Addressing the Life Expectancy Gap in Pension Policy

Jorge M. Bravo[1] Mercedes Ayuso[2] Robert Holzmann[3]
Edward Palmer[4]

August 25, 2020

[1]Universidade Nova de Lisboa - NOVA IMS & Université Paris-Dauphine PSL & MagIC & CEFAGE-UE; Corresponding author: Email: jbravo@novaims.unl.pt. Address: Campus de Campolide, 1070-312 Lisbon-Portugal.

[2]Full professor in Actuarial Statistics at University of Barcelona (Department of Econometrics, Statistics and Applied Economy, Riskcenter-UB). Director of the Master in Management of Insurance and Financial Organizations at University of Barcelona.

[3]Governor of the Austrian National Bank and Elected Fellow of the Austrian Academy of Sciences (Vienna); robert.holzmann@oeaw.ac.at

[4]Professor, Uppsala Center for Labor Studies and Department of Economics, Uppsala.

Abstract

Understanding the systematic relationship between period and cohort life expectancy and how the relationship evolves over time are critical issues in formulating the design of retirement income products, evaluating the actuarial balance of pension schemes, and more generally for all analyses where demographic projections are involved. In this study, estimates of the life expectancy gap at all ages are performed using data for 1960-2018 from the Human Mortality Database and projections are generated through 2050 for the 42 national populations, disaggregated by gender. Contrary to previous research that often uses a single deemed to be «best» model to forecast mortality rates, we use a novel adaptive Bayesian Model Ensemble of heterogeneous parametric generalized age-period-cohort stochastic mortality models, principal component methods, and smoothing approaches. The procedure involves both the selection of the model confidence set and the determination of optimal weights. Model-averaged Bayesian credible prediction intervals are derived accounting for both the uncertainty arising from model error and parameter uncertainty. With intergenerational actuarial fairness and neutrality as the guiding principles the study then explores potential policy interventions to address the consequences of the life expectancy gap - spanning over adjustments in the accumulation, benefit determination, and payout stages. Comprehensive numerical results are provided for two policy options: (i) introducing a sustainability factor; and (ii) conditional pension indexation. The results show that: (i) the life expectancy gap is positive and significant for almost all countries and years studied, (ii) it will continue to increase, (iii) the magnitude of the subsidy rates between generations can be sizeable demanding important initial pension benefit reduction and/or a gradual diminution in the annual indexation rate of pensions to correct them.

JEL Code: H55, G22, C63, C53, H23.

Keywords: Period and cohort life expectancy; Pension reforms; Bayesian Model Ensemble; Mortality forecasting; Pension indexation; Sustainability factor; Intergenerational fairness.

1 Introduction

Steadily increasing longevity and population ageing constitute a challenge to the financial outcomes of all pension schemes. In the end the financial outcome is all about getting the projection of life expectancy right. To attain the sustainability and/or the solvency of pension systems, in recent decades most OECD countries have responded to continuous longevity improvements either with systemic or parametric pension reforms or a combination of both. For public national pension schemes, a common denominator of most – if not all – reforms has been to create an automatic link of future pensions to changes in population life expectancy, usually employing the projections of national statistical agencies. The link between life expectancy and pension benefits has been tightened in at least seven different ways (Whitehouse, 2007; OECD, 2017; Ayuso, Bravo and Holzmann, 2020): (i) by linking the life expectancy and/or other demographic markers to initial pension benefits, for instance through sustainability factors and/or reduction coefficients (e.g., Germany, Finland, Portugal, Spain, Japan); (ii) by linking the normal retirement age to life expectancy (so far 10 countries including Denmark, Italy, the Netherlands, and Portugal); (iii) by linking the qualifying conditions to life expectancy (e.g., years of contributions needed for a full pension), targeting a constant ratio between the contribution period for a full pension and the duration of retirement (e.g., France); (iv) by linking penalties (bonuses) for early (late) retirement to years of contributions and normal retirement age (e.g., Portugal); (iv) by replacing traditional Notionally Defined Benefit (NDB) public PAYG schemes with Nonfinancial (Notional) Defined Contribution (NDC) schemes that replicate the main features of Financial Defined Contribution (FDC) plans, namely the way in which the computation of the benefit (annuity) is based on, using the projected average life expectancy of the retiree's birth cohort at the year the benefit is claimed, automatically adjusting benefit payments to changes in life expectancy (e.g., Sweden, Poland, Latvia, Italy, Norway); (vi) by phasing in national FDC plans as a (usually partial) substitute in the reform of their Nonfinancial Defined Benefit (NDB) pension schemes (e.g., Chile, Latvia, Mexico, Poland, and Sweden); (vi) by conditioning pension indexation on the pension funds' solvency position (e.g., the Netherlands) or modifying the annual account indexation rate through the use of a solvency-based "extra" balancing mechanism in NDC schemes (e.g., Sweden).

Previous research has also emphasized the role that can be played by automatic balancing mechanisms to improve the solvency of a pension system (Turner, 2009; Vidal-Meliá et al., 2009). The introduction of these automatic balance mechanisms represents a fundamental change in the way aggregate longevity risk is shared between contributors and retirees, and between current and future generations, making it more explicit and based on automatic rules. Although designed to maintain financial solvency and through this sustainability, balancing mechanisms also impact on intergenerational fairness and equity. And, depending on other parameters of the scheme (especially the method of valorisation of accumulated pension rights) they may also affect adequacy. The way these automatic longevity risk sharing mechanisms have been introduced in pension schemes suffers, however, from various weaknesses. Although the precise outline of pension reforms differs among countries, they typically share one common feature: in almost all cases and countries, unisex life expectancy measures computed from period and not cohort life tables have been used to automatically link ex-ante longevity improvements and pension benefits. Abstracting from the well-known sex gradient in life expectancy, in a scenario of continuous decline in age-specific mortality rates, the use of period life expectancy results in a systematic underestimation of the remaining lifetime at retirement, incorrectly signalling solvency prospects and, as a result, delaying or watering-down pension reforms.

The underestimation of future longevity generates unintended and potentially sizable implicit subsidies from future to current generations. By giving a false signal and an unfair actuarial link between contributions and pension entitlements, underestimation also distorts labor supply decisions and leads to macroeconomic inefficiency. In this way, underestimated life expectancy counteracts the objectives of recent reform approaches targeting a strengthening of contributory principles and actuarial fairness on a lifetime basis. Besides, for pension policy, the issue of the correct estimation of the future life expectancy is also amplified when ex-ante differences in mortality are observed, the longevity improvements are not homogenous across socioeconomic groups, there is high lifespan inequality at retirement and subjective mortality beliefs differ significantly from objective longevity measures, translating into implicit intragenerational tax/subsidy effects and ineffective financial planning (NASEM, 2015; Ayuso, Bravo and Holzmann, 2017a,b; Chetty et al., 2016; Sasson, 2016; Cairns et al., 2019; Alaminos and Ayuso, 2019; Sánchez-Romero et al., 2020; Chen et al., 2020; Palmer and Zhao de Gosson de Varennes, 2020; Holzmann et al., 2020a,b). In recent decades observed longevity improvements result mostly from a continuous linear shift from younger to older ages in the distribution of mortality reductions (Alho, Bravo and Palmer, 2013). In the future, this will lead to a further underestimation of the longevity of retired cohorts with an increasing impact on social welfare systems.

The goal of the life annuity is to secure an adequate, stable and predictable lifelong income stream and providing a cost effective and efficient risk pooling mechanism that addresses the individual uncertainty of the timing of death. In fulfilling this objective pension schemes redistribute income in a welfare-enhancing manner. By construction, annuity contracts generate ex-post redistribution, as some individuals die early and forfeit their resources to those who die later. This type of inequality is not perceived as a negative aspect of the system since it is the consequence of the longevity insurance (pooling) device.

Understanding the relationship between period and cohort life expectancy measures and quantifying the size of the gradient and how it evolves over time are all critical issues for pension design and reform, for pricing retirement income products and more generally for all analyses contingent on demographic development of the population. Ayuso, Bravo and Holzmann (2020) introduced the concept of life expectancy gap to measure the systematic difference between cohort and period life expectancy at a given age and time and expressed it in terms of the Lee and Carter (1992) stochastic mortality model. The life expectancy gap expresses (in years of life) by how much period life expectancy at age x in year t differs from the life expectancy of the cohort attaining age x in year t. It represents, when positive, an estimate of the extra years of life a given cohort will enjoy as a result of expected future mortality improvements. For pension policy, the life expectancy gap is a proxy, at the individual level, of the scope of unfunded pension liabilities due to the use of an incorrect measure of remaining lifetime, i.e., of the amount by which pension wealth exceeds the value of the accumulation, a measure of the implicit debt transferred to future generations unless corrective actions are undertaken. In their work, the authors explore the limited comparable country data on period and cohort life expectancy available from official sources for selected countries (Australia, UK, U.S.) and own estimates obtained for Portugal and Spain to conclude that the gap at birth exceeds 10 years for the countries and periods analysed. The gap is projected to decrease gradually over time (suggesting a reduced margin for survival improvements) and greater differences are observed for men compared to women. The magnitude of the life expectancy gap at retirement (age 65) is naturally smaller but the size of the implicit subsidy rate is still sizable (with an average of nearly 20% in 2010), comparable to that created by longevity heterogeneity by socioeconomic status.[1]

In this paper we provide detailed comparable cross-country estimates of the life expectancy gap at retirement ages for 42 homogeneous national populations (countries or areas), disaggregated by sex, from 1960 to 2050. The analysis covers medium- and high-income countries in North America, Central and Western Europe, Asia and the Pacific, as well as Chile in Latin America. Assuming that the benefit indexation equals the discount rate, which broadly holds for wage-indexed pensions, the life expectancy gap estimates are then used to quantify the size of the underfunded pension liabilities attributed to the adoption of an inappropriate life expectancy measure. Taxes/subsidies between current and future generations are thus determined by the extent to which pension liabilities are not backed by the accumulated contribution assets. The subsidy (tax) rate measures the rate at which own accumulations would need to increase (decrease) to achieve the same benefit level as that resulting from the use of period

[1] The authors note, however, that further research is needed to have a comparable and comprehensive cross-country measure of the implicit tax/subsidies, given the methodological differences in life expectancy computation detected in previous studies.

life expectancy in computing the actuarial factor to convert the accumulated financial/notional contributions into pension benefits at retirement.

We then explore several potential policy interventions to address the life expectancy gap (and longevity heterogeneity) in three stages of the process in creating a pension: the accumulation, the benefit determination, and the disbursement stage. Policy interventions involve adapting the key parameters of the pension system in line with life expectancy developments. These can consist of adjusting the initial pension benefits, modifying the qualifying conditions and the statutory retirement age and diverse risk-sharing mechanisms. Examples of the latter are: longevity-linked life annuities, conditioning pension indexation to observed deviations between expected and observed life expectancy, deferred annuities with a sharing of common and asymmetric longevity development between annuity calculation and disbursement. In this paper numerical results are provided for two policy options: (i) introducing a sustainability factor; (ii) conditioning pension indexation.[2] The principle guiding the choice and calibration of policy interventions is the criterion of intergenerational actuarial fairness.

To compute the life expectancy gap by age, sex and calendar year, period and cohort life expectancy must be estimated. In addition to scenario and expert-based approaches and explanatory (e.g., cause of death decomposition) models, mortality is usually forecasted by employing one of three broad classes of models: parametric models, principal component methods, and smoothing approaches. The number of models examining the structure of mortality rates across the dimensions of age, period and cohort (and other covariates) has been increasing. In the actuarial and demographic literature, a significant number of single population discrete-time extrapolative stochastic mortality models have been proposed for modelling the dynamics of mortality rates.[3] More recently, several multi-population discrete-time mortality models have been suggested to identify the underlying long-term mortality trends.[4] Additionally, a number of continuous-time affine jump diffusion models have also been proposed for modelling the dynamics of mortality rates and for insurance and capital market applications.[5] Contrary

[2]For alternative policy options to cope with the life expectancy gap (for instance, adjusting the statutory retirement age) see Bravo et al. (2020).

[3]See, e.g., Lee and Carter (1992), Brouhns et al. (2002), Renshaw and Haberman (2003, 2006), Currie, Durban and Eilers (2004), Cairns et al. (2006a, 2009), Hyndman and Ullah (2007), Booth and Tickle (2008), Plat (2009), Hunt and Blake (2014, 2020), Currie (2016), Hunt and Villegas (2017), Villegas, Millossovich and Kaishev (2018), Palmer, Alho, and Zhao de Gosson de Varennes (2019), Blake et al. (2019) and references therein.

[4]See, e.g., Li and Lee (2005), Cairns et al. (2011), Dowd et al. (2011), Li, Zhou and Hardy (2015), Zhu,Tan, and Wang (2017), Villegas et al. (2017) and references therein.

[5]See, e.g., Milevsky and Promislow (2001), Dahl (2004), Biffis (2005), Cairns, Blake and Dowd (2006a,b, 2008), Biffis and Millossovich (2006), Schrager (2006), Bravo (2011), Blackburn and Sherris (2013), Bravo and Nunes (2020) and references therein.

to static period life expectancy markers, cohort life expectancy computations are normally subjected to a considerable amount of model risk (i.e., risk from inappropriate model selection) as well as to parameter risk (i.e., risk from inaccurate estimation of model parameters) since past trends do not necessarily hold for the future. Because of this, most countries shy away from offering official cohort life tables. This restriction ultimately influences the policymaker's choice of a life expectancy measure for pension policy.

In this paper we employ a relatively new method in the context of actuarial studies involving multiple competing models – "ensemble learning methods". We (use the same data to) estimate models "in parallel" and then weight the model results together to obtain a single projection. The approach is based on the rationale that no specific model alone can consistently achieve best forecasts for a class of time series, but a combination of models can best approximate the actual data generation process. When compared to a single model, ensemble learning has demonstrated to improve traditional and machine learning forecasting. Despite the popularity and long tradition of ensemble learning methods in the statistical and forecasting literature (see, e.g., Makridakis and Winkler (1983)), model combination has received little attention in the actuarial, demographic or pension literature with some noticeable exceptions (see, e.g., Kontis et al. (2017) and Shang and Haberman (2018)). This approach offers an additional advantage in that it reduces the inherent uncertainty in the choice of the appropriate projection model (model risk), benefiting from the strengths of the constituent models and effectively reducing the errors resulting from flawed assumptions, bias or errors in the data.

The approach we employ in this paper is based on an adaptive Bayesian Model Ensemble (BME) of several heterogeneous models, including six well-known parametric single population Generalised Age-Period-Cohort (GAPC) stochastic mortality models, one single-population univariate functional demographic time-series model (the weighted Hyndman-Ullah method), one bivariate functional data model (the Regularized singular value decomposition model) and the recently proposed two-dimensional smooth constrained P-splines model, all of which can probabilistically contribute towards projecting future period and cohort life expectancy measures and the life expectancy gap. We consider a range of both existing and new models that have proven to perform well in fitting and forecasting empirical studies. Ensemble learning is the field of machine learning that combines different predictive models to address a given predictive task. The novel adaptive BME strategy adopted in this paper is motivated by the model confidence set procedure developed by Hansen et al. (2011). First, we select the subset of GAPC superior models among the set of nested candidates using a fixed-rule trimming scheme and considering the model's out-of-sample forecasting performance in the validation period. Second, after estimating the individual models we linearly combine the forecasts of the subset of superior GAPC models and the non-parametric and smoothing approaches by assigning weights conditional on the model's out-of-sample predictive accuracy. We combine individual forecasting models by weighting according to the inverse of each model's forecasting accuracy. In contrast with previous approaches focusing either on the selection of optimal combination schemes and weights (see, e.g., De Menezes et al. (2000), Jose and Winkler (2008), Andrawis et al. (2011), Hsiao and Wan (2014)) or assigning equal weights to the set of superior models (see, e.g., Stock and Watson (2004), Samuels and Sekkel (2017)), our BME procedure involves, for each subpopulation, both the selection of superior models (accounting for the existence of nested candidates) and the determination of optimal weights. To derive BME prediction intervals for the quantities of interest, we use the Model-Averaged Tail Area (MATA) construction proposed by Turek and Fletcher (2012) accounting for both the uncertainty arising from the error in the forecast of the individual stochastic mortality model parameters and parameter uncertainty resulting from model fitting.

We project that period and cohort life expectancy will continue to increase in 37 of the 42 countries analysed in this study for both the male and female populations, and that the sex differences in cohort life expectancy are expected to shrink in most countries. We show that the life expectancy gap is positive for almost all countries and years studied, confirming that period life expectancy measures tend to systematically underestimate human remaining lifetime, and that the gap at retirement ages is persistent and expected to increase in the future. Our results also show that the magnitude of the subsidy rates between generations can be sizeable, and approach 20 percent for several countries and periods analyzed. With the objective of eliminating these, we estimate that important initial pension benefit reduction is required and/or a gradual diminution in the annual indexation rate of pensions.

The remainder of the paper is organized as follows. Section 2 outlines the key concepts and research methods used in the paper, presents the adaptive Bayesian Model Ensemble approach for mortality modelling and life expectancy gap computation and describes the data used in fitting the models. Section 3 reports the results for the observed and forecasted size of the life expectancy gap by age, gender, country and across years, and estimates the implicit subsidy rates between generations introduced by policy reforms linking future pension benefits to period life expectancy. Section 4 discusses and explores alternative policy interventions at the initial benefit computation and disbursement stages addressing the intergenerational unfairness problems introduced by the life expectancy gap. Numerical illustrations for two of these new policy options are presented. Finally, in Section 5 the main conclusions and policy implications are presented and discussed. All accessory results are relegated to the appendix.

2 Materials and methods

This section outlines the key concepts and statistical methods used in the paper to estimate period and cohort life expectancy, to determine the size and the dynamics of the life expectancy gap and to assess the magnitude of the implicit tax/subsidy across time and between countries. The framework comprises an adaptive Bayesian Model Ensemble of stochastic mortality models to forecast the survival prospects and an actuarially fair and neutral setting for policy evaluation.

2.1 Life expectancy gap

Let $T_x(t)$ be the remaining lifetime of an individual aged x on his/her last birthday at time t. We consider the time interval $[0, \omega]$, with ω denoting the highest attainable age. Let ${}_\tau p_x\,(t)$ denote the τ-year survival rate of a reference population cohort aged x at time t:

$$ {}_\tau p_x\,(t) := \exp\left(-\int_0^\tau \mu_{x+s}\,(s)\,ds\right), \tag{1} $$

where $\mu_x\,(t)$ denotes the stochastic force of mortality process of an individual aged x at time t. For the discretized stochastic process, we assume that the age-specific forces of mortality are constant within yearly bands of time and age, i.e., within each square of the Lexis diagram. Formally, given any integer age x and calendar year t, we assume that $\mu_{x+\xi}\,(t+\epsilon) = \mu_x\,(t)$ for any $0 \leq \xi, \epsilon < 1$. Under this assumption, the mortality intensity is approximated by the central death rate $m_x\,(t)$ and the one-year survival probability is given by $p_x\,(t) = \exp\left(-m_x\,(t)\right)$. The complete cohort life expectancy for an x-year old individual belonging to population g in year t, $\dot{e}^C_{x,g}(t)$, is given by the expected remaining lifetime at time t, i.e.,

$$ \dot{e}^C_{x,g}(t) := \mathbb{E}\left(T_x(t)\right) = \frac{1}{2} + \sum_{k=1}^{\omega-x} \exp\left(-\sum_{j=0}^{k-1} m_{x+j,g}\,(t+j)\right), \tag{2} $$

whereas the corresponding complete period life expectancy, $\dot{e}^P_{x,g}(t)$, is given by

$$ \dot{e}^P_{x,g}(t) := \frac{1}{2} + \sum_{k=1}^{\omega-x} \exp\left(-\sum_{j=0}^{k-1} m_{x+j,g}\,(t)\right). \tag{3} $$

Given $\dot{e}^C_{x,g}(t)$ and $\dot{e}^P_{x,g}(t)$, the concept of life expectancy gap, $\dot{e}^{Gap}_{x,g}\,(t)$ corresponds to the systematic difference between the (complete) cohort and period life expectancy measures for a given population at age x in year t and is defined by (Ayuso et al., 2020):

$$ \dot{e}^{Gap}_{x,g}\,(t) := \dot{e}^C_{x,g}\,(t) - \dot{e}^P_{x,g}\,(t). \tag{4} $$

The gap in equation (4) can be positive (in most cases) or negative, depending on expected future mortality developments, and expresses, in years of life, by how much period life expectancy at age x in year t differs from the life expectancy of the cohort attaining age x in year t, i.e., it is an estimate of the additional (reduced) years of life a given cohort will enjoy as a result of expected future mortality improvements (deterioration). From (4) it is also clear that period life expectancy will only match cohort life expectancy if age-specific mortality rates do not change over time. This means that the life expectancy gap can also be interpreted as a sort of leading indicator of future trends in longevity and of the existence of maximum lifespans, contributing also to the discussion on lifespan inequality. A positive but declining (increasing) gap signals a deceleration (acceleration) in expected mortality improvements. A zero gap at very old ages suggests that a given population may be reaching the frontier of human survival. A negative gap suggests that life expectancy is expected to fall in the future. When $\dot{e}^{Gap}_{x,g}(t) > 0$, period life expectancy is a lagged cohort life expectancy indicator, systematically underestimating the remaining lifetime of individuals. Note also that contrary to $\dot{e}^{C}_{x,g}(t)$ that is a random variable and requires mortality projections, the computation of $\dot{e}^{P}_{x,g}(t)$ for past t is objective, i.e., it is not subject to model or parameter risk, except if graduation methods are used to smooth mortality rates and/or life table closing methods are adopted to estimate mortality rates for the oldest-old. The life expectancy gap changes over time since mortality improvement rates fluctuate due to, for instance, a slowing down or acceleration of the aging process, turning points in the long-term trend of mortality or simply a (postponement) deceleration of aging in the younger cohorts as the ageing process progresses to older ages.

2.2 Actuarial fairness and implied tax/subsidies

The estimated magnitude of the life expectancy gap at the retirement age and its ratio against period life expectancy can be given a welfare economic interpretation through the concept of pension wealth in an intergenerational actuarial fairness setting (Ayuso et al., 2017b, 2020). For an actuarially fair pension scheme, the accumulation at retirement equals the pension wealth, i.e., the present value of lifetime contributions equals the actuarial present value of lifetime pension benefits. The pension wealth at the retirement age $x_r(t)$ can be computed as

$$PW_t^{x_r(t)} = B_t^{x_r(t)} a^{\pi,y}_{x_r(t)}, \tag{5}$$

where $B_t^{x_r(t)}$ denotes the initial annual pension benefit and $a^{\pi,y}_{x_r(t)}$ is the life annuity factor, computed using the (period or cohort) survival probabilities, the uprating rate for pensions (π) and the discount rate (y), i.e.,

$$a^{\pi,y}_{x_r(t)} = \sum_{t=1}^{\omega - x_r} \left(\frac{1+\pi_t}{1+y_t}\right)^t \, {}_tp_{x_r}(t). \tag{6}$$

In equation (5) the "normal" retirement age changes with time (e.g., according to life expectancy developments) as is presently a component for policy for many countries. In a defined contribution (FDC/NDC) pension scheme, the initial pension benefit is computed dividing the accumulation (notional/financial capital) $V_t^{x_r(t)}$ by the unisex annuity factor (6), i.e., $B_t^{x_r(t),DC} = V_t^{x_r(t)}/a_{x_r(t)}^{\pi,y}$. The accumulation is a function of contributions and of a notional/financial interest rate.[6] In a defined benefit (NDB/FDB) pension scheme the initial pension benefit usually depends on the contribution history, on the accrual rate and on some demographic and/or early/late retirement adjustment factors (see, e.g., Bravo and Herce, 2020), i.e.,

$$B_t^{x_r(t),DB} = RE_t^{x_r(t)} \times AC_t^{x_r(t)} \times RF_t^{x_r(t)} \times \left(1 \pm b_{x_r(t)}^{\%}\right), \tag{7}$$

where $RE_t^{x_r(t)}$ denotes the monthly reference earnings or pensionable salary, $AC_t^{x_r(t)}$ is the global accrual rate, $RF_t^{x_r(t)}$ is the demographic (sustainability) reduction factor and $b_{x_r(t)}^{\%}$ is the percentage penalty (bonus) for early (postponed) retirement, all at time t,

$$RE_t^{x_r(t)} = \frac{1}{n_w}\left[\frac{1}{x_r(t)-x_0}\left(W_t^{x_r(t)} + \sum_{x=x_0}^{x_r(t)-1} W_{t-x_r(t)+x}^{x_r(t)} \prod_{j=t-x_r(t)+x+1}^{t}(1+\upsilon_j)\right)\right], \tag{8}$$

where x_0 is the contributory career entry age, W_t^x denotes the annual age and time-dependent salary for ages $x = x_0, ..., x_r$ at time $t > 0$, υ_t is the annual valorisation rate at time t (i.e., the rate by which each year's contributions are revalued for the calculation of the first pension) and n_w is the number of salaries per year received (typically a number between 12 and 14).

Pension wealth is an appropriate summary measure of pension entitlements to be transferred in the future to the pensioner as pension income, i.e., the liabilities of the pension provider to the scheme participants. Assuming that the benefit indexation equals the discount rate, i.e., $\pi_t = y_t \ \forall t$, an assumption that broadly holds for wage-indexed pensions, the pension wealth at retirement can be estimated multiplying the initial pension benefit by life expectancy. Now, if the initial benefit is computed using an inappropriate longevity measure (e.g., period life expectancy), then as long as cohort longevity is increasing the actual pension wealth of the birth cohort exceeds the liability calculated using the period method by an amount equal to

[6] The generic rate of return used to index individual accounts in NDC is the rate of growth of the contribution wage base. In Sweden – which has a large demographic reserve fund and continuously positive labor force growth since the 1960s – the rate of growth of the per capita wage (assumed to represent productivity growth in the long-run) is the rate of return applied. In Sweden, in addition, the accumulation in the NDC scheme includes a third factor: the redistribution of the pension balance of those who die before retirement among the survivors born in the same year (a universal survivor's bonus).

the initial pension benefit multiplied by the life expectancy gap, i.e.,[7]

$$\Delta^d PW_{x_r,g}(t) = B_t^{x_r(t)}\left[\dot{e}^C_{x_r,g}(t) - \dot{e}^P_{x_r,g}(t)\right] = B_t^{x_r(t)}\dot{e}^{Gap}_{x_r,g}(t). \tag{9}$$

Equation (9) is a proxy, at the individual level and under the above assumptions, of the amount of unfunded pension liabilities of retired workers attributed to the adoption of an incorrect measure of remaining lifetime in the computation of initial pension benefits. The larger the gap, the more significant the implicit debt burden transferred to future generations.

The actual implications of the life expectancy gap for current and future pensioners are country-specific and depend on the type and design of the pension scheme. For instance, for nonfinancial defined benefit (NDB) schemes i.e., the design of the majority of countries in the OECD, it is likely that the underestimation of actual liabilities will be larger due to weaker – non-transparent actuarial link between contributions and benefits of contributors. In other words, there is an efficiency loss. In addition to this, the use of an inadequate longevity measure directly influences the statutory retirement age and the bonus (penalties) for late (early) retirement. Expressed as a fraction of (3), the life expectancy gap amounts to a ex-ante tax/subsidy, $S_{x_r,g}(t)$, i.e.,

$$S_{x_r,g}(t) := \frac{\dot{e}^{Gap}_{x_r,g}(t)}{\dot{e}^P_{x_r,g}(t)} \times 100 = \left(\frac{\dot{e}^C_{x,g}(t)}{\dot{e}^P_{x_r,g}(t)} - 1\right) \times 100, \tag{10}$$

that a given generation would pay/receive unless (initial or subsequent) benefit adjustments are undertaken to make the system actuarially fairer. In equation (10) negative values represent a tax rate and positive values a subsidy rate to current generations. Stated differently, the larger the life expectancy gap, the more pension benefits depart from an intergenerationally fair and neutral pension scheme.

2.3 Mortality forecasting: An adaptive Bayesian Model Ensemble approach

2.3.1 Bayesian model-averaging

Traditionally, mortality rate estimation and forecasting have been based on a single model selected from the set of candidate models using some method or criteria (e.g., information criteria, forecasting accuracy measure, hypothesis testing for the selection between nested models, cross-validation, bootstrapping, construction of confidence intervals), often with no allowance made for model uncertainty. The selected model is then assumed to be the «true» or «best»

[7] Stated differently, the use of period life expectancy implicitly generates higher accruals when compared to the actually used in the formulae to calculate pensions.

model for statistical inference purposes. However, this implicit conditioning does not reveal the uncertainty in the model selection process (model risk), producing biases in the resulting inferences and overly narrow confidence intervals. To address both this problem and the general lack of comparable longitudinal mortality data, in this paper we adopt a novel approach based on an adaptive Bayesian Model Ensemble (BME) or averaging of heterogeneous models, including six well-known parametric single population Generalised Age-Period-Cohort (GAPC) stochastic mortality models, one single-population univariate functional demographic time-series model (the weighted Hyndman-Ullah method), one bivariate functional data model (the Regularized singular value decomposition model) and the recently proposed two-dimensional smooth constrained P-splines model, all of which can probabilistically contribute towards deriving point forecasts and confidence intervals for period and cohort life expectancy measures and the life expectancy gap.

Bayesian model-ensemble or averaging is an application of Bayesian theory to model selection and inference under model uncertainty. The approach overcomes the problem of drawing conclusions based on a single deemed to be “best” model by conditioning the statistical inference on the entire ensemble of statistical models initially considered in the analysis (or a subset of them). The adaptive BME approach adopted in this paper is inspired in the concept of model confidence set proposed by Hansen et al. (2011), which builds on the idea that, prior to model averaging, the subset of best models from a collection of candidates should be determined based on a criterion that is user-specified. Theoretically, any potential model carrying useful information should be used for forecasting. However, the benefits of adding one additional forecast to the BME should be weighed against the cost of estimating additional parameters, particularly when the number of models is large relative to the sample size and/or the set of candidates includes nested models. In such circumstances, trimming models could lead to better estimates of each model’s weight in the combined forecast (Aiolfi and Timmermann, 2006; Samuels and Sekkel, 2017).

Let each candidate model be denoted by M_l, $l = 1, ..., K$ representing a set of probability distributions comprehending the likelihood function $L(y|\theta_l, M_l)$ of the observed data y in terms of model specific parameters θ_l and a set of prior probability densities for said parameters $p(\theta_l|M_l)$. Consider a quantity of interest Δ present in all models, such as the future observation of y. The law of total probability tells us that its marginal posterior distribution across all models is given by

$$p(\Delta|y) = \sum_{k=1}^{K} p(\Delta|y, M_k)\, p(M_k|\, y), \tag{11}$$

where $p(\Delta|y, M_k)$ denotes the forecast PDF based on model M_k alone, and $p(M_k|\, y)$ is the posterior probability of model M_k given the observed data ("lookforward window" in case of time series), thus reflecting how well model M_k fits the training data. The posterior probability for model M_k is given by

$$p\left(M_k \middle| y\right) = \frac{p\left(y|M_k\right)p\left(M_k\right)}{\sum_{l=1}^{K} p\left(y|M_l\right)p\left(M_l\right)}, \tag{12}$$

where

$$p\left(y|M_k\right) = \int L\left(y|\theta_k, M_k\right)p\left(\theta_k|M_k\right)d\theta_k \tag{13}$$

is the integrated likelihood of model M_k. The posterior model probabilities add up to one, i.e., $\sum_{k=1}^{K} p\left(M_k \middle| y\right) = 1$ and can be interpreted as weights. The Bayesian Model Ensemble PDF is a weighted average of the PDFs given the individual models, weighted by their posterior model probabilities (Kass and Raftery 1995; Raftery et al., 2005).

To determine which models received a greater or lesser weight in the final forecast (and to construct the model confidence set), for each subpopulation we first ranked the models according to their out-of-sample predictive accuracy. We carry out a backtesting exercise in the spirit of Dowd et al. (2010) and use the forecast error in mortality rates as measured by the symmetric mean absolute percentage error (SMAPE) to assess the forecasting accuracy. The SMAPE for model k and subpopulation g is defined by

$$SMAPE_k, g = \frac{1}{n_{x,t}} \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} \frac{\left|\dot{\mu}_{x,t,g} - \mu_{x,t,g}\right|}{0.5 \times \left(\dot{\mu}_{x,t,g} + \mu_{x,t,g}\right)}, \tag{14}$$

where $\dot{\mu}_{x,t,g}$ and $\mu_{x,t,g}$ denote the point forecast and observed mortality rates, respectively, and $n_{x,t} = \left(x_{\max} - x_{\min} + 1\right)\left(t_{\max} - t_{\min} + 1\right)$. The historical "lookback window" and the forecasting horizon ("lookforward window") are set such that a 5-year forecasting horizon for all models and populations is possible. This represents a compromise between the data availability for some of the countries investigated and the horizon over which we will make our forecasts. To compute the model weights, we use the normalized exponential function (also known as the softmax function or softargmax) as follows

$$p\left(M_k \middle| y\right) = \frac{\exp\left(-\left|\xi_k\right|\right)}{\sum_{l=1}^{K} \exp\left(-\left|\xi_k\right|\right)}, \quad \xi_k = \frac{S_k}{\max\left\{S_k\right\}_{k=1,\dots,K}}, \quad k = 1, ..., K, \tag{15}$$

where S_k is the value of the SMAPE for model k. The normalized exponential function assigns larger weights to models with smaller forecasting error, with the weights decaying exponentially the larger the error. Given the model confidence set and the model weights, the BME point forecast of the life expectancy measures is obtained probabilistically combining the individual models using equation (11).

The sampling distribution of the quantity of interest is more complex than in the single-model setting, being a mixture of the individual model sampling distributions. This leads to complications in the calculation of a BME confidence interval for Δ. To derive model-averaged confidence intervals for the quantity of interest, we use the Model-Averaged Tail

Area (MATA) construction proposed by Turek and Fletcher (2012). The idea underlying this construction is similar to that of a model-averaged Bayesian credible interval, with confidence limits defined such that the weighted sum of error rates (utilizing model weights $w_k \equiv p\left(M_k \middle| y\right)$) under each single-model interval will produce the desired overall error rate. The method assumes the sampling distribution for the estimator is asymptotically normal, possibly after a transformation. Let $\varphi = g\left(\Delta\right)$ be a transformation of the variable of interest for which the sampling distribution of $\hat{\varphi}_k = g\left(\hat{\Delta}_k\right)$ is approximately normal, given that M_k is true. The $\left(1-2\alpha\right)100\%$ MATA-Wald confidence interval for Δ is given by the values Δ_L and Δ_U which satisfy the pair of equations: (i) $\sum_{l=1}^{K} w_k\left(1-\Phi_{L,k}\right) = \alpha$ and (ii) $\sum_{l=1}^{K} w_k\left(\Phi_{U,k}\right) = \alpha$, where $z_{L,k} = \left(\hat{\Delta}_k - \Delta_L\right)/se\left(\hat{\Delta}_k\right)$, $z_{U,k} = \left(\hat{\Delta}_k - \Delta_U\right)/se\left(\hat{\Delta}_k\right)$, $\Phi_L = g\left(\Delta_L\right)$, $\Phi_U = g\left(\Delta_U\right)$ and $\Phi\left(\cdot\right)$ is c.d.f. of the standard normal distribution.

2.3.2 Generalised Age-Period-Cohort stochastic mortality models

Generalised Age-Period-Cohort (GAPC) mortality models are a class of parametric models that link a response variable with a linear or bilinear predictor structure consisting of a series of factors dependent on age of the individual, x; period effects, t; and year of birth (or cohort) effects, $c = t - x$. GAPC models fit into the general class of generalized nonlinear models (GNM), with a structure that includes a random component, a systematic component, a link function, a set of parameter constraints to ensure identifiability and time series methods for forecasting and simulating the period and cohort indexes (see, e.g., Hunt and Blake (2014, 2020) and Villegas, Millossovich and Kaishev (2018) for reviews). The random component specifies whether the number of deaths recorded at age x during calendar year t, $D_{x,t}$, follows a Poisson distribution $D_{x,t} \sim \mathcal{P}\left(\mu_{x,t}E^c_{x,t}\right)$, with $\mathbb{E}\left(D_{x,t}/E^c_{x,t}\right) = \mu_{x,t}$, or a Binomial distribution $D_{x,t} \sim \mathcal{B}\left(q_{x,t}E^0_{x,t}\right)$, with $\mathbb{E}\left(D_{x,t}/E^0_{x,t}\right) = q_{x,t}$, where $E^0_{x,t}$ and $E^c_{x,t}$ denote, respectively, the population initially or centrally exposed-to-risk, and $q_{x,t}$ is the one-year death probability for an individual aged x last birthday in year t. The systematic component links a response variable (e.g., $q_{x,t}$ or $\mu_{x,t}$) to an appropriate linear predictor $\eta_{x,t}$, capturing the general shape of mortality across all ages and other time invariant features of the mortality curve, the determinants of the evolution of mortality rates through time, the pattern of mortality change across ages and systematic year-of-birth effects,

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)}\kappa_t^{(i)} + \beta_x^{(0)}\gamma_{t-x}, \tag{16}$$

where $\exp\left(\alpha_x\right)$ denotes the general shape of the mortality schedule across age, $\beta_x^{(i)}\kappa_t^{(i)}$ is a set of N age-period terms describing the mortality trends, with each time index $\kappa_t^{(i)}$ contributing in specifying the general mortality trend and $\beta_x^{(i)}$ modulating its effect across ages, and the term $\gamma_{t-x} \equiv \gamma_c$ accounts for the cohort effect c with $\beta_x^{(0)}$ modulating its effect across ages. The age modulating coefficients $\beta_x^{(i)}$ can be preset or non-parametric terms to be estimated, e.g., as in the Lee-Carter model. The period indexes $\kappa_t^{(i)}$ and the cohort index γ_{t-x} are stochastic processes. The link function g connects the random component and the systematic component, i.e., $g\left(\mathbb{E}\left(D_{x,t}/E_{x,t}\right)\right) = \eta_{x,t}$, typically following the canonical link, i.e., matching the Poisson distribution with the log-link function and the Binomial distribution with the logit-link function. The specification is complemented with a set of parameter constraints to ensure unique parameter estimates.

Table 1 summarizes the structure of the six candidate GAPC mortality models considered in this study. The set of models includes: [LC] the standard age-period Lee-Carter model under a Poisson setting for the number of deaths and using the log-link function with respect to $\mu_{x,t}$ (Brouhns et al., 2002; Renshaw and Haberman, 2003); [APC] the age-period-cohort model proposed by Currie (2006); [RH] the generalization of the Lee-Carter model by incorporating cohort effects with particular substructure obtained by setting $\beta_x^{(0)} = 1$ (Renshaw and Haberman, 2006; Haberman and Renshaw, 2011) and additional approximate identifiability constraint on the parameters of the model to improve the robustness and convergence rate (Hunt and Villegas, 2015); [CBD] the Cairns-Blake-Dowd model considering a predictor structure with two age-period terms, pre-specified age-modulating parameters $\beta_x^{(1)} = 1$ and $\beta_x^{(2)} = (x - \bar{x})$, with $\bar{x}$ the average age in the data, no cohort effects, assuming a Binomial distribution of deaths and using a logit-link function targeting one-year death probabilities $q_{x,t}$ (Cairns, Blake and Dowd, 2006a); [M7] an extension of the original CBD model with cohort effects and a quadratic age effect (Cairns et al., 2009), called M7 in Dowd et al. (2010); and [Plat] the three period factor model proposed by Plat (2009) incorporating the dependence between ages with particular substructure obtained by setting $\kappa_t^{(3)} = 0$ since we are interested only in older ages. These models were selected to incorporate the age, period and cohort features of mortality rates across different countries and ages and because they have proven to perform well in fitting and forecasting exercises (see, e.g., Dowd et al., 2010).

Parameter estimates are obtained using maximum-likelihood (ML) methods. To forecast age-specific mortality rates, we first calibrate the models using each country population (male, female, total) data from 1960 to the most recent year available and for ages in the range $60 - 95$. We focus on retirement/decumulation ages since our objective is to discuss the policy implications of the life expectancy gap in pension design and reform. To forecast and simulate mortality rates, we assume the age vectors α_x and $\beta_x^{(i)}$ remain constant over time and model period indices $\kappa_t^{(i)}$ using a multivariate random walk with a drift. Cohort indices γ_{t-x} were modelled with general univariate ARIMA(p, d, q) models with drift. Box-Jenkins methodology (identification-estimation-diagnosis) is used to estimate the appropriate ARIMA model. To

Table 1: Structure of the GAPC stochastic mortality models used in this study

Model	Predictor	Original reference
LC	$\eta_{x,t} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)}$	Brouhns et al. (2002)
APC	$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \gamma_{t-x}$	Currie (2006)
RH	$\eta_{x,t} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} + \beta_x^{(0)} \gamma_{t-x}$	Renshaw and Haberman (2006)
CBD	$\eta_{x,t} = \kappa_t^{(1)} + (x - \bar{x})\, \kappa_t^{(2)}$	Cairns, Blake and Dowd (2006a)
M7	$\eta_{x,t} = \kappa_t^{(1)} + (x - \bar{x})\, \kappa_t^{(2)} + \left((x - \bar{x})^2 - \sigma\right) \kappa_t^{(3)} + \gamma_{t-x}$	Cairns et al. (2009)
Plat	$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + (x - \bar{x})\, \kappa_t^{(2)} + (\bar{x} - x)^+ \kappa_t^{(3)} + \gamma_{t-x}$	Plat (2009)

close the life tables at high ages and to establish the highest attainable age ω, we use the simple and efficient method proposed by Denuit and Goderniaux (2005) setting the ultimate age at $\omega = 125$ for all years, countries and subpopulations investigated to ensure comparable and comprehensive cross-country life expectancy gap estimates.

Some models described in Table 1 are nested within one of the others, i.e., they are particular cases of more general models. For instance, model LC is nested within model RH, with $\beta_x^{(0)} = 0$ for all x, and $\gamma_{t-x} = 0$ for all c, being a special case of APC with $\beta_x^{(1)} = 1$ or all x and no cohort effects. Model APC is a special case of RH with $\beta_x^{(1)} = \beta_x^{(0)} = 1$ for all x. The CBD model is a restricted version of M7 with $\kappa_t^{(3)} = 0$ for all t and $\gamma_{t-x} = 0$ for all c. In such circumstances, trimming models could lead to better estimates of each model's weight in the combined forecast. For that, we can use for instance the likelihood ratio test to investigate the null hypothesis that the nested (restricted) model delivers lower in-sample forecast errors versus the alternative hypothesis that the more general model is superior (Dowd et al., 2010).

In this paper we follow a different route inspired in the concept of model confidence set. To implement the adaptive BME procedure, we use a fixed-rule trimming scheme in which the number of GAPC models to be discarded is fixed exogenously and determine the statistically superior set of best models among the nested candidates conditional on the model's out-of-sample performance in the validation period. We construct the model confidence set by ranking the models according to their SMAPE, discarding a fixed proportion of models (three out of the six GAPC candidate models), and using the remaining ones to form the BME set of best forecasts. We note that while the fixed trimming approach presets exogenously the number of models to be discarded (and hence the number to be considered in the ensemble), there is nothing constraining the procedure to discard the same models in each forecast period (or even in each country and subpopulation when we account for parameter uncertainty in estimating bootstrap prediction intervals). Moreover, the selected models contribute differently to the

BME forecast if they exhibit different out-of-sample predictive accuracy.

To derive prediction intervals for the mortality rates (and thus for the life expectancy measures) accounting for both uncertainty arising from the error in the forecast of the period and cohort indexes and parameter uncertainty arising from the estimation of the GAPC model parameters, we follow Brouhns et al. (2005) and adopt a semiparametric bootstrap approach. For each of the candidate models and subpopulation, B samples of the number of deaths $d^B_{x,t,g}$, $b = 1, ..., B$, are generated by sampling from the Poisson Distribution with mean $d_{x,t,g}$. Each bootstrapped sample $d^b_{x,t,g}$ is then used together with the corresponding exposure-to-risk to re-estimate the model to obtain B bootstrapped parameter estimates α^b_x, $\beta^{(0),b}_x, \beta^{(1),b}_x, ..., \beta^{(N),b}_x$, $\kappa^{(1),b}_t, ..., \kappa^{(N),b}_t$, γ^b_{t-x}, $b = 1, ..., B$. In this paper we considered 5000 semiparametric bootstrap samples for each model. The model fitting, forecasting and simulation procedures have been implemented using the R package StMoMo (Villegas et al., 2018).

2.3.3 Single-population functional time-series models

Hyndman and Ullah (2007) proposed a method combining nonparametric penalized regression spline with functional principal component analysis for modelling and forecasting log mortality rates, the so-called Functional Demographic Model (FDM). The method extends the principal components approach by adopting a functional data paradigm, utilizing second and higher order principal components to capture additional variation in mortality rates and implementing a smoothing technique to graduate the mortality rates and to reduce or eliminate measurement error. The authors assume that the logarithm of the observed mortality rate at age $x \in [x_1, x_p]$ in year $t \in [t_1, t_n]$, $\log m_{x_i,t} \equiv y_t(x_i)$ is a realization of an underlying continuous and smooth function $f_t(x_i)$ that is observed with error at discrete ages. Expressed mathematically,

$$y_t(x_i) = f_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}, \; i = 1, .., p \quad t = 1, ..., n, \tag{17}$$

where $\sigma_t(x_i)$ allows the amount of noise to vary with x_i in year t, thus rectifying the assumption of homoscedastic error in the LC model, and $\varepsilon_{t,i}$ is an independent and identically distributed standard normal random variable.

The log mortality rates are smoothed prior to modelling using penalized regression splines with a partial monotonic constraint. Here, we smooth age-specific mortality rates in the range $60 - 95$. Using functional principal component analysis, the smoothed mortality curves $\mathcal{I} = \{y_1(x), .., y_n(x)\}$ are then decomposed into orthogonal functional principal components and their uncorrelated principal component scores where, differently from the original LC model, more than one principal component is kept to capture non-random patterns that are not explained by the first principal component. In the original Hyndman-Ullah (HU) method the authors use equal weights in the estimation of the model parameters. In this paper we prefer to use the weighted Hyndman-Ullah (HUw) extension proposed in Shang, Booth, and Hyndman (2011) which differs from the HU method in that uses geometrically decaying weights in the estimation of the model parameters, thus allowing more recent data to weigh more in the estimation of these quantities when compared to data from the distant past. This is justified by the higher predictive accuracy of HUw when compared to HU in empirical studies and its ability to reduce jump-off error. Formally,

$$f_t(x) = \hat{a}^*(x) + \sum_{j=1}^{J} b_j^*(x)\, k_{t,j} + e_t(x), \tag{18}$$

where $\hat{a}^*(x)$ is the weighted functional mean age function estimated by

$$\hat{a}^*(x) = \frac{1}{n}\sum_{j=1}^{J} w_t f_t(x), \quad \sum_{j=1}^{J} w_t = 1, \tag{19}$$

where $\left\{w_t = \pi(1-\pi)^{n-t},\ t = 1, ..., n\right\}$ denotes a set of weights, and $\pi \in (0,1)$ refers to the geometrically decaying weight parameter, with optimal value chosen so as to minimize an overall forecast error measure within the validation data; $\mathcal{B}^* = \left\{b_j^*(x)\right\}\ j = 1, ..., J$ is a set of weighted first J functional principal components with uncorrelated principal component scores $\{k_{t,j}\}$ derived by functional principal component analysis from the set of weighted curves $\{w_t[f_t(x) - \hat{a}^*(x)];\ t = 1, ..., n\}$; $e_t(x)$ is the residual function with mean zero and variance $v(x)$ estimated by averaging $\{e_1^2(x), ..., e_n^2(x)\}$, $e_t(x) \sim \mathcal{N}(0, v(x))$; and $J < n$ is the number of principal components used. Following Shang et al. (2011), we chose $J = 6$ as the maximum number of components required. By conditioning on $\mathcal{I}$ and $\mathcal{B}^*$, the h-step-ahead point forecast of $y_{n+h}(x)$ can be obtained by

$$\hat{y}_{n+h|n}(x) = E\left[y_{n+h}(x)\,|\mathcal{I}, \mathcal{B}^*\right] = \hat{a}^*(x) + \sum_{j=1}^{J} b_j^*(x)\, \hat{k}_{n+h|n,j}, \tag{20}$$

where $\hat{k}_{n+h|n,j}$ denotes the h-step-ahead forecast of $k_{n+h,j}$ obtained using a univariate time-series model (e.g., ARIMA model). By assuming that all sources of uncertainty in the model are uncorrelated and have a normal distribution, the $100(1-\alpha)\%$ prediction interval of $y_{n+h}(x)$ is constructed as $\hat{y}_{n+h|n}(x) \pm z_\alpha \sqrt{var\left[y_{n+h}(x)\,|\mathcal{I}, \mathcal{B}^*\right]}$, where z_α is the $(1-\alpha/2)$ standard normal quantile. To account for parameter uncertainty in the construction of prediction intervals for $y_{n+h}(x)$, we adopt a semiparametric bootstrap approach as described in Section 2.3.2. The model fitting, forecasting and simulation procedures have been implemented using the R package demography (Hyndman, 2019).

2.3.4 Two-dimensional smooth constrained P-splines

Several bivariate models have been proposed to smooth mortality over age and time without any specific model structure. For instance, Currie et al. (2004) and Currie(2006) extended the

P-splines (penalized B-splines) method of smoothing for generalized linear models to the two dimensional case of mortality data with Poisson errors, treating the forecasting of mortality rates as a missing-value problem, estimating the fitted and forecast values simultaneously but not for the whole age range. Although the approach performs well in in-sample empirical studies, its out-of-sample forecasting accuracy tends to low, generating unreasonable trends from a demographic perspective. Camarda (2019) recently enhanced the two-dimensional P-splines model through incorporating demographic constraints to ensure that future mortality over the whole age range follows a plausible and well-behaved demographic profile when estimated from past data. The proposed approach is called CP-spline model.

Consider a mortality dataset comprising deaths and exposure-to-risk arranged in two $m \times n$ matrices, $\boldsymbol{Y} = (d_{ij})$ and $\boldsymbol{E} = (E_{ij})$, respectively, with rows and columns classified by single age at death, x, $m \times 1$ and single year of death, t, $n \times 1$, respectively. The approach assumes that the number of deaths d_{ij} at age i in year j is Poisson-distributed with mean $\mu_{ij}E_{ij}$, i.e., $d_{ij} \sim \mathcal{P}\left(\mu_{ij}E_{ij}\right)$. The goal is to model and forecast mortality over both age and time combining (fixed knot) B-splines with a roughness penalty to achieve a compromise between fitting accuracy and smoothness. Let $\boldsymbol{B}_x$, $m \times k_x$ and $\boldsymbol{B}_t$, $n \times k_t$ be the B-splines over ages and years, respectively. The regression matrix for the two-dimensional model is given by the Kronecker product of the k equally spaced B-splines bases for age x and year t

$$\boldsymbol{B} = \boldsymbol{B}_t \otimes \boldsymbol{B}_x, \tag{21}$$

where $\otimes$ denotes the Kronecker product of two matrices. The two-dimensional penalty is given by

$$\boldsymbol{P} = \lambda_x \left(\boldsymbol{I}_{k_t} \otimes \boldsymbol{D}_x' \boldsymbol{D}_x\right) + \lambda_t \left(\boldsymbol{D}_t' \boldsymbol{D}_t \otimes \boldsymbol{I}_{k_x}\right), \tag{22}$$

where λ_x and λ_t are the smoothing parameters used for age and year, respectively; $\boldsymbol{I}_{k_x}$ and $\boldsymbol{I}_{k_t}$ are identity matrices of dimension k_x and k_t, respectively; and $\boldsymbol{D}_x$ and $\boldsymbol{D}_t$ are difference matrices over the rows (ages) and columns (years) of the coefficient matrix. The author adds shape constraints and asymmetric penalties on the rate of aging (relative derivatives of the age mortality profile), $\boldsymbol{D}_x^t$, and on the rate of change of mortality rates over time, $\boldsymbol{D}_t^t$, to enforce mortality patterns over age and time, with

$$\boldsymbol{D}_x^t = \boldsymbol{B}_t \otimes \boldsymbol{C}_x, \tag{23}$$

where

$$\boldsymbol{C}_x = \frac{1}{h}\left[{}^{q-1}\boldsymbol{B}_x^{\nu} - {}^{q-1}\boldsymbol{B}_x^{\nu-1}\right], \tag{24}$$

with h, q and ν being knot distance, degree, and positions of the original B-spline basis, $\boldsymbol{B}_x$; Similarly, matrix $\boldsymbol{D}_t^t$ computes the first difference of B-spline basis estimated coefficients over years for each age. Similar to Camarda (2019), we derive confidence intervals for forecasted mortality rates by carrying out a residual bootstrap of the fitted model (Koissi et al., 2006).

2.3.5 Regularized SVD model

Huang et al. (2009) and Zhang et al. (2013) extend one-way functional principal component analysis (PCA) to two-way functional data by introducing regularization of both left and right singular vectors in the singular value decomposition (SVD) of the data matrix. It is based on minimization of a regularized sum of squared reconstruction errors of a low-rank matrix approximation. The authors assume the regularized SVD (RSVD) fits the following model for explaining the mortality rate in terms of period t and age x

$$m(x,t) = d_1 U_1(t) V_1(x) + \ldots + d_q U_q(t) V_q(x) + \varepsilon(x,t), \tag{25}$$

where d_q is the singular value, $U_i(\cdot)$ and $V_j(\cdot)$ are smooth functions of period and age, respectively, and $\varepsilon(x,t)$ is a mean zero random noise. The authors stress that the fitted $U_1(\cdot)$ and $V_1(\cdot)$ are components in the best fitting model for mortality with product terms and should not be interpreted as mean functions. Similarly, $U_q(\cdot)$ and $V_q(\cdot)$ should not be interpreted as the $(k-1)$th principal components. The model is fitted iteratively. The first pair of singular vectors of a data matrix $\boldsymbol{X} = (m_{x,t})_{n\times p}$, $U_1(t)$ and $V_1(x)$, whose discretized realizations are, respectively, denoted as $\boldsymbol{u}_1 \equiv (U_1(t_1), \ldots, U_1(t_n))^T$ and $\boldsymbol{v}_1 \equiv (v_1(x_1), \ldots, V_1(x_p))^T$, is obtained by solving a least squares problem as

$$(\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}) = \underset{(\boldsymbol{u},\boldsymbol{v})}{\arg\min} \left\|\boldsymbol{X} - \boldsymbol{u}\boldsymbol{v}^T\right\|_F^2, \tag{26}$$

where $\|\cdot\|_F$ is the Frobenius norm (sometimes also called the Euclidean norm) of a matrix. Subsequent pairs are extracted sequentially by removing the effect of preceding pairs. For two-way functional data, the RSVD of Huang et al. (2009) defines the regularized singular vectors as

$$(\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}) = \underset{(\boldsymbol{u},\boldsymbol{v})}{\arg\min} \left\{\left\|\boldsymbol{X} - \boldsymbol{u}\boldsymbol{v}^T\right\|_F^2 + \mathcal{P}_\lambda(\boldsymbol{u},\boldsymbol{v})\right\}, \tag{27}$$

where $\mathcal{P}_\lambda(\boldsymbol{u},\boldsymbol{v})$ is a regularization penalty

$$\mathcal{P}_\lambda(\boldsymbol{u},\boldsymbol{v}) = \lambda_{\boldsymbol{u}} \boldsymbol{u}^T \boldsymbol{\Omega}_u\, \boldsymbol{u} \cdot \|\boldsymbol{v}\|^{\mathbf{2}} + \lambda_{\boldsymbol{v}} \boldsymbol{v}^T \boldsymbol{\Omega}_v v \cdot \|\boldsymbol{u}\|^{\mathbf{2}} + \lambda_{\boldsymbol{u}} \boldsymbol{u}^T \boldsymbol{\Omega}_u \boldsymbol{u} \cdot \lambda_{\boldsymbol{v}} \boldsymbol{v}^T \boldsymbol{\Omega}_v v, \tag{28}$$

where $\boldsymbol{\Omega}_u$ $(n \times n)$ and $\boldsymbol{\Omega}_v$ $(p \times p)$ are symmetric and nonnegative definite domain-specific penalty matrices, whose purpose is to balance goodness-of-fit against smoothness, and λ is a vector of regularization parameters optimally estimated based on generalized cross-validation (GCV) criterion. To forecast mortality rates and derive confidence intervals, we treat the time functions $U_i(t)$ as time series and model them using general univariate ARIMA processes, rescaling the pairs in (25) by the ratio d_i/d_1, $i = 2, \ldots, q$. We account for parameter uncertainty in the construction of prediction intervals adopting the semiparametric bootstrap approach described in Section 2.3.2. The model fitting, forecasting and simulation procedures have been implemented using the R package RobRSVD (Zhang and Pan, 2013).

2.4 Data

The paper uses historical mortality $D_{x,t}$ and population $E_{x,t}$ data for 42 homogeneous national populations (countries or areas) in different regions of world obtained from the Human Mortality Database (2019). The set of procedures applied by HMD to input data from vital statistics, census counts, birth counts, and population estimates from various sources, particularly national statistical offices, facilitates the cross-national comparability of the information. Table 2 lists the countries considered in this study together with details about data availability in the defined historical "lookback window". The analysis covers medium - and high - income countries in North America, central and western Europe and Asia and the Pacific, as well as Chile in Latin America.

Table 2: Selected HMD countries and available data period used.

Available data	Countries and Regions
1960 – 2016	Australia (AUS), Belarus (BLR), Canada (CAN), Denmark (DNK), Iceland (ISL), Netherlands (NDL), Poland (POL),Spain (ESP), England & Wales (ENW), Scotland (SCO), Northern Ireland (NIR)
1960 – 2017	Austria (AUT), Bulgaria (BGR), Czech Republic (CZE), Estonia (EST), France (FRA), Hungary (HUN), Ireland (IRL), Japan (JPN), Latvia (LVA), Lithuania (LTU), Slovakia (SVK), Luxembourg (LUX), Sweden (SWE), Switzerland (CHE), U.S.A. (USA)
1960 – 2018	Belgium (BEL), Finland (FIN), Norway (NOR)
1992 – 2008	Chile (CHL)
2002 – 2017	Croatia (HRV)
1990 – 2017	Germany (DEU)
1981 – 2013	Greece (GRC)
1983 – 2016	Israel (ISR)
1960 – 2014	Italy (ITA), Russia (RUS)
1960 – 2013	New Zealand (NZL), Ukraine (UKR)
1960 – 2015	Portugal (PRT)
2003 – 2018	Republic of Korea / South Korea (KOR)
1983 – 2017	Slovenia (SLV)
1970 – 2014	Taiwan (TWN)

We can observe that data availability is not the same for all populations investigated. Most countries have data from 1960 onwards. Given the purpose of our study, we have focused on a subset of data from 1960 (or the more distant year available) to 2018 (or the more recent year available) and the 60-110 age range, totaling 126 populations (male, female, total). The datasets provided, for all countries, a sufficient number of years to estimate parameters and to implement the backtesting exercise.

3 Results

Figure 1 plots, for the total population of each country, the adaptive BME model confidence set (vertical axis) and respective estimated model weights (horizontal axis). The set of selected models and respective weights were determined based on the individual model out-of-sample predictive accuracy considering the SMAPE criteria. The higher the weight, the more significant is the contribution of a given individual model to the BME forecast combination. We can observe that the model confidence set varies between countries and that their predictive accuracy is population specific. Although there are a couple of stochastic mortality models that consistently perform well in all countries, we find that no single model dominates on the basis of the predictive accuracy criteria.

Table 3 summarizes, for the total population, the individual model rankings of the methods selected by the adaptive BME procedure to be part of the model confidence set. We can observe, for instance, that in the 5-year lookforward window the CP-Splines and the RSVD methods outperform the remaining methods 13 times each out of 42, followed by the APC model that ranks first 9 times. On the flip side, the RH is selected only once to integrate the model confidence set and the Plat model ranks last 32 out of 42 times.

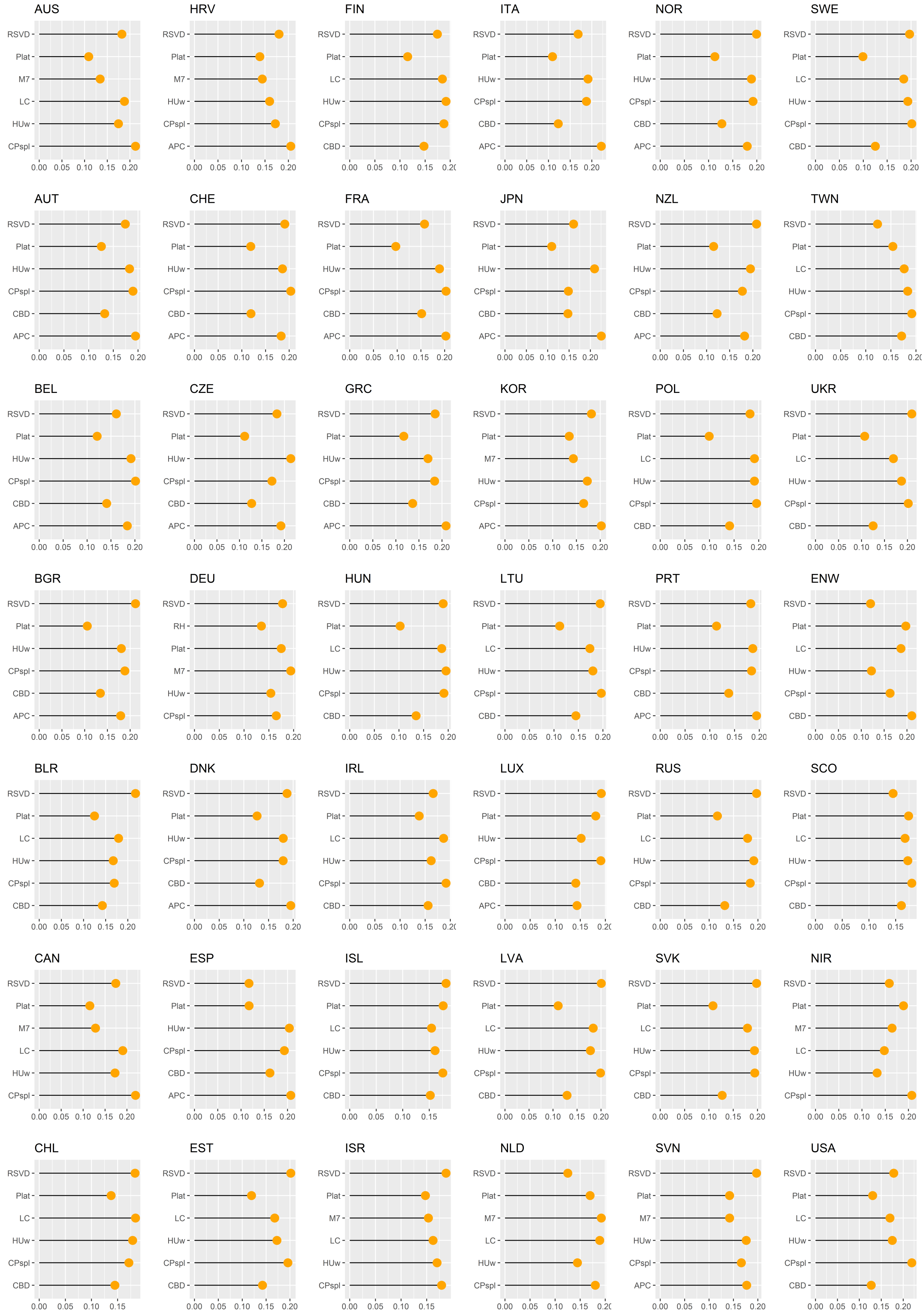
Table 3: Ranking of individual models included in the BME model confidence set

	Model								
Rank	**LC**	**APC**	**RH**	**CBD**	**M7**	**Plat**	**HUw**	**CPspl**	**RSVD**
(1)	1	9		1	2		3	13	13
(2)	5	3				4	9	11	10
(3)	5	2			1	2	17	9	6
(4)	10	3		2		1	9	9	8
(5)	2	1		27	6	3	3		
(6)			1	3		32	1		5

Note: The values represent the number of times models selected to integrate the model confidence set rank first, second and so on, based on the SMAPE forecasting accuracy criterion and using total population data. The column sum indicates how many times a given model was chosen to be part of the model confidence set.

Figure 2 shows, for each country analysed, the observed and forecasted male (M) and female (F) period (P) and cohort (C) life expectancy at age 60 from 1960 to 2050. According to the estimation results, period and cohort life expectancy will continue to increase in 37 of the 42 countries investigated for both the male and female populations. However, the rate at which

Figure 1: BME model confidence set and estimated weights per country (Total population)



we project longevity will increase varies across countries, by sex and by longevity measure. For five countries (Bulgaria, Belarus, Lithuania, Russia and Ukraine), we project a decline in the male period and cohort life expectancy in the next decades, following the recent trends observed in these countries. In three of these five countries (Belarus, Russia and Ukraine), we also project a decline or a stagnation in the female period and cohort life expectancy.

The size of the projected life expectancy improvements for women and men has been changing over time for all countries, altering the ranking of the leading cohort life expectancy countries. In 1960, the ranking was dominated by Scandinavian countries (See Tables 6 and 7 in the Appendix A for details). The leading countries in terms of female cohort life expectancy were Iceland (22.51 years), Canada (21.80), the Netherlands (21.56), Sweden (21.49) and Norway (21.48), whereas for men the leading countries were Iceland (18.79), Sweden (17.63), Norway (17.45), the Netherlands (17.27) and Denmark (17.02).

In 2019, the female ranking was dominated by South Korea (34.16), Japan (31.98), Spain (30.62), France (30.52) and Italy (29.62). For men, the 2019 ranking was dominated by South Korea (27.47), Austria (27.20), Canada (26.50), Switzerland (26.34) and New Zealand (26.25).

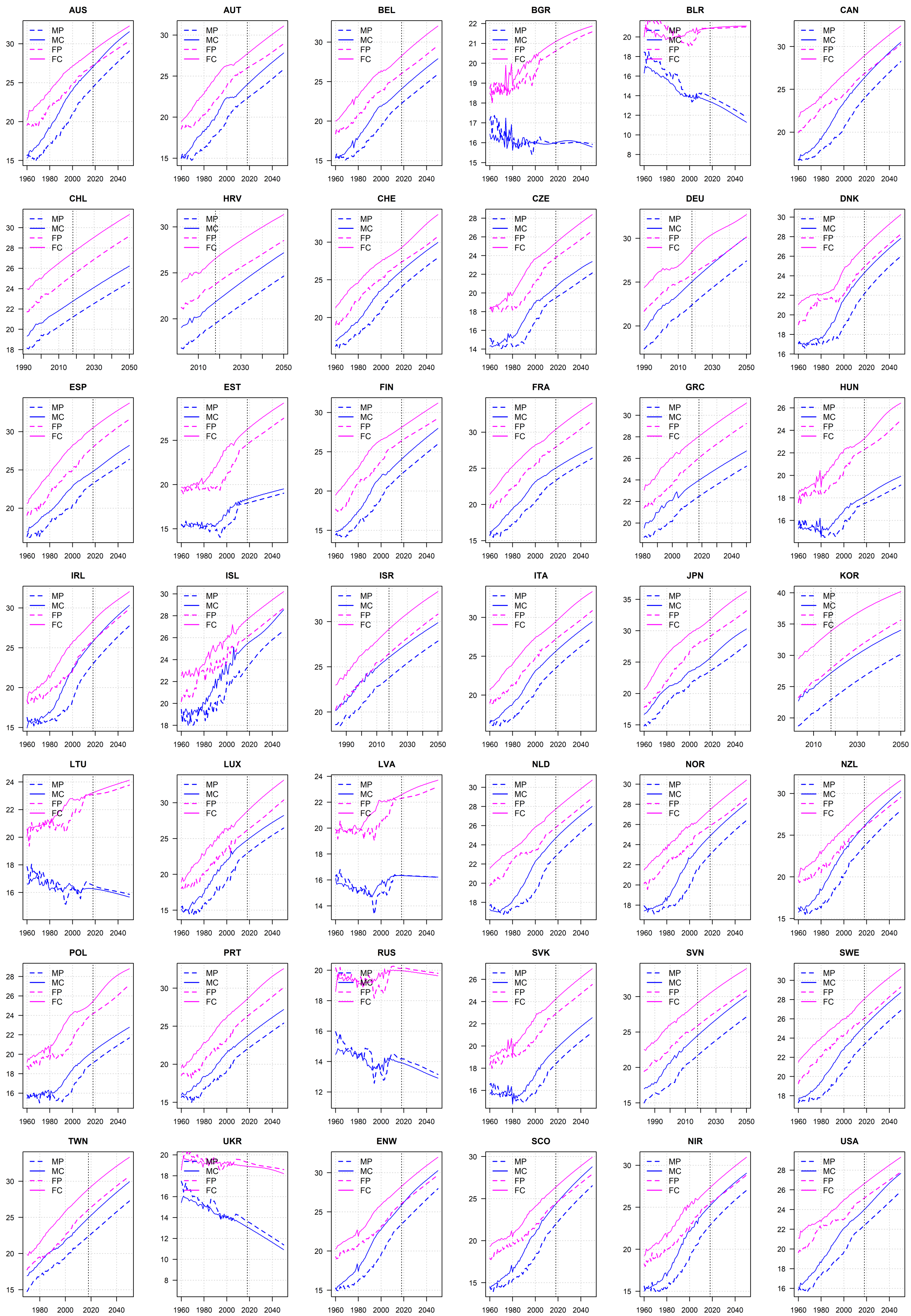
According to the model projections the leading countries in terms of female cohort life expectancy in 2050 are South Korea (40.19), Japan (36.24), France (34.03), Slovenia (33.95) and Spain (33.74), whereas for the male population the leading countries are South Korea (34.04), Austria (31.56), Canada (30.51), Ireland (30.33) and Japan (30.30).

The rapid increase in life expectancy in South Korea has been documented in several studies (see, e.g., Yang et al., 2010; Kontis et al., 2017) and were mostly achieved by reductions in infant mortality and in diseases related to infections and blood pressure.[8] On the flip side of the mean longevity marker, we project the worst performers in 2050 will be Bulgaria (15.79), Lithuania (15.67), Russia (12.91), Belarus (11.29) and Ukraine (10.93) for the male population, and Latvia (23.70), Bulgaria (21.88), Belarus (21.12), Russia (19.64) and Ukraine (18.20) for the female population, with enormous cohort life expectancy differences in comparison with the leading countries. The study also finds that observed and forecasted life expectancy is higher for women than for men in all the 42 countries, with sex differences in cohort life expectancy expected to shrink in most countries by 2050. Some exceptions are Bulgaria, Belarus, Chile, Switzerland, Czech Republic, Estonia, Russia, and Ukraine, which are essentially explained by unhealthier lifestyle (e.g. smoking and alcohol consumption), by biological and environmental differences that include behavioral, cultural, and social factors, by sex differences in the age pattern of mortality and by sex differences in survival after onset of heart disease and cancer.

The female (male) expected age at death computed using a cohort approach $x + \mathring{e}^{C}_{x,g}(t)$

[8] We note, however, that the magnitude of the forecasted longevity improvements should be read with some caution due to the relatively short dataset available for South Korea at HMD.

Figure 2: Actual and forecasted period and cohort life expectancy at age 60, by sex



is forecasted to break the 90-year barrier in 29 (9) countries by 2050, a threshold considered unattainable at the beginning of this century (e.g. Olshansky et al., 2001), albeit this trend in thought was challenged immediately and convincingly by Oeppen and Vaupel (2002) and analyzed in depth in Alho, Bravo and Palmer (2013). The male (female) cohort life expectancy at age 60 is forecasted to increase by an average of 1.13 (1.14) years per decade between 1960 and 2050, with significant disparity across countries - the average increase per country ranges between -0.55 (-0.04) and 2.36 (2.24) years.

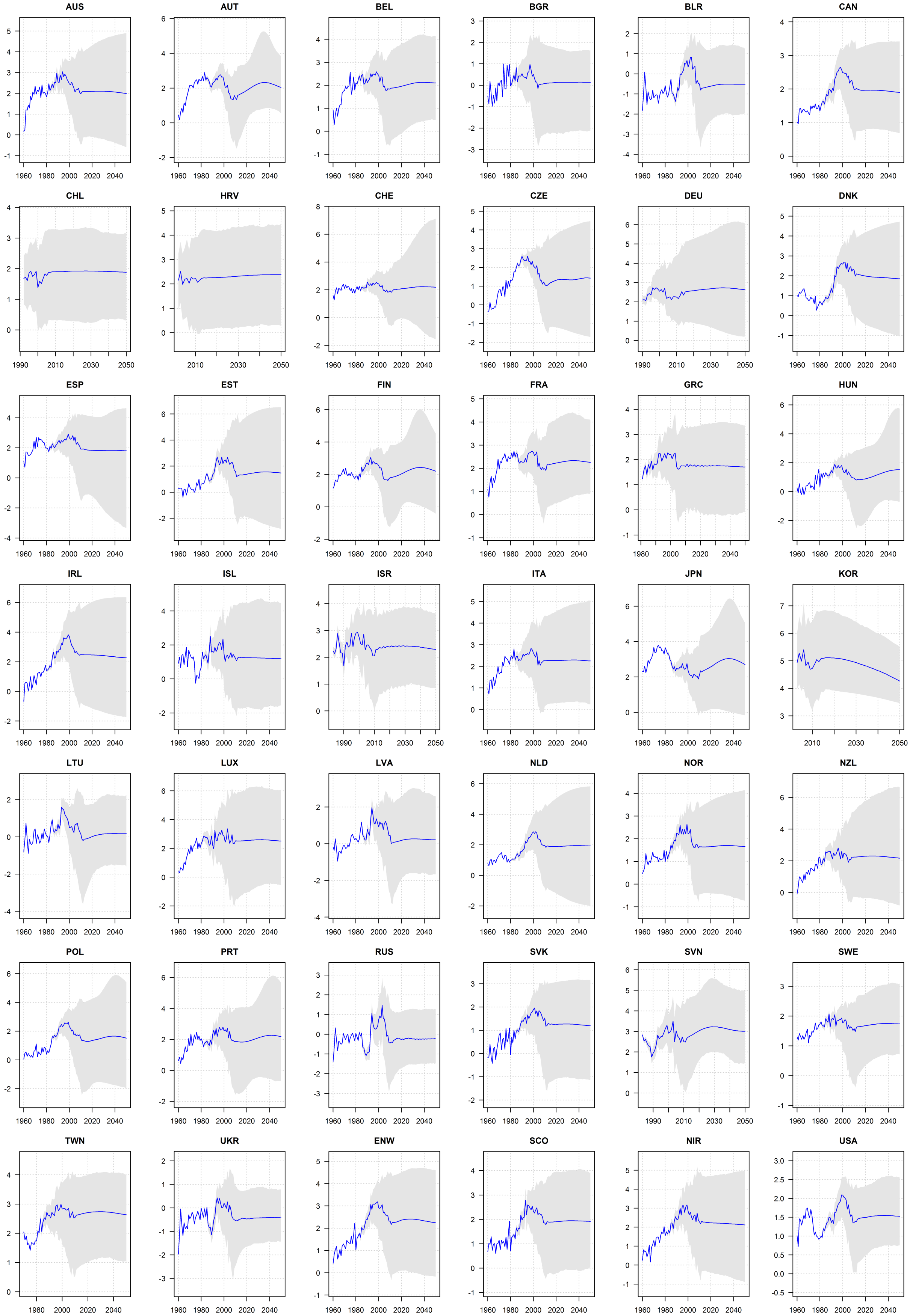
Figure 3 shows the actual and BME point forecast of the life expectancy gap at age 60 for the total population from 1960 to 2050 by country, together with the 95% Model-Averaged Tail Area (MATA) prediction intervals.[9] In the computation of the prediction intervals we account for both the uncertainty arising from the error in the forecast of the individual stochastic mortality model parameters and parameter uncertainty resulting from model fitting. Similar to Turek and Fletcher (2012), we find that the Bayesian MATA interval is slightly more asymmetric than the corresponding frequentist Wald interval.

The results show that for most countries (37 out of 42) and years, the gap is positive confirming that period life expectancy measures tend to systematically underestimate human remaining lifetime. The exceptions are again Bulgaria, Belarus, Lithuania, Russia and Ukraine, for which the gap is either negative or close to zero, signalling a deterioration in the longevity prospects in these countries. The life expectancy gap at age 60 exceeds 5 years for several countries and periods analysed (see Figure 8 in Appendix A for details on the average unisex life expectancy gap per year and age).

The largest life expectancy gap values are observed in South Korea (5.41 years in 2006), followed by Ireland, Japan, Slovenia, Luxembourg, England and Wales, Finland, Australia and Northern Ireland, all with gaps above 3 years at some moment in time. The high life expectancy gap values estimated in South Korean are expained by the rapid decline in mortality rates at all ages observed in the country in the last decades (Yang et al., 2010). Negative life expectancy gap values of up to a two-year difference are observed in the five above mentioned countries. The gap is naturally smaller at high ages but still significant, suggesting that the trend of longevity improvements due to a continuous shift from younger to older ages in the distribution of mortality reductions observed in recent decades is expected to continue. Figure 3 also evidences that the unisex life expectancy gap at age 60 is expected to increase or remain roughly the same in a large majority of countries between 2019 and 2060 - up to an additional difference between period and cohort life expectancy of 0.65 years in Hungary-, with a notable exception in South Korea where a decline of 0.84 years is anticipated. In sum,

[9]Due to space constraints, the full set of results by age in the range 60-125 and by sex and country is not displayed in the paper but can be obtained from the authors upon request.

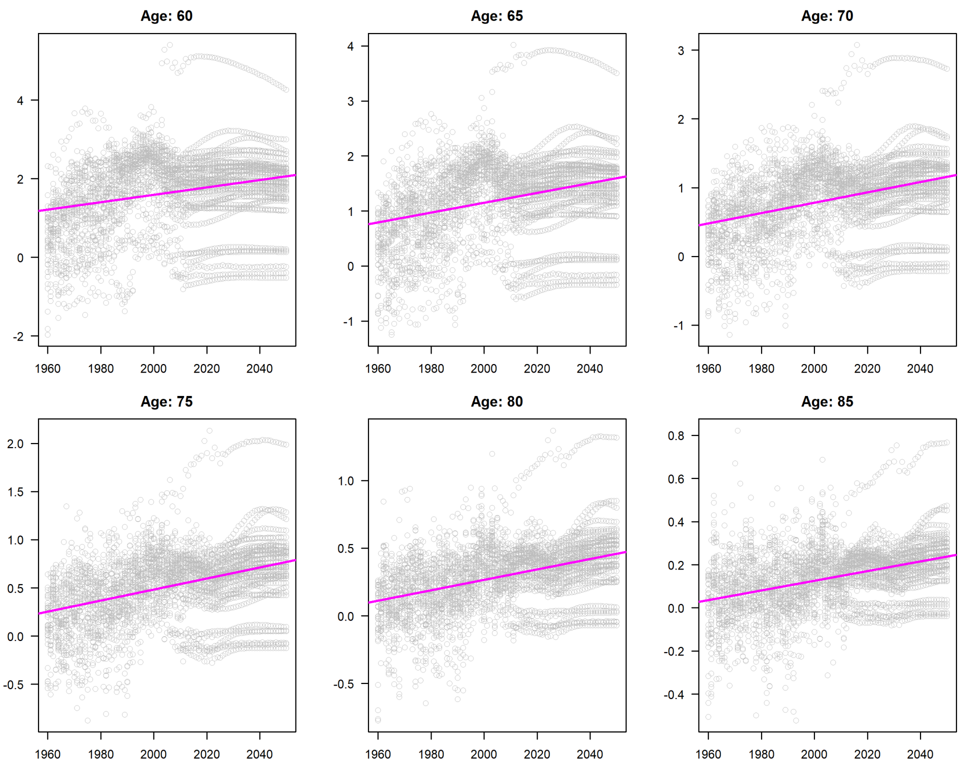
Figure 3: Life expectancy gap: Point forecast and MATA prediction intervals (Age 60, Total)



the results suggest that, in most countries, life expectancy underestimation for social policy and other purposes is persistent over time and that, despite the high survival probabilities already reached at advanced ages, there is still some margin (perhaps smaller) for longevity improvements.

Absolute differences between countries are also observed by gender and age. In general, greater life expectancy gap differences are seen in men than in women, which on an aggregate basis also show steeper positive trends (see Figures 4, 5 and 6).

Figure 4: Life expectancy gap by age and year (all countries, Total population)



Although the differences are minor, the steeper positive slopes observed in the linear regression between the life expectancy gap at older ages and time confirm that the rates of mortality improvement at these ages are still increasing as the age distribution of mortality decline changes, with longevity gains increasingly concentrated among the aged (e.g., due to a reduction in cardiovascular disease mortality).

A non-zero life expectancy gap means pension benefits computed using period over cohort life expectancy measures generate an implicit tax (when negative) or subsidy (when positive) a given generation will pay or receive which is not backed by the lifetime accumulated contribution effort. Table 4 reports the estimated taxes and subsidies cohorts aged 65 in selected years will pay/receive because of the use of unisex period life expectancy measures to determine

Figure 5: Life expectancy gap by age and year (all countries, Male)

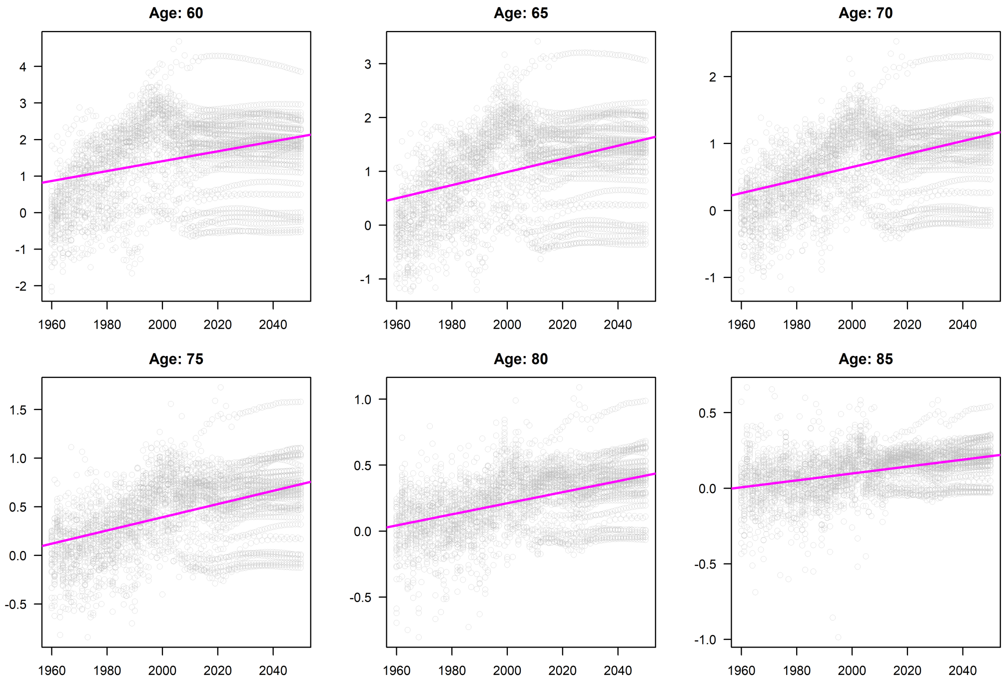
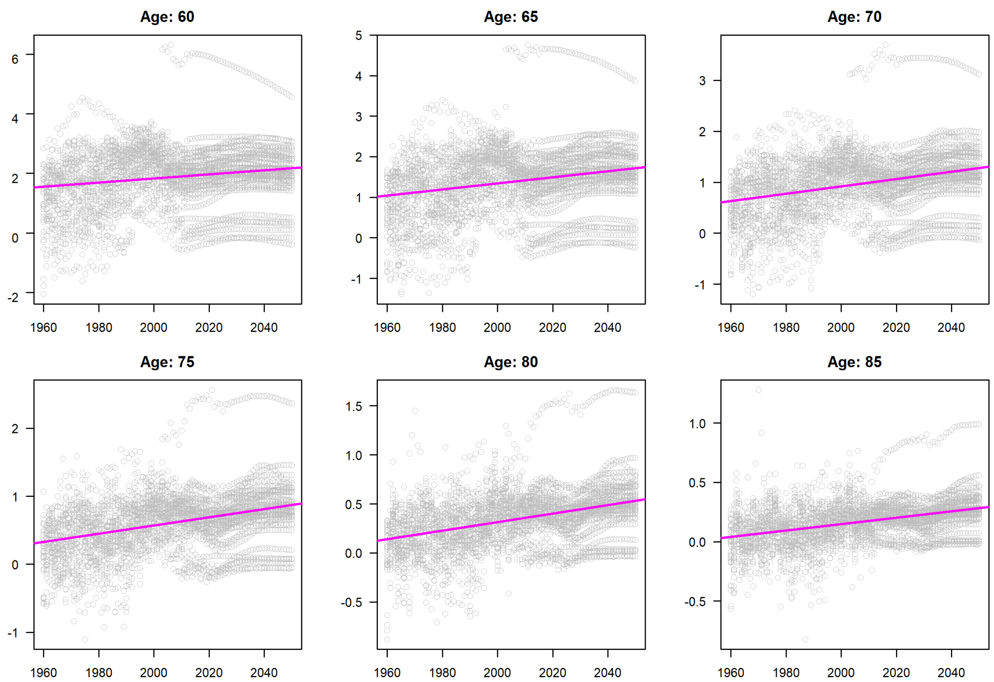


Figure 6: Life expectancy gap by age and year (all countries, Female)



pension benefits. The values were computed based on equation (10) using the mean estimated life expectancy for the total population and are expressed in percentage points. We note that taxes and subsidies reported in Table 4 refer only to the cohort/period life expectancy gap discussed in this paper and not to the cross-subsidy from men to women resulting from the sex gradient in life expectancy observed in a unisex life table context.

Table 4: Implicit tax/subsidy rates of applying unisex period life expectancy measures

Country	Year					Country	Year				
	1960	1980	2000	2019	2050		1960	1980	2000	2019	2050
AUS	2.5	7.1	10.0	7.2	6.3	ITA	3.9	12.8	10.1	8.0	7.1
AUT	2.5	12.0	10.2	6.1	7.1	JPN	10.4	17.0	8.1	8.1	8.5
BEL	4.3	12.1	10.9	7.0	7.0	KOR	NA	NA	NA	18.0	12.3
BGR	-3.7	-1.2	4.0	0.6	0.7	LTU	-3.4	0.7	3.1	-0.4	0.8
BLR	-5.0	-2.8	1.8	-3.4	-2.6	LUX	2.5	12.5	13.3	9.2	8.1
CAN	6.7	5.0	10.4	6.8	6.0	LVA	-1.9	1.4	3.8	-0.2	1.0
CHL	NA	NA	5.3	7.2	6.3	NLD	4.0	3.6	11.6	6.9	6.4
HRV	NA	NA	NA	9.1	8.3	NOR	2.4	4.5	9.5	5.9	5.5
CHE	7.5	9.4	9.8	7.1	6.9	NZL	2.3	10.5	8.2	8.1	7.0
CZE	-1.6	6.2	11.0	4.7	5.5	POL	0.9	4.2	11.6	5.2	5.9
GER	NA	NA	8.8	9.2	8.6	PRT	1.9	8.5	11.5	6.6	7.5
DNK	6.6	4.1	9.9	7.6	6.3	RUS	-5.4	-0.2	3.5	-1.5	-1.2
ESP	5.7	9.1	10.8	6.4	5.8	SVK	0.2	3.2	9.1	5.2	4.7
EST	1.5	3.1	12.1	5.3	5.6	SVN	NA	NA	12.2	10.5	9.4
FIN	5.1	8.4	11.7	6.9	7.5	SWE	6.4	7.4	7.0	5.9	5.7
FRA	6.7	11.5	10.8	7.6	7.1	TWN	NA	6.0	11.8	9.9	8.4
GRC	NA	NA	9.6	6.4	5.7	UKR	-5.5	-0.9	0.3	-2.3	-2.0
HUN	2.9	3.4	7.0	3.7	6.2	ENW	3.1	6.4	12.9	8.1	7.2
IRL	0.1	6.7	17.6	9.1	7.4	SCO	4.1	4.6	11.3	7.2	6.5
ISL	1.6	-0.9	5.3	4.5	4.0	NIR	2.0	9.2	13.9	8.2	6.9
ISR	NA	NA	11.5	8.3	7.2	USA	6.0	3.7	8.9	5.4	5.1

Note: Implicit tax (when negative) and subsidy (when positive) rates cohorts aged 65 in a given year will pay/receive because of the use of unisex period over cohort life expectancy measures to compute pension benefits. Values in percentage points computed based on equation (10).

Table 4 indicates both major differences and commonalities in the implicit tax/subsidy rate across time. First, the magnitude of the subsidy rates between generations can be sizeable, and approximate 20% for several countries and periods analysed. For instance, for 21 out of the 42 countries investigated the subsidy rate exceeded 10% for the cohort aged 65 in 2000, with a global average of 8.8% that year. We forecast the subsidy rate will decline from that year and the 2050 cohort. However, when comparing the most recent estimates and the 2050 cohort, in 15 out of the 42 countries the subsidy rate is expected to slightly increase in the future. This is an important finding for pension systems financial results and projected expenditures as it means that the life expectancy gap continues to widen at an age that is still a benchmark retirement age in the pension systems of most developed economies. The negative life expectancy gap observed and forecasted for most of the selected cohorts in Bulgaria, Belarus, Lithuania, Russia and Ukraine translates into a tax: cohorts retiring at age 65 will be paying to future generations due to an overestimation of their remaining lifetime. The implicit taxes and subsidies due to the life expectancy gap add to those generated by socioeconomic gaps in life expectancy and lifespan inequality at retirement.

4 Policy options

This section discusses and explores several potential policy interventions to address the life expectancy gap on pension schemes' objectives and outcomes, many of which are also viable policy approaches to correct for the unintended ex-ante redistributive effects caused by longevity heterogeneity (see, e.g., Ayuso et al. (2017b) for a detailed discussion). Policy options are better identified and selected if the pension scheme's ultimate goals are clearly stated. In this sense, we adopt an intergenerational actuarial fairness and neutrality principle to pension design and reform for policy evaluation.[10]

The starting position is a pension scheme with no ex-ante redistributive objectives, i.e., the suggested interventions aim to eliminate (or at least mitigate) the wealth redistribution effects, and the distortions created by the life expectancy gap on individual labour supply and savings decisions. Any other intended well-targeted redistributive features of the pension scheme, notably safety-nets to protect lower-income groups from poverty or specific rules designed to positively discriminate women or families with children, typically do not fulfil the conditions of actuarial fairness and/or actuarial neutrality but are considered essential and should not be part of the proposed correction. For clarity of presentation, the latter considerations are ignored in the remainder of the paper. Moreover, this paper discusses policy interventions designed across the total population, regardless of the well-known longevity heterogeneity between women and men or between or within socioeconomic groups. Many countries mandate unisex life tables (e.g., for countries in the European Union the use of unisex life expectancy is mandated) to be applied in annuity calculations, making it (currently) politically difficult to implement legislation allowing for gender differentiation.[11]

[10] Queisser and Whitehouse (2006) use the concept of actuarial neutrality relating to the effect of working an additional year on pension entitlements, i.e., it is a marginal concept used to compare individuals with common life expectancy but different (anticipated/postponed) effective retirement age. The concept of intergenerational actuarial neutrality is commonly used to compare individuals of succeeding generations with different life expectancies (see, e.g., Meneu et al. (2016)).

[11] The combined analysis of longevity heterogeneity and the life expectancy gap is beyond the scope of this paper, but it is in the agenda for future research.

The scope of the unfunded pension liabilities or, equivalently, of the intergenerational tax/subsidy effects of the life expectancy gap in the pension scheme before and after the policy intervention are suggested as a performance measure. A successful pension scheme redesign should be able to correct for the ex-ante redistributive effects and reduce the aggregate unfunded pension liabilities substantially if not totally. A zero ex-ante distortion takes place if account balances in DC schemes and accumulated rights in DB schemes at the time of retirement are converted into an annuity based on cohort survival probabilities at retirement, if the means of obtaining the life expectancy projection is free of systematic bias. This means that the expected value of the difference between projections and outcomes is essentially zero.

Conceptually, however, the policy interventions/scheme redesign employed to counteract the implications of the life expectancy gap can take place at the accumulation, annuitization and decumulation phases. In practice, this might encompass mixed interventions that combine elements of all three stages. Given the nature of the distortion addressed in this paper, we believe that redesign is best approached if implemented at the annuitization and/or decumulation phases. We note, however, that in a traditional NDB scheme the natural adjustment would – as is customary in practice – come through an update in the contribution rate to achieve fiscal balance, redistributing risk from pensioners to contributors.

Policy options at the annuitization stage include: (i) adjusting the initial pension benefit through an actuarially designed sustainability/reduction factor based on the relationship between period and cohort life expectancy at the retirement age, in the spirit of the reforms adopted, for instance, in Finland, Portugal and Spain; (ii) Adjusting the statutory retirement age x_r and the contribution period along with cohort life expectancy (maintaining the accrual rate per year constant or keeping constant the total replacement rate by reducing the accrual rate per year); (iii) Updating the early (late) retirement bonus-malus coefficients to restore actuarial fairness and neutrality; (iv) modifying the eligibility conditions for retirement, namely by requiring additional (reduced) contributions years if the gap is positive (negative); (v) Adjusting the valorisation (pre-retirement indexation) of past earnings when calculating the initial benefit; (vi) Linking the minimum pension age to cohort life expectancy and not to period life expectancy.

Policy interventions at benefit disbursement stage include: (i) Linking annual pension indexation to actual cohort-specific life expectancy developments; (ii) Updating benefits periodically based on the dynamics of both a longevity index, defined as the ratio between the expected and the observed cohort survival probability, and of an interest rate adjustment factor, defined as the ratio between actual and frontloaded/guaranteed interest rate (see, e.g., Alho et al. (2013), Bravo and El Mekkaoui de Freitas (2018), Bravo (2019, 2020) and Palmer and Zhao de Gosson de Varennes (2020)); (iii) a change in the annual account indexation rate (in NDC schemes);

(iv) a reduction in the nominal benefit level. In what follows, for illustration we elaborate on a subset of two of these promising interventions providing numerical results on the magnitude of the adjustments required. We select one policy option at the annuitization stage (adjusting the initial pension benefit to life expectancy) and one at the benefit disbursement stage (conditional pension indexation). The analysis of the remaining policy interventions is left for a forthcoming paper for sake of clarity, coherence and conciseness.

4.1 Adjusting the initial pension benefit to life expectancy

For a positive (negative) life expectancy gap, reducing (increasing) the initial pension in response to a longer (shorter) expected disbursement period through a reduction factor (called sustainability factor in some countries and life expectancy coefficient in others) based on life expectancy is one of the redesign options available at the annuitization stage to restore fairness and neutrality across generations. These factors have been introduced, for instance, in Finland, Portugal and Spain with some design differences. They are typically computed as a simple ratio between period life expectancy observed at some reference age (e.g., 65 in Portugal and 67 in Spain) in some (past) reference year (e.g., 2000 in Portugal and 2012 in Spain) and period life expectancy observed at the time of retirement (in Spain according to life tables for the pensioner population as they are designed by the social security system), irrespectively of the birth cohort or actual retirement age.

This policy option maintains the retirement age and the contribution period but implicitly reduces (at an increasing rate) the accrual rate (in NDB schemes) and the pension entitlements of new pensioners only, i.e., it transfers all longevity risk to future generations (pensioners). To the extent that trends in period and cohort life expectancy differ between past and current generations, the system could redistribute in favor of older cohorts and have a negative impact on the public pension system's sustainability. From equation (9), one way to eliminate the implied tax/subsidy and the unfunded pension liabilities attributed to the adoption of a period life expectancy estimate in computing initial benefits is to introduce an age-specific correction factor $RF_t^{x_r(t)}$ for each birth cohort such that $\Delta^d PW_{x_r,g}(t)$ is zero, i.e., such that

$$\Delta^d PW_{x_r,g}(t) = B_t^{x_r(t)} \left[\dot{e}^C_{x_r,g}(t) \times RF_t^{x_r(t)} - \dot{e}^P_{x_r,g}(t) \right] = 0, \tag{29}$$

from which we obtain

$$RF_t^{x_r(t)} = \frac{\dot{e}^P_{x_r,g}(t)}{\dot{e}^C_{x_r,g}(t)}, \tag{30}$$

for all x_r or, equivalently,

$$RF_t^{x_r(t)} = 1 - \frac{\dot{e}^{Gap}_{x_r,g}(t)}{\dot{e}^C_{x_r,g}(t)}. \tag{31}$$

For countries a with positive life expectancy gap (most cases as concluded in Section 3), equations (30) and (31) state that to restore actuarial fairness and neutrality the initial benefit has to be reduced by a factor equal to the ratio between the period and the cohort life expectancy at the retirement age. The larger the life expectancy gap, the larger the correction factor and thus the larger the benefit reduction needed to compensate for a longer disbursement period (the retirement age remains constant in this case).

Table 5 reports the estimated reduction factors for cohorts retiring at selected ages in 2019 for all countries investigated in this study. The factors were computed using equation (31) and total population mean estimate life expectancy measures. As expected, in all countries the reduction factors are slightly higher (lower) for cohorts retiring at older (younger) ages, i.e., those who retire early (later) must accept a larger (smaller) benefit cut so as to the pension system to restore actuarial fairness and intergenerational neutrality. We find that in countries with positive life expectancy gap the initial benefit cut required to eliminate the ex-ante subsidy can be significant. For instance, for a 60-year old cohort retiring in 2019 in South Korea the initial pension must be reduced by 16.45% to eliminate the implicit intergenerational subsidy, whereas for an equivalent cohort retiring in Slovenia the initial pension must be reduced by 11.04%. Benefit cuts of close to 10% are also estimated for 60-year old cohorts retiring in 2019 in countries such as Japan, Taiwan, Luxembourg, Germany, Croatia, Ireland, England and Wales or Israel, with an average pension cut of 6.64%.

We note that the reduction factor does not vary significantly with the retirement age (the maximum absolute difference between the benefit adjustment required for cohorts retiring at 60 and at age 66 in 2019 is 1.86% in Slovenia, with an average value of 0.96% across all countries). The result is not surprising since, although there are age- and cohort-specific factors influencing future mortality rates, the cohort life expectancy at all ages (and, thus, both the numerator and the denominator in equation (31)) is also determined by common time trends. These results mean that in countries in which strict intergenerational actuarial neutrality is not a key issue a policy response based on a common reduction factor computed for some representative age (e.g., the statutory retirement age) may be considered.

Our results show that in countries with negative life expectancy gap the initial benefit should be increased to eliminate the ex-ante tax current generations are paying to future pensioners. For instance, for a 66-year old cohort retiring in 2019 in Belarus the initial pension must be increased by 3.47% to eliminate the implicit tax. Benefit increases of up to 3% are estimated for cohorts retiring in 2019 in Belarus, Lithuania, Russia, Latvia and Ukraine, with an average pension raise of 1.45%. Of course, if a differentiation of the correction factors by gender were to be permitted, in most countries the correction would be larger for women than for men and the number of countries requiring positive benefit adjustments due to a negative

Table 5: Estimated reduction factors per country for selected retirement ages, 2019

Country	Age				Country	Age			
	60	62	64	66		60	62	64	66
AUS	0.9252	0.9282	0.9314	0.9351	ITA	0.9186	0.9214	0.9245	0.9279
AUT	0.9300	0.9347	0.9396	0.9447	JPN	0.9129	0.9179	0.9229	0.9275
BEL	0.9259	0.9291	0.9328	0.9368	KOR	0.8355	0.8397	0.8448	0.8510
BGR	0.9930	0.9934	0.9938	0.9942	LTU	1.0003	1.0018	1.0035	1.0056
BLR	1.0369	1.0359	1.0350	1.0347	LUX	0.9072	0.9105	0.9141	0.9181
CAN	0.9286	0.9314	0.9344	0.9376	LVA	1.0002	1.0090	1.0160	1.0026
CHL	0.9245	0.9278	0.9313	0.9349	NLD	0.9282	0.9312	0.9339	0.9368
HRV	0.9061	0.9101	0.9145	0.9191	NOR	0.9373	0.9400	0.9429	0.9462
CHE	0.9260	0.9288	0.9329	0.9358	NZL	0.9164	0.9195	0.9229	0.9266
CZE	0.9436	0.9472	0.9516	0.9564	POL	0.9430	0.9461	0.9490	0.9518
GER	0.9043	0.9083	0.9131	0.9188	PRT	0.9294	0.9328	0.9362	0.9396
DNK	0.9222	0.9249	0.9279	0.9312	RUS	1.0141	1.0153	1.0151	1.0146
ESP	0.9334	0.9357	0.9383	0.9411	SVK	0.9434	0.9460	0.9490	0.9523
EST	0.9418	0.9452	0.9484	0.9537	SVN	0.8896	0.8954	0.9018	0.9082
FIN	0.9241	0.9285	0.9330	0.9375	SWE	0.9367	0.9394	0.9423	0.9455
FRA	0.9211	0.9241	0.9274	0.9311	TWN	0.9004	0.9041	0.9081	0.9126
GRC	0.9331	0.9366	0.9389	0.9397	UKR	1.0280	1.0268	1.0252	1.0236
HUN	0.9589	0.9613	0.9634	0.9653	ENW	0.9142	0.9182	0.9228	0.9275
IRL	0.9088	0.9117	0.9149	0.9186	SCO	0.9246	0.9277	0.9311	0.9349
ISL	0.9522	0.9539	0.9558	0.9580	NIR	0.9164	0.9192	0.9222	0.9263
ISR	0.9135	0.9182	0.9222	0.9254	USA	0.9414	0.9441	0.9470	0.9501

Note: Estimated reduction factors for cohorts retiring in 2019 computed using equation (18) and total population (unisex) mean estimated life expectancy measures.

life expectancy gap would increase.

4.2 Conditioning pension indexation

The pension indexation rules are an important design feature of public and private pension systems, which affects not only the sustainability but also the long-term income adequacy and security of pensions. The way indexation impacts income security at old ages depends critically on the actual design and structure of pension systems (e.g., number and coverage of the different pillars, flat-rate or earnings-related scheme, significance of occupational and personal pension schemes, differentiated indexation rates according to the pension level). In most countries nowadays, the typical earnings-related pension scheme provides a DB or a DC pension in which only the purchasing power of benefits is "guaranteed" – however if finances are threatened so is the guarantee.

This said, price and/or wage indexation has long been considered a guaranteed right by pension beneficiaries, a claim that resurfaced in recent years as a result of sustainabilitydriven pension reforms (e.g., extension of the contribution-wage assessment period from the last few best years to full career, tightening of the eligibility conditions, cuts in initial benefit levels as a consequence of various automatic balancing mechanisms, reductions of the accrual rate) which reduced pension levels and the replacement rate. In DB schemes, the way a benefit is commonly computed and indexed is often justified as a mechanism to share risk intergenerationally, following the development of both real wages or prices. This is in fact one of the key generic features of the relatively new NDC schemes. Although current inflation levels are at an historical low in many countries around the world, nevertheless, benefit indexation below price increases erodes the purchasing power of pensions and raises long-term adequacy concerns. In political reality, pension indexation is conditional on the pension fund's (current and/or long-term financial position), as the recent experience showed with a significant number of countries either nominally freezing benefits in payment or limiting automatic indexation mechanisms (European Commission, 2015).

Conditioning the indexation of pensions in payment to the actual survival experience of a given cohort of pensioners is one of the policy options to cope with the life expectancy gap at the pay-out phase of pensions. Assume that, for a given preset discount rate y_t in (6), policy-makers consider to adjust the annual pension benefit uprating rate in order to eliminate the intergenerational tax/subsidy generated by a non-zero gap $\dot{e}^{Gap}_{x_r,g}(t)$, maintaining the other parameters unchanged. Denote by π^P_t the promised (exogenous or endogenous) pension indexation rate at time t and by π^C_t the intergenerationally fair and neutral indexation rate, i.e., the rate consistent with zero unfunded pension liabilities. The estimate of the unfunded pension liabilities of retired workers in equation (9) can now be written as

$$\Delta^d PW_{x_r,g}(t) = B_t^{x_r(t)} \left[\sum_{t=1}^{\omega - x_r} \left({}_tp^{[C]}_{x_r}(t) \left(1+\pi^C_t\right)^t - {}_t\, p^{[P]}_{x_r}(t) \left(1+\pi^P_t\right)^t \right) (1+y_t)^{-t} \right], \qquad (32)$$

where ${}_tp^{[P]}_{x_r}$ and ${}_tp^{[C]}_{x_r}$ denote the t-year survival probability computed using a period and cohort approach respectively. Eliminating the implicit tax/subsidies generated by the life expectancy gap requires setting $\Delta^d PW_{x_r,g}(t)$ in equation (32) to zero. Assuming countries continue to use a period approach to compute initial pension benefits, this would require adopting the following pension indexation rule

$$\pi^C_t = \left[\left(1+\pi^P_t\right)^t \times \frac{{}_tp^{[P]}_{x_r}(t)}{{}_tp^{[C]}_{x_r}(t)} \right]^{\frac{1}{t}} - 1. \qquad (33)$$

From equation (33) it is clear that if the actual survival rates of a given retired cohort observed at time t are higher (lower) than the ones anticipated at the retirement age, then the annual indexation rate of pensions must decrease (increase) against the promised rate to be consistent with the actuarial present value of lifetime benefits computed at the retirement age using the period annuity factor. Moreover, «ceteris paribus» the higher the promised pension indexation rate the higher the correction required to cope with the life expectancy gap. If the actual cohort survival rates match the expected survival prospects, i.e., if ${}_tp_{x_r}^{[C]} ={}_t p_{x_r}^{[P]}\ \forall t$, then $\pi_t^C = \pi_t^P$.

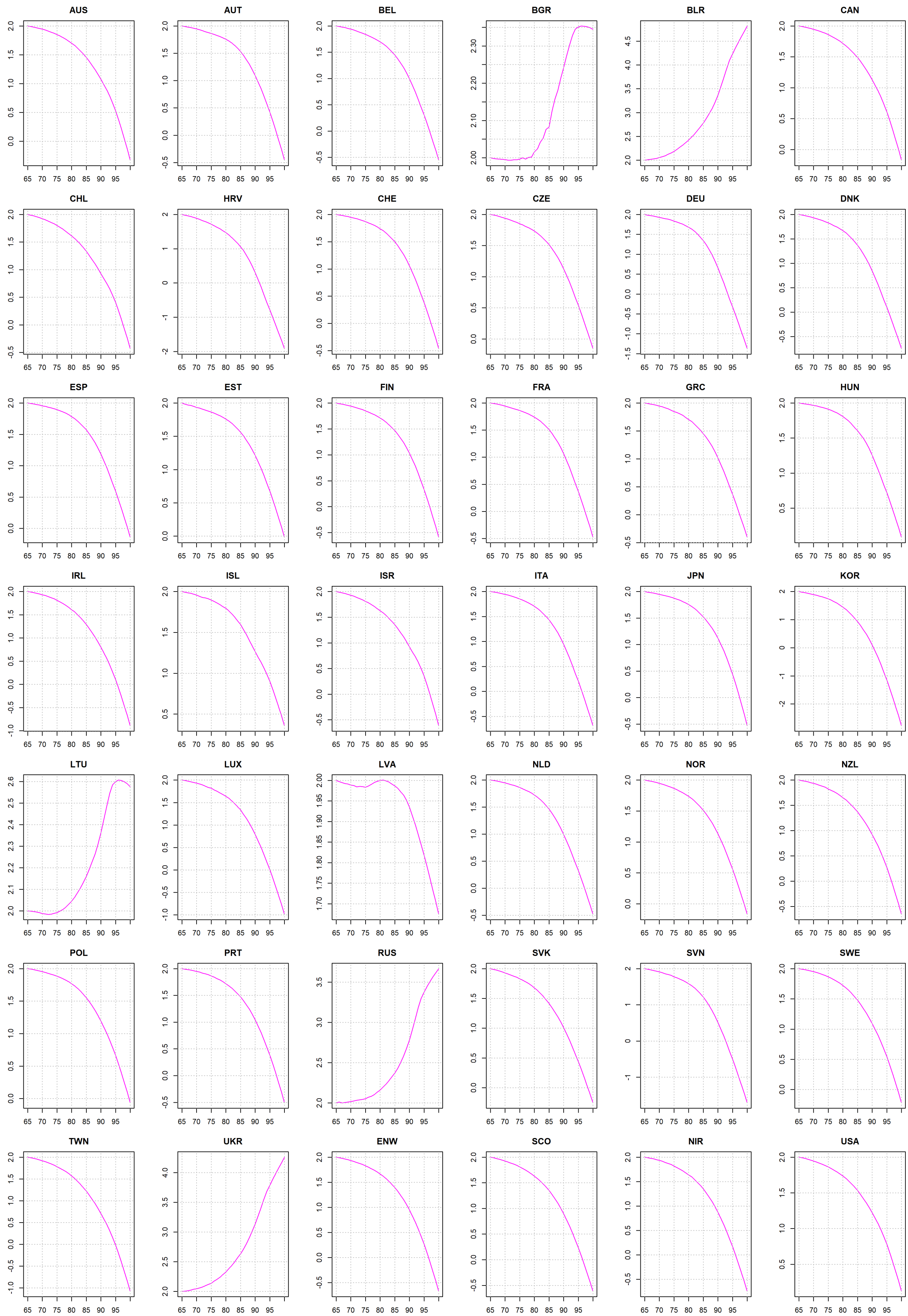
Note that strict actuarial neutrality would require indexing pensions differently depending on the pensioner's birth cohort, a solution that would be politically sensitive and likely difficult to implement in practice. Most countries have some price, wage or GDP (or a mix) benefit indexation mechanism after retirement including, in some cases, progressive (redistributive) indexation which grants higher indexation to smaller pensions, i.e., they are dependent on income (adequacy) considerations and not on birth cohort. To mitigate adequacy and poverty concerns, countries can use different types of complementary provision mechanisms (e.g., minimum pensions, basic pensions or minimum income provision, social assistance-like, means tested benefits), funded outside of the earnings-related scheme. This is in fact the generic approach of national NDC and FDC schemes. Additionally, to prevent significant real income losses and income insecurity at the very old-age the indexation rule (33) could be applied up to a certain ceiling age (e.g., 85 years old). For illustration, Figure 7 plots the adjusted annual pension indexation rate for 65-year old cohorts retiring in 2019, considering unisex survival probabilities and assuming a flat 2% promised indexation rate per year for all countries, i.e., $\pi_t^P = 2\%$ $(t = 2019, 2020, ...)$.

The results illustrate that for countries with a positive (negative) estimated life expectancy gap, the indexation rate would have to be reduced (increased) gradually from the promised 2% per year in order to restore actuarial fairness among generations. The correction varies between countries (and cohorts) and tends to increase with age – as acceleration in the decline in mortality is moving up in the age scale (with the exception of a few countries already identified above), and may even determine negative indexation (i.e., a nominal reduction in pensions) in countries that significantly underestimate the remaining lifetime at retirement. On the contrary, countries which over-estimate life expectancy at retirement (i.e., with negative life expectancy gap) can afford to increase the annual benefit indexation without jeopardizing the pension system's solvency and actuarial fairness. The average annual correction in the pension indexation rate against the promised rate varies between countries, ranging between −135.5 basis points (b.p.) in South Korea, −119.6 b.p. in Croatia, and −105.9 b.p. in Slovenia, and +52.1 in Russia, +75.4 b.p. in Ukraine and +93.1 b.p. in Belarus.[12]

Conditioning directly pension indexation annually on actual life expectancy developments is impractical due to the random outcomes surrounding any time series of outcomes. So, although in a world without random "noise" this transfers all longevity risk to current surviving

[12] Average values per country computed in the 65-100 age range for 65-year old cohorts retiring in 2019.

Figure 7: Adjusted annual indexation rate for 65-year old cohorts retiring in 2019



pensioners, in practice the policymaker would implement a smoothing rule based e.g. on three-year or longer moving averages of the most recent years. If this is the model employed in practice, then the weight of the changing indexation would be successively deferred to the survivors in the cohort. This is an option that to our knowledge has never been employed and could be politically difficult to implement in practice. Several countries have in fact adopted automatic balance mechanisms conditioning the indexation rate of pensions to restore the solvency of the system but not linked exclusively and directly to life expectancy developments.

For instance, in NDC schemes the effect of actual longevity developments may be embedded in the internal rate of return, via the overestimation of the PAYG asset and through a turnover time of a unit of money liquidity effect (Palmer, 2013). In Sweden, the NDC scheme benefit indexation includes an exogenous correction mechanism that accounts for the difference between the norm (the frontloaded long-term expected value of the internal rate of return) and the actual yearly outcome. In Germany, the pension-point system links the annual update of the pension point value to the changes in the statutory pension scheme's dependency ratio. Japan computes the indexation of pensions based on the rate of the normal indexation minus a modifier, where the modifier is the sum of the annual decrease rate of the number of active participants in the social security pension schemes and the annual increase rate of life expectancy at the age of 65 (this increase is fixed at 0.3% in order to avoid fluctuation), with a floor at 0%. In the Netherlands indexation is conditional on the pension fund's financial position, which is affected by, among other things, the aging of the population and the increase in the ratio of pension liabilities to the total wage bill. In private insurance contracts, linking benefits to actual longevity and financial market developments has been proposed, for instance, in longevity-linked life annuities.[13]

Leaving the indexation (totally or partially) outside the creation of the annuity and instead indexing benefits exogenously on a regular (e.g., yearly) basis based on actual longevity and return developments has some advantages, including indexing benefits with the rate of growth in the real wage of all contributors in order to maintain the relative value of benefits to wages, a persistent challenge in old-age protection (Palmer and Zhao de Gosson de Varennes, 2020).

5 Discussion and policy implications

The paper confirms very affirmatively the deficiency of period estimates of life expectancy compared to cohort life expectancy. If mortality across ages improves, period life expectancy substantially underestimates the cohort life expectancy that differs for each birth cohort and

[13]See, for instance, Bravo and El Mekkaoui de Freitas (2018), Bravo (2019, 2020) and Palmer and Zhao de Gosson de Varennes (2019).

over ages. The degree of underestimation increases with improvement in mortality rates. To make this point in a convincing manner the paper (i) uses a global database (the Human Mortality database) for 42 countries – the most extensive illustration of the phenomenon yet produced; (ii) proposes a promising "model-ensemble" approach to estimate the period and cohort life expectancy based on the weighted outcomes of several projection models that generally speaking performs better than any of them do if applied individually; and (iii) translates the mis-estimation of the period vs cohort life expectancy into a statistical measure of the life expectancy gap that serves as a direct measure of the impact of the gap on the pension scheme liabilities and consequently the financial status of a pension scheme.

The new ensemble model estimation approach for life expectancy is a promising approach to addressing model and parameter uncertainty (or risk) that – by nature – surrounds any projections of future mortality developments. The pooling and weighting of different (Bayesian) estimates reduce the modelling and, thus, forecasting risk. The approach uses extrapolative methods, which focus on extending into the future the regularity of patterns and trends observed in the past without recourse to other current knowledge about actual and prospective developments in key variables (e.g., lifestyles, medicine, new diseases). This is their advantage, but it is also their essential limitation.

The numerical results across the 42 investigated countries confirm the importance of the topic and methodological approach. For most countries (37 out of 42) and years, the life expectancy gap is positive confirming that period life expectancy measures typically and systematically tend to underestimate human remaining lifetime. Four of the five exceptions are former Soviet states (Bulgaria, Belarus, Lithuania, Russia and Ukraine), for which the gap is either negative or close to zero and can be claimed to be the result of issues emerging from an incomplete transition process, including especially serious life-style public-health related issues.

Translating the estimated life expectancy gap into a tax/subsidy framework offers a better grasp of the numerical relevance of the results and implications for pension policy. Using the unisex data at a retirement age of 65 in 2019, we find that in 21 of the 42 countries investigated in our study, the use of period life expectancy underestimates life expectancy by an amount surpassing 10 % of the cohort's accumulated pension wealth. This is happening at the cost of future generations. In Korea, it reaches 18.0 percent in 2019. As a result of the gender difference in life expectancy, the financial pooling results indicate a larger subsidy for women since as a group they live longer than men; for men the subsidy can be interpreted to be an implicit tax.

In order to reduce or neutralize the tax/subsidy effects of underestimated life expectancy the paper presents multiple policy options and explores empirically two policy interventions. Examples are, a correction factor at the time of retirement and at the level of the annuity calculated, and a correction factor at the time of disbursement (after retirement and till death) and taking place periodically at the time of benefit indexation. The most promising intervention options are around benefit determination phase and disbursement/benefit indexation phase; interventions during the accumulation phase would be difficult to justify and apply in practice.

Ideally both interventions should be differentiated not only with respect to the retirement age but also by the birth cohort. However, applying differentiated correction factors are likely to meet with political resistance. Furthermore, while a one-time correction at retirement may seem plausible even if it may amount to 5 and more percent in order to avoid a burdening of future generations, a smaller but annual correction index for benefits under disbursement may be more inclined to receive greater political acceptance. It would come on top of an often-perceived under-indexation of pension benefits where nowadays price indexation is typical. As a conclusion, the best approach is to get the benefit estimation correct the first time by applying the best estimate of the cohort life expectancy to the accumulated funds at retirement. In a (F & N) DC world this requires simply to have estimates of cohort life expectancy that are best provided by the national statistical office and based on a methodology that is developed and compared jointly with the relevant international organizations. The statistical offices of developed economies are capable of and technically prepared to do job. Although emerging economies may often require technical support in the preparation of the data and estimates. Finally, we can note that the development of heterogeneity in longevity attributable to socio-economic characteristics in conjunction with benefit creation – in conjunction with lifetime earnings – creates a more complicated technical hurdle. But this is a topic of a different paper.

Acknowledgements

The authors would like to express his gratitude to the editors and one anonymous referee for his or her careful review and insightful comments, that helped strengthen the quality of the paper. We thank also the suggestions and remarks from participants at the Fifteenth International Longevity Risk and Capital Markets Solutions Conference, Washington, September 2019. J. Bravo was supported by Portuguese national funds through FCT under the project UIDB/04152/2020 - Centro de Investigação em Gestão de Informação (MagIC). M. Ayuso is grateful to the Spanish Ministry of Science and Innovation for funding received under grant PID2019-105986GB-C21.

References

Aiolfi, M., and Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. Journal of Econometrics, 135 (1–2): 31–53.

Alaminos, E., and Ayuso, M. (2019). Marital status, gender, mortality, and pensions: the disadvantages of being single in old age. Revista Española de Investigaciones Sociológicas, 165: 3–24.

Alho, J., Bravo, J. M., and Palmer, E. (2013). Annuities and life expectancy in NDC. In Holzmann, R., Palmer, E. and Robalino, D. (Eds.) Nonfinancial defined contribution Pension Schemes in a Changing Pension World, Volume 2, Gender, Politics, and Financial Stability: 395–436.

Andrawis R.R., Atiya, A.F., and El-Shishiny, H. (2011). Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition. International Journal of Forecasting 27(3): 672–688.

Ayuso, M., Bravo, J. M., and Holzmann, R. (2017a). On the heterogeneity in longevity among socioeconomic groups: Scope, trends, and implications for Earnings-Related Pension Schemes. Global Journal of Human Social Sciences - Economics, 17(1): 31–57.

Ayuso, M., Bravo, J. M., and Holzmann, R. (2017b). Addressing longevity heterogeneity in pension scheme design. Journal of Finance and Economics, 6(1): 1–21.

Ayuso, M., Bravo, J. M., and Holzmann, R. (2020). Getting Life Expectancy Estimates Right for Pension Policy: Period versus Cohort Approach. Journal of Pension Economics and Finance, 1–20. doi:10.1017/S1474747220000050.

Biffis, E. (2005). Affine processes for dynamic mortality and actuarial valuations. Insurance: Mathematics and Economics, 37(3): 443–468.

Biffis, E. and Millossovich, P. (2006). The Fair Value of Guaranteed Annuity Options. Scandinavian Actuarial Journal, 19: 23–41.

Blackburn, C., and Sherris, M. (2013). Consistent dynamic affine mortality models for longevity risk applications. Insurance: Mathematics and Economics, 53: 64–73.

Blake, D., Cairns, A. J. G., Dowd, K. and Kessler, A. R. (2019). Still living with mortality: The longevity risk transfer market after one decade. British Actuarial Journal, 24: 1–80.

Booth, H., and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. Annals of Actuarial Science, 3(1-2): 3–43.

Bravo, J. M. (2011). Pricing longevity bonds using Affine-Jump Diffusion Models. CEFAGE– UE Working Paper 2011/29. Available at: https://ideas.repec.org/p/cfe/wpcefa/2011_29.html.

Bravo, J. M. (2019). Funding for longer lives: Retirement wallet and risk-sharing annuities. Ekonomiaz, 96(2): 268–291.

Bravo, J. M. (2020). Pricing participating longevity-linked life annuities: A Bayesian Model Ensemble approach. Revised and resubmitted to European Actuarial Journal.

Bravo, J. M., and Nunes, J. P. V. (2020). Pricing Longevity Derivatives via Fourier Transforms. Revised and resubmitted to Insurance: Mathematics and Economics.

Bravo, J. M., and El Mekkaoui de Freitas, N. (2018). Valuation of longevity-linked life annuities. Insurance: Mathematics and Economics, 78: 212–229.

Bravo, J. M., and Herce, J. A. (2020). Career breaks, Broken pensions? Long-run effects of early and late-career unemployment spells on pension entitlements. Journal of Pension Economics and Finance 1–27. https://doi.org/10.1017/S1474747220000189

Bravo, J. M., Ayuso, M., Holzmann, R., and Palmer, E. (2020). Coping with the Life Expectancy Gap: Amending the Retirement Age to Restore Actuarial Neutrality Across Generations. Preprint presented at the SMTDA 2020 - 6th Stochastic Modeling Techniques and Data Analysis International Conference, Barcelona 2-5 June.

Brouhns, N., Denuit, M., and Vermunt, J. (2002). A Poisson Log-Bilinear Regression approach to the construction of projected life tables. Insurance: Mathematics and Economics, 31: 373–393.

Brouhns, N., Denuit, M., and Van Keilegom, I. (2005). Bootstrapping the Poisson Log-Bilinear Model for Mortality Forecasting. Scandinavian Actuarial Journal, 3: 212–224.

Cairns, A. J. G., Blake, D. and Dowd, K. (2006a). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. Journal of Risk and Insurance, 73: 687–718.

Cairns, A. J. G., Blake, D. and Dowd, K. (2006b). Pricing death: Framework for the valuation and securitization of mortality risk. ASTIN Bulletin, 36: 79–120.

Cairns, A. J. G, Blake, D. and Dowd, K. (2008). Modelling and Management of Mortality Risk: A Review. Scandinavian Actuarial Journal, 2-3: 79–113.

Cairns, A., Blake, D., Dowd, K., Coughlan, G., and Khalaf-Allah, M. (2011). Bayesian stochastic mortality modelling for two populations. ASTIN Bulletin, 41: 29–59.

Cairns, A., Blake, D., Dowd, K., Coughlan, G., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. North American Actuarial Journal, 13(1): 1–35.

Cairns, A., Kallestrup-Lamb, M., Rosenskjold, C., Blake, D., and Dowd, K. (2019). Modelling socio-economic differences in mortality using a new affluence index. ASTIN Bulletin, 49(3): 555–590.

Camarda, C. G. (2019). Smooth constrained mortality forecasting. Demographic Research, 41(38): 1091–1130.

Chen, A., Hieber, P. and Rach, M. (2020). Optimal retirement products under subjective mortality beliefs. Insurance: Mathematics and Economics. https://doi.org/10.1016/j.insmatheco.2020.07.002.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., and Cutler, D. (2016). The association between income and life expectancy in the United States, 2001-2014. The Journal of the American Medical Association, 315 (14): 1750–1766.

Currie, I. (2006). Smoothing and forecasting mortality rates with P-Splines. DP Heriot Watt University.

Currie, I. (2016). On fitting generalized linear and non-linear models of mortality. Scandinavian Actuarial Journal, 4: 356–383.

Currie, I., Durban, M., and Eilers, P. (2004). Smoothing and forecasting mortality rates. Statistical Modelling, 4: 279–298.

Dahl, M. (2004). Stochastic mortality in life insurance: Market reserves and mortality-linked insurance contracts. Insurance: Mathematics and Economics, 35(1): 113–136.

Denuit, M., and Goderniaux, A. C. (2005). Closing and projecting life tables using log-linear models. Bulletin of the Swiss Association of Actuaries, 1: 29–48.

De Menezes, L. M., Bunn, D. W., and Taylor, J. W. (2000). Review of guidelines for the use of combined forecasts. European Journal of Operational Research, 120 (1): 190–204.

Dowd, K., Cairns, A., Blake, D., Coughlan, G., and Khalaf-Allah, M. (2011). A gravity model of mortality rates for two related populations. North American Actuarial Journal, 15: 334–356.

Dowd, K., Cairns, A., Blake, D., Coughlan, G., Epstein, D. and Khalaf-Allah, M. (2010). Backtesting stochastic mortality models. North American Actuarial Journal, 14(3): 281–298.

European Commission (2015). The 2015 pension adequacy report: Current and future income adequacy in old age in the EU, Vol. I, Luxembourg: Publications Office of the European Union.

Haberman S., and Renshaw A. (2011). A Comparative Study of Parametric Mortality Projection Models. Insurance: Mathematics and Economics, 48(1): 35–55.

Hansen, P., Lunde, A., and Nason, J. (2011). The model confidence set. Econometrica, 79: 453–497.

Holzmann, R., Alonso-García, J., Labit-Hardy, H., and Villegas, A.M. (2020). NDC schemes and heterogeneity in longevity: Proposals for redesign. In Holzmann, R., Palmer, E., Palacios, R., and Sacchi, S. (Editors). Progress and challenges of Nonfinancial Defined Contribution Pension (NDC) Schemes, Volume 1: Addressing marginalization, polarization, and the labor market. Washington, DC: World Bank, 307–332.

Holzmann, R., Ayuso, M., Alaminos, E., and Bravo, J. M. (2020). Life Cycle Saving and Dissaving Revisited across Three-tiered Income Groups: Starting hypotheses, refinement through literature review, and ideas for empirical testing. Submitted to Work, Aging and Retirement.

Hsiao, C. and Wan, S. (2014). Is there an optimal forecast combination? Journal of Econometrics, 178 (2): 294–309.

Huang, J. Z., Shen, H. and Buja, A. (2009). The analysis of two-way functional data using two-way regularized singular value decompositions. Journal of the American Statistical Association 104 (488): 1609–1620.

Human Mortality Database (2019). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 15-August-2019).

Hunt, A., and Blake, D. (2014). A general procedure for constructing mortality models. North American Actuarial Journal, 18(1): 116–138.

Hunt, A., and Blake, D. (2020). On the structure and classification of mortality models. North American Actuarial Journal. doi: 10.1080/10920277.2019.1649156

Hunt, A., and Villegas, A. M. (2015). Robustness and convergence in the Lee-Carter model with cohort effects. Insurance: Mathematics and Economics 64: 186–202.

Hunt, A., and Villegas, A. M. (2017). Mortality improvement rates: Modeling and parameter uncertainty. In Living to 100, Society of Actuaries International Symposium. Retrieved from https://www.soa.org/essays-monographs/2017-living-to-100/2017-living-100-monograph-hunt-villegas-paper.pdf

Hyndman, R., and Ullah, S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. Computational Statistics & Data Analysis, 51: 4942–4956.

Hyndman, R.J. (2019). demography: Forecasting mortality, fertility, migration and population data. https://CRAN.R-project.org/package=demography.

Jose, V.R., and Winkler, R. L. (2008). Simple robust averages of forecasts: some empirical results. International Journal of Forecasting 24(1): 163–169.

Kass, R. E., and Raftery, A. E. (1995). Bayes factors. Journal of the American Statistical Association, 90: 773–795.

Koissi, M.C., Shapiro, A.F., and Högnäs, G. (2006). Evaluating and extending the Lee–Carter model for mortality forecasting: Bootstrap confidence interval. Insurance: Mathematics and Economics 38(1): 1–20.

Kontis, V., Bennett, J., Mathers, C., Li, G., Foreman, K. and Ezzati, M. (2017). Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. Lancet, 389 (10076): 1323–1335.

Lee, R., and Carter, L. (1992). Modeling and forecasting U.S. mortality. Journal of the American Statistical Association, 87(419): 659–671.

Li, J., Zhou, R., and Hardy, M. R. (2015). A step-by-step guide to building two-population stochastic mortality models. Insurance: Mathematics and Economics, 63: 121–134.

Li, N., and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter Method. Demography, 42: 575–594.

Makridakis, S., and Winkler, R. (1983). Averages of forecasts: Some empirical results. Management Science, 987–996.

Meneu, R., Devesa, E., Devesa, M., Domínguez, I., and Encinas, B. (2016). Adjustment mechanisms and intergenerational actuarial neutrality in pension reforms. International Social Security Review, 69(1): 87–107.

Milevsky, M., and Promislow, S. (2001). Mortality derivatives and the option to annuitise. Insurance: Mathematics and Economics, 29: 299–318.

NASEM (National Academy of Sciences, Engineering and Medicine) (2015). The growing gap in life expectancy by income: Implications for federal programs and policy responses. Committee on the Long-Run Macroeconomic Effects of the Aging U.S. Population Phase II. Washington, DC: National Academies Press.

Oeppen, J., and Vaupel, J. (2002). Broken limits to life expectancy. Science, 296 (5570): 1029–31.

OECD (2017). Pensions at a Glance 2017: OECD and G20 Indicators. Paris: OECD Publishing.

Olshansky, S. J., Carnes, B. A., and Desesquelles, A. (2001). Demography - prospects for human longevity. Science, 291: 1491–92.

Palmer, E. (2013). Generic NDC: Equilibrium, valuation, and risk sharing with and without NDC bonds. In Holzmann, R., Palmer, E., and Robalino, D. (Eds). Nonfinancial defined contribution Pension Schemes in a changing pension world, Vol. 2, Gender, politics, and financial stability, Washington D.C., World Bank: 309–333.

Palmer, E., Alho, J., and Zhao de Gosson de Varennes, Y. Z. (2019). Projecting cohort life expectancy from the changing relationship between period and cohort mortalities. University of Uppsala/University of Helsinki (mimeo), January.

Palmer, E., and Zhao de Gosson de Varennes, Y. Z. (2020). Annuities in (N)DC Pension Schemes: Design, heterogeneity, and estimation issues. In Holzmann, R., Palmer, E., Palacios, R., and Sacchi, S. (Eds). Progress and challenges of Nonfinancial Defined Contribution Pension Schemes, Vol. 1, Addressing marginalization, polarization, and the labor market. Washington, DC, World Bank, 281–306.

Plat, R. (2009). On stochastic mortality modeling. Insurance: Mathematics and Economics, 45(3): 393–404.

Queisser, M., and Whitehouse, E. R. (2006). Neutral or fair? Actuarial concepts and pension-system design. OECD social, employment and migration working paper, 40, Paris, OECD.

Raftery, A., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). Using Bayesian Model Averaging to calibrate forecast ensembles. Journal of American Meteorological society, 133: 1155–1174.

Renshaw, A., and Haberman, S. (2003). Lee-Carter mortality forecasting with Age-Specific Enhancement. Insurance: Mathematics and Economics, 33(2): 255–272.

Renshaw, A., and Haberman, S. (2006). A cohort-based extension to the Lee-Carter Model for mortality reduction factors. Insurance: Mathematics and Economics, 38(3): 556–570.

Samuels, J.D., and Sekkel, R.M. (2017). Model confidence sets and forecast combination. International Journal of Forecasting, 33(1): 48–60.

Sánchez-Romero, M., Lee, R., and Prskawetz, A. (2020). Redistributive effects of different pension systems when longevity varies by socioeconomic status. The Journal of the Economics of Ageing 17, https://doi.org/10.1016/j.jeoa.2020.100259.

Sasson, I. (2016). Trends in life expectancy and lifespan variation by educational attainment: United states, 1990–2010. Demography, 53(2): 269–293.

Schrager, D. (2006). Affine stochastic mortality. Insurance: Mathematics and Economics, 35(1): 81–97.

Shang, H., and Haberman, S. (2018). Model confidence sets and forecast combination: an application to age-specific mortality. Genus, 74, 19.

Shang, H.L., Booth, H., and Hyndman, R.J. (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. Demographic Research 25: 173–214.

Stock, J., and Watson, M. (2004). Combination forecasts of output growth in a seven-country data set. Journal of Forecasting, 23: 405–430.

Turek, D. and Fletcher, D. (2012). Model-Averaged Wald Confidence Intervals. Computational Statistics and Data Analysis, 56(9): 2809–2815.

Turner, J. A. (2009). Social security financing: Automatic adjustments to restore solvency (Research report, 2009-1). Washington, DC: AARP Public Policy Institute.

Vidal-Meliá, C., Boado-Penas, M. C., and Settergren, O. (2009). Automatic balance mechanisms in pay-as-you-go pension systems. The Geneva Papers on Risk and Insurance: Issues and Practice, 34(2): 287–317.

Villegas, A. M., Haberman, S., Kaishev, V. V., and Millossovich, P. (2017). A comparative study of two-population models for the assessment of basis risk in longevity hedges. ASTIN Bulletin, 47(3): 631–679.

Villegas, A. M., Millossovich, P., and Kaishev, V. V. (2018). StMoMo: An R Package for stochastic mortality modelling. Journal of Statistical Software, 84(3): 1-38.

Yang, S., Khang, Y. H., Harper, S., Davey Smith, G., Leon, D. A., and Lynch, J. (2010). Understanding the rapid increase in life expectancy in South Korea. American journal of public health, 100(5): 896–903.

Whitehouse, E. (2007). Life-expectancy risk and pensions: Who bears the burden?. OECD Social, Employment and Migration Working Papers, 60, OECD Publishing, Paris.

Zhang, L. and Pan, C. (2013). RobRSVD: Robust Regularized Singular Value Decomposition. https://CRAN.R-project.org/package=RobRSVD.

Zhang, L., Shen, H., and Huang, J. Z. (2013). Robust regularized singular value decomposition with application to mortality data. The Annals of Applied Statistics, 7(3): 1540-1561.

Zhu, W., Tan, K., and Wang, C-W. (2017). Modeling multicountry longevity risk with mortality dependence: A Lévy Subordinated Hierarchical Archimedean Copulas approach. The Journal of Risk and Insurance, 84(S1): 477-494.

Appendix A Supplementary results

Figure 8: Average unisex life expectancy gap per year and age, 1960-2050

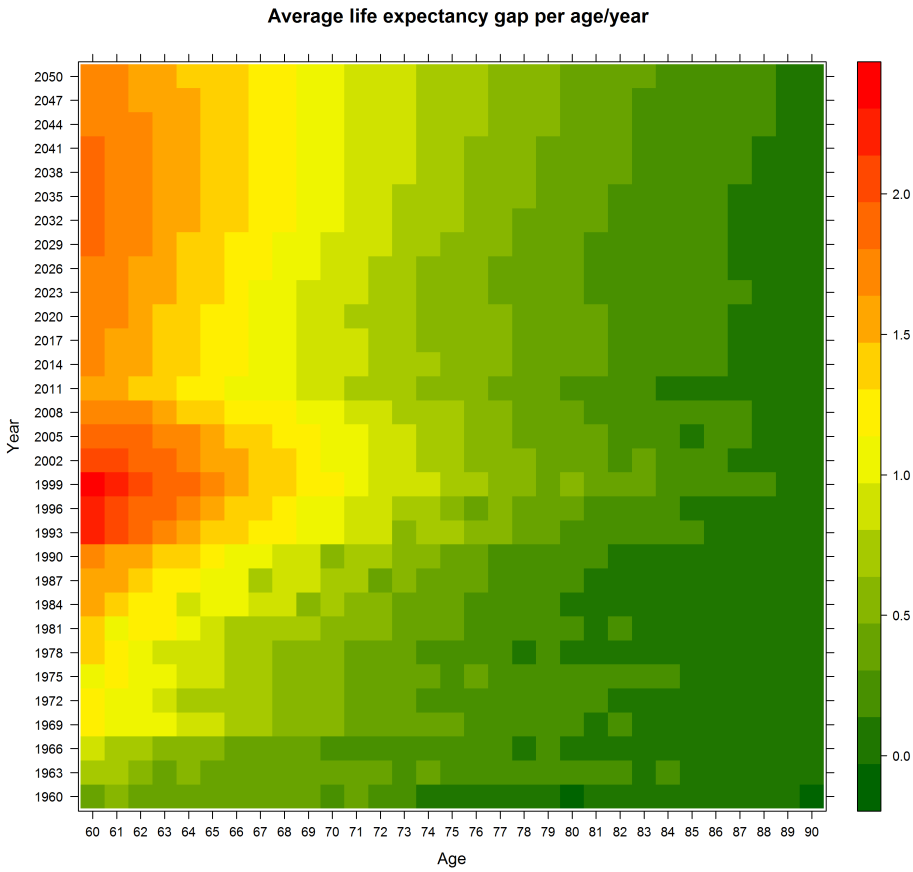


Table 6: Ranking of Cohort life expectancy at age 60 in selected years: Men

Rank	Country	1960	Country	2019	Country	2050
1	ISL	18.91	KOR	27.47	KOR	34.04
2	SWE	17.63	AUS	27.20	AUS	31.56
3	NOR	17.45	CAN	26.50	CAN	30.51
4	NLD	17.27	CHE	26.34	IRL	30.33
5	DNK	17.02	NZL	26.25	JPN	30.30
6	CHE	16.92	ISR	26.19	ENW	30.28
7	ESP	16.90	ENW	26.05	NZL	30.25
8	CAN	16.83	IRL	25.96	DEU	30.16
9	JPN	16.67	JPN	25.82	SVN	30.10
10	LTU	16.64	ITA	25.78	CHE	29.96
11	ITA	16.58	ISL	25.52	TWN	29.93
12	BGR	16.43	SWE	25.40	ISR	29.88
13	BLR	16.30	NIR	25.34	ITA	29.45
14	FRA	16.18	DEU	25.22	NIR	29.07
15	USA	16.13	FRA	25.15	SCO	28.79
16	LVA	16.13	TWN	25.11	SWE	28.72
17	PRT	15.91	NOR	24.98	ISL	28.55
18	NZL	15.76	LUX	24.93	LUX	28.22
19	SVK	15.69	ESP	24.91	ESP	28.20
20	EST	15.53	NLD	24.83	NOR	28.10
21	BEL	15.50	SVN	24.70	NLD	28.01
22	POL	15.44	SCO	24.52	FIN	27.98
23	UKR	15.43	DNK	24.43	FRA	27.91
24	AUS	15.30	BEL	24.20	BEL	27.90
25	HUN	15.29	FIN	24.20	AUT	27.82
26	ENW	15.15	USA	24.12	DNK	27.82
27	LUX	15.15	AUT	24.11	USA	27.68
28	NIR	15.01	GRC	24.07	PRT	27.20
29	IRL	14.96	PRT	23.82	HRV	27.19
30	AUT	14.95	CHL	22.88	GRC	26.70
31	FIN	14.82	HRV	22.01	CHL	26.23
32	RUS	14.47	CZE	20.80	CZE	23.36
33	SCO	14.36	POL	20.38	POL	22.79
34	CZE	14.31	SVK	19.83	SVK	22.55
35			EST	18.35	HUN	19.92
36			HUN	18.16	EST	19.52
37			LVA	16.33	LVA	16.22
38			LTU	16.27	BGR	15.79
39			BGR	16.02	LTU	15.67
40			RUS	13.90	RUS	12.91
41			BLR	13.34	BLR	11.29
42			UKR	13.03	UKR	10.93

Table 7: Ranking of Cohort life expectancy at age 60 in selected years: Women

Rank	Country	1960	Country	2019	Country	2050
1	ISL	22.51	KOR	34.16	KOR	40.19
2	CAN	21.80	JPN	31.98	JPN	36.24
3	NLD	21.56	ESP	30.62	FRA	34.03
4	SWE	21.49	FRA	30.52	SVN	33.95
5	NOR	21.48	ITA	29.62	ESP	33.74
6	FRA	21.43	CHE	29.40	CHE	33.66
7	CHE	21.30	ISR	29.33	ITA	33.33
8	USA	21.13	SVN	29.29	TWN	33.32
9	DNK	21.04	AUS	29.24	ISR	33.32
10	ITA	20.67	TWN	29.10	LUX	33.15
11	JPN	20.64	CAN	29.04	DEU	32.72
12	ESP	20.62	LUX	28.96	PRT	32.57
13	ENW	20.24	PRT	28.73	CAN	32.40
14	AUS	20.20	DEU	28.66	AUS	32.28
15	LTU	20.05	IRL	28.51	IRL	32.05
16	BLR	20.00	ENW	28.40	BEL	32.04
17	BEL	19.99	BEL	28.39	ENW	31.99
18	NZL	19.97	NZL	28.22	NZL	31.60
19	EST	19.68	FIN	28.22	HRV	31.34
20	LVA	19.60	GRC	28.19	CHL	31.30
21	PRT	19.53	SWE	28.18	SWE	31.21
22	FIN	19.46	AUT	27.84	FIN	31.21
23	AUT	19.44	NLD	27.84	GRC	31.14
24	SCO	19.31	ISL	27.78	AUT	31.11
25	NIR	19.23	CHL	27.70	NIR	30.96
26	POL	19.19	NOR	27.56	NLD	30.75
27	LUX	18.94	NIR	27.48	NOR	30.39
28	SVK	18.85	DNK	27.04	DNK	30.26
29	RUS	18.61	HRV	26.75	ISL	30.20
30	UKR	18.53	USA	26.73	SCO	29.92
31	BGR	18.51	SCO	26.51	USA	29.32
32	CZE	18.39	EST	26.35	EST	29.21
33	HUN	18.38	POL	25.59	POL	28.81
34	IRL	18.29	CZE	25.49	CZE	28.40
35			SVK	24.33	SVK	26.96
36			HUN	23.37	HUN	26.43
37			LTU	23.26	LTU	24.13
38			LVA	22.63	LVA	23.70
39			BGR	21.13	BGR	21.88
40			BLR	20.93	BLR	21.12
41			RUS	19.96	RUS	19.64
42			UKR	18.90	UKR	18.20