Smooth Constrained Mortality Forecasting: Supplementary material

Carlo G. Camarda
Institut National d'Études Démographiques
carlo-giovanni.camarda@ined.fr

Introduction

In the following document, in order to further assess the performance of the suggested CP-spline model, we perform an out-of-sample forecast exercise and test model stability by changing the time-window over which the model is estimated.

In Section A, we compare CP-splines with other five alternative forecasting methods. We perform this study on 8 populations. Effects of the change in time-windows (Sec. B) are presented for the suggested model and for a variant of the model proposed by Hyndman and Ullah (2007) which results as runner-up model after the previous comparison.

Finally, in Section C, we evaluate the impact and consequences of the choice of level of confidence on future rate-of-change over time. In other words, we check the effects of either reducing or increasing the 50% confidence interval we adopted for computing $\boldsymbol{\delta}_L^{t_1}$ and $\boldsymbol{\delta}_U^{t_1}$ in Section 3.2 of the paper.

A Comparison and out-of-sample forecast

In this section, we used mortality data for four countries (USA, Denmark, Japan and France) for both males and females. Data were taken from the Human Mortality Database (2018). We estimate six alternative and comparable forecasting methods:

- the proposed CP-splines: CP-S
- the smooth Lee-Carter variant by Delwarde et al. (2007): LCsmo
- the Lee-Carter variant by Lee and Miller (2001): LM
- the Lee-Carter variant by Booth et al. (2002): BMS
- the functional data approach by Hyndman and Ullah (2007): HU
- a robust version of the HU model: HUrob

We selected these approaches for two main reasons: they all model mortality in a single population by single year of age, and routines for estimating these models are freely available (e.g. Hyndman et al., 2018).

We first model all populations over the period 1960-2014 and we forecast up to 2050. The first three columns in Table A.1 and A.2 presents outcomes in terms of accuracy to model observed data. Goodness-of-fit is measured by Deviance (Dev, McCullagh and

Nelder, 1989) and effective dimension (ED) gives a measures of the complexity of the model. The Bayesian Information Criterion (BIC) balances these two statistics and it allows us to evaluate precision in describing data. In all instances, CP-splines outperforms its competitors.

Based on previous estimations, we present life expectancy at birth (e_0) and average number of life years lost at birth ($e_0^\dagger$) in 2050. See associated columns in Table A.1 and A.2. For females, outcomes in e_0 are similar with exception of HU: this model is either too pessimistic (USA, Japan and France) or too optimistic (Denmark). Similar results can be seen for males. Lifespan disparity in 2050 is more model-dependent, especially for females and Danish males.

We assess accuracy and performance of the all six models by an out-of-sample forecast against the observed trends in life expectancy at birth (e_0), average number of life years lost at birth ($e_0^\dagger$) and log-mortality over all ages and years ($\boldsymbol{\eta}$). Specifically, we fit all models from 1960 to 2004, and we forecast mortality 10 years ahead (2005-2014), comparing these forecast values to those observed over that decade.

We measure the accuracy of the models by computing mean absolute error (MAE), root mean square error (RMSE) and mean error (ME). Whereas ME describes the forecast bias due to the model, the other measures assess the performance of the model on the same units as the measured variable (Chatfield, 2000).

Right panels in Table A.1 and A.2 present the outcomes of this test. Given 4 populations, 2 sexes, 3 accuracy measures and 3 demographic indicators, we compare six alternative approaches over 72 values: the proposed CP-spline model outperforms its competitor 51 times (71%), followed by HU and LM with 8 and 4 times, respectively.

To give a graphical overview of these results, Figure A.1 shows three selected ages over time for US males mortality data from the presented out-of-sample forecast. This plot gives an immediate picture of the goodness-of-fit and forecasting reliability for both CP-spline and Hyndman-Ullah approach. To ensure a fair comparison, a residual bootstrap approach is applied to acquire predictive intervals on both models. In particular, the Hyndman-Ullah merely extrapolates a linear pattern which is already inaccurate to describe the mortality patterns. On the contrary, the proposed CP-spline approach is able to fully capture observed mortality developments, including erratic trends in the latest years.

Finally, an out-of-sample forecast exercise as presented in this section could also be performed to set the confidence interval used to compute $\boldsymbol{\delta}_L^{t_1}$ and $\boldsymbol{\delta}_U^{t_1}$ (cf. Section C).

B Changing the time-window

To test the stability of the CP-splines and of the second classified model by Hyndman and Ullah (2007), we apply both of them to different periods of time. We select progressively smaller data periods, going from t_0 to 2014. We start with a value of t_0 equal to 1935 and end with t_0 equal to 1985. The models applied to different time-windows are used to produce forecasts of life expectancy at birth and $e_0^\dagger$ in the year 2050, for both USA males and Danish females. Results are plotted in Figure B.1.

Whereas life expectancy at birth for Danish females is barely affected by the changes in the modelled periods when forecast is obtained by CP-splines, outcomes from Hyndman-Ullah model are very erratic and time-window dependent. The observed intervals of e_0 in 2050 obtained by changing the modelled periods are $[86.32, 87.29]$ for the CP-splines

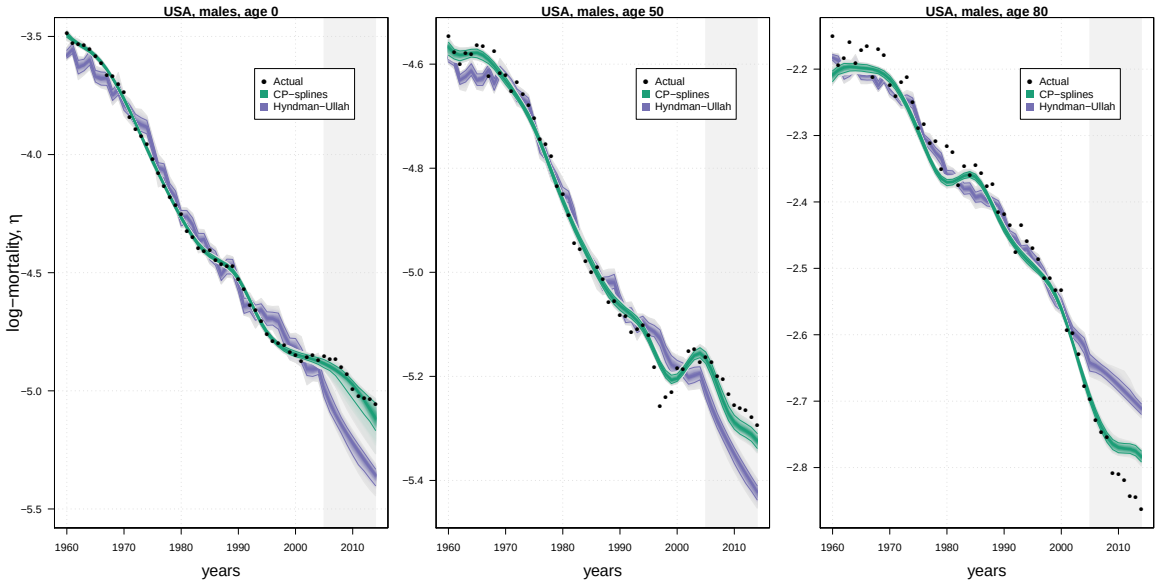

Figure A.1: Illustration of the outcomes from the out-of-sample forecast exercise. Actual, model and forecast mortality along with their bootstrapped distributions for selected ages (0, 50 and 80) over years. The CP-spline approach is compared with a variant of the Hyndman-Ullah model. USA males, ages 0-105. Fitting period 1960-2004, forecast up to 2014 and compared to actual values in 2005-2014.

and $[85.02, 88.96]$ years for the Hyndman-Ullah model. For US males, the values of life expectancy at birth in 2050 are sensitive to the time-windows used in modelling the data, for both models, although outcomes for the Hyndman-Ullah model are more stable with respect to Danish case. The difference between the largest and smallest value of e_0 estimated by different time-windows is 2.43 (2.4) years for the CP-spline (Hyndman-Ullah) model.

Moreover, for USA males, the Hyndman-Ullah tends to regularly provide higher values than those obtained by the CP-spline model. This feature is likely due to the ability of the proposed method to capture recent mortality stagnation among USA males, as portrayed in the middle panel of Figure A.1 and Figure 9 in the paper. On the contrary, although Hyndman-Ullah model generalizes a Lee-Carter model using six principal components to describe mortality development, it is still too rigid to incorporate further fluctuations in its forecasts.

Concerning lifespan variability, the proposed model gives a rather constant value of $e_0^\dagger$ in 2050, whatever the modelled period and for both datasets. Conversely, the values of $e_0^\dagger$ in 2050 obtained by the Hyndman-Ullah model are highly sensitive to the choice of the time-window selected to model the data, with exceptional peaks and troughs over t_0, especially for Danish females.

C Confidence level in rate-of-change over time

In Section 3.2 of the paper we suggest using a 50% confidence interval of the smooth rate-of-change over time to compute $\boldsymbol{\delta}_L^{t_1}$ and $\boldsymbol{\delta}_U^{t_1}$. In other words, we constrain future mortality at each age to lie within the interquartile range of the observed mortality improvement.

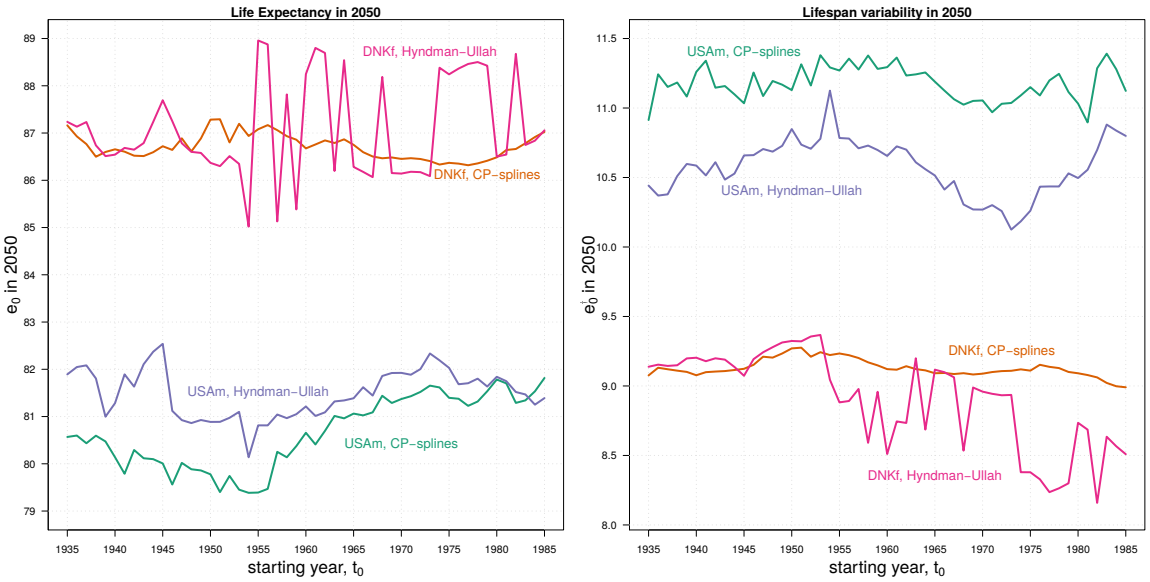

Figure B.1: Illustration of the outcomes from different time-windows for modelling the data. Life expectancy at birth (left panel) and a measure of lifespan variability ($e_0^\dagger$, right panel) in 2050. Modelled period t_0 to 2014. Year t_0 given on horizontal axis. The proposed CP-spline approach is compared with a variant of the Hyndman-Ullah model. USA males and Denmark females, ages 0-105.

In this section we assess the consequences of this choice, showing that larger percentages may lead to unreasonable outcomes, especially when large fluctuations are observed in the past. We already pointed out in the paper that extremely small values for computing $\boldsymbol{\delta}_L^{t_1}$ and $\boldsymbol{\delta}_U^{t_1}$ leads to an excessively rigid model, similar to a Lee-Carter model.

Figure C.1 presents the results for the two illustrative datasets when different confidence intervals are used to compute $\boldsymbol{\delta}_L^{t_1}$ and $\boldsymbol{\delta}_U^{t_1}$. In the top panels, we show both life expectancy and a measure of lifespan variability ($e_0^\dagger$). The time trend for a specific age (35) is portrayed in the bottom panels for Danish females (left) and USA males (right). We select this age since it displays relatively large fluctuations for USA males and therefore differences in changing confidence levels are more visible.

It is clear that for Danish females, selecting different percentages only slightly changes future trends in summary measures and log-mortality. On the other hand, future mortality development for USA males depends upon the choice of confidence interval for computing $\boldsymbol{\delta}_L^{t_1}$ and $\boldsymbol{\delta}_U^{t_1}$. Whereas values below 50% produce similar outcomes in terms of e_0 and $e_0^\dagger$, allowing a high degree of flexibility in future mortality rate-of-change leads to unreasonably low mortality rates (see purple and green lines in the bottom-right panel), and consequently high future life expectancy and an odd pattern of lifespan variability over a short-term forecasting horizon. This is a consequence of the highly erratic pattern of US mortality in the recent decades. Taking age 35 as an example, allowing future rate-of-change to lie between the 95% confidence intervals of past rate-of-change means that the rapid mortality improvements experienced by US males in the late 1990s would be conceivable for future years. However this fast improvement at age 35 was due mainly to the introduction of antiretroviral therapies to treat HIV-AIDS. Meanwhile, by choosing 50%, we also consider as implausible for the future a mortality increase such as that experienced by US males at age 35 in the 80s due to the HIV-AIDS epidemic.

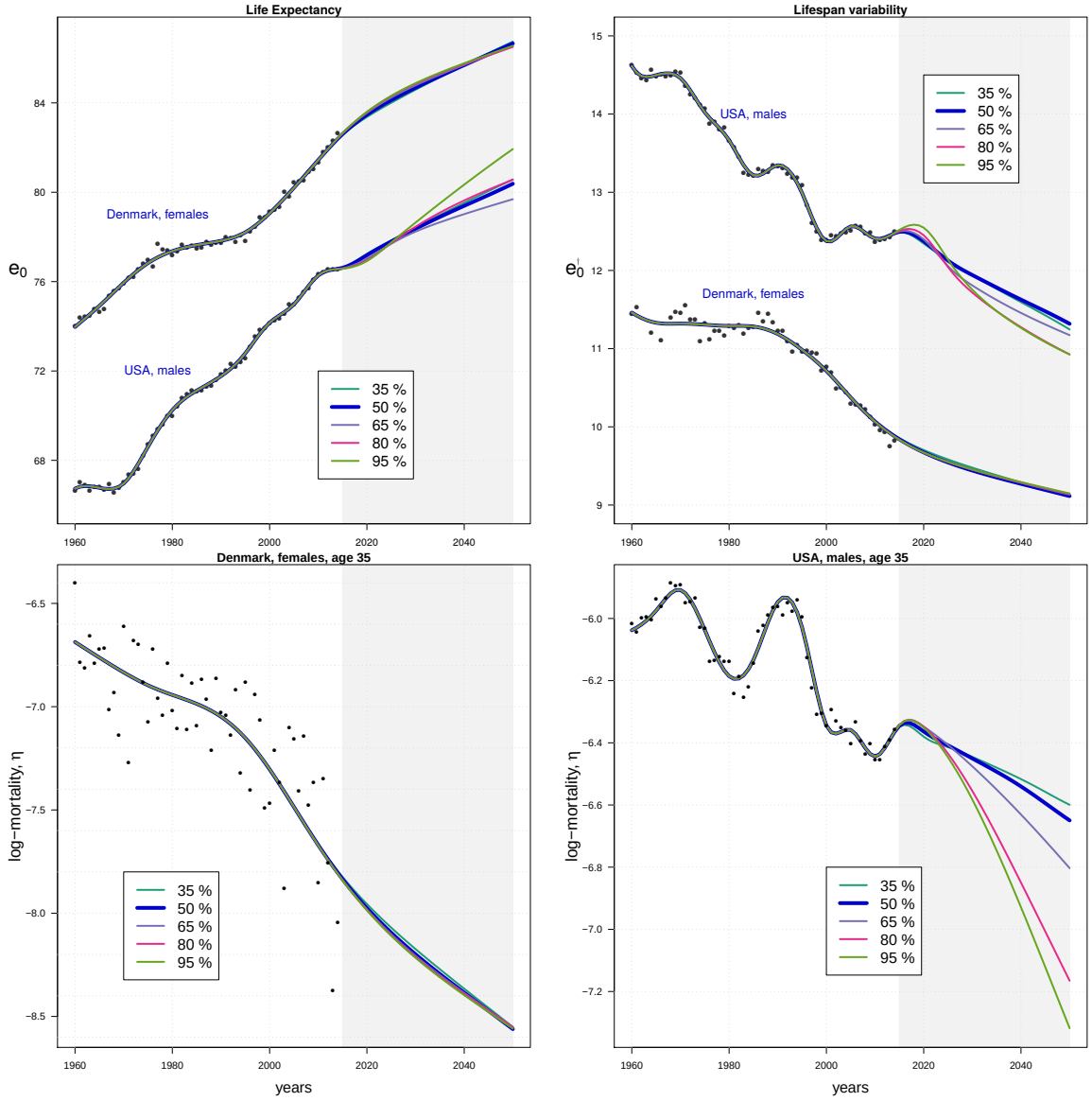

Figure C.1: Illustration of the outcomes of different levels of confidence for computing $\boldsymbol{\delta}_L^{t_1}$ and $\boldsymbol{\delta}_U^{t_1}$. We used 35, 50, 65, 80 and 95% confidence intervals from the smooth observed rate-of-change over time. Top panels: Life expectancy at birth (left) and a measure of lifespan variability ($e_0^\dagger$, right). Bottom panels: death rates for age 35 over years. *CP*-spline model on Danish females and US males, ages 0-105, years 1960-2014, forecast up to 2050.

References

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

Chatfield, C. (2000). *Time-series forecasting*. CRC Press.

Delwarde, A., M. Denuit, and P. H. C. Eilers (2007). Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach. *Statistical Modelling 7*, 29–48.

Human Mortality Database (2018). *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).* Available at `www.mortality.org`. Data downloaded on January 2018.

Hyndman, R. J., H. Booth, L. Tickle, and J. Maindonald (2018). *demography: Forecasting Mortality, Fertility, Migration and Population Data.* . R package version 1.21.

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

Lee, R. D. and T. Miller (2001). Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. *Demography 38*, 537–549.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* (2nd ed.). Monographs on Statistics Applied Probability. London: Chapman & Hall.

Table A.1: Numerical outcomes from the estimated and forecast models as well as for the out-of-sample forecast exercise. Females. Models estimated over years 1960-2014, ages 0-105 and forecast up to 2050. Deviance (Dev), Effective Dimension (ED) and Bayesian Information Criterion (BIC) are presented to compare performances of the models in the years 1960-2014. Lower values of BIC (in bold) indicates a better fit. Values for on e_0, $e_0^\dagger$ in 2050 are also provided. Out-of-sample forecast is carried out with fitting period 1960-2004, forecast up to 2014 and compared to actual values in 2005-2014. Models are compared using Mean absolute error (MAE), root mean square error (RMSE) and mean error (ME). Accuracy measures are computed on e_0, $e_0^\dagger$ and $\boldsymbol{\eta}$. Lower values of the MAE and the RMSE, and ME closer to zero (in bold) correspond to greater accuracy. Populations: USA, Denmark, Japan and France, females. Models: CP-splines (CP-S), smooth Lee-Carter (LCsmo), Lee-Miller (LM), Booth-Maindonald-Smith (BMS), Hyndman-Ullah (HU) and a robust version of Hyndman-Ullah (HUrob).

		Observed years + Forecast					Out-of-sample: 1960-2004 → 2005-2014								
		1960-2014			2050		e_0			$e_0^\dagger$			$\boldsymbol{\eta}$		
		Dev	ED	BIC	e_0	$e_0^\dagger$	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE	ME
USAf	CP-S	27330	258	**29567**	84.96	10.29	**0.062**	**0.069**	**0.031**	**0.079**	0.121	0.055	**0.046**	**0.071**	**-0.007**
	LCsmo	84424	249	86579	85.66	9.57	0.245	0.272	0.242	0.091	0.116	0.005	0.093	0.122	-0.027
	LM	87236	301	89845	85.46	9.65	0.177	0.200	0.158	0.089	**0.110**	**-0.003**	0.088	0.115	-0.011
	BMS	87236	301	89845	85.46	9.65	0.252	0.280	0.251	0.094	0.114	-0.030	0.089	0.116	-0.019
	HU	24952	1276	36016	84.50	9.57	0.390	0.428	0.390	0.152	0.189	0.151	0.081	0.103	-0.027
	HUrob	35912	1276	46976	85.05	10.11	0.292	0.323	0.291	0.113	0.146	0.106	0.090	0.117	-0.018
DNKf	CP-S	6605	100	**7470**	86.68	9.12	**0.830**	**0.922**	**0.830**	**0.258**	**0.278**	**-0.258**	**0.208**	**0.328**	-0.135
	LCsmo	11059	135	12233	86.80	9.22	1.592	1.661	1.592	0.872	0.889	-0.872	0.309	0.444	-0.280
	LM	12065	301	14674	87.04	9.26	0.951	1.057	0.951	0.716	0.736	-0.716	0.264	0.372	**-0.125**
	BMS	11247	301	13857	86.94	9.28	1.573	1.644	1.573	0.755	0.774	-0.755	0.280	0.395	-0.207
	HU	6584	1276	17648	88.25	8.51	1.476	1.592	1.476	0.574	0.600	-0.574	0.255	0.370	-0.202
	HUrob	7195	1276	18259	84.14	9.51	1.880	1.970	1.880	0.875	0.900	-0.875	0.283	0.396	-0.234
JPNf	CP-S	14441	228	**16420**	93.32	8.32	**0.597**	**0.676**	**-0.597**	**0.102**	**0.116**	**-0.079**	**0.101**	0.138	0.088
	LCsmo	68415	239	70488	93.94	7.22	1.122	1.197	-1.122	0.400	0.408	0.400	0.257	0.326	0.255
	LM	82310	301	84920	93.91	7.27	0.809	0.907	-0.809	0.521	0.527	0.521	0.219	0.280	0.206
	BMS	73420	301	76030	93.89	7.27	1.042	1.115	-1.042	0.570	0.576	0.570	0.243	0.305	0.235
	HU	10987	1276	22051	84.68	6.85	0.900	1.019	-0.900	0.202	0.224	-0.202	0.104	**0.137**	0.062
	HUrob	25398	1276	36462	93.72	7.29	0.823	0.927	-0.823	0.108	0.133	-0.090	0.106	0.139	**0.061**
FRAf	CP-S	13061	216	**14936**	90.55	8.75	**0.127**	**0.144**	**0.040**	**0.043**	**0.052**	**0.010**	**0.074**	**0.108**	-0.021
	LCsmo	20759	170	22229	90.89	7.89	0.342	0.368	-0.342	0.281	0.290	0.281	0.134	0.184	-0.030
	LM	23262	301	25872	90.82	7.91	0.245	0.281	-0.245	0.253	0.263	0.253	0.120	0.164	-0.026
	BMS	21599	301	24209	90.90	7.88	0.350	0.376	-0.350	0.280	0.290	0.280	0.121	0.164	**-0.015**
	HU	8515	1276	19578	89.69	8.42	0.237	0.263	0.237	0.073	0.080	0.067	0.096	0.136	-0.040
	HUrob	12850	1276	23914	89.98	8.20	0.458	0.475	0.458	0.160	0.173	0.160	0.098	0.127	-0.040

Table A.2: Numerical outcomes from the estimated and forecast models as well as for the out-of-sample forecast exercise. Males. Models estimated over years 1960-2014, ages 0-105 and forecast up to 2050. Deviance (Dev), Effective Dimension (ED) and Bayesian Information Criterion (BIC) are presented to compare performances of the models in the years 1960-2014. Lower values of BIC (in bold) indicates a better fit. Values for on e_0, $e_0^\dagger$ in 2050 are also provided. Out-of-sample forecast is carried out with fitting period 1960-2004, forecast up to 2014 and compared to actual values in 2005-2014. Models are compared using Mean absolute error (MAE), root mean square error (RMSE) and mean error (ME). Accuracy measures are computed on e_0, $e_0^\dagger$ and $\boldsymbol{\eta}$. Lower values of the MAE and the RMSE, and ME closer to zero (in bold) correspond to greater accuracy. Populations: USA, Denmark, Japan and France, males. Models: *CP*-splines (*CP*-S), smooth Lee-Carter (LCsmo), Lee-Miller (LM), Booth-Maindonald-Smith (BMS), Hyndman-Ullah (HU) and a robust version of Hyndman-Ullah (HUrob).

		Observed years + Forecast					Out-of-sample: 1960-2004 → 2005-2014								
		1960-2014			2050		e_0			$e_0^\dagger$			$\boldsymbol{\eta}$		
		Dev	ED	BIC	e_0	$e_0^\dagger$	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE	ME
USAm	*CP*-S	30977	305	**33621**	80.66	11.29	**0.128**	**0.146**	**0.070**	**0.227**	**0.261**	**0.227**	**0.057**	**0.079**	**-0.011**
	LCsmo	158590	227	160555	81.71	10.56	0.231	0.269	0.205	0.471	0.489	0.471	0.100	0.124	-0.033
	LM	168394	301	171003	81.62	10.47	0.226	0.263	0.196	0.449	0.467	0.449	0.095	0.118	-0.026
	BMS	164014	301	166624	81.66	10.46	0.213	0.249	0.180	0.453	0.472	0.453	0.095	0.118	-0.025
	HU	30348	1276	41412	81.21	10.66	0.364	0.422	0.339	0.418	0.452	0.418	0.089	0.115	-0.031
	HUrob	42532	1276	53596	81.41	10.93	0.299	0.341	0.288	0.535	0.548	0.535	0.100	0.125	-0.025
DNKm	*CP*-S	6887	110	**7843**	83.00	9.82	**1.270**	**1.459**	**1.270**	0.265	0.324	-0.251	**0.232**	**0.359**	-0.177
	LCsmo	10885	124	11960	82.47	9.35	1.451	1.615	1.451	0.227	0.275	-0.204	0.261	0.367	-0.160
	LM	10993	301	13603	82.61	10.25	1.503	1.667	1.503	0.395	0.446	-0.395	0.290	0.408	**-0.091**
	BMS	10940	301	13550	82.69	10.27	1.749	1.892	1.749	0.427	0.473	-0.427	0.289	0.395	-0.128
	HU	7140	1276	18204	83.36	8.29	1.445	1.553	1.445	**0.095**	**0.112**	**-0.027**	0.261	0.402	-0.218
	HUrob	9670	1276	20734	83.92	8.48	2.082	2.239	2.082	0.415	0.479	-0.415	0.306	0.437	-0.262
JPNm	*CP*-S	16583	233	**18607**	86.75	9.45	**0.336**	**0.374**	-0.336	0.116	0.133	-0.116	**0.083**	0.121	0.059
	LCsmo	47278	242	49374	87.31	8.87	0.525	0.556	-0.525	**0.099**	**0.113**	**0.077**	0.118	0.167	0.085
	LM	50000	301	52610	87.54	9.61	0.409	0.448	-0.409	0.117	0.137	0.109	0.113	0.175	0.081
	BMS	49024	301	51634	87.48	9.61	0.515	0.546	-0.515	0.129	0.151	0.126	0.120	0.182	0.092
	HU	17988	1276	29052	86.98	9.16	0.388	0.448	-0.388	0.330	0.356	-0.330	0.084	**0.115**	0.042
	HUrob	50826	1276	61889	86.68	9.09	0.324	0.373	**-0.324**	0.299	0.330	-0.299	0.095	0.129	**0.032**
FRAm	*CP*-S	12449	223	**14380**	84.38	9.94	0.483	0.571	0.483	**0.090**	0.110	**-0.090**	**0.082**	**0.113**	-0.054
	LCsmo	39132	181	40706	85.10	9.65	0.533	0.593	0.533	0.149	0.156	-0.149	0.152	0.210	-0.099
	LM	42491	301	45100	85.25	9.61	0.327	0.390	0.299	0.098	**0.108**	-0.098	0.132	0.181	-0.062
	BMS	41619	301	44229	85.21	9.62	0.459	0.523	0.459	0.131	0.140	-0.131	0.135	0.184	-0.076
	HU	10584	1276	21647	85.96	9.11	**0.127**	**0.145**	**0.069**	0.146	0.158	0.146	0.097	0.133	-0.029
	HUrob	23714	1276	34778	85.21	9.46	0.754	0.775	0.754	0.124	0.143	-0.124	0.145	0.211	**-0.024**