Decisions in Economics and Finance
https://doi.org/10.1007/s10203-023-00417-x

Lee–Carter model: assessing the potential to capture gender-related mortality dynamics

Giovanna Apicella[1] · Emilia Di Lorenzo[2] · Gabriella Piscopo[2] · Marilena Sibillo[3]

Received: 25 May 2023 / Accepted: 20 October 2023

Abstract

We investigate the ability of the Lee–Carter model to effectively estimate the gender gap ratio (GGR), the ratio between the male death rates over the female ones, by using a Cox–Ingersoll–Ross (CIR) process to provide a stochastic representation of the fitting errors. The novelty consists in the fact that we use the parameters characterizing the CIR process itself (long-term mean and volatility), in their intrinsic meanings, as quantitative measures of the long-term fitting attitude of the Lee–Carter model and synthetic indicators of the overall risk of this model. The analysis encompasses 25 European countries, to provide evidence-based indications about the goodness of fit of the Lee–Carter model in describing the GGR evolution. We highlight some stylized facts, namely systematic evidence about the fitting bias and the risk of the model across ages and countries. Furthermore, we perform a functional cluster analysis, allowing to capture similarities in the fitting performance of the Lee–Carter model among countries.

Giovanna Apicella, Emilia Di Lorenzo, Gabriella Piscopo and Marilena Sibillo have contributed equally to this work.

✉ Giovanna Apicella
giovanna.apicella@uniud.it

Emilia Di Lorenzo
diloremi@unina.it

Gabriella Piscopo
gabriella.piscopo@unina.it

Marilena Sibillo
msibillo@unisa.it

[1] Department of Economics and Statistics, University of Udine, Via Tomadini 30/A, 33100 Udine, Italy

[2] Department of Economic and Statistical Science, University of Naples Federico II, Via Cintia, Complesso Monte S. Angelo, 80126 Naples, Italy

[3] Department of Economics and Statistics, University of Salerno, Campus Universitario, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy

Published online: 15 November 2023

Keywords Gender gap ratio · Lee–Carter model · Fitting analysis · Functional cluster analysis

Mathematics Subject Classification C63 · G22 · G52

1 Introduction

The empirical evidence reveals a higher life expectancy for females compared to males, albeit with worsening gaps for females in terms of quality of life (cf. MEF 2021) due to the impact, amplified in adults, of disparities linked to labour market, wage levels and choices related to work–life balance.

Implementing economic and social policies should be informed by quantitative analysis of differences in male and female mortality and their evolution over time. Just consider the design of insurance/pension contracts and, in general, the definition of legislative measures that incorporate a fair gender differentiation (cf. Ayuso et al. 2021), in light, for instance, of healthy life expectancy at birth, that is lower for females (MEF 2021).

The literature on the gender gap in mortality is far-reaching and frequently highlights descriptive numerical features of the trend. Glei and Horiuchi (2007) analyse the sex differential in the expected future lifetime for several high-income national populations; Bergeron-Boucher et al. (2018) focus on evolutionary trends in male and female life expectancy characterized by a pattern convergence in mortality; Jallbjørn and Jarner (2022) select multi-population mortality models through their capability of representing the gender differences in mortality. The analysis by Zarulli et al. (2021) focuses in detail on the dynamics of absolute and relative differences in male and female mortality across populations. Ayuso et al. (2021) compare the gap in life expectancy in different populations, disaggregated by gender, to define adequate pension policies. Apicella et al. (2023) deepen the measure of the gender gap in the probability of death across a wide range of ages, emphasizing, in specific geographical areas, some characterizing aspects both in the historical trend and in future projections. Moreover, recent literature has considered the divergences between the correct representation of the gender gap in death probabilities and the subjective perception of life expectancy, which guides financial choices and pension planning (cf. Apicella and De Giorgi 2022). This issue is closely related to investment choices and the effective ability to plan them in line with life expectancy and quality of life; in this sense, Aristei and Gallo (2022) assess gender differences in objective financial knowledge and effective self-confidence about their financial expertise.

The architecture of lifecycle-based financial policies and financial solutions for ageing populations therefore underlies the quantitative design of the evolutionary trend of the gender gap in death probabilities.

In this paper, we investigate the ability of the Lee–Carter (LC) model to effectively estimate the gender gap ratio (GGR), the ratio of the male death rates to the female ones. Such investigation is based on the analysis of the discrepancies between the realized values of the GGR and the ones estimated by the LC model.

Apicella et al. (2019) provide a quantitative background, based on the Cox–Ingersoll–Ross (CIR) process, for the analysis of the fitting errors of stochastic mortality models. The optimal parameters of the CIR process calibrated to the fitting errors provide important directions to improve the forecasting performance of the considered mortality models.

In this paper, we focus on a different phenomenon than mortality rates per se, since we study the gender gap ratio. In particular, we intend to assess the ability of the LC model to effectively describe the GGR. We apply the CIR-error process to model the ratio between the GGR observed values and the GGR fitted values, according to the LC estimates of male and female mortality rates. The CIR process thus describes a new phenomenon, namely the relationship between the gender-related fitting performances of the LC model, and loses a demographic connotation. We study the parameters of the CIR process under an original perspective, by assigning to them a specific role in the quantitative description of the fitting performance. Indeed, we provide an extensive analysis of the age-specific behaviour of the fitting bias and of the riskiness of the LC model as revealed, respectively, by the long-term mean and by the volatility of the CIR error process. Such parameters convey intrinsically dynamic information, other than providing a static representation of the goodness of fit of the LC model, as they reveal also the long-run potential of the LC model to approximate the trend of the GGR.

Empirical analyses and functional cluster methods allow us to capture stylized facts of the fitting performance of the LC model that are systematically verified on the GGR data of different countries. The multi-country approach gives generality and robustness to our study and extends the scope of application of the LC model. Indeed, the quantitative study of the gender gap in mortality is relevant within the decision-making process on socioeconomic, health and welfare issues, especially in the aftermath of the COVID-19 pandemic (Flor et al. 2022). The assessment of the LC performance for 25 European countries can provide guidance to decision-makers on the extent of the applicability of this model for quantitative analyses on the gender gap in mortality at supranational level.

The paper is organized according to the following layout. In Sect. 2, we present the GGR forecasting topic, with its implications in demographic, economic and educational relevance. Section 3 is dedicated to the mathematical models representing the dynamics of mortality and the error stochastic process, focusing on the meaning of its parameters (long-term mean and volatility) and their role in the LC performance valuation. A numerical application follows in Sect. 4, where some stylized evidence about the two parameters under study is highlighted in a detailed analysis. Within the analysis of the parameters of the error process, a new approach for investigating the goodness of fit of the Lee–Carter model in describing the evolution of the GGR is considered in Sect. 5, in which we outline the basics of the functional cluster analysis and apply it in a cross-country perspective. Section 6 closes the paper with conclusions.

2 Gender mortality gap forecasting: demographic, economic and educational relevance

It is well established that females, on average, live longer compared to men. Indeed, in the twenty-first century, the gender gap in longevity has become a universal phenomenon (Barford et al. 2006). Nonetheless, the size of the discrepancies between female and male longevity varies across time and countries (Schünemann et al. 2017), being affected also by non-biological factors, such as environmental and socioeconomic factors.

Not only the drivers, but also the implications of the gender gap in mortality go beyond the demographic domain. Indeed, different survival prospects for females and for males have remarkable economic consequences, for instance related to the adequacy of the normal retirement age and of pension incomes. As stressed by Coppola et al. (2022), accounting for the future evolution of the gender longevity gap enables policy-makers to devise retirement schemes that preserve the principles of equality and solidarity, while reducing the gender pension gap. Apicella et al. (2023) argue that assessing the future path of the gender gap in mortality has both a socio-economic relevance and an educational value. According to Eurostat (2020), a larger share of women than of men faces old-age poverty risk. Peeters and De Tavernier (2015) explain the enduring financial vulnerability of women based on the interactions between work history, family history and pension regulations. Nevertheless, also the ability and the skills to proactively plan for the own retirement acquire a particular relevance in the new retirement income provision paradigm, where individuals' forward-looking behaviour (with respect to investments, savings and annuitization) is given a more prominent role (OECD 2018; Lusardi and Mitchell 2008; Kalmi and Ruuskanen 2018). Such decisions reverberate their effects in the long term and are crucially affected by the understanding of complex risks, such as financial and longevity risks. Consequently, besides financial illiteracy, a lack of longevity risk consciousness can undermine individuals' retirement readiness (Hurwitz et al. 2022; Yakoboski et al. 2022), including women's financial security at the silver ages. Apicella et al. (2023) show that a "Gender Gap Ratio" (namely the ratio between male and female mortality rates) ranging between 1.5 and 2.5, according to the age and country, implies a reduction of up to 25% in the benefits from a temporary life annuity contract for females compared to men, against the same amount invested in the annuity. The mis-estimation of the time span that is likely to be spent in retirement can thus have significant economic implications. On the top of this, a vast literature documents that the mis-estimation of survival prospects may arise from behavioural biases (Heimer et al. 2019; Grevenbrock et al. 2021; Apicella and De Giorgi 2022). It is thus crucial to address the lack of demographic literacy, to allow individuals to take more informed and, possibly, less biased forward-looking economic decisions involving the estimation of longevity. To assess the extent of demographic illiteracy, it is useful to devise quantitative methods to obtain sound and reliable future patterns of male and female longevity and, accordingly, of their prospective discrepancy, acting as an objective benchmark for longevity evolution.

As already mentioned, the gender gap in mortality, likewise the female and male mortality phenomena per se, is impacted by dynamic variables, such as economic and biological factors. Mortality models based on stochastic time-series methods do not assume a dynamic evolution for the interactions between biological and economic drivers with mortality over time. They instead exploit the regularity in both age patterns and trends over time (Booth and Tickle 2008). As stressed in Apicella et al. (2019), the deviations of the mortality rates fitted by the stochastic mortality models from the observed mortality rates reflect the implications of such an underlying assumption. Measures of the fitting errors of the stochastic mortality models can thus provide extra-information to set a mathematical framework for the assessment of the mortality models' potential to catch, in the long run, the dynamic effects of the main contributors of mortality, as revealed by the real data. Following Apicella et al. 2019, in relation to the gender gap in mortality, we measure, in the affine diffusion framework, the ability of the Lee–Carter model (Lee and Carter 1992) to catch the dynamics of the system consistently across ages.

3 Mathematical framework and key ideas

3.1 Empirical mortality and GGR observations

In actuarial mathematics, $m(t, x)$ commonly denotes the crude, unsmoothed, death rate (see, for instance, Cairns et al. 2009). The empirical estimate of the crude death rate is obtained as follows:

$$\widehat{m}(t, x) = D(t, x)/E^c(t, x), \tag{1}$$

namely as the recorded number of deaths at age x last birthday in calendar year t, $D(t, x)$, over the central number of exposed at risk, $E^c(t, x)$. As stressed by Cairns et al. (2009), the underlying death rate, $m(t, x)$, is obtained by dividing the expected deaths by the exposure.

As a measure of the crude gender gap in mortality, we adopt the ratio between male and female death rates. This ratio identifies in the unit value the parity between male and female death rates and in a higher (lower) value than 1 a larger (smaller) magnitude of the male death rate compared to the female one. The state of the art already addresses the analysis of this measure of discrepancy between male and female death rates, for instance Hyndman et al. (2013). Following Apicella et al. (2023), we denominate such a ratio "Gender Gap Ratio" (GGR), and we empirically compute it as follows:

$$\widehat{\mathrm{GGR}}(t, x) = \hat{m}^M(t, x)/\hat{m}^F(t, x), \tag{2}$$

where $\hat{m}^M(t, x)$ and $\hat{m}^F(t, x)$ denote the crude death rate, respectively, for males and females aged x in calendar year t.

3.2 Description of the Lee–Carter model

Lee and Carter (1992) developed the following model for death rates:

$$\log\ m(t, x) = \beta_x^{(1)} + \beta_x^{(2)} k_t, \tag{3}$$

where $\beta_x^{(1)}$ describes the averaged age-pattern of mortality over time, k_t captures the change in the general level of mortality and $\beta_x^{(2)}$ reflects the age-specific deviations of mortality from the averaged pattern when k_t varies. The Lee–Carter model exploits standard time-series methods to derive forecasts of the single time index k_t, that, combined with the estimates of the age-related parameters, allow to obtain age-specific mortality forecasts. As emphasized by Basellini et al. (2022), in their thirtieth anniversary review of Lee and Carter (1992), the Lee–Carter model was a breakthrough in stochastic mortality modelling. Its success and very wide application, among national and international statistical offices (ISTAT 2018), practitioners in the private sector and academics, rely on characteristics such as simplicity. Furthermore, the LC model is based on a purely extrapolative approach, requiring little subjective expert judgement or external information, uniquely related to the choice of the fitting period. As emphasized by Booth et al. (2006), determining the most appropriate fitting period cannot disregard this judgement, since mortality patterns have experienced remarkable structural changes over the twentieth century, thus making the more distant past less informative for current forecasts. Apicella et al. (2023) test the accuracy of models M1 (Lee–Carter), M5 (or CBD, Cairns et al. 2006) and M8 (Cairns et al. 2009) in fitting the female and male death rates characterizing the national populations of Italy, France, Sweden and USA. Based on the Bayesian information criterion (BIC), for any choice of the mortality dataset and of the fitting period (20, 30 or 40 years of data), the Lee–Carter model turns out to be the best-fitting model. Furthermore, extending the fitting period to more than 20 years of data leads to no significant gain in the fitting accuracy. Analogously, within our paper, we exploit time series of historic mortality data made up by 20 years when performing the calibration of the Lee–Carter model.

Booth et al. (2006) show that the Lee–Carter model tends, for males, to underestimate mortality between ages 30 and 40 and to overestimate mortality above age 45. Analogously, the Lee–Carter model tends to underestimate mortality for females aged 20–45. Indeed, as shown in Eq. (3), the Lee–Carter model assumes a constant age-specific pace of mortality change in response to the time index, as described by the time-invariant parameter $\beta_x^{(2)}$. Such assumption can lead to underestimate the more accelerated decline experienced by mortality for some age groups, for instance at advanced ages (Rau et al. 2008). Given the relevance of mortality estimation to predict retirement income provisions and healthcare expenditures, inter alia, investigating the ability of the widely applied Lee–Carter model to accurately estimate mortality across ages has its peculiar importance, in a context where ageing and new welfare are considered high concern emerging risks (Generali Group 2022). Bergeron-Boucher and Kjærgaard (2022) evaluate the accuracy, bias and robustness of the Lee–Carter model for age 65 and above based on an out-of-sample analysis concerning four countries and both sexes. It is shown that the Lee–Carter model is not substantially biased, but the assumption of a constant rate of mortality change does not fit all ages and all the populations.

As explained in Sect. 2, our analysis of the performance of the Lee–Carter model in catching the GGR dynamics builds on the methodological approach developed in Apicella et al. (2019).

3.3 Lee–Carter estimates of the GGR

The GGR fitted by the Lee–Carter model, for age x and calendar year t, is defined as:

$$\overline{\mathrm{GGR}}(t,x) = \overline{m}^M(t,x)/\overline{m}^F(t,x), \tag{4}$$

where $\overline{m}^M(t,x)$ and $\overline{m}^F(t,x)$ are, respectively, the male and female death rate that the Lee–Carter model assigns to age x and calendar year t, based on the information extrapolated from the underlying data. Accordingly, the estimate of the gender gap ratio reflects the (in)-effectiveness of the Lee–Carter model in extrapolating the gender-related past mortality dynamics. As already mentioned, there exists an age-dependent bias of the Lee–Carter model (mortality under- or over-estimation) and such a bias differs by gender (Booth et al. 2006). Furthermore, as stressed by Zarulli et al. (2021), some ages are more decisive in moulding the sex gap in life expectancy. Indeed, the higher mortality among men aged 60 and above plays a dominant role in shaping such a gap. Analysing the accuracy of the estimation of the gender gap ratio at each age allows to focus on a measure that summarizes the overall estimation bias of the Lee–Carter model, accounting for both genders.

3.4 The CIR-error process for the GGR

We define the fitting error of the Lee–Carter model, with respect to the GGR, $Y^{\mathrm{GGR}}(t,x)$, as follows:

$$Y^{\mathrm{GGR}}(t,x) = \frac{\widehat{\mathrm{GGR}}(t,x)}{\overline{\mathrm{GGR}}(t,x)} = \frac{\widehat{m}^M(t,x)}{\widehat{m}^F(t,x)}\frac{\overline{m}^F(t,x)}{\overline{m}^M(t,x)} = \frac{Y^M(t,x)}{Y^F(t,x)}, \tag{5}$$

where $\widehat{\mathrm{GGR}}(t,x)$ is the empirical estimate of the GGR obtained from the data, for age x and calendar year t, while $\overline{\mathrm{GGR}}(t,x)$ is the GGR resulting from the calibration of the Lee–Carter model, as shown in Eq. (4). $Y^M(t,x)$ and $Y^F(t,x)$ represent the errors made by the Lee–Carter model in fitting the male and the female death rates, respectively, for each considered age x and calendar year t:

$$Y^M(t,x) = \widehat{m}^M(t,x)/\overline{m}^M(t,x), \quad Y^F(t,x) = \widehat{m}^F(t,x)/\overline{m}^F(t,x).$$

$Y^{\mathrm{GGR}}(t,x)$ in (5) quantifies how much the estimated GGR is far from the realized GGR observed in the data, in relation to each single age x and calendar year t. Values of $Y^{\mathrm{GGR}}(t,x)$ lower than 1 reflect GGR overestimation, while values of $Y^{\mathrm{GGR}}(t,x)$ larger than 1 denote GGR underestimation.

$Y^{\mathrm{GGR}}(t,x)$ compares the fitting performance of the Lee–Carter model on male mortality data against female ones and expresses which of the two is more accurate (or, equivalently, less biased), in relation to a given age and calendar year. For instance, a $Y^{\mathrm{GGR}}(t,x)$ value being larger than 1 denotes the long-run attitude of the Lee–Carter model to underestimate the empirical GGR; it reveals that the source of such estimation inaccuracy is the larger error made by the LC model with respect to the male mortality data, compared to the female ones. $Y^{\mathrm{GGR}}(t,x)$ thus establishes a comparison between two errors, $Y^M(t,x)$ and $Y^F(t,x)$, that allows to understand which of the two gender-related model performances is delivering the larger accuracy and returning the best result in the global representation of the GGR.

Following Apicella et al. 2019, we assume that, for any fixed age x^f, $Y^{\mathrm{GGR}}(t,x^f)$ is governed by a Cox–Ingersoll–Ross (CIR) process (Cox et al. 1985), as follows:

$$dY^{\mathrm{GGR}}(t,x^f) = \alpha(\zeta - Y^{\mathrm{GGR}}(t,x^f))dt + \sigma\sqrt{Y^{\mathrm{GGR}}(t,x^f)}dW(t), \tag{6}$$

where $W(t)$ is a Wiener process and α, ζ and σ represent the parameters. ζ denotes the long-term mean towards which $Y^{\mathrm{GGR}}(t,x^f)$ is elastically pulled, at a speed α. The stochastic component $\sigma\sqrt{Y^{\mathrm{GGR}}(t,x^f)}dW(t)$, instead, continuously moves the process from its deterministic path, with constant intensity σ.

$Y^{\mathrm{GGR}}(t,x^f)$ cannot become negative if its starting point is nonnegative and the coefficients fulfil the condition:

$$2\alpha\zeta \geq \sigma^2. \tag{7}$$

As explained by Apicella et al. (2019), the CIR process holds empirically relevant properties making it suitable to model the ratios between mortality rates, such as the preservation of nonnegativity, under condition (7). Such property does not characterize, for instance, the Ornstein–Uhlenbeck process. Furthermore, given the probability density of the state variable at time t, conditional on its value at the current time s, for some $s < t$, straightforward computations give closed form formulas for its expected value and variance, thus ensuring computational tractability. In order to properly exploit the extra-information coming from the fitting errors, it is important that the chosen model incorporates mean reversion. The presence of the long-term mean and of a stochastic term being source of uncertainty, together with the nonnegativity, fits well our analysis, as described in the remainder of this section. The CIR process has these characteristics and, compared to more complex models, is simple and parsimonious, holding the required information in a few parameters.

α, ζ and σ provide the coordinates of the evolution over time of $Y^{\mathrm{GGR}}(t,x^f)$ or, in other terms, they characterize the dynamics of the goodness of fit of the Lee–Carter model, with respect to the GGR. Accordingly, such parameters convey intrinsically dynamic information, other than providing a static representation of the performance of the Lee–Carter model. In this respect, in Apicella et al. (2019), the optimal values of α, ζ and σ allow to obtain the best estimate of the CIR process trajectory over the forecasting time horizon, acting as an effective correction factor of the longevity projections of the mortality model.

The estimated parameter $\hat{\zeta}$ acts also as a measure of the fitting bias. Indeed, $\hat{\zeta}$ not only describes the historical attitude of the Lee–Carter model to under- or over-estimate the GGR, but also summarizes the long-run potential of the mortality model to accurately approximate the trend of the GGR, as extrapolated from the data. For instance, in Apicella et al. (2023), $\hat{\zeta}$ is used as a further quantitative selection criterion among competing mortality models.

The CIR process of the in-sample errors $Y^{\mathrm{GGR}}(t,x^f)$ deviates from its deterministic path towards $\hat{\zeta}$, because of a diffusion term whose intensity is constant and equal to

$\hat{\sigma}$. Such a parameter represents a measure of risk, namely the risk that the information provided by $\hat{\zeta}$ is unreliable. Indeed, the larger $\hat{\sigma}$, the more powerful the diffusion term and, thus, the more unstable, or volatile, over time the goodness of fit of the Lee–Carter model in relation to a specific age. $\hat{\sigma}$ plays an important role, given that stability is widely recognized as an important feature of stochastic mortality models (Coelho and Nunes 2011). Since instability can derive from the characteristics of the underlying data, in Apicella et al. (2023), the described methodological approach based on the CIR description of the in-sample errors $Y^{\mathrm{GGR}}(t, x^f)$ and $\hat{\sigma}$ is used to infer statistical information about the volatility of the time series of mortality data concerning females and males, and, by construction, of the gender gap ratio.

We leverage the intrinsic meanings of the long-term mean and the volatility of the CIR-error process as revealing the fitting bias and the riskiness of the LC model in the description of the GGR. We provide an extensive study of these two parameters to uncover their behaviour by age. We shed light on how the magnitude of the fitting bias and of the volatility vary over age and show if the uncovered age patterns have systematic features, consistently verified across several mortality datasets. This analysis corroborates the long-term mean and the volatility parameters as quantitative criteria for the global assessment of the LC model performance with respect to the GGR; it offers novel evidence on the persistence of the parameters of the CIR-error process over age. Our analysis allows also to discriminate the age intervals for which the LC model systematically shows the best-fitting performance, according to the outlined criteria.

4 Numerical application

For our numerical application, we use data from the Human Mortality Database (HMD 2023) related to the recorded number of deaths and the exposed at risk for females and males of the following 25 European national populations: Austria, Belgium, Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Lithuania, Luxembourg, the Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, and United Kingdom (UK).

The observed data allow us to compute the empirical estimates of the female and male death rates $\widehat{m}^i(t, x)$, $i = M, F$, as in Eq. (1), where x is the age, t represents one calendar year between 2000 and 2019, M denotes males and F denotes females. We relate such estimates to obtain the empirical values of the GGR for each considered age and calendar year, $\widehat{\mathrm{GGR}}(t, x)$, according to Eq. (2).

Rectangular arrays of deaths and exposures, including ages from 18 to 90 on the rows and calendar years from 2000 to 2019 on the columns, are the input data for the estimation of the Lee–Carter model, for each gender and considered country. The choice of the time interval of 20 years is coherent with the results obtained in Apicella et al. (2023) and recalled in Sect. 3.2. We implement the Lee–Carter model in the **R** software (https://www.r-project.org/) through the **StMoMo** package. The StMoMo function allows to create an object representing the LC model, based on information about the link function, the predictor structure and the set of parameter constraints.

We assume that deaths follow a binomial distribution and use a logit link function targeting 1-year death probabilities $q(t, x)$, that are then transformed in $m(t, x)$. The StMoMo package also includes the function *lc* that facilitates the definition of Lee–Carter model. The generic function *fit* can be used for estimating the parameters of the Lee–Carter model, by maximization of the model log-likelihood, as described in Villegas et al. (2016). The outputs of the described implementation procedure are the fitted male and female death rates $\overline{m}^M(t, x)$ and $\overline{m}^F(t, x)$, for $x = 18, ..., 90$ and $t = 2000, ..., 2019$, for each country under study. We compute the ratio of $\overline{m}^M(t, x)$ to $\overline{m}^F(t, x)$ to obtain the fitted values of the GGR, $\overline{GGR}(t, x)$, according to Eq. (4), for the same age and time intervals. For our further analysis, we select the interval of integer ages between 45 and 85, because this age interval encompasses the life-cycle phases when individuals undertake relevant financial decisions and retirement planning strategies. This age interval is the period in life when these choices ultimately reverberate their effects, that combine with the outcomes of longevity risk.

For each integer age in [45–85], we focus on the time series of the in-sample fitting errors of the GGR, $Y^{\text{GGR}}(t, x^f)$, $t = 2000, \ldots, 2019$, as in Eq. (5). We assume that these are sample paths of a Cox–Ingersoll–Ross process and obtain calibrations of such a process returning the optimal parameters $\hat{\alpha}$, $\hat{\zeta}$ and $\hat{\sigma}$ for each age under study, independently.

We estimate the parameter vector of the CIR process, by minimizing the negative of the log-likelihood function for the CIR process (6) over its parameter space, in **MATLAB** as in Kladivko (2012). In particular, we use the **MATLAB** function *fminsearch* for performing unconstrained optimization and verify ex post that all the parameters are positive and fulfil condition (7). We obtain initial estimates of the parameter vector, by applying ordinary least squares on the discretized version of (6), as explained in Kladivko (2012). The optimization routine requires the $Y^{\text{GGR}}(t, x^f)$ time series for each integer age under study and a time step equal to 1, as the error time series are built from yearly empirical values of the GGR and corresponding LC model estimates. The optimal parameters convey valuable information about the error dynamics over time and thus on the fitting performance of the Lee–Carter model, for each considered age. As explained in Sect. 3, we analyse $\hat{\zeta}$ and $\hat{\sigma}$, representing, respectively, a measure of the fitting bias and a measure of the associated risk.

4.1 Stylized evidence about the fitting bias: $\hat{\zeta}$

The optimal parameter $\hat{\zeta}_x$ denotes the long-term mean of the Cox–Ingersoll–Ross process governing the fitting errors $Y^{\text{GGR}}(t, x)$ for age x.

We use such an optimal parameter to assess the sign of the fitting bias of the Lee–Carter model across ages, namely whether we observe underestimation or overestimation of the gender gap ratio among the countries under study. It is systematically verified, across countries, that, on the full age sample, GGR underestimation occurs more frequently than GGR overestimation. In this respect, we compute the empirical probability of GGR underestimation, namely the ratio of the number of ages for which GGR underestimation occurs to the total number of ages in the sample. This empirical probability ranges from the minimum value of 54% for France to the maximum

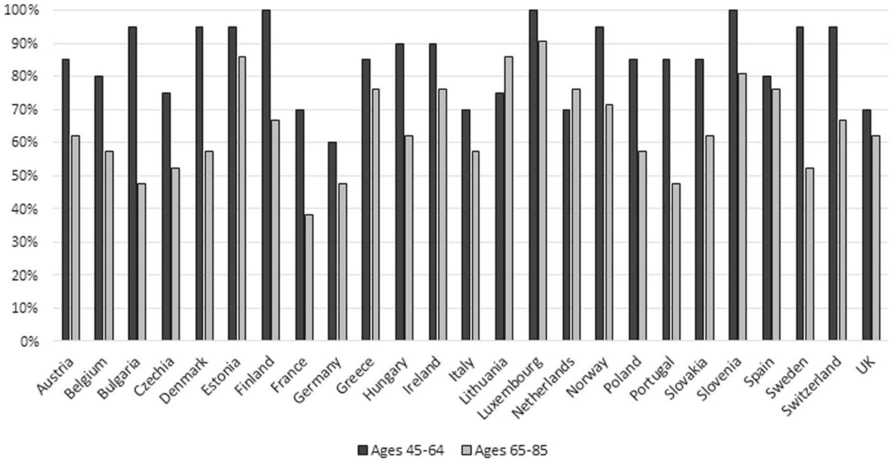

Fig. 1 Empirical probability that the Lee–Carter model underestimates the realized gender gap ratio in the age groups 45–64 (black bars) and 65–85 (grey bars), across countries (Color figure online)

value of 95% for Luxembourg. Since we do not observe cases of a perfect match between GGR LC estimates and GGR empirical estimates (namely no $Y^{\mathrm{GGR}}(t, x^f)$ equals 1), the complementary probability represents the empirical probability of GGR overestimation, accounting for all the ages in [45–85].

In Fig. 1, we display the empirical probabilities of GGR underestimation, specifically concerning two age subgroups: [45–64] (black bars) and [65–85] (grey bars). Except for Lithuania and the Netherlands, it is consistently observed that younger ages in the age group [45–64] are more likely to be affected by the attitude of the LC model to underestimate the realized GGR than the older ages.

We complement this evidence, by studying the magnitude of the fitting bias, namely by assessing how much each $\hat{\zeta}_x$ differs from the desirable value of 1, representing the tendency of the LC model to make no estimation errors.

For each country, we denote by $\Delta\hat{\zeta}_x$ such a discrepancy, considered in its absolute value, for each single age x in the sample [45–85]. Overall, for each country, we thus observe 41 values of $\Delta\hat{\zeta}_x$, namely one per age. We rank $\Delta\hat{\zeta}_x$ in ascending order, so that to find the 15 smallest values ("top 15", $\{\Delta\hat{\zeta}^i : i = 1, ..., 15\}$) and the 15 largest values of the fitting bias ("bottom 15", $\{\Delta\hat{\zeta}^j : j = 27, ..., 41\}$). In this respect, in Table 1, for each country, we report the lower and the upper bounds of each set of values, and the respective ages at which they occur.

Furthermore, for each country, we gain some insights into the cross-age persistence of the fitting bias $\Delta\hat{\zeta}_x$. Indeed, the smallest recorded value of $\Delta\hat{\zeta}_x$, that is $\Delta\hat{\zeta}^1$, corresponds to the least biased estimate of the GGR, with respect to the realized one, achieved by the Lee–Carter model, over the age range [45–85] and for a given country. The closer the $\Delta\hat{\zeta}_x$ values among them and to $\Delta\hat{\zeta}^1$, the more persistent the lowest fitting bias of the of the LC model throughout the considered ages.

We compute what we call $Spread^i_{\zeta} = \Delta\hat{\zeta}^{(i+1)} - \Delta\hat{\zeta}^i, i = 1, ..., 40$, namely the difference between each of the 40 pairs of consecutive ordered values. The average

Table 1 Evidence about $\Delta\hat{\zeta}_x$

Country	$\Delta\hat{\zeta}_x$								$\overline{Spread}_\zeta$ (Unit 10^{-4})
	Top 15: bounds (Unit 10^{-4})				Bottom 15: bounds (Unit 10^{-4})				
	Lower	Age	Upper	Age	Lower	Age	Upper	Age	
Austria	1.5	82	19.4	74	53.9	69	223.2	46	5.5
Belgium	3.7	72	20.6	70	44.3	68	157.9	48	3.9
Bulgaria	0.1	73	22.3	69	54.4	63	133.5	61	3.3
Czech Republic	0.5	68	17.9	79	35.9	75	142.7	51	3.6
Denmark	0.0	76	32.2	69	61.1	56	331.9	45	8.3
Estonia	0.7	83	55.0	74	128.0	54	804.7	45	20.1
Finland	3.1	75	28.4	62	56.9	65	289.3	46	7.2
France	0.0	72	4.6	75	16.9	60	43.1	66	1.1
Germany	0.1	78	11.3	81	20.2	72	61.1	52	1.5
Greece	0.5	83	30.7	68	58.0	82	120.8	48	3.0
Hungary	0.3	58	14.9	71	31.7	48	109.4	45	2.7
Ireland	1.8	78	38.9	66	102.4	45	621.0	49	15.5
Italy	0.5	48	6.1	63	11.9	67	37.3	85	0.9
Lithuania	3.2	64	31.9	82	53.9	51	377.5	46	9.4
Luxembourg	6.8	78	259.3	85	702.7	56	2065.5	54	51.5
Netherlands	0.4	85	12.4	73	32.0	64	190.1	45	4.7
Norway	0.1	80	25.1	69	83.2	73	383.7	46	9.6
Poland	0.2	50	13.2	53	26.9	62	82.3	68	2.1
Portugal	0.5	70	14.3	67	33.2	83	198.1	52	4.9
Slovakia	4.8	66	26.9	69	53.7	82	181.5	50	4.4
Slovenia	1.2	73	62.0	72	123.4	63	1073.7	46	26.8
Spain	0.6	84	7.8	74	23.7	45	65.5	59	1.6
Sweden	1.7	84	27.9	64	74.6	62	264.5	46	6.6
Switzerland	2.2	65	17.0	59	48.4	82	200.5	46	5.0
UK	0.3	80	6.7	84	17.1	47	64.3	52	1.6

In this table, we report the upper and lower bounds of the 15 smallest and the 15 largest values of $\Delta\hat{\zeta}_x$, expressing the magnitude of the fitting bias, and the ages at which such values occur. We also report, in column 10, $\overline{Spread}_\zeta$, namely the average distance between the observed values of $\Delta\hat{\zeta}_x$, for each country

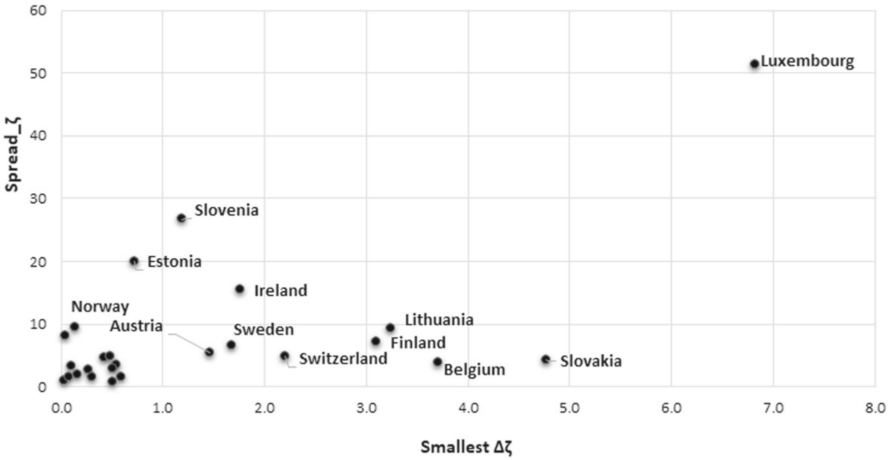

Fig. 2 In this scatter plot, we display, by countries, the pairs of the smallest value attained by $\Delta\hat{\zeta}_x$ ($\Delta\hat{\zeta}^1$, x-axis), and the average distance between all the $\Delta\hat{\zeta}_x$s values, as a measure of persistence of $\Delta\zeta^1$ across ages ($\overline{Spread}_\zeta$, y-axis). The dots in the bottom-left corner of the figure are not labelled, due to the tight closeness of such dots

difference $\overline{Spread}_\zeta$, reported in column 10 of Table 1, measures how much $\Delta\hat{\zeta}_x$ is consistent, in terms of its magnitude, across ages. The higher the $\overline{Spread}_\zeta$, the more unstable the fitting bias in our age sample.

The smallest value $\Delta\hat{\zeta}^1$, along with the average bias spread $\overline{Spread}_\zeta$, allows to gain preliminary insights into the fitting attitudes of the LC model, in terms of $\hat{\zeta}$, across countries. In Fig. 2, we provide graphical evidence about the outcomes reported in Table 1. Indeed, in such figure, we display, on the x-axis, the values taken on by $\Delta\hat{\zeta}^1$ and, on the y-axis, the values of $\overline{Spread}_\zeta$; accordingly, each dot identifies the pair of $\Delta\hat{\zeta}^1$ and $\overline{Spread}_\zeta$ that characterizes a given country. We see that, when calibrated to the mortality data concerning the population of Luxembourg, the Lee–Carter model returns the worst performance in terms of the fitting bias, since it is characterized by the highest value of $\Delta\hat{\zeta}^1$ (6.8, unit 10^{-4}) and by the largest instability (with $\overline{Spread}_\zeta$ attaining the value of 51.5, unit 10^{-4}). By contrast, the countries for which the Lee–Carter model shows the best performance are displayed in the bottom-left corner of the plot and are France, Germany, Bulgaria, Poland, Hungary and UK, with values of $\Delta\hat{\zeta}^1$ in the range $[0.0, 0.3]$ and values of $\overline{Spread}_\zeta$ in the range $[1.1, 3.3]$. This group of countries is closely followed by Italy, Spain, Greece, Czech Republic, The Netherlands and Portugal with values of $\Delta\hat{\zeta}^1$ in the range $[0.4, 0.6]$ and values of $\overline{Spread}_\zeta$ in the range $[0.9, 4.7]$. With respect to the mentioned countries, the other ones are characterized by either higher values of $\Delta\hat{\zeta}^1$ or larger values of $\overline{Spread}_\zeta$ or both. In particular, for Slovenia and Estonia the Lee–Carter model shows a markedly higher instability in the magnitude of the fitting bias, with respect to the other countries.

We assess the age group for which the Lee–Carter model shows a more marked attitude to make the smallest errors $\Delta\hat{\zeta}_x$ observed throughout the age sample. In this respect, we compute the empirical probability that an age belonging to [65–85]

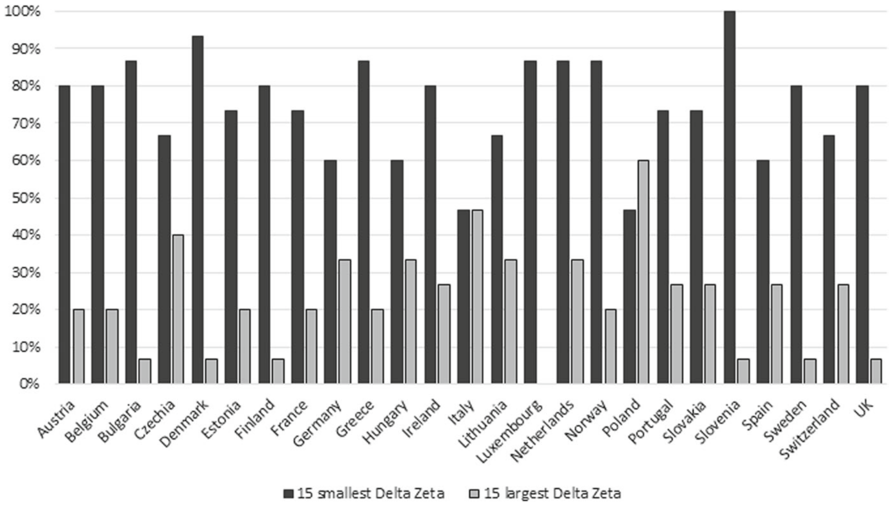

Fig. 3 Evidence about how the magnitude of the fitting bias of the LC model, as measured by $\hat{\zeta}_x$ differs by age group. For each country, the black bars display the empirical probability that an age belonging to [65–85] is associated with one of the 15 smallest values of $\Delta\hat{\zeta}_x$. The grey bars display the empirical probability that an age belonging to [65–85] is associated with one of the 15 largest values of $\Delta\hat{\zeta}_x$ (Color figure online)

is associated with one of the 15 smallest values of $\Delta\hat{\zeta}_x$. For each country, these probabilities are displayed with the black bars in Fig. 3 and are compared against the empirical probabilities that an age belonging to [65–85] is associated with one of the 15 largest values of $\Delta\hat{\zeta}_x$ (grey bars). The complementary probabilities relate to the other age group [45–64]. It is consistently verified for all countries, except for Italy and Poland, that the "top 15" largely pertains to ages in the interval [65–85]. Indeed, if we exclude Italy and Poland, we find that 60% up to 100% of the 15 smallest values of $\Delta\hat{\zeta}_x$ is associated with an age x higher than 64. By contrast, we observe much more frequently that the ages in the interval [45, 64] are characterized by the largest values of $\Delta\hat{\zeta}_x$. Indeed, as shown by the grey bars in Fig. 3, there are a few older ages out of the total being characterized by the greatest values of $\Delta\hat{\zeta}_x$ (from the 0% of Luxembourg to the 40% of Czech Republic, with the exclusion of Italy of Poland that both deviate from this stylized evidence).

Our analysis suggests that, for almost all the countries under study, the Lee–Carter model tends to estimate the gender gap ratio more accurately for the ages we could conventionally refer to as "post-retirement ages".

4.2 Stylized evidence about the risk: $\hat{\sigma}$

The optimal parameter $\hat{\sigma}_x$ denotes the constant intensity with which the Cox–Ingersoll–Ross process governing the fitting errors $Y^{\mathrm{GGR}}(t, x)$ for age x deviates from its long-term mean.

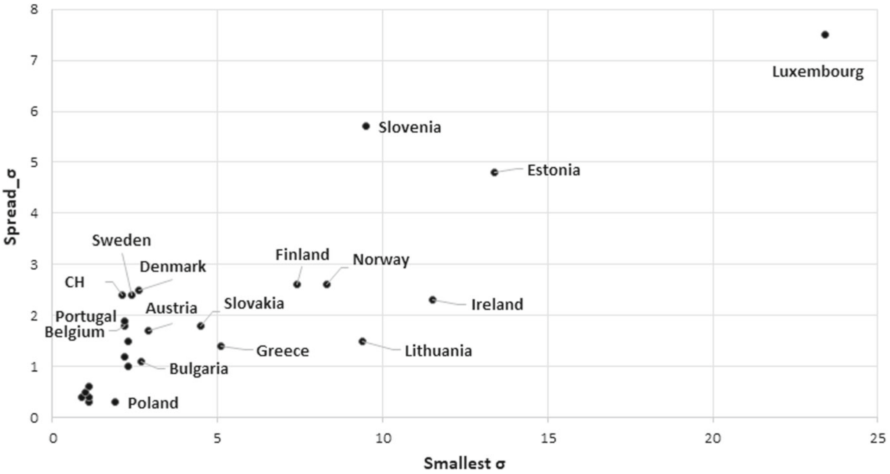

Fig. 4 In this scatter plot, we display, by countries, the pairs of the smallest value attained by $\hat{\sigma}_x$ ($\hat{\sigma}^1$, x-axis), and the average distance between all the $\hat{\sigma}_x$s values, as a measure of persistence of $\hat{\sigma}_x$ across ages ($\overline{Spread}_\sigma$, y-axis). The dots in the bottom-left corner of the figure are not labelled, due to the tight closeness of such dots

We use such an optimal parameter as a measure of the risk for the Lee–Carter model to produce inaccurate, volatile, estimates of the gender gap ratio in the long run, for each age, by country. As for $\Delta\hat{\zeta}_x$, we rank $\hat{\sigma}_x$ in ascending order so that to detect the 15 smallest values ("top 15", $\{\hat{\sigma}^i : i = 1, ..., 15\}$) and the 15 largest values of this measure of risk ("bottom 15", $\{\hat{\sigma}^j : j = 27, ..., 41\}$). In this respect, in Table 2, for each country, we report the lower and the upper bounds of each set of values and the respective ages at which they occur.

Furthermore, for each country, we investigate the cross-age persistence of $\hat{\sigma}_x$. Indeed, the smallest recorded value of $\hat{\sigma}_x$, that is $\hat{\sigma}^1$, expresses the lowest risk, among the ages under study, that the Lee–Carter model gives rise to inaccurate estimates of the GGR, in relation to a given country. The closer the $\hat{\sigma}_x$ values among them and to $\hat{\sigma}^1$, the more persistent this low risk over the age sample.

We compute what we call $Spread^i_\sigma = \hat{\sigma}^{(i+1)} - \hat{\sigma}^i, i = 1, ..., 40$, namely the difference between each of the 40 pairs of consecutive ordered values of $\hat{\sigma}_x$. The average difference $\overline{Spread}_\sigma$, reported in column 10 of Table 2, measures how much $\hat{\sigma}_x$ is consistent, in terms of its magnitude, across ages. The higher the $\overline{Spread}_\sigma$, the more unstable the risk in our age sample.

As for $\Delta\hat{\zeta}_x$, we use the smallest value $\hat{\sigma}^1$ and the average risk spread $\overline{Spread}_\sigma$ to assess the risk inherent to the fitting performance of the LC model, in terms of $\hat{\sigma}$, across countries. In Fig. 4, we display countries, according to the outcomes reported in Table 2. Indeed, in such figure, we display, on the x-axis, the values taken on by $\hat{\sigma}^1$ (column 2 of Table 2) and, on the y-axis, the values of $\overline{Spread}_\sigma$ (column 10 of Table 2). Accordingly, each dot identifies the pair of $\hat{\sigma}^1$ and $\overline{Spread}_\sigma$ that characterizes a given country. In line with our findings related to the fitting bias $\Delta\hat{\zeta}_x$, we see that Luxembourg stands out sharply from the other countries because of the high level of riskiness of the LC estimations of the GGR. Indeed, for Luxembourg, the lowest

Table 2 Evidence about $\hat{\sigma}_x$

Country	$\hat{\sigma}_x$								$\overline{Spread}_\sigma$ (Unit 10^{-2})
	Top 15: bounds (Unit 10^{-2})				Bottom 15: bounds (Unit 10^{-2})				
	Lower	Age	Upper	Age	Lower	Age	Upper	Age	
Austria	2.9	83	13.4	78	23.8	57	69.6	49	1.7
Belgium	2.2	84	11.9	60	18.7	72	74.5	45	1.8
Bulgaria	2.7	81	8.4	61	13.8	54	48.0	45	1.1
Czech Republic	2.3	79	8.1	75	18.1	66	60.5	50	1.5
Denmark	2.6	73	15.3	85	26.3	74	101.1	45	2.5
Estonia	13.4	81	38.4	52	59.5	69	205.1	45	4.8
Finland	7.4	84	18.1	60	36.0	45	112.2	48	2.6
France	1.1	84	2.3	75	5.8	61	13.7	51	0.3
Germany	0.9	84	2.4	74	4.1	62	17.1	45	0.4
Greece	5.1	79	11.7	80	24.7	64	60.7	48	1.4
Hungary	2.2	71	6.1	73	12.5	59	51.1	50	1.2
Ireland	11.5	79	20.5	81	45.5	49	103.1	47	2.3
Italy	1.1	85	2.1	80	6.2	62	26.8	47	0.6
Lithuania	9.4	66	18.8	74	30.9	72	69.4	61	1.5
Luxembourg	23.4	76	70.8	60	153.9	66	321.6	45	7.5
Netherlands	2.3	84	8.3	62	12.7	59	41.8	52	1.0
Norway	8.3	81	23.1	60	41.0	65	112.4	45	2.6
Poland	1.9	81	3.1	78	5.1	64	13.8	46	0.3
Portugal	2.2	74	9.0	78	29.8	60	77.4	46	1.9
Slovakia	4.5	85	13.3	65	25.9	50	78.0	46	1.8
Slovenia	9.5	80	41.6	65	70.8	58	239.0	46	5.7
Spain	1.1	85	3.0	82	8.1	58	19.0	48	0.4
Sweden	2.4	82	12.8	64	22.8	68	97.3	46	2.4
Switzerland	2.1	85	17.5	82	31.8	66	96.2	46	2.4
UK	1.0	83	1.9	71	5.2	62	21.6	45	0.5

In this table, we report the upper and lower bounds of the 15 smallest and the 15 largest values of $\hat{\sigma}_x$, expressing a measure of the risk inherent to the fits of the LC model, and the ages at which such values occur. We also report, in column 10, $\overline{Spread}_\sigma$, namely the average distance between the observed values of $\hat{\sigma}_x$, for each country

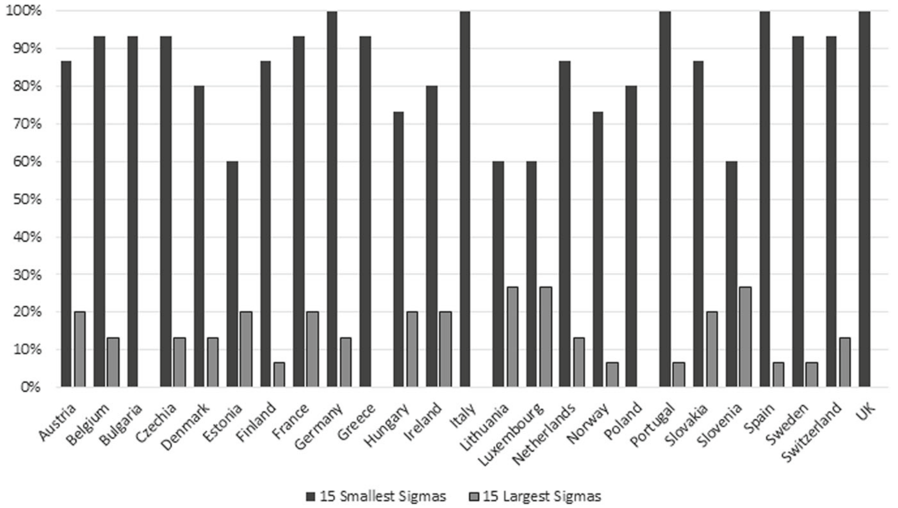

Fig. 5 Evidence about the magnitude of the risk of the LC model, as measured by $\hat{\sigma}_x$, by age group. For each country, the black bars display the empirical probability that an age belonging to [65–85] is associated with one of the 15 smallest values of $\hat{\sigma}_x$. The grey bars display the empirical probability that an age belonging to [65–85] is associated with one of the 15 largest values of $\hat{\sigma}_x$ (Color figure online)

attained risk, among all the ages under study, is 374% higher than the average risk over all the considered nations; furthermore, such a risk is dramatically volatile across ages. By contrast, the group of countries appearing in the bottom-left corner of Fig. 4 show the most moderate level of riskiness. Such countries are France, Germany, Italy, Poland, Spain and the UK, with values of $\hat{\sigma}^1$ between 0.9 and 1.9 and values of $\overline{Spread_\sigma}$ in the range [0.3, 0.6]. With respect to the mentioned countries, the other ones are characterized by either higher values of $\hat{\sigma}^1$ or larger values of $\overline{Spread_\sigma}$ or both. As for $\Delta\hat{\zeta}_x$, we find that the LC fitting performance is distinctive also in terms of riskiness for Slovenia and Estonia.

We assess the age group for which the Lee–Carter model shows the lowest risk. In this respect, we obtain the empirical probability that an age in the interval [65–85] is associated with one of the 15 smallest values of $\hat{\sigma}$. For each country, these probabilities are displayed with the black bars in Fig. 5, while the grey bars represent the empirical probabilities to observe one the 15 largest values of $\hat{\sigma}$ for the same age group. In both cases, the complementary probabilities measure the likelihood of the two events (occurrence of one of the 15 smallest values or otherwise) for the other age group [45–64]. The evidence about $\hat{\sigma}_x$ demonstrates, even more markedly than the findings about $\Delta\hat{\zeta}_x$, that the age group [65–85] is noticeably characterized by the best fitting performance of the Lee–Carter model. Not only such an evidence is consistent across all countries, but also the occurrences of the smallest $\hat{\sigma}_x$ among the ages higher than 64 are dramatically high. Indeed, for 16 countries out of 25 we find that more than 80% of the 15 smallest values of $\hat{\sigma}_x$ is associated with an age in the interval [65–85]. By contrast, we observe that the ages lower than 65 are more frequently associated with the largest values of $\hat{\sigma}_x$. Indeed, as displayed the grey bars in Fig. 5, for a very few older ages out of the total we observe values of $\hat{\sigma}_x$ being among the 15 greatest ones (never more than 27% of the total cases).

4.3 Assessment of the LC model based on $\hat{\zeta}$ and $\hat{\sigma}$

In this section, we combine the two goodness-of-fit assessment criteria, $\hat{\zeta}$ and $\hat{\sigma}$, to uncover evidence about the age range for which the Lee–Carter model shows an overall better attitude to accurately approximate the Gender Gap Ratio phenomenon.

In Sect. 4.1, we detect the 15 smallest values of the $\Delta\hat{\zeta}_x s$ obtained over the age sample [45–85] (top 15 of $\Delta\hat{\zeta}_x s$). Indeed, in relation to age x, the lower $\Delta\hat{\zeta}_x$, in absolute value, the smaller the distance from the desirable value of 1, representing the tendency of the LC model to make no estimation errors, as inferred from observing the model's performance on the time horizon [2000–2019]. Analogously, in Sect. 4.1, we detect the 15 smallest values of the $\hat{\sigma}_x s$, for $x \in$ [45–85] (top 15 of $\hat{\sigma}_x s$). With respect to age x, the lower $\hat{\sigma}_x$, the more stable the tendency toward $\hat{\zeta}_x$, this denoting the reliability of the model to catch the dynamics of the underlying data.

Accordingly, when assessing the LC fitting performance in relation to the GGR, the lowest values of both $\Delta\hat{\zeta}_x s$ and $\hat{\sigma}_x$ guide the assessment towards the ages for which the potential of the LC model to precisely capture the dynamics of the GGR is at its highest levels. Based on this reasoning, for each country, we detect the ages for which both one of the 15 smallest values of the $\Delta\hat{\zeta}_x s$ and one of the 15 smallest values of the $\hat{\sigma}_x s$ contextually occur. These ages are shown in Table 3. The 25 countries under study vary by the number of simultaneous occurrences of a $\Delta\hat{\zeta}_x$ and of a $\hat{\sigma}_x$ in the top 15, whose maximum attainable value is 11, according to our evidence. Only two countries, the Netherlands and Sweden, achieve the result of 11 matches, that are associated, in the vast majority of cases, to ages in the interval [65–85]. The fact that the Lee–Carter model performs better for more advanced ages than for younger ones is quite stylized over our sample of countries. Indeed, higher ages than 64 account for the 67% up to the 100% (for 14 countries out of 25) of the found matches. We remark that this sharp prevalence of more advanced ages has a different meaning and impact depending on the country under study. Indeed, it has a greater relevance for those countries being characterized by a weak persistence of both the fitting bias and the riskiness across ages, as shown in Sects. 4.1 and 4.3. If, otherwise, the magnitude of the fitting bias and of the riskiness are stable over ages, ages in the range [65, 85] stand out less remarkably from the younger ones.

To enhance the robustness of our findings, we perform the analysis on the fitting bias and the riskiness of the Lee–Carter model, when this model is calibrated to a 30-year time span, starting from 1990. Also under this new setting, we obtain very strong evidence that the 15 smallest values of $\Delta\hat{\zeta}_x$s and of $\hat{\sigma}_x$ are more likely to be associated with ages between 65 and 85, consistently across countries. When we combine the two goodness-of-fit assessment criteria, $\hat{\zeta}$ and $\hat{\sigma}$, by computing the number of simultaneous occurrences of a $\Delta\hat{\zeta}_x$ and of a $\hat{\sigma}_x$ in the top 15, we find that higher ages than 64 account for the 71% up to the 100% of the found matches across countries. Furthermore, Luxembourg stands out sharply from the other countries, as it is characterized by the

Table 3 Goodness of fit of the LC model by ages

Country	Ages										
Austria	63	76	77	78	79	80	82	84	85		
Belgium	60	69	74	76	77	79	80	81	85		
Bulgaria	66	68	72	73	74	75	76	78	80	81	
Czech Republic	67	71	79	80	83	84					
Denmark	73	76	78	79	80	82	83	84			
Estonia	61	62	72	74	80	84					
Finland	75	79	80	81	83	84	85				
France	65	71	74	75	81	83	84	85			
Germany	75	77	78	80	81	84	85				
Greece	69	72	74	75	76	77	79	83	84	85	
Hungary	60	64	71	74	75	78	81	83			
Ireland	64	74	76	77	78	79	81	82	85		
Italy	69	70	73	74	75						
Lithuania	64	74	75	76	79	82					
Luxembourg	59	61	70	74	76	81	83	84			
Netherlands	62	67	71	77	79	80	81	82	83	84	85
Norway	64	67	74	76	79	80	82	85			
Poland	53	57	58	72	76	78	81	85			
Portugal	70	76	77	78	79	80	81	82			
Slovakia	65	75	76	77	78	80	81	85			
Slovenia	73	77	78	80	81	83	84				
Spain	73	74	75	78	80	82	84	85			
Sweden	64	67	75	76	77	78	79	82	83	84	85
Switzerland	68	71	73	75	76	78	79				
UK	67	73	75	76	79	80	81	83	84	85	

In relation to each country, we report the ages for which both one of the 15 smallest values of the $\Delta\hat{\zeta}_x s$ and one of the 15 smallest values of the $\hat{\sigma}_x s$ contextually occur

highest fitting bias and the largest riskiness of the Lee–Carter model in the description of the GGR.

5 Further investigations: functional cluster analysis of the GGR

The goodness of fit, already studied on the basis of the discrete trend of ζ and σ (by ages), can be explored according to a different perspective, aimed at jointly capturing the evolutionary and magnitude characteristics of the two parameters. Within this context, we use a procedure based on the technique of the functional multi-country clustering applied to the parameters ζ and σ. Through this procedure, the previous discrete analysis is now developed by means of a continuous representation, obtained by suitable interpolating curves. The goodness of fit of the stochastic GGR error process is involved in the trend of the interpolating curves, that jointly highlight magnitude and evolutionary features of the parameters ζ and σ, respectively.

5.1 A brief outline on the fundamentals of the functional cluster analysis

In this section, we describe the functional multi-country clustering of the parameters ζ and σ as described in Sect. 3, formula (6). In particular, for each fixed age, we obtain the optimal values of ζ and σ, reflecting the dynamics of the in-sample errors $Y^{\text{GGR}}(t, x^f)$, as extrapolated over the 20-year time period [2009–2019]. We analyse the evolution of ζ and σ over ages in different countries and produce a comparative analysis, by implementing a functional data clustering of the trends through data interpolation, which returns a curve over ages for each country. Through the spline functions, it is possible to identify suitable mathematical functions to describe these curves. Once these functions have been identified, the clustering algorithm allows us to highlight groups of countries that have shown similarities in the trend of the two parameters over ages.

In the statistical literature, functional data clustering has been introduced for univariate time series (James and Sugar 2003; Tarpey and Kinateder 2003; Chiou and Li 2007; Bouveyron and Jacques 2011). Afterwards, multivariate time-series functional data clustering has been proposed by Jacques and Preda (2014b), aiming at taking into account the dependence among curves and catching the similarities between them. In Jacques and Preda (2014a), the authors describe four methods to implement the functional clustering of time series. In the raw-data approach, the clustering is performed directly on the discretization of the functions at some time points, working on a very large amount and being time-consuming. In the two-stage approach, firstly data are approximated by curves through basis expansion or principal component tool (James and Sugar 2003), and then, the clustering is performed. The reducing dimension step generally consists in approximating the curves into a finite basis of functions. In Abraham et al. (2003), the two-step method consists in a k-means clustering performed on the B-spline coefficients. Wahba (1990) discusses the attractiveness of B-spline expansions, able to grasp the nonlinear effects in multivariate dataset. Kayano et al. (2010) implement a functional cluster analysis via orthonormalized Gaussian basis expansions.

As in Ferraty and Vieu (2006), in the nonparametric approach, hierarchical clustering is applied to assigned measures of distances or dissimilarities among functional curves. Finally, in the model-based clustering, an assumption of density probability generating data is considered.

In this paper, the values of ζ and σ referred to each age belonging to the interval 45–85 have been obtained as the optimal values of these parameters, calibrating the CIR process in the time interval 2000–2019. They are fixed for each age and synthetize the in-sample errors process dynamics revealed by that data on that time horizon.

In this paper, we have selected the two-step method to cluster the multivariate series of $\hat{\zeta}$ and $\hat{\sigma}$, respectively, with respect to the ages. Each of these parameters can be described through its functional form, as a function of the age, considering the values of the observed parameters as its realizations. Consequently, it exists a stochastic process underlying the family of random variables that generates the observed values at each age point. In a $n-$dimensional setting, the analysis deals with n curves.

Let $\gamma_t^j(x)$ (where $\gamma = \hat{\sigma}$ or $\hat{\zeta}$) be the observed value of the parameter γ for the population $j \in [1, ..., n]$ at age $x \in [x_1, ..., \omega]$ in a given year t.

Let us consider the following basis expansion:

$$\gamma_t^j(x) = \sum_{l=1}^{L} \alpha_{jl} \cdot \phi_l(x) \tag{8}$$

where $\phi_l(x)$ with $L \geq 1$ is the selected number of basis functions, $\alpha_{jl} \in \mathbb{R}$ are the coefficients and $\phi_l(x)$ are the basis functions. The basis coefficients are estimated from the observed values using the least-square estimation.

Given the discrete observations γ_t^{jk} of each sample path $\gamma_t^j(x)$ at a finite set of knots $\{x_{jk} : k = 1, ..., m_j\}$, the functional predictor becomes:

$$\gamma_t^{jk} = \gamma_t^j(x_{jk}) + \varepsilon_{jk} \quad k = 1, ..., m_j$$

with ε_{jk} independent and identically zero mean distributed errors.

The basis coefficients of each sample path $\sigma_t^j(x)$ are then approximated by:

$$\hat{\alpha}_j = (\Theta_j' \Theta_j)^{-1} \Theta_j' \tilde{\sigma}_t^j$$

with $\hat{\alpha}_j = (\hat{\alpha}_{j1}, ..., \hat{\alpha}_{jL})'$, $\Theta_j = (\phi_l(x_{jk}))_{1 \leq k \leq m_j, 1 \leq l \leq L}$ and $\tilde{\gamma}_t^j = (\gamma_t^{j1}, ..., \gamma_t^{jm_j})'$.

Equation (8) provides a functional representation of the curve $\gamma_t^j(x)$ allowing to handle the rough surfaces.

Once the functional form of each curve is derived, in the second step we implement the k-means method to cluster the coefficients of the expansions of the curves of the parameter γ, in order to capture the dependence on the entire curves among the considered populations. The k-means is an iterative clustering algorithm, that assigns a data point to a cluster working on the Euclidean distance, so that the sum of the squared distance between each data point and the arithmetic mean of all data points in the cluster is minimized.

5.2 Cross-country analysis

We use functional cluster analysis to assess cross-country similarities in the dynamics of ζ over age, that is $\zeta(x)$, as described in Sect. 5.1. Through B-splines, we obtain, from the 41 $\hat{\zeta}_x s$ we observe by age, an approximating functional form, whose parameters are calibrated, according to the OLS methods, at a finite set of knots (one for each couple of consecutive ages). This allows us to obtain, for each country under study, a functional representation of the curve of $\zeta(x)$ with respect to age, for which we know the coefficients expressing the magnitude and the slope for each knot. Once the functional form of each curve is derived, in a second step, we implement the k-means method to cluster the coefficients of the expansions of the curves of $\zeta(x)$, in order

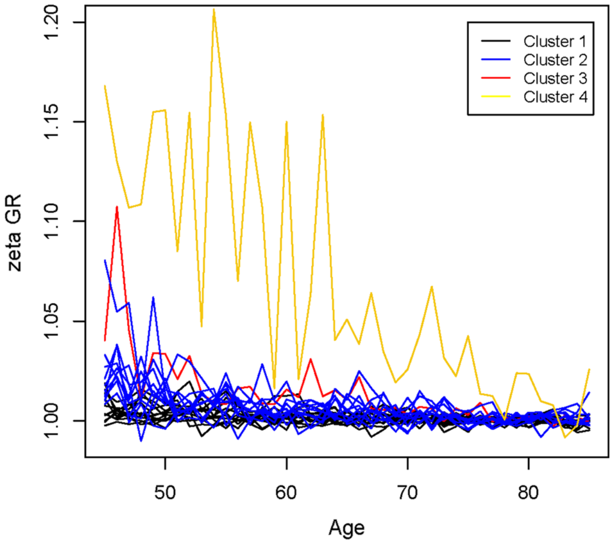

Fig. 6 The figure displays the curves describing how $\zeta(x)$ evolve over age, for each of the 25 countries under study. The black colour identifies countries that, according to the performed functional cluster analysis, belong to cluster 1, whereas colours blue, red and yellow are associated, respectively, with clusters 2, 3 and 4 (Color figure online)

to capture the dependence on the entire curves among the considered countries. The k-means is an iterative clustering algorithm, that, through minimization of within-cluster variances specifically, squared Euclidean distances), allows to partition the obtained curves $\zeta(x)$ into a given number of clusters, based on their similarities, in terms of magnitude and shape. Our outcomes, based on the Elbow method, suggest to partition the 25 curves (one per country) describing how $\zeta(x)$ evolve over age into four clusters, depicted in Fig. 6 and listed in Table 4. In Fig. 6, the black colour identifies countries that, according to the performed functional cluster analysis, belong to cluster 1, whereas colours blue, red and yellow are associated, respectively, with clusters 2, 3 and 4. Clusters differ by the magnitude of parameter $\zeta(x)$ and by how it evolves over age. The major differences among countries relate to the oscillation that $\zeta(x)$ experiences for younger ages in the group [45–65]. Because of this more marked oscillation, Slovenia (cluster 3) is set apart from the countries belonging to cluster 2. As it turned out from the empirical analysis developed in Sect. 4.1, Luxembourg stands out sharply from the other countries under study, because of the higher magnitude of $\zeta(x)$ for each x and the higher oscillation age by age.

We proceed analogously to assess cross-country similarities in the dynamics of σ over age, that is $\sigma(x)$, as described in Sect. 5.1. Our outcomes, based on the Elbow method, suggest to partition the 25 curves (one per country) describing how $\sigma(x)$ evolve over age into four clusters, depicted in Fig. 7 and listed in Table 5. As observed for $\zeta(x)$, we find that Slovenia and Luxembourg stands out from the other clusters, pooling larger groups of nations. The functional form of the curve, other than the

Table 4 Cross-country comparison of the dynamics of ζ over age: partition into clusters

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Belgium	Austria	Slovenia	Luxembourg
Bulgaria	Denmark		
Czech Republic	Estonia		
France	Finland		
Germany	Ireland		
Greece	Lithuania		
Hungary	Norway		
Italy	Slovakia		
Netherlands	Sweden		
Poland	Switzerland		
Portugal			
Spain			
UK			

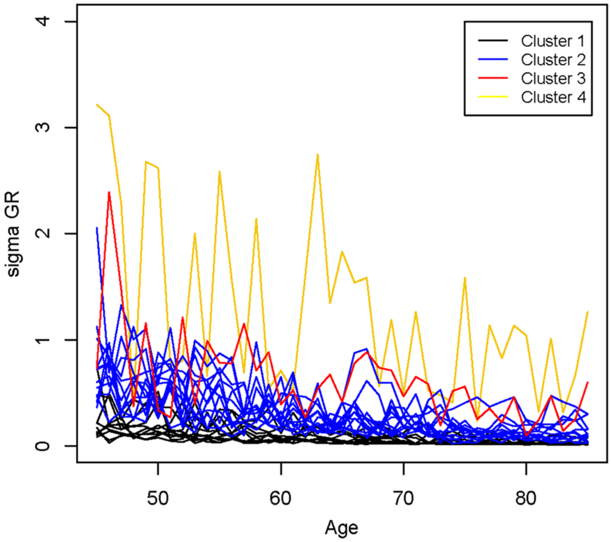

Fig. 7 The figure displays the curves describing how $\sigma(x)$ evolve over age, for each of the 25 countries under study. The black colour identifies countries that, according to the performed Functional Cluster Analysis, belong to cluster 1, whereas colours blue, redf and yellow are associated, respectively, with clusters 2, 3 and 4 (Color figure online)

magnitude, plays the major role in discriminating cluster 3 (Slovenia) from cluster 2, especially with respect to younger ages.

The study performed through functional cluster analysis is more refined in the unfolding of the fitting bias and model risk characterizing the Lee–Carter model in the description of the gender gap ratio. Furthermore, by virtue of the peculiarities of this procedure and of the underlying algorithm, under a cross-country perspective, we

Table 5 Cross-country comparison of the dynamics of σ over age: partition into clusters

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Belgium	Austria	Slovenia	Luxembourg
Bulgaria	Denmark		
Czech Republic	Estonia		
France	Finland		
Germany	Greece		
Hungary	Ireland		
Italy	Lithuania		
Netherlands	Norway		
Poland	Portugal		
Spain	Slovakia		
UK	Sweden		
	Switzerland		

can grasp similarities in the performance of ζ and σ, according to a joint view of the evolutionary and the magnitude variations for each parameter.

Identifying such similarities allows to categorize the goodness of fit of the LC model, providing clusters of countries identified along with the benchmarks for achieving the "best-fitting" results; each cluster shows similarities in the trend of the two parameters over age. Cluster 1, namely the black curves in the figures, thus groups those nations for which there is a better goodness of fit of the LC model for describing the GGR, while for Slovenia and for Luxembourg there is evidence of a poor adequacy of the LC model to represent the true trend of the GGR.

6 Conclusions

The study of gender differences in survival, in this paper expressed and studied as the gender gap ratio (GGR), the ratio between male and female mortality rates, is the core of the work. The aim is to assess the ability of the Lee–Carter model to accurately represent the GGR real data trends in a multi-country perspective, in order to provide application guidelines. The results provide elements to infer which are the geographical areas where the LC is more effective in representing the GGR.

The paper focuses on the description and analysis of errors between real data and GGR values estimated with the LC model, by using the stochastic dynamic Cox–Ingersoll and Ross model and in particular by studying the parameters of this process. The parameters are, in our opinion, particularly interesting and explanatory with respect to the scope of the work: the estimated parameter ζ, the long-term average of the error process, acts as a measure of the fitting bias. It not only describes the historical attitude of the Lee–Carter model to under or over-estimate the gender gap ratio, but also summarizes long-run potential of the mortality model to accurately approximate the trend of the GGR, as extrapolated from the data. In fact, the idea of studying the error process between real GGRs and GGRs estimated through the LC model implies that the closer the long-term mean of the CIR process represented by the parameter ζ is to 1, the higher the quality of the LC model in describing the real GGR data. Hence, the significance of the long-term mean parameter as a model benchmark and thus as a further quantitative selection criterion among competing mortality models.

Noting then that the CIR of the in-sample errors deviates from its deterministic path towards its long-term mean because of a diffusion term, whose intensity is constant and equal to the parameter σ, such a parameter represents a measure of risk, understood as the risk that the information provided by the estimated long-term mean is unreliable. Thus, we can state that the higher σ, the more powerful the diffusion term and, thus, the more unstable, or volatile, over time the goodness of fit of the Lee–Carter model in relation to a specific age. It follows that σ plays an important role, given that stability is widely recognized as an important feature of stochastic mortality models.

To summarize, the LC model is better at representing actual GGR data the closer ζ approaches 1 and σ approaches 0. The overall analysis developed on these parameters (long-term mean and volatility) for 25 European countries provides an interpretation of the extent of the fitting bias from which a model risk analysis arises. We perform the study under two different approaches. The first consists in a widespread exploration of the evidence emerging from the comparison of fitted and observed data in order to capture some significant stylized facts. The second is based on a functional cluster analysis, carried out on the same countries, which allows to grasp similarity characters in the performance of the considered parameters, according to a joint view of evolutionary and magnitude variations for each of them.

The work focuses on the adult ages, believing it particularly significant to investigate the phenomenon where it is realized in a greater number of survivors of one gender than another and therefore has a greater economic, political and social impact. The value of the information the work aims to provide is important to correctly guide political and social choices in certain areas related to the life cycle. As is well known, a longer life expectancy for females is characterized by a compromised state of health, which will cause not only a higher number of female elderly to live, but also a higher number of females with fragile health. This requires a more cautious approach to protecting old age women under difficult health conditions, for example by implementing prevention programmes targeted at diseases characterizing one gender rather than another. In addition, other things being equal, women generally have lower retirement income than their male counterparts. Furthermore, the results obtained can support the role of insurance over the life course: from an insurer's point of view, it is also important to know, also in terms of internal control, whether a portfolio of pension annuities is biased to one gender or another. The interplay among these elements thus requires preventive action from a health point of view and prudent planning from a financial point of view. Active and healthy ageing is indeed a primary objective for national and supranational authorities, and the methodological approach in this paper can be useful for identifying what public and private resources are capable of addressing the disparities in the ageing process. But also, in a variety of other fields, gender differences in survival can cause orientations in social behaviour: for example in electoral choices, in the greater or lesser propensity to use the media and the social media in particular, in marketing strategy and in many other fields of human behaviour in which scientific branches such as sociology, health policies, economics and economic policy find their application.

Also in this context, we have paid attention to the circumstance that currently supranational organizations draw up regulatory structures with the aim of more nations adopting them and making them their own, which is why the work extends the analysis to 25 European countries. The paper provides a "local" picture of the GGR and of the greater or lesser ability of the Lee–Carter model to well capture actual GGR data in a certain geographic area, where specific clinical, economic, and anthropogenic disparities are to be grasped. This in-depth analysis, carried out using cluster analysis methodology, made it possible to isolate geographical areas for which the reliability of the Lee–Carter model as a description of mortality appears questionable.

In light of this, the trend and the size of the GGR take on an important role of providing information, being a basic tool to build new welfare structures and financially sustainable policies; a deeper level of understanding of these elements allows to define solutions aimed at senior protection and well-being. The estimation of GGR through the most effective and popular model, the Lee–Carter model, is therefore of substantial importance in this framework of interest.

Future research lines can be developed throughout a interdisciplinary perspective, deepening the specific determinants from which the quantitative results originate. In particular, this analysis could inspire and innovate financial/insurance tools and social policies which support the ageing process.

Last but not least, within a financial education program, knowing the quantitative dimension of the GGR is culturally significant, increasing the awareness of elderly people (and not only) to take informed decisions based on the life-cycle assessment.

Acknowledgements We thank the anonymous reviewers for their very thoughtful comments.

Funding Open access funding provided by Università degli Studi di Udine within the CRUI-CARE Agreement.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article. Emilia Di Lorenzo and Gabriella Piscopo were partially supported by the Grant Agreement "Progetto Partenariati Estesi - PNRR Project "Age-It - Ageing Well in an ageing society "– Codice MUR: PE0000015- CUP E63C22002050006".

References

Abraham, C., Cornillon, P.A., Matzner-Løber, E., Molinari, N.: Unsupervised curve clustering using B-splines. Scand. J. Stat. **30**(3), 581–595 (2003)

Apicella, G., Dacorogna, M., Di Lorenzo, E., Sibillo, M.: Improving the forecast of longevity by combining models. N. Am. Actuar. J. **23**(2), 298–319 (2019)

Apicella, G., De Giorgi, E.: Gender attitudes toward longevity and retirement planning: theory and evidence. In: Corazza, M., Perna, C., Pizzi, C., Sibillo, M. (eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance. MAF 2022, pp. 19–24. Springer, Cham (2022)

Apicella, G., De Giorgi, E. G., Di Lorenzo, E., Sibillo, M.: Gender-inclusive financial and demographic literacy: lessons from the empirical evidence. Swiss Finance Institute Research Paper No. 23-02. Available at SSRN: https://ssrn.com/abstract=4331841 or https://doi.org/10.2139/ssrn.4331841 (2023)

Aristei, D., Gallo, M.: Assessing gender gaps in financial knowledge and self-confidence: evidence from international data. Financ. Res. Lett. **46**, 102200 (2022)

Ayuso, M., Bravo, J.M., Holzmann, R.: Getting life expectancy estimates right for pension policy: period versus cohort approach. J. Pension Econ. Finance **20**(2), 212–231 (2021)

Barford, A., Dorling, D., Smith, G.D., Shaw, M.: Life expectancy: women now on top everywhere. Br. Med. J. **332**(7545), 808 (2006)

Ugofilippo, B., Camarda, C.G., Booth, H.: Thirty years on: A review of the Lee-Carter method for forecasting mortality. Int. J. Forecast. **39**(3), 1033–1049 (2023). https://doi.org/10.1016/j.ijforecast.2022.11.002

Bergeron-Boucher, M.-P., Canudas-Romo, V., Pascariu, M., Lindahl-Jacobsen, R.: Modeling and forecasting sex differences in mortality: a sex-ratio approach. Genus **74**(20), 1–28 (2018)

Bergeron-Boucher, M.-P., Kjærgaard, S.: Mortality forecasting at age 65 and above: an age-specific evaluation of the Lee–Carter model. Scand. Actuar. J. **2022**(1), 64–79 (2022)

Booth, H., Tickle, L.: Mortality modelling and forecasting: a review of methods. Ann. Actuar. Sci. **3**(1–2), 3–43 (2008)

Booth, H., Hyndman, R.J., Tickle, L., de Jong, P.: Lee–Carter mortality forecasting: a multi-country comparison of variants and extensions. Demogr. Res. **15**, 289–310 (2006)

Bouveyron, C., Jacques, J.: Model-based clustering of time series in group-specific functional subspaces. Adv. Data Anal. Classif. **5**(4), 281–300 (2011)

Cairns, A.J.G., Blake, D., Dowd, K.: A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. J. Risk Insur. **73**(4), 687–718 (2006)

Cairns, A.J.G., Blake, D., Down, K., Coughlan, G.D., Epstein, D., Ong, A., Balevich, I.: A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. N. Am. Actuar. J. **13**(1), 1–35 (2009)

Chiou, J.M., Li, P.L.: Functional clustering and identifying substructures of longitudinal data. J. R. Stat. Soc. Ser. B Stat. Methodol. **69**(4), 679–699 (2007)

Coelho, E., Nunes, L.C.: Forecasting mortality in the event of a structural change. J. R. Stat. Soc. A. Stat. Soc. **174**(3), 713–736 (2011)

Coppola, M., Russolillo, M., Simone, R.: On the evolution of the gender gap in life expectancy at normal retirement age for OECD countries. Genus **78**, 27 (2022)

Cox, J.C., Ingersoll, J.E., Ross, S.A.: A theory of the term structures of interest rates. Econometrica **53**(2), 385–407 (1985)

Eurostat: Ageing Europe—statistics on pensions, income and expenditure. Available at https://ec.europa.eu/eurostat/statistics-explained/index.php?_income_and_expenditure&oldid=500271. ISSN:2443-8219 (2020)

Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics. Springer, Berlin (2006)

Flor, L.S., Friedman, J., Spencer, C.N., Cagney, J., Arrieta, A., Herbert, M.E., Stein, C., Mullany, E.C., Hon, J., Patwardhan, V., Barber, R.M., Collins, J.K., Hay, S.I., Lim, S.S., Lozano, R., Mokdad, A.H., Murray, C.J.L., Reiner, R.C., Sorensen, R.J.D., Haakenstad, A., Pigott, D.M., Gakidou, E.: Quantifying the effects of the COVID-19 pandemic on gender equality on health, social, and economic indicators: a comprehensive review of data from March, 2020, to September, 2021. The Lancet **399**(10344), 2381–2397 (2022)

Generali Group: Emerging Risks Booklet. 2022 Update Group Risk Management. Available at https://www.generali.com/what-we-do/emerging-risks (2022)

Glei, D.A., Horiuchi, S.: The narrowing sex differential in life expectancy in high-income populations: effects of differences in the age pattern of mortality. Popul. Stud. **61**(2), 141–159 (2007)

Grevenbrock, N., Groneck, M., Ludwig, A., Zimper, A.: Cognition, optimism and the formation of age-dependent survival beliefs. Int. Econ. Rev. **62**(2), 887–918 (2021)

Heimer, R.Z., Myrseth, K.O.R., Schoenle, R.S.: YOLO: mortality beliefs and household finance puzzles. J. Finance **74**(6), 2957–2996 (2019)

HMD: Human Mortality Database, Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at www.mortality.org (data downloaded on February 8, 2023) (2023)

Hurwitz, A., Mitchell, O.S., Sade, O.: Testing methods to enhance lonhevity awareness. J. Econ. Behav. Org. **204**, 466–475 (2022)

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

ISTAT: Il futuro demografico del paese. previsioni regionali della popolazione residente al 2065 (base 1.1.2017). Available at: https://www.istat.it/it/files//2018/05/previsioni_demografiche.pdf (2018)

Jacques, J., Preda, C.: Functional data clustering: a survey. Adv. Data Anal. Classif. **8**(3), 231–255 (2014a)

Jacques, J., Preda, C.: Model-based clustering for multivariate functional data. Comput. Stat. Data Anal. **71**, 92–106 (2014b)

Jallbjørn, S., Jarner, S.F.: Sex differential dynamics in coherent mortality models. Forecasting **4**, 819–844 (2022)

James, G., Sugar, C.: Clustering for sparsely sampled functional data. J. Am. Stat. Assoc. **98**(462), 397–408 (2003)

Kalmi, P., Ruuskanen, O.-P.: Financial literacy and retirement planning in Finland. J. Pension Econ. Finance **17**(3), 335–362 (2018)

Kayano, M., Dozono, K., Konishi, S.: Functional cluster analysis via orthonormalized Gaussian basis expansions and its application. J. Classif. **27**, 211–230 (2010)

Kladivko, K.: Maximum likelihood estimation of the Cox–Ingersoll–Ross process: THE MATLAB implementation. https://it.mathworks.com/matlabcentral/fileexchange/37297-maximum-likelihood-estimation-of-the-cox-ingersollross (2012)

Lee, R.D., Carter, L.R.: Modeling and forecasting U.S. mortality. J. Am. Stat. Assoc. **87**(419), 659–671 (1992)

Lusardi, A., Mitchell, O.: Planning and financial literacy: how do women fare? Am. Econ. Rev. **98**(2), 413–417 (2008)

MEF: Le disuguaglianze di genere in Italia e il potenziale contributo del Piano Nazionale di Ripresa e Resilienza per ridurle. Ministero dell'Economia e delle Finanze. Dipartimento di Ragioneria generale dello Stato. Studi e pubblicazioni. Available at https://www.rgs.mef.gov.it/_Documenti/VERSIONE-I/Comunicazione/Workshop-e-convegni/Seminario_Il_Piano_Nazionale_di_Ripresa_e_Resilienza_e_le_diseguaglianze_di_genere/PNRR-GM-ex-ante.pdf (2021)

OECD: Improving retirement incomes considering behavioural biases and limited financial knowledge. OECD Pensions Outlook 2018, OECD Publishing Paris (2018)

Peeters, H., De Tavernier, W.: Lifecourses, pensions and poverty among elderly women in Belgium: interactions between family history, work history and pension regulations. Ageing Soc. **35**(6), 1171–1199 (2015)

Rau, R., Soroko, E., Jasilionis, D., Vaupel, J.W.: Continue reductions in mortality at advanced ages. Popul. Dev. Rev. **34**(4), 747–768 (2008)

Schünemann, J., Strulik, H., Trimborn, T.: The gender gap in mortality: how much is explained by behavior? J. Health Econ. **54**, 79–90 (2017)

Tarpey, T., Kinateder, K.: Clustering functional data. J. Classif. **20**(1), 93–114 (2003)

Villegas, A.M., Millossovich, P., Kaishev, V.K.: Stmomo: an R package for stochastic mortality modelling. https://cran.r-project.org/web/packages/StMoMo/vignettes/StMoMoVignette.pdf (2016)

Wahba, G.: Spline Models for Observational Data. Society for Industrial and Applied Mathematics. SIAM, Philadelphia (1990)

Yakoboski, P.J., Lusardi, A., Hasler, A.: Financial literacy, longevity literacy and retirement readiness. The 2022 TIAA Institute-GFLEC Personal Finance Index. TIAA Institute Research Paper. Available at SSRN: https://ssrn.com/abstract=4338957 or https://doi.org/10.2139/ssrn.4338957 (2022)

Zarulli, V., Kashnitsky, I., Vaupel, J.W.: Death rates at specific life stages mold the sex gap in life expectancy. PNAS **118**(20), e2010588118 (2021)