Projecting maximum country life expectancy using Provincial data only

Anthony Medford
Biodemography Unit, Department of Public Health
University of Southern Denmark

March 19, 2018

1 Aims

In this paper we introduce a new approach to modelling and projecting life expectancy for a region using only information from subregions within this larger region by applying principles from the statistical theory of Extreme Values.

The most popular mortality forecasting models, the Lee Carter Model (Lee and Carter, 1992) and its numerous extensions and variants e.g Li et al. (2004); Renshaw and Haberman (2003); Hyndman et al. (2007) fit trends to age-standardized (log) death rates. However, there is a strong argument for using life expectancy in forecasting. White (2002) found that linear trends in life expectancy give a better empirical fit to the experience of individual countries than linear trends in age-standardized (log) death rates in his study of 21 developed countries. The large and increasing family of Lee Carter type models have exhibited variable performance with respect to forecasting ability. In addition the time parameter (κ_t) in these models can be unstable (some models have multiple time parameters) over time, making forecasting even more difficult.

Among those who have forecast life expectancy are Alho and Spencer (2005); Andreev and Vaupel (2006); Lee (2006); Torri and Vaupel (2012). Extreme value theory has previously applied to the study of maximum life expectancies over a selection of developed countries by Medford (2015) (forthcoming) and this paper builds upon ideas presented therein.

2 Methods

2.1 Basic Extreme Value Theory and the GEV

The application of the statistical theory of extreme values facilitates the study of a random processes at very high or low levels. The limiting distributions of these extremes give rise to the extreme value distributions. In this paper, the parameterisations and notation of Coles (2001) will be used.

Formally, suppose that $X_1, X_2, \ldots, X_n$ is a sequence of independent random variates all having a common distribution function $F(x)$. Let the maximum of this sequence of n variables be M_n. We would like to find the distribution of M_n as n becomes large. Now,

$$\begin{aligned} P(M_n \leq z) &= P(X_1 \leq z, X_2 \leq z, \ldots, X_n \leq z) \\ &= P(X_1 \leq z)P(X_2 \leq z)\ldots P(X_n \leq z) \\ &= F^n(z) \end{aligned}$$

This result however is not particularly useful as the distribution of $F(x)$ is unknown. However, it is possible to find the distribution of M_n, say G, without any reference to F.

The distribution of M_n is degenerate since as n tends to infinity, the distribution function F converges with certainty to a single point. To avoid the difficulty of the degenerate limit a linear rescaling of M_n is applied - a result known as the Extremal Types Theorem (Fisher and Tippett, 1928; Gnedenko, 1943; Coles, 2001).

If there exists sequences of constants $\{a_n > 0\}$ and $\{b_n\}$, such that as $n \to \infty$,

$$P\Big(\frac{M_n - b_n}{a_n} \leq z\Big) \to G(z) \tag{1}$$

where $G(z)$ is a non-degenerate distribution function, then G must be a member of the Generalized Extreme Value (GEV) family of distributions. This is a remarkable result because regardless of the underlying distribution, the distribution of the maxima (or minima) converges to one of the Generalized Extreme Value family of distributions.

The GEV distribution function is given by,

$$G(z) = \exp\Big\{ - \Big[1 + \xi(\frac{z-\mu}{\sigma})\Big]^{\frac{-1}{\xi}}\Big\}, \tag{2}$$

defined on $\{z : 1+\xi(z-\mu)/\sigma > 0\}$. The model is described by three parameters: $\mu(-\infty < \mu < \infty)$, $\sigma(\sigma > 0)$ and $\xi(-\infty < \xi < \infty)$ referred to as the location, scale and shape parameters respectively. The location parameter indicates the center of the distribution; the scale parameter the size of deviations around the location parameter; and the shape parameter governs the tail behavior of the GEV distribution.

The shape parameter, ξ determines the heaviness of the right tail and this leads to three types of distributions. When $\xi < 0$, the distribution has a bounded upper finite end point and is short-tailed leading to the Weibull Distribution. When $\xi > 0$, there is polynomial tail decay leading to heavy tails and the GEV is of the Fréchet type. The case where $\xi = 0$ is taken to be the limit of Eq. 2 as $\xi \to \infty$ and there is exponential tail decay leading to light tails and the GEV is of the Gumbel type with distribution function,

$$G(z) = \exp\Big\{-\exp\Big[- (\frac{\mathrm{z}-\mu}{\sigma})\Big]\Big\}.$$

In practice, for sufficiently large n, $G(z)$ can be calculated without the need to know the normalising constants $\{a_n > 0\}$ and $\{b_n\}$ (Coles, 2001). This has motivated an approach to GEV modelling known as the Block Maxima approach, where for large enough n, $P(M_n < z)$ can be approximated by using an appropriate member of the GEV family.

In summary, the Block Maxima approach works as follows. Suppose we have independent observations $X_1, X_2, \ldots$. Let these observations be divided into blocks of length n for sufficiently large n. Then, take the maximum of each of these blocks to obtain a series of block maxima and fit a GEV distribution to these maxima in order to obtain parameter estimates $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$. Specific to our analysis, we have observed life expectancies from various countries over periods of length one year. The largest from among these annual period life expectancies is extracted and a GEV distribution is fitted to these annual maxima in order to obtain parameter estimates.

For inference, estimates of extreme quantiles of the maxima are obtained by solving for z in equation 2:

$$z_p = \mu - \frac{\sigma}{\xi}\Big[1 - \{-log(1-p)\}^{-\xi}\Big], \tag{3}$$

where the distribution function of the GEV, $G(z_p) = 1 - p$ and p is the tail probability or the probability of realising a value at least as large as z_p. In extreme value terminology, the quantiles of the distribution, z_p are termed return levels and are associated with the so called return period $1/p$. If we are considering annual maxima, which is usually the case, then on average the quantile z_p is expected to be exceeded with probability p or on average once every $1/p$ years (Coles, 2001). For example, if $p = 0.01$ then the return level, z_p is the 99th percentile, corresponds to the $1/(1-0.01) = 100$-year return period, and is the amount which one expects to see once every 100 years, on average.

The second way to model extremes is to consider exceedances over high thresholds or put differently, all data points over some suitably chosen high value - not just the maxima. The Peak Over Thresholds model e.g. Balkema and De Haan (1974); Pickands (1975); Davison and Smith (1990), adopts this approach. Both the Peak Over Thresholds model and the Largest Order Statistic Model are special cases of the Point Process representation (Coles, 2001) of extreme values, under which we approximate the exceedances over a threshold by a two-dimensional Poisson process (e.g. Leadbetter et al. (1983)).

2.2 Application to Canada

From the CHMD life expectancy data is available for the various Canadian provinces over the period 1921- 2011. In line with the theoretical outline in Subection 2.1, we take the maximum life expectancy from among the various provinces over each of the years for which the data is available.

Figure 1 shows the province which has the highest period life expectancy in Canada by male and female, at birth and age 65 and covering the period 1921 to 2011. Over much of that period one can observe very strong linear trends over time. Oeppen and Vaupel (2002) and Vallin and Meslé (2009) studied the best practice life expectancy, albeit at a global level rather than within any particular country and observed strong linear trends similar to what is observed here.

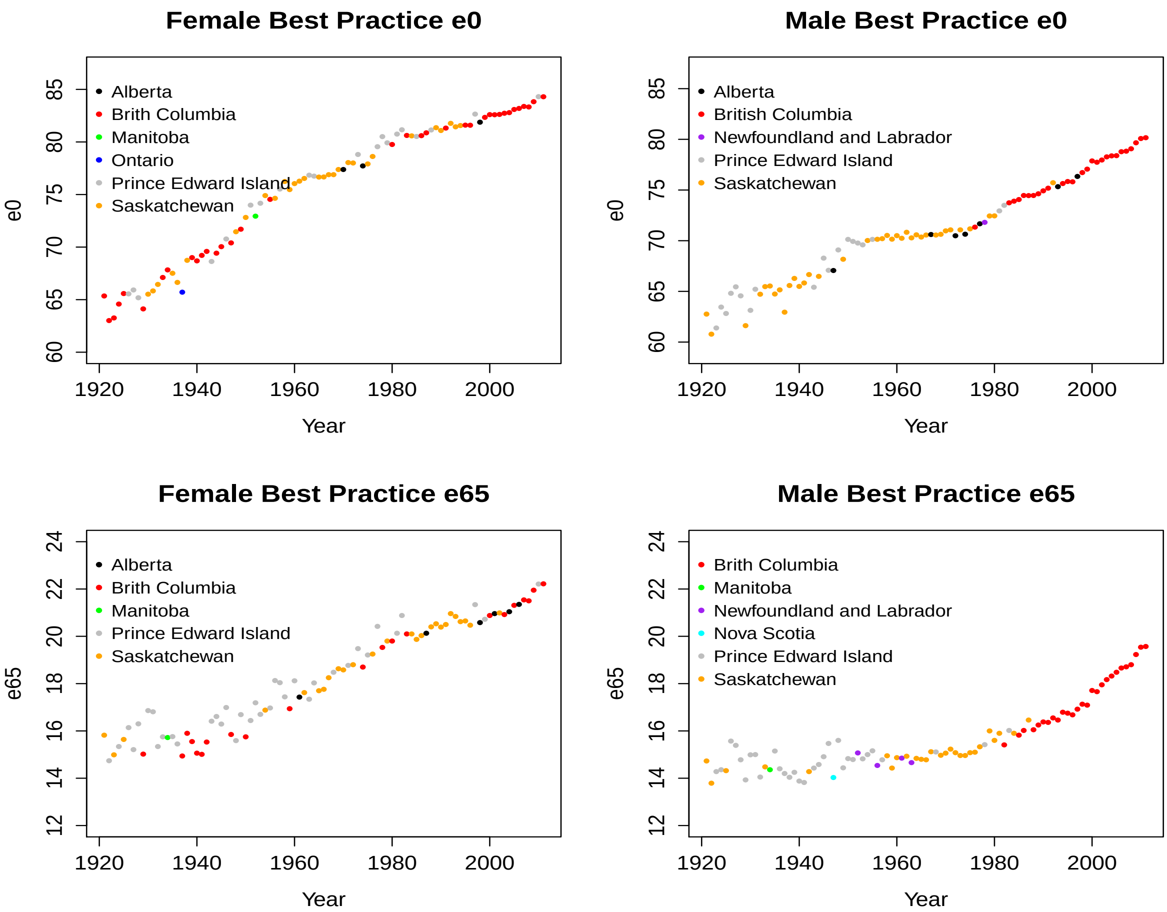

Figure 1: Provinces with the highest life expectancies at birth and age 65, males and females separately, from 1921 - 2011.

Let us define e_x^* to be Best Practice Life Expectancy at given age x where " $*$ " indicates maximum. Further let $e_{x,f}^*$ be Best Practice Life Expectancy at given age x for females and $e_{x,m}^*$ for males.

It is evident that the speed of increase in e_x^* has not been constant. For example, the rate of increase in $e_{0,f}^*$ has been slowing while $e_{65,m}^*$ has been accelerating. Therefore, rather than assume a constant linear increase, we formally investigate the presence of differential rates of increase. This is done by testing the null hypothesis of a non-zero difference in slope parameter of a segmented relationship using the Davies Test (Davies, 2002) and then finding the break points and allowing the slope parameter of a fitted linear regression to vary between these break points. The GEV model is fitted to the data beginning at the most recent breakpoint, thus ensuring that the correct speed of life expectancy increase is captured as accurately as possible. These segmented relationships are presented in Figure 2.

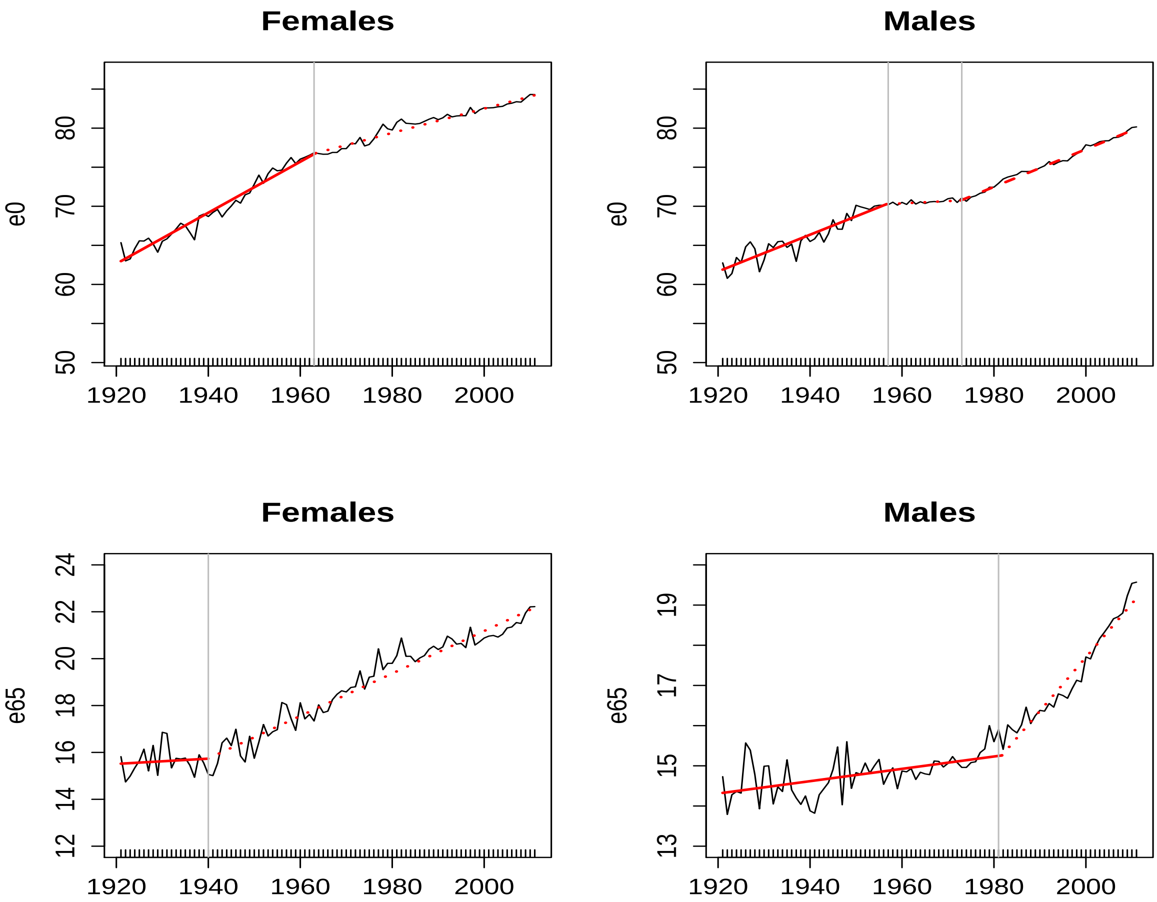

Figure 2: Breakpoints in the trend of the highest provincial life expectancies at birth and age 65, males and females separately, from 1921 - 2011

The linear trend over time in e_x^* $(x = 0, 65)$ is accounted for by allowing the location parameter of the fitted GEV model to vary linearly with time such that, instead of a fixed location parameter μ, a more flexible parameter is adopted. Time is introduced as a covariate into the parametrisation of the GEV distribution by assuming a location parameter of the form, $\mu_t = \beta_0 + \beta_1 t$ where t represents calendar time. More specifically, t is an index commencing at 1 in the first year of the data (e.g. 1964 for $e_{0,f}^*$) and increases by one unit per consecutive year.

In summary, we fit a time-dependent GEV model which can be represented succinctly as:

$$GEV(\mu_t, \sigma, \xi)$$

where the location parameter, $\mu_t = \beta_0 + \beta_1 t$, the scale parameter is σ, the shape parameter is ξ and $t = 1 \ldots t_{max}$, where t_{max} is the last year of the data.

The parameters of the GEV distribution $\mu_t, \beta_0, \beta_1, \sigma$ and ξ are found using maximum likelihood estimation. Once calculated, the parameter estimates $\hat{\mu_t}, \hat{\beta_0}, \hat{\beta_1}, \hat{\sigma}$ and $\hat{\xi}$ can then used in the computation of return levels (quantiles), probabilities and other items of interest.The estimated parameters are presented in Table 1.

	Neg. Likelihood	$\hat{\beta_0}$	$\hat{\beta_1}$	$\hat{\sigma}$	$\hat{\xi}$
Female e_0	30.8	76.4(0.11)	0.16 (0.003)	0.37 (0.030)	0.10
Male e_0	3.9	70.6 (0.10)	0.24 (0.004)	0.27 (0.03)	-0.34 (0.15)
Female e_{65}	45.2	15.4 (0.11)	0.09 (0.002)	0.42 (0.040)	-0.15 (0.08)
Male e_{65}	-3.60	15.1 (0.11)	0.13 (0.006)	0.27 (0.04)	-0.29 (0.14)

Table 1: Maximized negative log-likelihoods, parameter estimates and standard errors (in parentheses) of the Block Maxima Model; e_0 and e_{65} for males and females shown separately

We can take advantage of the time dependence by projecting forward values from the fitted model. This is done simply by updating the time-varying location parameter of the model μ_t for future values of time, t. In order to make projections, we first fit the time varying GEV model and then calculate the desired quantiles of maximum life expectancy and finally project forward in time. Of course, the other parameters could be allowed to also vary with time but here most of the time dependence is driven by the location parameter.

This simple projection methodology is enabled by the linear evolution of life expectancy commented upon by e.g. Oeppen and Vaupel (2002); Vallin and Meslé (2009). Provided that the observed linear trends in life expectancy in Canada continue, then this method presents an effective way to project life expectancy. Figure 3 presents forecasts for female life expectancy at birth using this methodology. The fitted model predicts a median life expectancy in 2035 of 88.2 years and a predicted 95% confidence interval of (87.9 years - 88.5 years).

Besides quantiles, another useful way of making inferences for GEV models is in the calculation of probabilities (see Table 2). This is a simple consequence of having fit a full parametric probability distribution to the data and can be easily accomplished by using Equation 2 and the estimated parameters. Hence, given quantiles one can calculate probabilities and vice versa.

For example the probability that the maximum female life expectancy at birth for some province in Canada will exceed 89 years in 2035 is approximately 11% whereas the chance that in 2030 some province will have a life expectancy at birth greater than 87.5 years is approximately 44%.

Year	$P(e^*_{0,f} > 87.5)$	$P(e^*_{0,f} > 89)$
2030	0.44	0.02
2035	0.99	0.11

Table 2: Probability of the maximum female life expectancies at birth, $e^*_{0,f}$, exceeds certain levels for the years 2030 and 2035.

3 Conclusion and Ongoing Work

We present a model which takes advantage of the past linear trends in life expectancy to make predictions about future medium-term life expectancy trajectories. The model is rooted in the statistical theory of Extreme Values, in particular the method of Block maxima. As mentioned, this methodology is based on asymptotic results and in pratice, sample sizes are limited. The fit of the model can be assessed using various tools.

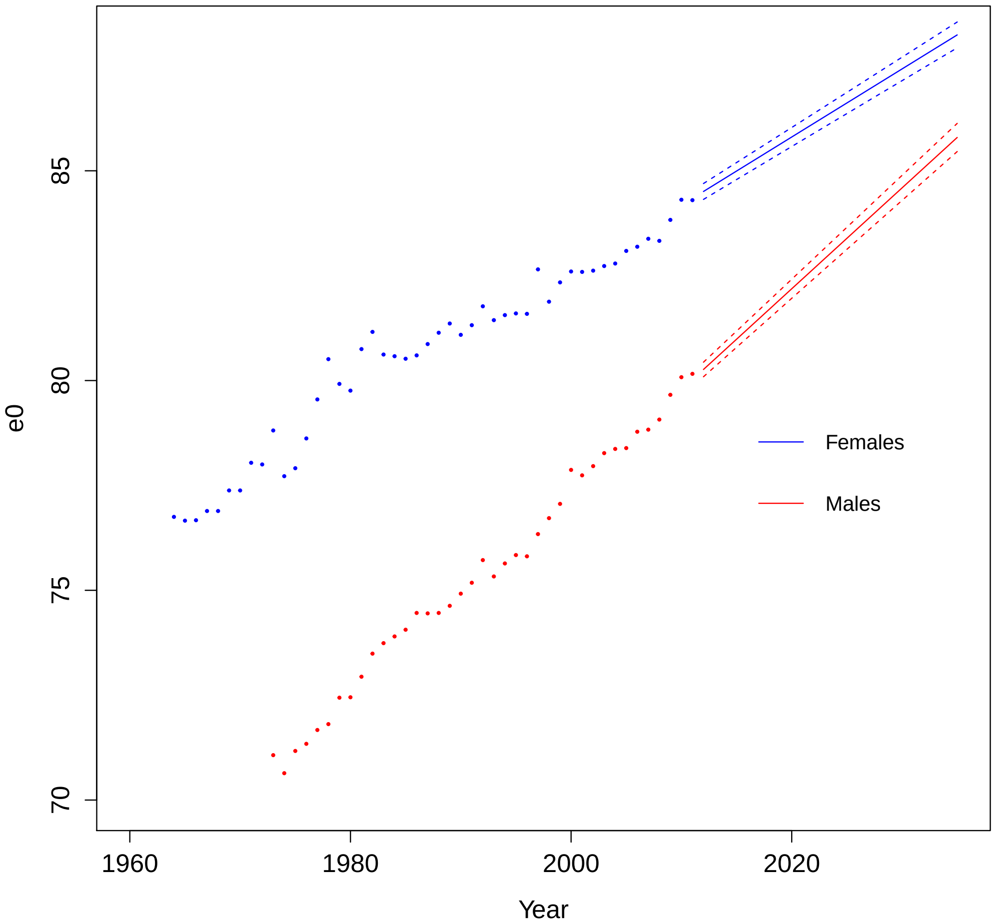

Figure 3: Forecasts and 95% confidence intervals for Canadian female and male life expectancy at birth.

One advantage of this model is that we are able to make probabilistic statements, based on a reasonable set of assumptions about the future evolution of Canadian provincial life expectancy. Another is that we are able to make statements about the evolution of maximum life expectancy for Canada by using only life expectancy at the provincial level. This method could be applied in other situations where regional-only data is available but we would like to have an idea about (maximum)life expectancy at a supra-regional level. An ancillary benefit is that if the maxima are dominated by regions which are also heavily weighted in terms of population size then the method can give a decent estimate of future median life expectancy. A further feature of the method is that it allows us to make probability statements about the maximum future life expectancy.

Further work remains to be done in assessing forecast accuracy of the model against others models and in refining the methodology surrounding forecast error bounds. The model is also being fit to Japanese data using the Japanese Mortality Database. Finally, a more sophisticated approach which allows the GEV parameters to vary stochastically in the spirit of Huerta and Sansó (2007) is being investigated to project the time varying parameter through the use of Dynamic Linear Models.

References

Alho, J. and B. D. Spencer (2005). *Statistical demography and forecasting.* Springer.

Andreev, K. F. and J. W. Vaupel (2006). Forecasts of cohort mortality after age 50. *Max Planck Institute for Demographic Research Working Paper 12*, 2006.

Balkema, A. A. and L. De Haan (1974). Residual life time at great age. *The Annals of Probability*, 792–804.

Coles, S. (2001). *An Introduction to statistical modeling of extreme values*, Volume 208. Springer.

Davies, R. B. (2002). Hypothesis testing when a nuisance parameter is present only under the alternative: Linear model case. *Biometrika*, 484–489.

Davison, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 393–442.

Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 24, pp. 180–190. Cambridge Univ Press.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics 44*, 423–453.

Huerta, G. and B. Sansó (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics 14*(3), 285–299.

Hyndman, R. J., S. Ullah, et al. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis 51*(10), 4942–4956.

Leadbetter, M., G. Lindgren, and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer, Berlin.

Lee, R. (2006). *Perspectives on Mortality Forecasting. III. The Linear Rise in Life Expectancy: History and Prospects*, Volume III of *Social Insurance Studies.* Swedish Social Insurance Agency, Stockholm.

Lee, R. D. and L. R. Carter (1992). Modeling and forecasting us mortality. *Journal of the American statistical association 87*(419), 659–671.

Li, N., R. Lee, and S. Tuljapurkar (2004). Using the lee–carter method to forecast mortality for populations with limited data*. *International Statistical Review 72*(1), 19–36.

Medford, A. (2015). Best practice life expectancy and extreme value theory. Unpublished.

Oeppen, J. and J. W. Vaupel (2002). Broken limits to life expectancy. *Science 296*(5570), 1029–1031.

Pickands, J. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, 119–131.

Renshaw, A. E. and S. Haberman (2003). Lee–carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics 33*(2), 255–272.

Torri, T. and J. W. Vaupel (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting 28*(2), 519–531.

Vallin, J. and F. Meslé (2009). The segmented trend line of highest life expectancies. *Population and Development Review 35*(1), 159–187.

White, K. M. (2002). Longevity advances in high-income countries, 1955–96. *Population and Development Review 28*(1), 59–76.