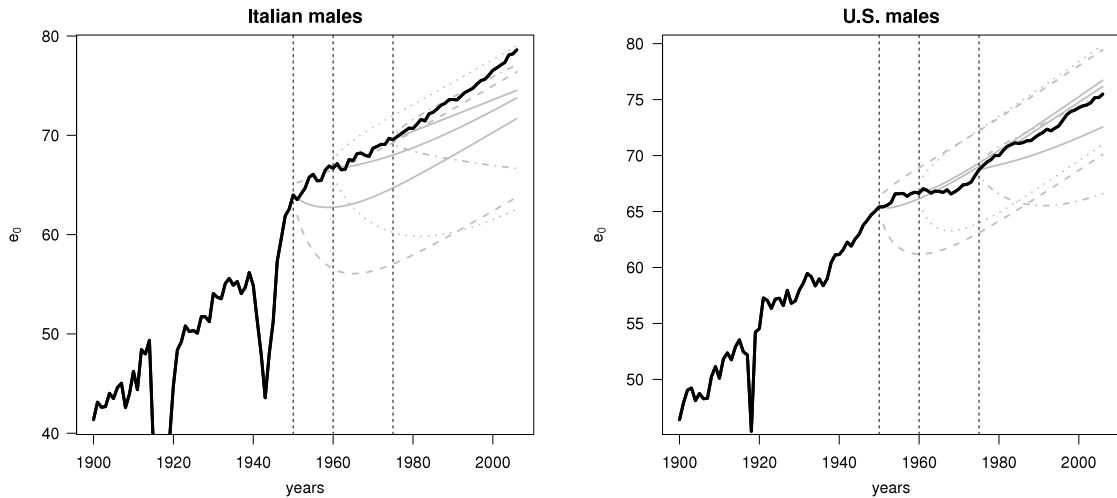


Fig. 13. Actual and forecast life expectancies at birth, e_0 , using a MRP model, with associated 95% prediction intervals, for Italy and the United States. Forecast periods: 1951–2006, 1961–2006, 1976–2006. Males.

life expectancy. They differed mainly in their estimation of the uncertainty. The discrete geometric Brownian motion model sometimes produced erratic and implausible prediction intervals, while the discrete geometric mean-reverting process model generally yielded more conservative and reasonable prediction intervals.

Although similar, the discrete mean-reverting process model proved to perform better than the discrete geometric Brownian motion. It provided more reasonable prediction intervals, returned future values of life expectancies in 2050 which were closer to the official forecasts, and proved not to yield biased mean errors in the out-of-sample evaluations.

In summary, we were able to forecast future values of the life expectancy, building on the solid foundation of the concept of the best-practice level and the gap between a population's performance and best-practice. We generated stochastic forecasts of the life expectancy, with reasonable

error bands, coherent with the trend of the best-practice levels.

References

- Alho, J. M., Jensen, S. E. H., & Lassila, J. (2008). *Uncertain demographics and fiscal sustainability*. Cambridge University Press.
- Alho, J. M., & Spencer, B. D. (2005). *Statistical demography and forecasting*. Springer.
- Andreev, K. F., & Vaupel, J. W. (2006). *Forecasts of cohort mortality after age 50*. Working paper. Max Planck Institute for Demographic Research. Rostock, Germany.
- Berkeley Mortality Database (1995). University of California, Berkeley, USA. Available at: www.demog.berkeley.edu/bmd/states.html (accessed January 2010).
- Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22, 547–581.
- Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control*. San Francisco: Holden-Day.
- Brockwell, P. J., & Davis, R. A. (1996). *Introduction to time series and forecasting*. Springer.

- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427–431.
- Human Mortality Database (2010). University of California, Berkeley, USA; and Max Planck Institute for Demographic Research, Germany. Available at: www.mortality.org (accessed January 2010).
- Keilman, N. (2001). Uncertain population forecasts. *Nature*, 412, 490–491.
- Lee, R. (2006). Mortality forecasts and linear life expectancy trends. In T. Bengtsson (Ed.), *Perspectives on mortality forecasting. III. The linear rise in life expectancy: history and prospects* (pp. 19–40). Stockholm: National Social Insurance Board.
- Lee, R., & Carter, L. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87, 659–671.
- Li, N., & Lee, R. (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee–Carter method. *Demography*, 42, 575–594.
- Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296, 1029–1030.
- Vallin, J., & Meslé, F. (2009). The segmented trend line of highest life expectancies. *Population and Development Review*, 35, 159–187.
- Vaupel, J. W. (2010). Biodemography of human ageing. *Nature*, 464, 536–542.
- Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., Yashin, A. I., Holm, N. V., et al. (1998). Biodemographic trajectories of longevity. *Science*, 280, 855–860.
- Wilmoth, J. R. (1998). The future of human longevity: a demographer's perspective. *Science*, 280, 395–397.
- Tiziana Torri**, Ph.D., is a post-doctoral fellow at the Max Planck Institute for Demographic Research in Rostock, Germany. She wrote her doctoral dissertation on the subject “Assessment and Transfer of the Longevity Risk: New methods and applications”, with the main focus of the work being on mortality forecasting.
- James W. Vaupel**, Ph.D. (Harvard, 1978), is Director of the Max Planck Institute for Demographic Research in Rostock, Germany. He is a member of the US National Academy of Sciences and the author of 54 articles which have been cited at least 54 times each. His research related to forecasting has been reported widely.