Probabilistic Projection of Subnational Life Expectancy

Hana Ševčíková, Adrian E. Raftery
University of Washington

Abstract

Projecting mortality for subnational units, or regions, is of great interest to practicing demographers. We seek a probabilistic method for projecting subnational life expectancy that is based on the national Bayesian hierarchical model used by the United Nations, and at the same time is easy to use. We propose three methods of this kind. Two of them are variants of simple scaling methods. The third method models life expectancy for a region as equal to national life expectancy plus a region-specific stochastic process which is a heteroskedastic first-order autoregressive process (AR(1)), with a variance that declines to a constant as life expectancy increases. We apply our models to data from 29 countries. In an out-of-sample comparison, the proposed methods outperformed other comparative methods and were well calibrated for individual regions. The AR(1) method performed best in terms of crossover patterns between regions. Although the methods work well for individual regions, there are some limitations when evaluating within-country variation. We identified four countries for which the AR(1) method either underestimated or overestimated the predictive between-region within-country standard deviation. However, none of the competing methods works better in this regard than the AR(1) method. In addition to providing the full distribution of subnational life expectancy, the methods can be used to obtain probabilistic forecasts of age-specific mortality rates.

1 Introduction

In 2015 for the first time, the United Nations Population Division issued official probabilistic population projections for all countries (United Nations 2015), using the methodology described by Raftery et al. (2012). The two key components that are forecasted probabilistically are the total fertility rate (TFR) and life expectancy at birth (e_0). They are both based on Bayesian hierarchical models (BHM) that are tailored to work well at the national level.

At the subnational level, e.g. provinces, states, counties, regions (which we will refer to as regions for the purposes of this article), probabilistic population projections are of great interest to national and local governments for planning, policy and decision-making (Rayer et al. 2009). In Ševčíková et al. (2018) a generic method was developed for projecting subnational TFR probabilistically, based on the BHM for national TFR. Here, we present a methodology for probabilistic projection of subnational e_0 which is based on the BHM for national life expectancy.

In demographic research over the past decade, a great deal of attention has been given to forecasting mortality. Most of these studies have focused on forecasting age-specific mortality rates, mostly at the national level, see e.g. Booth et al. (2006), Booth and Tickle (2008), Cairns et al. (2009), Stoeldraijer et al. (2013) and Janssen (2018) for reviews. Most of

the national models cannot be easily applied in subnational or small area settings due to various challenges, including missing data and measurement error. However, several recent studies have focused on forecasting mortality for subnational units, e.g. Bennett et al. (2015), Cairns et al. (2011), Bohk-Ewald and Rau (2017), proposing somewhat complex methods. Complexity, however, poses challenges to the adoption of a method by practitioners.

Wilson (2018) provided an excellent review of current methods from the point of view of practicing demographers in government and business. Stoeldraijer et al. (2013) evaluated differences in results when applying the most commonly used methods for mortality forecasting by statistical offices in Europe, of which the majority are simple extrapolation methods, either direct or Lee-Carter extrapolation (Lee and Carter 1992). They concluded that results are sensitive to the method and assumptions used, but that the sensitivity is small compared to uncertainty coming from different sources, which was however not assessed.

All these approaches focus on age-specific mortality rates from which life expectancy at birth or other measures are derived. A different approach, which is pursued in this paper, is to use life expectancy at birth as the main quantity to be forecasted, and then to disaggregate such forecasts into age-specific mortality rates. Oeppen and Vaupel (2002) showed stable trends in life expectancy over the course of 160 years with no signs of leveling off. Thus, forecasting life expectancy at birth directly, especially in subnational settings, could yield more accurate results and at the same time overcome various challenges, especially the data requirements.

A number of statistical agencies have adopted the approach of taking national mortality forecasts and creating regional forecasts via a simple relationship between them (Wilson 2018). These relationships are assumed to be time-invariant. However, a number of studies showed changes in regional mortality variation due to socioeconomic, environmental and other factors where the set of influencing factors might change over time as well (Bengtsson and van Poppel 2011; Kibele et al. 2015; Oosse 2003; Zajacova and Montez 2017).

Many statistical offices and planning agencies have not yet adopted probabilistic approaches in their mortality forecasts. Here we introduce three methods for forecasting regional mortality that are probabilistic, easy to implement, and are based on the established national BHM. They yield probabilistic forecasts of life expectancy at birth, each of which can be converted to age-specific mortality patterns, yielding a probability distribution of mortality rates. Such age-specific mortality foreasts can be further transformed, providing full probability distributions of any future life table quantity. Reporting mortality-related measures while taking account of uncertainty can become critical, especially in times of rapid population aging, for example for policy makers who deal with designing sustainable pension systems.

The paper is organized as follows. In Section 2 we describe our data, and in Section 3 the methods are presented. Section 4 compares the methods in terms of predictive distributions, out-of-sample validation, crossover patterns and within-country variance. An application of deriving age-specific mortality distribution is also presented here. The paper concludes with a discussion in Section 5.

2 Data

We use data on subnational female life expectancy at birth, e_0, for 447 regions from 29 countries spanning five continents; see Figure 1.

The data were collected from different sources, including EuroStat, Statistics Canada, the Australian Bureau of Statistics, and the Institute for Health Metrics and Evaluation (IHME). We put the data on a similar time scale to the national projections, resulting in a dataset

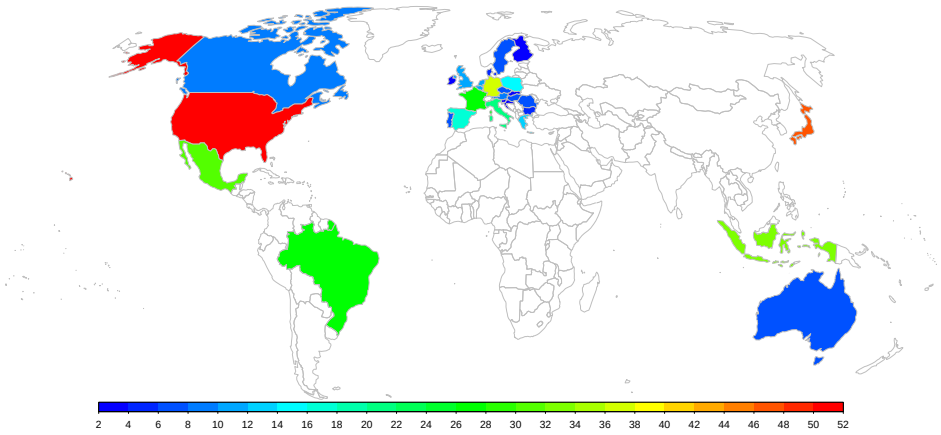

Figure 1: Countries from which we use data on subnational e_0. The color scale shows the number of subnational units.

for 5-year time periods, with the earliest being 1920-1925 and the latest being 2010-2015. Details of the data sources can be found in the Appendix. The majority of the regions are from European countries. However, the European time series are on average shorter than the time series from other continents, as seen in Table 1.

Table 1: Number of countries, regions and time points included in our estimation dataset of female e_0.

Area	**# countries**	**# regions**	**# time points**
Europe	22	238	1266
Asia	2	81	1134
Northern America	2	61	899
Latin America	2	59	826
Oceania	1	8	72
Total	29	447	4197

For national data, we use the time series from World Population Prospects (WPP) 2017 (United Nations 2017). We removed regions with volatile series and large variance from the estimation dataset, typically the result of small populations. For example, Yukon and the Northwest Territories were removed from the Canadian dataset.

Figure 2 shows examples of observed data for four countries, Brazil, Canada, Japan and the USA. Several features of the data are apparent. First, the series for the different regions in the same country tend to move in parallel, since life expectancy has been broadly increasing in all regions over time in each country. Second, the series for the different regions show a trend towards convergence. Third, after life expectancy has reached a certain point, the regional series tend to fluctuate and not converge any further. The convergence is particularly apparent for Brazil and Canada, where the series started at a lower level of life expectancy than for Japan and the USA.

This observation is reinforced by Figure 3, which shows the evolution over time of the within-country between-region standard deviation for each of the countries in our dataset. It can be seen clearly that for the countries with long time series, the variation between regions

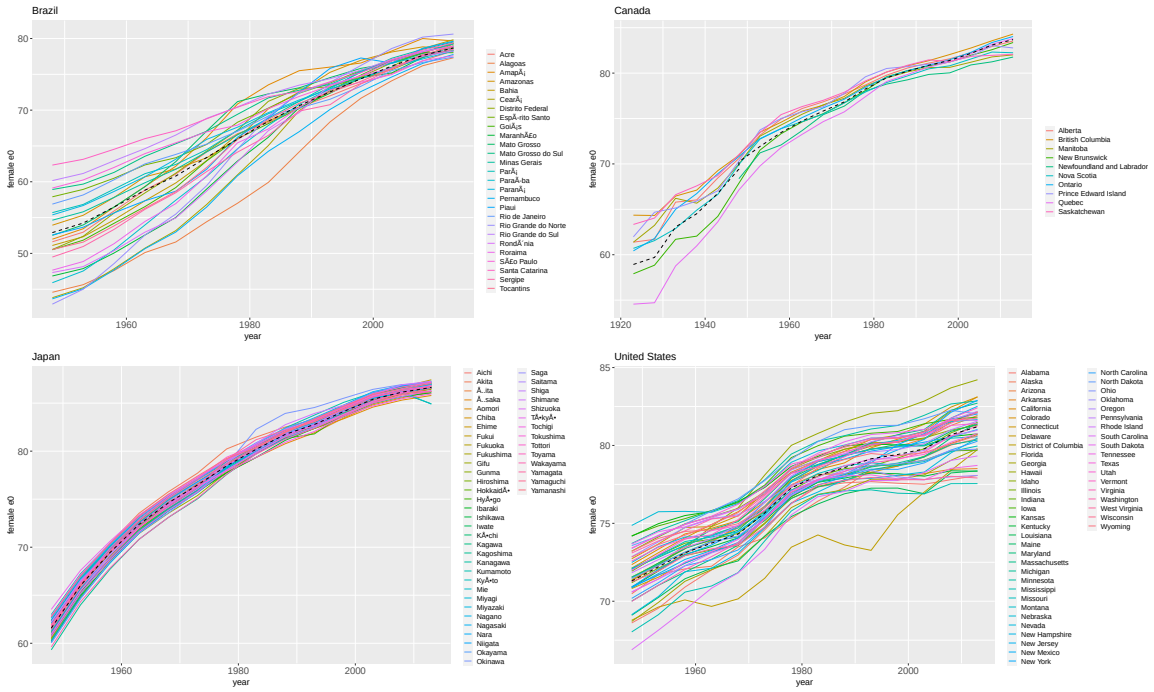

Figure 2: Observed time series for Brazil, Canada, Japan and the US. Each line represents one region. The dashed black lines represent the national estimates.

started high and declined steadily over the years, levelling off at a low level, in most cases around the 1990s, and then fluctuating.

3 Methods

3.1 Methods Overview

For generating projections of subnational e_0, we seek a method that is probabilistic, is based on the established national methodology used by the UN (Raftery et al. 2013), works well for most countries, is simple, and yields correlations between regions that are similar to the correlations in the observed data.

We introduce three methods that satisfy at least most of these requirements, and we compare them to two other baseline methods. Two of the new methods are adaptations of simple deterministic methods to a probabilistic framework. The third method is an extension to take into account correlation between regions.

The starting point for all three of the new methods is projections at the national level. First, a Bayesian hierarchical model is used to produce a set of trajectories of future female e_0 from its posterior predictive distribution (Raftery et al. 2013). Then a female-male gap model is used to generate trajectories of male e_0 (Raftery et al. 2014). To project at the subnational level, we apply the methods described below to each national trajectory, yielding probabilistic projections for each region.

In the following, $e_{c,t}^{(C)}$ denotes the life expectancy of country c at time t, and $e_{r,c,t}^{(R)}$ denotes the life expectancy of region r of country c at time t. In all three methods we model $e_{r,c,t}^{(R)}$ as

$$e_{r,c,t}^{(R)} \quad = \quad f(e_{c,t}^{(C)}, \alpha_{r,c}) \,, \tag{1}$$

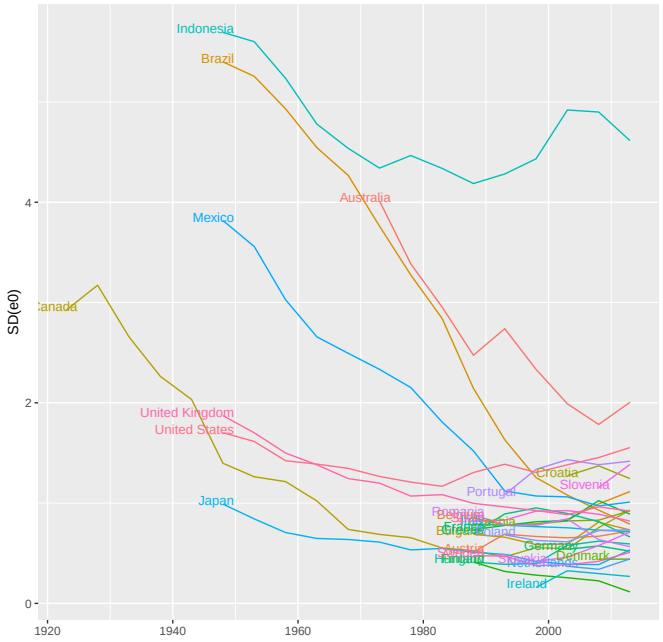

Figure 3: Within-country between-region standard deviation over time for the countries in our analysis.

namely that the regional life expectancy is a function of the national life expectancy and a regional term $\alpha_{r,c}$. The difference between the three methods lies in the definition of the function $f(\cdot,\cdot)$ and how the regional term $\alpha_{r,c}$ is modeled.

Scale method: The Scale method follows Wilson (2018) and defines $\alpha_{r,c}$ as a time-invariant factor, i.e.

$$e^{(R)}_{r,c,t} \quad = \quad \alpha_{r,c} e^{(C)}_{c,t} \,, \tag{2}$$

where $\alpha_{r,c} = e^{(R)}_{r,c,t_0} / e^{(C)}_{c,t_0}$ is derived from the last observed time period t_0.

The method is implemented by taking the i-th trajectory simulated from the posterior predictive distribution of future national life expectancy and multiplying it by the scale factor $\alpha_{r,c}$. This is done for each trajectory, yielding a set of trajectories of future regional life expectancy, representing its posterior predictive distribution. Thus if $e^{(C)}_{c,t,i}$ is the simulated value of future national life expectancy from the i trajectory, then the corresponding simulated value of regional life expectancy is $e^{(R)}_{r,c,t,i} = \alpha_{r,c} e^{(C)}_{c,t,i}$. Note that in this method, the scale factor $\alpha_{r,c}$ is constant across trajectories.

This method is simple and probabilistic. However, it has the drawback of yielding perfectly parallel trajectories with no crossovers between regions, whereas in fact such crossovers do occur.

Constant method: Here, the regional term is defined as a time invariant additive constant, i.e.

$$e^{(R)}_{r,c,t} \quad = \quad e^{(C)}_{c,t} + \alpha_{r,c} \,, \tag{3}$$

where $\alpha_{r,c} = e^{(R)}_{r,c,t_0} - e^{(C)}_{c,t_0}$ with t_0 being the last observed time period. It is implemented similarly to the Scale method. Like the Scale method, this method is simple and probabilistic, but yields parallel trajectories, not allowing for the possibility of crossovers.

AR(1) scale method: This method tries to rectify the drawback of the above methods that they produce perfectly parallel trajectories and thus do not account for the possibility of crossovers between regions. To do this, we leverage a perhaps surprising conclusion from Ševčíková et al. (2018) that subnational TFR can be well modeled using a simple scale method with a scale factor that is stochastic and follows a first-order autoregressive, or AR(1) model.

We propose a method for projecting subnational e_0 in a similar way, extending the Constant method. We take the regional term $\alpha_{r,c,t}$ to be time dependent and additive, i.e.

$$e^{(R)}_{r,c,t} \;=\; e^{(C)}_{c,t} + \alpha_{r,c,t}\,. \tag{4}$$

It is modeled using a first-order autoregressive, or AR(1), process:

$$\alpha_{r,c,t} \;=\; \rho\alpha_{r,c,t-1} + \varepsilon_{r,c,t}, \quad \text{with } \varepsilon_{r,c,t} \overset{\text{ind}}{\sim} N(0, \sigma^2_{c,t}), \tag{5}$$

where ρ is the autoregressive parameter, which is constant across all countries and regions and constrained to lie between 0 and 1, and $\sigma^2_{c,t}$ is the residual variance, which is allowed to vary between countries.

This is implemented similarly to the Constant method, with the addition of simulating the regional term, which is no longer taken to be constant in time but instead is allowed to vary randomly across time and between trajectories. Specifically, we simulate the i-th trajectory of the future values of the stochastic regional term, denoted by $\alpha_{r,c,t,i}$, as follows:

$$\begin{aligned} \alpha_{r,c,t_0,i} &= e^{(R)}_{r,c,t_0} - e^{(C)}_{c,t_0} \\ \alpha_{r,c,t,i} &= \rho\alpha_{r,c,t-1,i} + \varepsilon_{r,c,t,i} \text{ for } t > t_0, \end{aligned} \tag{6}$$

where the $\varepsilon_{r,c,t,i}$ are simulated as independently distributed random variables from a $N(0, \sigma^2_{c,t})$ distribution. Then the i-th trajectory from the posterior predictive distribution of regional life expectancy is $e^{(R)}_{r,c,t,i} = e^{(C)}_{c,t,i} + \alpha_{r,c,t,i}$.

The choice of an additive function in Equation 4 rather than a scale function is motivated by the fact that if using a scaling factor, with increasing $e^{(C)}_{c,t}$ the variance $\sigma^2_{c,t}$ would tend to increase over time because life expectancy does, while an additive term keeps the variance approximately constant. Using a scaling factor works well in the case of TFR (Ševčíková et al. 2018), where the national TFR is expected to decrease rather than increase over time. For life expectancy however, exploratory analyses indicated that the variance $\sigma^2_{c,t}$ decreases roughly linearly as $e^{(C)}_{c,t}$ increases up to a certain level of $e^{(C)}_{c,t}$ after which it stays approximately constant. This led us to use an additive term, and we set $\sigma^2_{c,t}$ to be a function of $e^{(C)}_{c,t}$, namely:

$$\sigma^2_{c,t} \;=\; \begin{cases} a + b(e^{(C)}_{c,t} - U) & e^{(C)}_{c,t} < U \\ a & e^{(C)}_{c,t} \geq U, \end{cases} \tag{7}$$

where U is the value of life expectancy at which the variance switches from being a linearly increasing function of the country-specific life expectancy $e^{(C)}_{c,t}$ with slope b, to a constant, a. The parameters a, b and U are taken as constant across countries and regions.

Using the observed data, we estimate $\hat{\rho} = 0.95$, $U = 82.5$, $a = 0.0482$, and $b = -0.0154$; see Section 3.2 for details.

3.2 Estimation of AR(1) Model

The parameters in Equations 5 and 7 are estimated using the observed data. The parameter ρ in Equation 5 is estimated by linear regression through the origin of $\alpha_{r,c,t}$ on $\alpha_{r,c,t-1}$, using 3,748 available observed pairs, $(\alpha_{r,c,t}, \alpha_{r,c,t-1})$. The regression line corresponding to the estimate $\hat{\rho} = 0.95$ is shown by the red line in the left panel of Figure 4.

The parameters in Equation 7 are estimated by nonlinear least squares, using the within-country variance of the residuals $\delta_{c,t} = \alpha_{r,c,t} - \hat{\rho}\alpha_{r,c,t-1}$, as shown in the right panel of Figure 4. For this estimation we excluded a small number of extreme outliers with observed variance larger than 1.2, yielding 182 values of $\delta_{c,t}$. The estimation yields $U = 82.5$, $a = 0.0482$, and $b = -0.0154$.

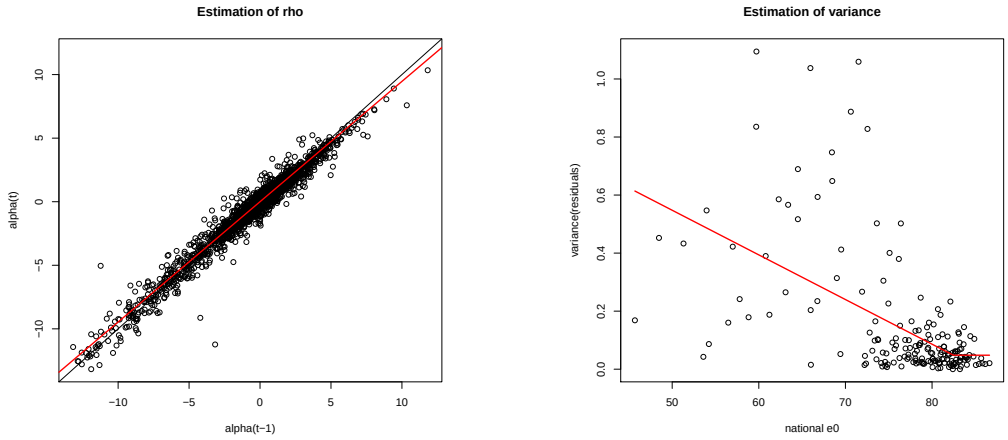

Figure 4: Left panel: Estimation of ρ from the relationship between $\alpha_{r,c,t}$ and $\alpha_{r,c,t-1}$. Each of the 3,748 dots corresponds to one region at one time period. The black line is the $x = y$ line, while the red line shows the estimate of $\hat{\rho} = 0.95$. Right panel: Estimation of U, a and b using the within-country variance of residuals $\delta_{c,t} = \alpha_{r,c,t} - \hat{\rho}\alpha_{r,c,t-1}$. Each of the 182 dots corresponds to one country at one time period. The red line corresponds to the resulting estimates of $U = 82.5$, $a = 0.0482$, and $b = -0.0154$.

4 Results

To predict female e_0 at the national level, we simulated a BHM, using the bayesLife R package (Ševčíková et al. 2019), resulting in 1,000 trajectories of life expectancy for each country, from 2020 to 2100. Then we applied the three methods above to each trajectory to obtain 1,000 future trajectories for each region in our data.

4.1 Predictive Distribution of Life Expectancy

We now present the marginal predictive distribution from the AR(1) and the Constant methods for three states in the USA, see Figure 5. Mississippi (on the left), currently has the lowest female life expectancy in the country. The distribution from the AR(1) method (red shaded area) is predicted to stay below the national distribution (grey area), with the gap between the medians slowly decreasing. The Constant method (yellow shaded area) keeps the gap between the medians constant, and thus the projection is lower than for AR(1). Virginia (in the middle), has female e_0 currently similar to the national one, and the projections

from both methods follow the national distribution with slightly higher uncertainty. Finally Hawaii, the state with the highest female e_0, is projected to stay mostly above the national distribution with a slow decrease of the gap between the two distributions for the AR(1), and a constant gap for the Constant method. These three states illustrate the types of results obtained for all regions in our data. Note that the Scale method yields very similar marginal distributions to the Constant method. Thus for clarity, we omit it in Figure 5.

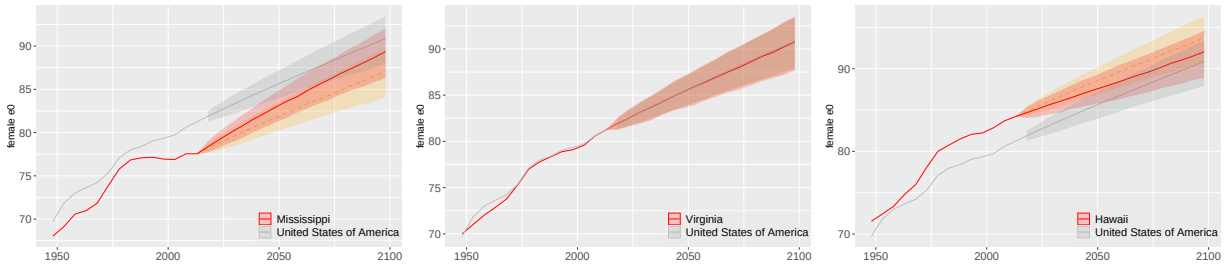

Figure 5: Predictive distribution of female e_0 for Mississippi, Virginia and Hawaii. Observed data and projected medians are shown as solid lines, while the 80% probability interval is shown as a shaded area. The subnational results from the AR(1) method are shown in red, while the national data and projections are shown in grey. Results from the Constant method are shown in light yellow with its median as dashed line.

4.2 Comparative methods

The remainder of this section compares results from the three methods described in Section 3, to two additional methods:

National: The regional projections are set to the national projections, so that $e^{(R)}_{r,c,t,i} = e^{(C)}_{c,t,i}$ for all trajectories i.

Persistence: The regional projections stay the same over time, i.e. $e^{(R)}_{r,c,t} = e^{(R)}_{r,c,t_0}$, yielding deterministic projections.

As in the case of the Scale and Constant methods, the National method is applied to each trajectory and thus yields probabilistic projections. However, the resulting trajectories are all parallel to one another within a region and within a country. Thus, all three methods, Scale, Constant and National, give probability zero of crossovers in life expectancy between regions in the same country, which is unsatisfactory because such crossovers are in fact observed in practice.

4.3 Validation

To assess the marginal predictions of e_0 made by our methods, we performed an out-of-sample validation exercise for the predicted $e^{(R)}$. We withheld data for two to five time periods, corresponding to 10 to 25 years. We did this for both the national model and the subnational model. We then predicted e_0 for these time periods, national and subnational, and compared the subnational prediction to the observed values. Table 2 includes validation measures such as the bias, the mean absolute error (MAE), the continuous ranked probability score (CRPS) (Gneiting and Raftery 2007), and the coverage of the 80% probability interval (PI80). For the bias and the MAE, the smaller the better, while for the coverage of the interval, the closer to 80% the better. The CRPS provides an overall assessment of the

Table 2: Out of sample validation of predicted $e^{(R)}_{r,c,t}$ for the proposed methods and other comparative methods. The "Nout" column shows how many time periods were withheld, "Nc" is the number of countries included, while "N" is the total number of data points included. The "MAE" column contains the mean absolute error, "CRPS" shows the continuous ranked probability score, and "PI80" shows the coverage of the 80% probability interval.

Nout	Method	Ncountries	N	BIAS	MAE	CRPS	PI80
2	persistence	27	874	1.457	1.463	−1.463	–
2	national	27	874	0.306	1.186	−1.473	56.2
2	constant	27	874	0.266	0.533	−0.947	89.2
2	scale	27	874	0.266	0.536	−0.950	88.8
2	AR(1)	27	874	0.265	0.534	−0.940	91.8
3	persistence	24	1152	2.011	2.027	−2.027	–
3	national	24	1152	0.496	1.485	−1.668	53.3
3	constant	24	1152	0.377	0.728	−1.020	88.0
3	scale	24	1152	0.377	0.726	−1.020	88.2
3	AR(1)	24	1152	0.383	0.749	−1.018	89.8
4	persistence	22	1504	2.490	2.514	−2.514	–
4	national	22	1504	0.633	1.660	−1.757	52.2
4	constant	22	1504	0.363	0.909	−1.058	86.4
4	scale	22	1504	0.358	0.909	−1.055	87.0
4	AR(1)	22	1504	0.391	0.936	−1.063	88.4
5	persistence	19	1680	2.912	2.928	−2.928	–
5	national	19	1680	0.640	1.748	−1.785	60.2
5	constant	19	1680	0.158	0.923	−0.985	92.1
5	scale	19	1680	0.146	0.927	−0.983	91.4
5	AR(1)	19	1680	0.221	0.948	−1.003	93.3

quality of a probabilistic forecast, including bias and variance. A better method corresponds to a larger value of CRPS.

It can be seen that the persistence method did not perform well. The second worst was the National method, which sets the regional projections to the national values. This underperformed in terms of both MAE and bias, as well as greatly underestimating the uncertainty. The Scale, Constant and AR(1) methods performed similarly, substantially outperforming the Persistence and National methods. The Scale and Constant methods performed slightly better than the AR(1) method. All three of the latter methods (Scale, Constant and AR(1)) slightly overpredicted uncertainty while yielding reasonable MAEs. As we have noted, the Constant and Scale methods have the undesirable property of excluding the possibility of crossovers, which the AR(1) method does not have.

4.4 Crossover Patterns

We now investigate whether the AR(1) method yields similar patterns of crossovers between regions in the same country as have been observed historically. This also gives an indication of whether the methods are representing the between-region within-country correlations well

enough. To illustrate this, we randomly selected one trajectory out of the 1,000, and viewed that trajectory for all regions of a country. Figure 6 contains such plots for the four countries from Figure 2, namely Brazil, Canada, Japan and the USA. It can be seen that the observed crossover pattern between regions (see lines to the left from the dotted vertical line) is similar to the predicted pattern (lines to the right) in each country.

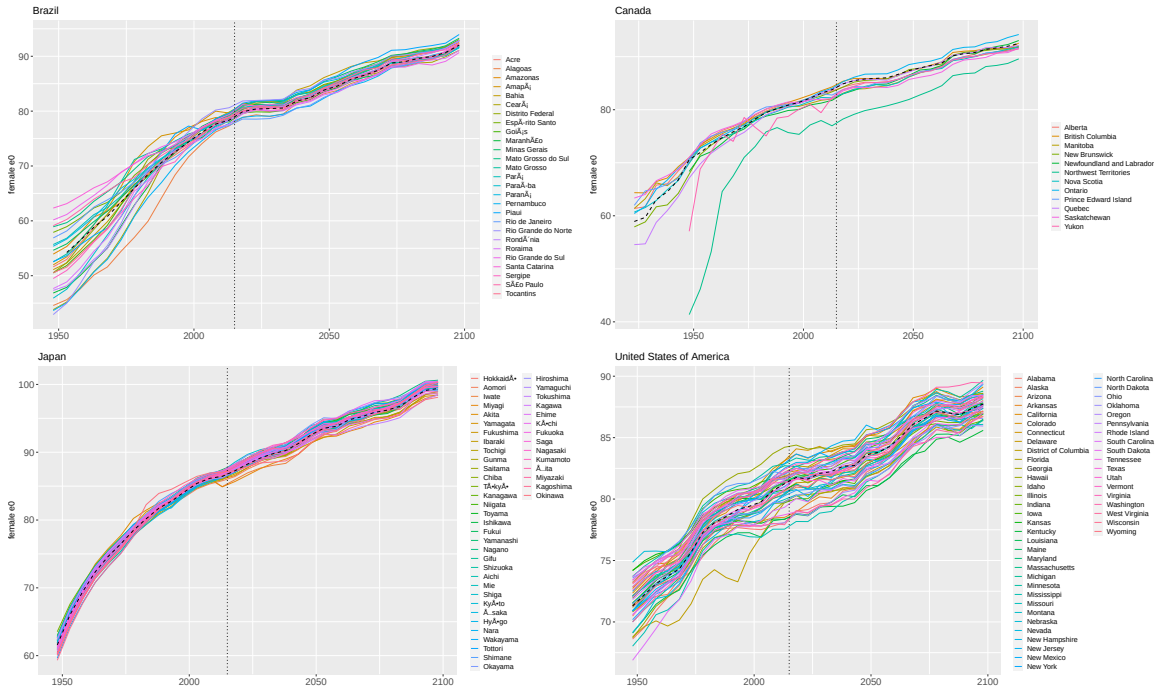

Figure 6: One-trajectory plot for Brazil, Canada, Japan and USA. Each line represents one region. Values to the left of the dotted vertical line are observed, while values to the right are results of one trajectory. National values from WPP 2017 are shown as a black dashed line.

Indeed, when quantifying the frequency of crossovers over 30 years, we found that crossovers between regions of the same country in the observed data happened in 23% of all possible combinations, while the predicted trajectories cross in 19% of the between-regions cases. Given that all the other probabilistic methods described above produce parallel trajectories and thus, yield 0% crossovers, we conclude that in terms of predicted crossovers between regions the AR(1) method performs best. This suggests also that the Scale and Constant methods are not representing between-region correlations as well as the AR(1) method.

4.5 Within-country Variation

Here the distribution of the predictive within-country between-region standard deviation, $sd(e^{(R)}_{\cdot,c,t})$, is of interest. To compare the predictive standard deviation to observed values, we withheld data from the past five time periods for both the national model and the subnational model. We then predicted e_0 for these time periods, national and subnational. Finally, $sd(e^{(R)}_{\cdot,c,t})$ was obtained for each of the 1,000 trajectories, yielding the distribution of the standard deviation for each c and t.

Figure 7 shows the distribution of the predictive standard deviation for the AR(1) method (red), the Scale method (green) and the Constant method (dashed line) for three countries,

namely Italy, United Kingdom and Austria. They represent a set of countries for which the observed standard deviation (shown as blue dots) is mostly captured by the span of the distribution of the AR(1) method. There are 19 countries for which we can do this comparison, i.e. countries with at least two regions with data starting in 1988 or earlier. Out of the 19 countries, the AR(1) model fits the predictive within-country variation well in 14. Note that the Constant method results in constant standard deviation, shown in the figure as a dashed horizontal line. The Persistence method yields the same standard deviation as the Constant method. Furthermore, since the National method results in parallel trajectories across regions, the standard deviation is zero (not shown in the figure).

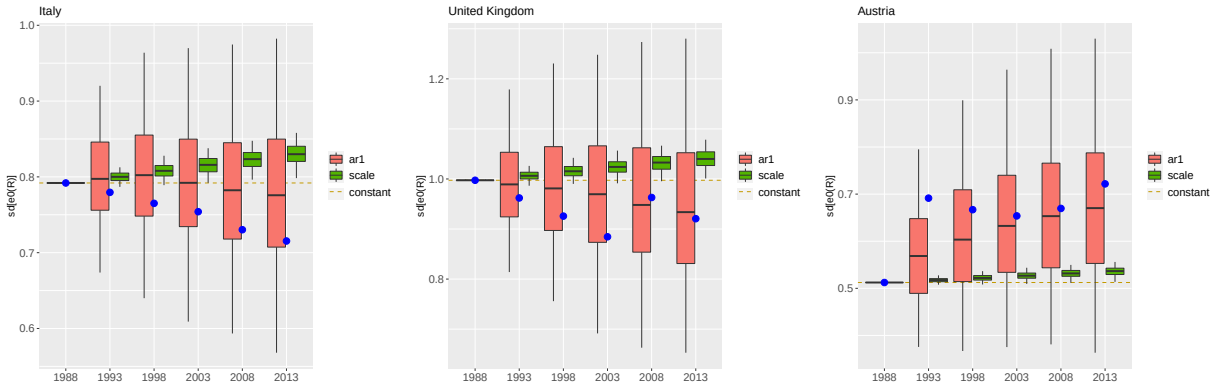

Figure 7: Predictive within-country standard deviation of $e^{(R)}_{\cdot,c,t}$ for the AR(1) method (red), the Scale method (green) and the Constant method (dashed orange line) for three of the 14 countries for which the AR(1) method works well. The thick horizontal lines are the medians, the boxes contain 50% of the distribution and the whiskers mark the 95% probability intervals. The blue dots show the observed standard deviation.

Figure 8 shows examples of two countries where the AR(1) model either underestimates the observed variance (United States on the left), or overestimates it (Japan on the right).

We identified four countries that belong to one of these two categories across multiple out-of-sample validation exercises. These are the USA and Indonesia, for which the observed standard deviation was higher than the predicted one, and Japan and Brazil, for which the observed standard deviation was lower than the predicted one[1]. For all remaining countries the within-country standard deviation is predicted well by the AR(1) model. Table 3 shows validation results similar to Table 2, but instead of validating the life expectancy directly, now our measure of interest is $sd(e^{(R)}_{\cdot,c,t})$ within each country c for each time period t.

The table reveals that the Scale, Constant and National methods greatly underpredict within-country variation, while the AR(1) method has reasonable coverage. The CRPS, which measures the overall performance of the probabilistic forecasts, is best for the AR(1) method.

Finally we note that under the AR(1) model and given the estimated parameters, the asymptotic standard deviation of $\alpha_{\cdot,c,t}$ is

$$sd_{t\to\infty}(\alpha_{\cdot,c,t}) \quad = \quad \sqrt{\frac{a}{1-\rho^2}} = 0.7. \tag{8}$$

[1]Note that the increase in Japan's variation in 2013 (Figure 8) was caused by the 2011 tsunami which resulted in a significant decrease of e_0 in two regions.

Table 3: Out of sample validation of predicted within-country between-region $sd(e^{(R)}_{\cdot,c,t})$ for the five comparative methods. Column "Nout" shows how many time periods were withheld, "Nc" is the number of countries included, "N" shows how many data points in total were included. Column "MAE" contains the mean absolute error, "CRPS" shows the continuous ranked probability score, and "PI80" shows the coverage of the 80% probability interval.

Nout	Method	Nc	N	BIAS	MAE	CRPS	PI80
2	persistence	27	54	0.016	0.109	−0.109	–
2	national	27	54	0.966	0.966	−0.966	0.0
2	constant	27	54	0.016	0.109	−0.109	0.0
2	scale	27	54	0.001	0.115	−0.109	5.6
2	AR(1)	27	54	0.025	0.111	−0.085	61.1
3	persistence	24	72	0.031	0.137	−0.137	–
3	national	24	72	0.987	0.987	−0.987	0.0
3	constant	24	72	0.031	0.137	−0.137	0.0
3	scale	24	72	0.009	0.132	−0.126	8.3
3	AR(1)	24	72	0.030	0.135	−0.112	68.1
4	persistence	22	88	−0.013	0.169	−0.169	–
4	national	22	88	1.032	1.032	−1.032	0.0
4	constant	22	88	−0.013	0.169	−0.169	0.0
4	scale	22	88	−0.045	0.165	−0.156	8.0
4	AR(1)	22	88	−0.005	0.189	−0.152	55.7
5	persistence	19	95	−0.054	0.210	−0.210	–
5	national	19	95	1.049	1.049	−1.049	0.0
5	constant	19	95	−0.054	0.210	−0.210	0.0
5	scale	19	95	−0.098	0.208	−0.198	7.4
5	AR(1)	19	95	−0.034	0.232	−0.185	60.0

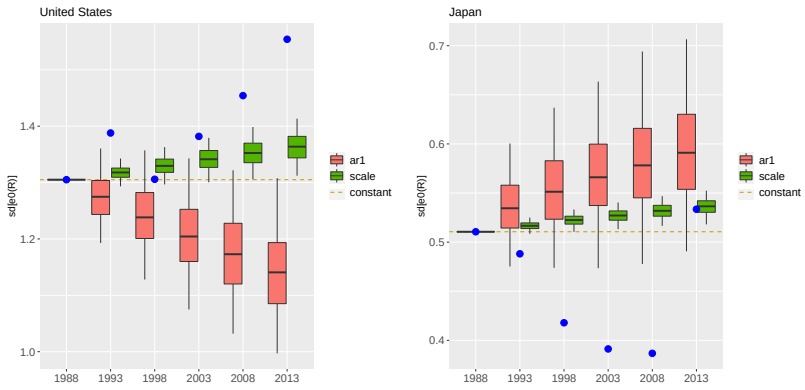

Figure 8: Predictive within-country standard deviation of $e^{(R)}_{\cdot,c,t}$ for the AR(1) method (red) and the Scale method (green) and the Constant method (dashed orange line) for two of the four countries for which the AR(1) method does not predict the standard deviation well. The thick horizontal lines are the medians, the boxes contain 50% of the distribution and the whiskers mark the 95% probability intervals. The blue dots show the observed variance.

Thus, we expect $sd(e^{(R)}_{\cdot,c,t})$ to converge to about 0.7 over time (either increasing or decreasing, depending on the current level), and stabilize thereafter. That is the reason why for example in Figure 8 we see a decreasing predictive standard deviation for the USA but an increasing one for Japan. The Scale method does not have this property and thus its standard deviation always increases, regardless of the current level.

4.6 Predictive Distribution of Subnational Mortality Rates

To project subnational age-specific mortality rates probabilistically, there are two main broad approaches one can take. One can develop a multivariate model for the age-specific rates jointly, as was done for example by Bohk-Ewald and Rau (2017). Alternatively, one can derive the distribution of the age-specific rates from the distribution of e_0.

Here, we use the method described by Ševčíková et al. (2016), which has been used by the UN in several revisions of the WPP, including the most recent 2019 revision (United Nations 2019). In that approach, each projected trajectory of future subnational e_0 values is converted into age-specific mortality rates using a modified Lee-Carter method. It is based on the basic Lee-Carter model (Lee and Carter 1992),

$$\log[m_x(t)] = a_x + b_x k(t) + \varepsilon_x(t), \quad \varepsilon_x(t) \sim N(0, \sigma^2_\varepsilon),$$

where $m_x(t)$ is the mortality rate for age x and time period t. To estimate the parameters a_x and b_x, one can use historical regional mortality rates. However in the absence of those, we will use the corresponding national mortality rates, as follows:

1. For all historical time periods we extend the national or regional mortality rates to higher ages (up to 130+) using the coherent Kannisto method as described in Ševčíková et al. (2016).
2. We use the extended data to estimate a_x and a coherent and rotated version of the b_x parameter, denoted by $B_x(t)$ (Li et al. 2013; Li and Lee 2005).
3. For each trajectory and future time period t, we find the value of $k(t)$ that matches the corresponding $e_0(t)$ using life tables (Li and Gerland 2011).

4. Finally, we project mortality rates using the equation $\log[m_x(t)] = a_x + B_x(t)k(t)$.

Applying the above steps to each projected trajectory of subnational e_0 yields a joint probabilistic distribution of age-specific mortality rates. The method is implemented in the MortCast R package (Ševčíková et al. 2020).

Figure 9 compares results from the AR(1) method for Alaska and British Columbia in two of the future time periods, namely 2020-2025 (left) and 2080-2085 (middle and right). As expected, in the first projected time period the uncertainty is quite narrow, while 60 years later the uncertainty increases for both states and most age groups. The joint distribution of mortality rates for these two states is shown in the right panel which suggests that the probability of crossovers between regions is low across all age groups.

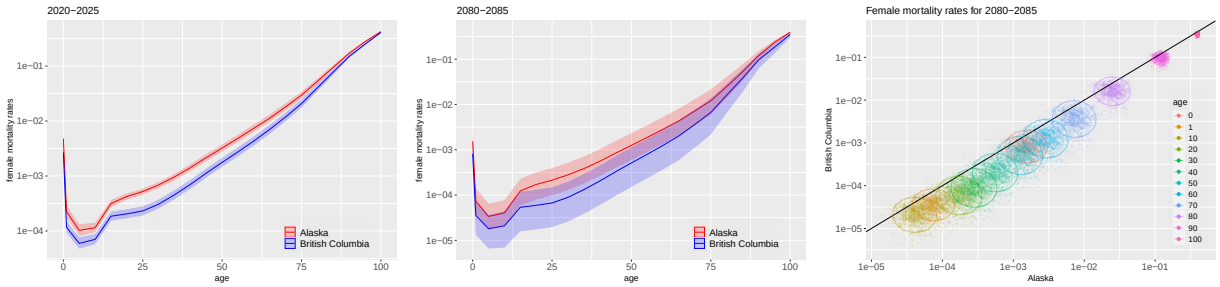

Figure 9: Predictive distribution of age specific mortality rates for Alaska and British Columbia (BC). Left and middle: Marginal distribution for Alaska (red) and BC (blue) in 2020-2025 (left) and 2080-2085 (middle). The shaded area represents the 95% probability interval with the solid lines being the medians. Right: Joint distribution between Alaska and BC in 2080-2085 for selected age groups. Each point corresponds to one trajectory. The ellipses show 95% probability intervals for each age group.

5 Discussion

We have introduced three methods for projecting subnational life expectancy that are probabilistic and build on the established Bayesian hierarchical framework for projecting national e_0 used by the United Nations. Once the national BHM has been estimated, the subnational methods are relatively simple to implement. Since trajectories from the national BHM for all countries are available for download, the additional effort required to produce subnational projections is modest. This may be an attractive feature for practitioners. All three methods can be used to generate probabilistic projections of age-specific mortality rates.

We have shown the strengths and weaknesses of the three methods. The Scale and Constant methods yield the best results in terms of the validation of the marginal predictive distribution. The biggest weakness of these two methods is that both yield perfectly parallel trajectories which is unrealistic.

The AR(1) method achieves reasonable validation of the marginal distribution, and also performs best in terms of within-country variance. It also yields crossover patterns between regions that are similar to those in the observed data, suggesting that the method reproduces the between-region correlation fairly well.

The biggest limitation of the AR(1) method is that for four countries of the countries analyzed, it either underestimated (USA, Indonesia) or overestimated (Japan, Brazil) the predictive between-region within-country variance. Note that for these countries it still gave well-calibrated probabilistic projections for the regions individually. In the case of the USA,

research has been done on the factors influencing the divergence of the mortality levels in recent years (Zajacova and Montez 2017). Oosse (2003) claimed to identify a cyclic pattern of convergence and divergence in the USA, but we did not find evidence of this in our data or analysis. To account for this explicitly would require complicating the model considerably, perhaps by incorporating stochastic volatility such as is used in modeling financial time series (Hull and White 1987).

We plan to extend the methods to work for both sexes, similarly to Raftery et al. (2014). We will also implement them in the bayesLife R package (Ševčíková et al. 2019), as we have done for the subnational projection of TFR which is now available in the bayesTFR R package (Ševčíková et al. 2011).

Adoption of the methods in practice would be part of the process that would allow agencies to issue probabilistic population forecasts for subnational regions. These could take the form, for instance, of upper and lower bounds for the future numbers of people in total and by age and sex. This could be useful for government planning and other purposes. For example, decisions about opening and closing schools depend on forecasts of the future number of children in the relevant age groups. Often what is needed is not a best guess, such as is provided by deterministic population forecasts of the kind widely used, but rather a number that will ensure that there are enough school places for all children with high probability. This would be an upper bound, such as the 90th percentile, of the forecast distribution of the future number of children. This is provided automatically by a probabilistic forecast, but not by a deterministic forecast.

This would add some complexity to current methods, but the main additional complexity is in the use of Bayesian forecasts of national life expectancy. These already exist and are now used routinely by the UN for their official population forecasts for all countries; software to produce them is freely available. Several short courses in the use of these methods have been held, and attended by demographers and statisticians working for national and regional agencies from many countries.

It has been suggested that another difficulty could be that people don't understand probabilistic forecasts. However there is considerable evidence from practice and cognitive experiments that this is not actually the case and that users do understand probabilities well enough to use them in decision-making; see Raftery (2016) for a review.

Most of the countries in our dataset are high-income countries, with some middle-income countries also. This raises the question of whether the methods could also be used in low-income countries. This requires two things: probabilistic forecasts of national life expectancy, and values of past regional life expectancy for at least one time point. Probabilistic forecasts of national life expectancy are now available for all countries, and are issued by the UN and updated on a regular basis (United Nations 2019). For most of the countries we have analyzed, the regional life expectancy values come from vital registration data, but many low-income countries lack high quality vital registration systems. However, for most countries without high quality vital registration systems, high quality surveys with national coverage have been carried out at least once, for example by the Demographic and Health Surveys (DHS) or the Multiple Indicator Cluster Surveys (MICS). These can provide regional life expectancy values for all regions for at least one time point, which is all that our methods require. Thus our methods could be extended to most low-income countries fairly easily.

References

Bengtsson, T. and F. van Poppel (2011, jul). Socioeconomic inequalities in death from past to present: An introduction. *Explorations in Economic History 48*(3), 343–356.

Bennett, J. E., G. Li, K. Foreman, N. Best, V. Kontis, C. Pearson, P. Hambly, and M. Ezzati (2015, jul). The future of life expectancy and life expectancy inequalities in England and Wales: Bayesian spatiotemporal forecasting. *The Lancet 386*(9989), 163–170.

Bohk-Ewald, C. and R. Rau (2017). Probabilistic mortality forecasting with varying age-specific survival improvements. *Genus 73*.

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

Booth, H. and L. Tickle (2008). Mortality Modelling and Forecasting: a Review of Methods. *Annals of Actuarial Science 3*(1-2), 3–43.

Cairns, A. J., D. Blake, K. Dowd, G. D. Coughlan, and M. Khalaf-Allah (2011). Bayesian Stochastic Mortality Modelling for Two Populations. *ASTIN Bulletin: The Journal of the IAA 41*(1), 29–59.

Cairns, A. J. G., D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, A. Ong, and I. Balevich (2009, jan). 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.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*, 359–378.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance 42*, 281–300.

Janssen, F. (2018). Advances in mortality forecasting : Introduction. *Genus 74*(21), 1–12.

Kibele, E. U. B., S. Klüsener, and R. D. Scholz (2015, sep). Regional Mortality Disparities in Germany: Long-Term Dynamics and Possible Determinants. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie 67*(S1), 241–270.

Lee, R. D. and L. Carter (1992). Modeling and forecasting the time series of US mortality. *Journal of the American Statistical Association 87*, 659–671.

Li, N. and P. Gerland (2011). Modifying the Lee-Carter method to project mortality changes up to 2100. Presented at the Annual Meeting of Population Association of America. `http://paa2011.princeton.edu/abstracts/110555`.

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

Li, N., R. D. Lee, and P. Gerland (2013). Extending the Lee-Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography 50*, 2037–2051.

Oeppen, J. and J. W. Vaupel (2002). Broken limits to life expectancy. *Science 296*, 1029–1031.

Oosse, M. (2003). Variations in State Mortality From 1960 to 1990. Working paper POP-WP049, `https://census.gov/content/census/en/library/working-papers/2003/demo/POP-twps0049.html`.

Raftery, A. E. (2016). Use and communication of probabilistic forecasts. *Statistical Analysis and Data Mining: The ASA Data Science Journal 9*(6), 397–410.

Raftery, A. E., J. L. Chunn, P. Gerland, and H. Ševčíková (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography 50*, 777–801.

Raftery, A. E., N. Lalic, and P. Gerland (2014). Joint probabilistic projection of female and male life expectancy. *Demographic Research 30*, 795–822.

Raftery, A. E., N. Li, H. Ševčíková, P. Gerland, and G. K. Heilig (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences 109*, 13915–13921.

Rayer, S., S. K. Smith, and J. Tayman (2009). Empirical prediction intervals for county population forecasts. *Population Research and Policy Review 28*, 773–793.

Ševčíková, H., L. Alkema, and A. E. Raftery (2011). bayesTFR: An R package for probabilistic projections of the total fertility rate. *Journal of Statistical Software 43*, 1–29.

Ševčíková, H., N. Li, and P. Gerland (2020). *MortCast: Estimation and Projection of Age-Specific Mortality Rates*. R package version 2.3-0.

Ševčíková, H., N. Li, V. Kantorová, P. Gerland, and A. E. Raftery (2016). Age-specific mortality and fertility rates for probabilistic population projections. In R. Schoen (Ed.), *Dynamic Demographic Analysis. The Springer Series on Demographic Methods and Population Analysis*, Volume 39. Springer, Cham.

Ševčíková, H., A. Raftery, and J. Chunn (2019). *bayesLife: Bayesian Projection of Life Expectancy*. R package version 4.1-0.

Ševčíková, H., A. E. Raftery, and P. Gerland (2018). Probabilistic projection of subnational total fertility rates. *Demographic Research 38*, 1843–1884.

Stoeldraijer, L., C. van Duin, L. van Wissen, and F. Janssen (2013, aug). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands. *Demographic Research 29*, 323–354.

United Nations (2015). *World Population Prospects: The 2015 Revision, Probabilistic Population Projections*. New York, NY: Population Division, Dept. of Economic and Social Affairs, United Nations.

United Nations (2017). *World Population Prospects: The 2017 Revision.* New York, NY: Population Division, Dept. of Economic and Social Affairs, United Nations.

United Nations (2019). *World Population Prospects: The 2019 Revision.* New York, NY: Population Division, Dept. of Economic and Social Affairs, United Nations. `http://esa.un.org/unpd/wpp`.

Wilson, T. (2018). Evaluation of simple methods for regional mortality forecasts. *Genus 74*(14), 1–22.

Zajacova, A. and J. K. Montez (2017). Macro-level perspective to reverse recent mortality increases. *Lancet 389*, 991–992.

Appendix

Data Sources and Preprocessing

In our study we used five-year data from 29 countries and 447 regions, totaling 4,197 data points. Table 4 shows the breakdown by countries. The time range column gives the maximum span of the time series. In cases when annual data were available, we used averages over the five-year time periods.

Table 4: Sources of the data used in the study.

Country	Geographic units	Units	Obs.	Time range	Subnational data source
Australia	States,Territ.	8	72	1970-2015	Australian Bureau of Statistics
Austria	States	9	54	1985-2015	EuroStat
Belgium	Provinces	11	66	1985-2015	EuroStat
Brazil	States	27	378	1945-2015	Global Burden of Disease Study (2017), IHME
Bulgaria	Regions	6	36	1985-2015	EuroStat
Canada	Provinces	10	185	1920-2015	Statistics Canada
Croatia	Regions	2	6	2000-2015	EuroStat
Czechia	Regions	8	40	1990-2015	EuroStat
Denmark	Regions	5	10	2005-2015	EuroStat
Finland	Frm.Provinces	4	24	1985-2015	EuroStat
France	Regions,Territ.	26	144	1985-2015	EuroStat
Germany	Regions	38	110	2000-2015	EuroStat
Greece	Regions	13	69	1985-2015	EuroStat
Hungary	Regions	7	42	1985-2015	EuroStat
Indonesia	Provinces	34	476	1945-2015	Global Burden of Disease Study (2017), IHME
Ireland	Regions	2	8	1995-2015	EuroStat
Italy	Regions	21	120	1985-2015	EuroStat
Japan	Prefectures	47	658	1945-2015	Global Burden of Disease Study (2017), IHME
Mexico	States	32	448	1945-2015	Global Burden of Disease Study (2017), IHME
Netherlands	Provinces	12	36	2000-2015	EuroStat
Poland	Provinces	16	80	1990-2015	Eurostat
Portugal	Regions	7	35	1990-2015	EuroStat
Romania	Regions	8	48	1985-2015	EuroStat
Slovakia	Macroregions	4	16	1995-2015	EuroStat
Slovenia	Macroregions	2	4	2005-2015	EuroStat
Spain	Auton.Communities	17	102	1985-2015	EuroStat
Sweden	National Areas	8	48	1985-2015	EuroStat
United Kingdom	Regions	12	168	1945-2015	Global Burden of Disease Study (2017), IHME
United States	States	51	714	1945-2015	Global Burden of Disease Study (2017), IHME

To prepare such or similar datasets for our analysis or for subnational projections, one can proceed as follows:

1. Create a dataset with four columns, one each for country, region, time period and regional $e^{(R)}_{r,c,t}$.
2. Add a column with the corresponding national $e^{(C)}_{c,t}$.
3. Add a column with the regional term $\alpha_{r,c,t}$ defined either as $e^{(R)}_{r,c,t} - e^{(C)}_{c,t}$ in case of the AR(1) and Constant method, or $e^{(R)}_{r,c,t}/e^{(C)}_{c,t}$ in case of the Scale method.
4. For the estimation described in Section 3.2, add a column of the lagged value of the regional term, $\alpha_{r,c,t-1}$.