Pure

Syddansk Universitet

Best Practice Life Expectancy

An Extreme value Approach

Medford, Anthony

Best-practice life expectancy: An extreme value approach

Anthony Medford

Table of Contents

Best-practice life expectancy: An extreme value approach

Anthony Medford[1]

Abstract

BACKGROUND
Whereas the rise in human life expectancy has been extensively studied, the evolution of maximum life expectancies, i.e., the rise in best-practice life expectancy in a group of populations, has not been examined to the same extent. The linear rise in best-practice life expectancy has been reported previously by various authors. Though remarkable, this is simply an empirical observation.

OBJECTIVE
We examine best-practice life expectancy more formally by using extreme value theory.

METHODS
Extreme value distributions are fit to the time series (1900 to 2012) of maximum life expectancies at birth and age 65, for both sexes, using data from the Human Mortality Database and the United Nations.

CONCLUSIONS
Generalized extreme value distributions offer a theoretically justified way to model best-practice life expectancies. Using this framework one can straightforwardly obtain probability estimates of best-practice life expectancy levels or make projections about future maximum life expectancy.

COMMENTS
Our findings may be useful for policymakers and insurance/pension analysts who would like to obtain estimates and probabilities of future maximum life expectancies.

[1] Max Planck Odense Center on the Biodemography of Aging and University of Southern Denmark. E-Mail: amedford@health.sdu.dk.

1. Introduction

Mortality projections are crucial in many areas. In life insurance and pensions where the financial well-being of millions of individuals is at stake, these projections are crucial. Globally, governments and other stakeholders depend on reliable mortality projections for management and administration of their financial liabilities.

Oeppen and Vaupel (2002) introduced the term 'best-practice life expectancy' (BPLE), referring to the maximum life expectancy observed among national populations during a particular year. Their best-practice life expectancies at birth have been increasing in a nearly linear fashion, beginning in Scandinavia around 1840 and continuing ever since at a pace of about 0.24 years per annum for females and 0.22 years per annum for males (Oeppen and Vaupel 2002). In a later study that covered a longer period (1750–2005) Vallin and Meslé (2009) indicated instead that life expectancy increased in a piece-wise linear fashion over four distinct periods. Nonetheless, their basic results, especially for recent decades, are generally consistent with the overall pattern of increase of about three months per year (Vaupel 2012). Additionally, Shkolnikov et al. (2011) showed that best practice cohort female life expectancy at birth increased across cohorts born from 1870 to 1920 by an average of about 0.43 years annually.

There is a strong argument for using life expectancy in forecasting. White (2002) found that linear trends in life expectancy give a better fit to the experience of individual countries than linear trends in age-standardized (log) death rates in his study of 21 developed countries. Among those who have forecast life expectancy are Alho and Spencer (2005), Andreev and Vaupel (2006), Lee (2006), and Torri and Vaupel (2012). Oeppen and Vaupel (2002) argue that since the increase in best-practice life expectancy is linear and regular then it could be used in forecasting by comparing country-specific performance with the best practice. This approach takes advantage of national mortality trends, which ought to be considered within a larger international context rather than being analyzed and projected individually (Lee 2006).

Predicting future life expectancy trends by extrapolating from only one country's mortality experience could be problematic, as the trend toward catch-up among countries tends to exaggerate differences in life expectancy between countries over time, because the fastest risers outstrip other countries by ever increasing amounts (Wilmoth 1998). White (2002) finds that nations experience more rapid life expectancy gains when they are farther below the international average and therefore tend to converge toward the average. In addition, the life expectancy levels in different countries tend to be positively correlated. Age-specific death rates are highly correlated between males and females, among countries, and across ages. By forecasting best-practice life expectancy and the gap from this best-practice forecast for males and females in various countries, then estimating age-specific mortality from life expectancy (e.g., Ševčíková et al. 2016), these

problems can be tackled. Torri and Vaupel (2012), Andreev and Vaupel (2006), and Lee (2006) proposed methods which attempt to do this.

Although the consistent evolution of BPLE for over a century and a half is remarkable, this is simply an empirical observation; in the original work of Oeppen and Vaupel (2002) no attempt was made to formally model BPLE. Torri and Vaupel (2012) employed classic univariate ARIMA techniques, but their emphasis was on forecasting country-specific life expectancy where they forecast the BPLE and, separately, the gap between the BPLE and country-specific life expectancy. In this paper we propose a formal theoretical model for BPLE using arguments from extreme value theory (EVT).

Extreme value statistical methodology has previously been used in a mortality demographic context. Some of the earliest work using EVT in a demographic application began with Gumbel (1937, 1958). More recently, a number of papers involving the use of EVT have been published. Aarssen and De Haan (1994) estimated a finite upper bound on the distribution of human life spans, contrary to Galambos and Macri (2000), who countered that such an upper bound could not exist. Thatcher (1999) modeled the highest attainable age by using classical EVT. Watts, Dupuis, and Jones (2006) modeled the highest attained age by using the generalized extreme value (GEV) distribution. Han (2005) uses EVT to model the death rates for the elderly. Li, Hardy, and Tan (2008) use some classical results from EVT to develop a model called the threshold life table in order to extrapolate survival distributions to the oldest ages and to close a life table. Hanayama and Sibuya (2015) estimate the upper limit of the lifetime probability distribution of the Japanese population.

The primary aim of this article is to explore the modeling of best-practice life expectancies by using extreme value theory. Secondarily, we investigate how fitted EVT models may be used to make inferences about future levels of life expectancy. We note that our application of the theory is somewhat different from previous work in that we attempt to fit a model to life expectancy directly, whereas previous work fitted models to high ages at death. Our key contribution is to show that BPLE may be modeled using EVT.

This paper is structured as follows: Section 2 presents the data used to construct the BPLE. Section 3 provides some motivation and gives an outline of basic extreme value theory and the block maxima approach to the analysis of extremes. Section 4 presents the results, including details of the fitted GEV models and how quantiles of extreme life expectancy can be projected. There is a discussion in Section 5, and Section 6 concludes the paper.

2. Data

The data for use in model implementation and testing comes from two sources. First, the Human Mortality Database (HMD 2015) covers the low-mortality countries that have the best data and the highest life expectancies. It contains life tables for 37 countries plus all the raw data used in constructing these tables. The specific data used was life expectancy at birth and at age 65, for both males and females, and covers the period 1900 to 2012. For a small number of years, some eastern European populations – former Soviet Republics – had the highest life expectancies. Among these are Poland, Lithuania, Belarus, Russia, and Ukraine. These life expectancies were likely to be overstated, so whenever this occurred the second highest value, which appeared more plausible, was taken to be the annual maximum.

The second source of data covers all other countries according to the standard United Nation definitions and classifications as of 6 November 2013 (UN Statistics Division 2013). Life expectancies at birth and age 65 from these countries cover the period 1950–2012 and amount to a further 204 countries. This data is available from the UN World Population Prospects: The 2015 Revision (United Nations, Department of Economic and Social Affairs, Population Division 2015). In total, there is data from 241 countries.

The following notation will be used throughout:

e_x^*: best-practice life expectancy at given age x where " $*$ " indicates maximum

$e_{x,f}^*$: best-practice life expectancy at given age x for females

$e_{x,m}^*$: best-practice life expectancy at given age x for males.

Oeppen and Vaupel (2002) first highlighted the remarkable linearity in the BPLE. Figure 1 presents a plot of these BPLEs, separately for males and females, at birth and age 65, from the year 1900 and identifies which countries were leaders. The (piecewise) linear nature of the BPLEs and how few countries managed to attain best-practice levels of life expectancy are clearly seen.

A closer inspection of the time series of life expectancies in Figure 1 reveals that the pace of increase is not constant but has slowed slightly for life expectancies at birth, with a more pronounced acceleration at age 65. Therefore, rather than assume a constant linear increase, the presence of differential rates of increase is formally investigated. This is done by first 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.

Figure 1: **Countries with the highest life expectancies at birth and age 65, males and females separately, from 1900–2012**
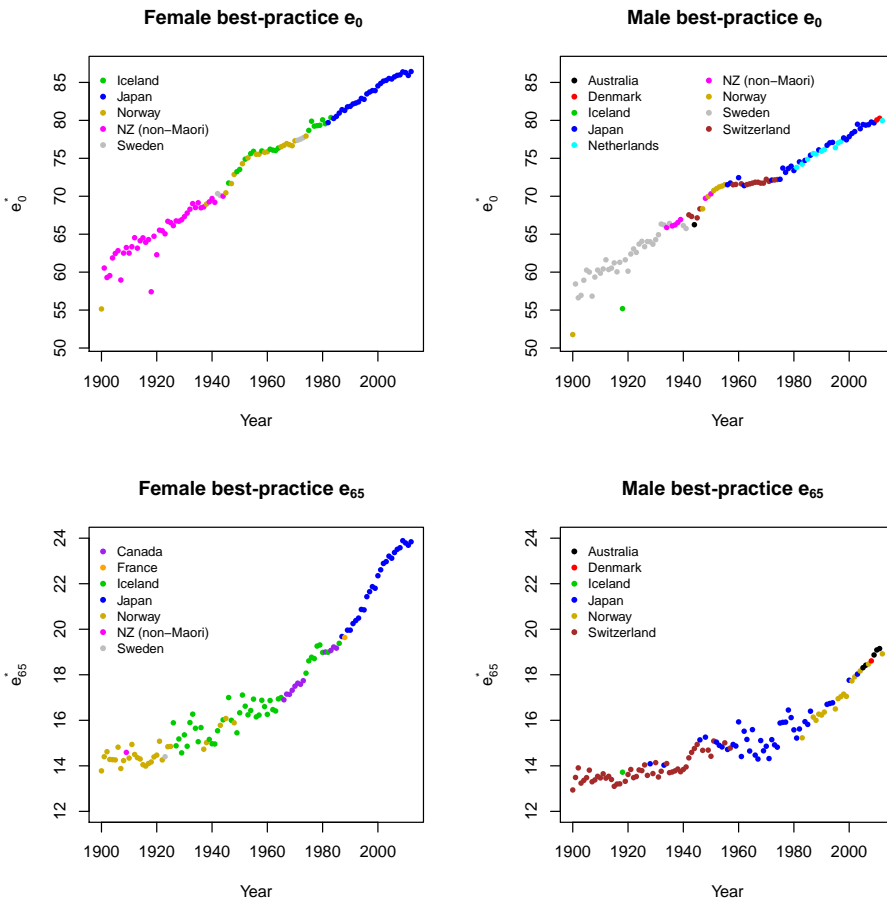
These segmented relationships are presented in Figure 2. For life expectancy at birth a structural breakpoint occurs around 1955 for females and 1950 for males. This is largely due to the pace of increase slowing down in the 1950s as the influence of the decline in life expectancy from reductions in infectious disease mortality during childhood and infancy begins to wane (Vallin and Meslé 2009) and the gains shift to the older ages. At age 65 the structural breaks occur later, around 1967 for females and 1984 for males, and this

is largely due to mortality reductions due to advances in the treatment and prevention of cardiovascular diseases that primarily affect the elderly.

Figure 2:   **Breakpoints in the trend of the highest life expectancies at birth and age 65, males and females separately, from 1900–2012**
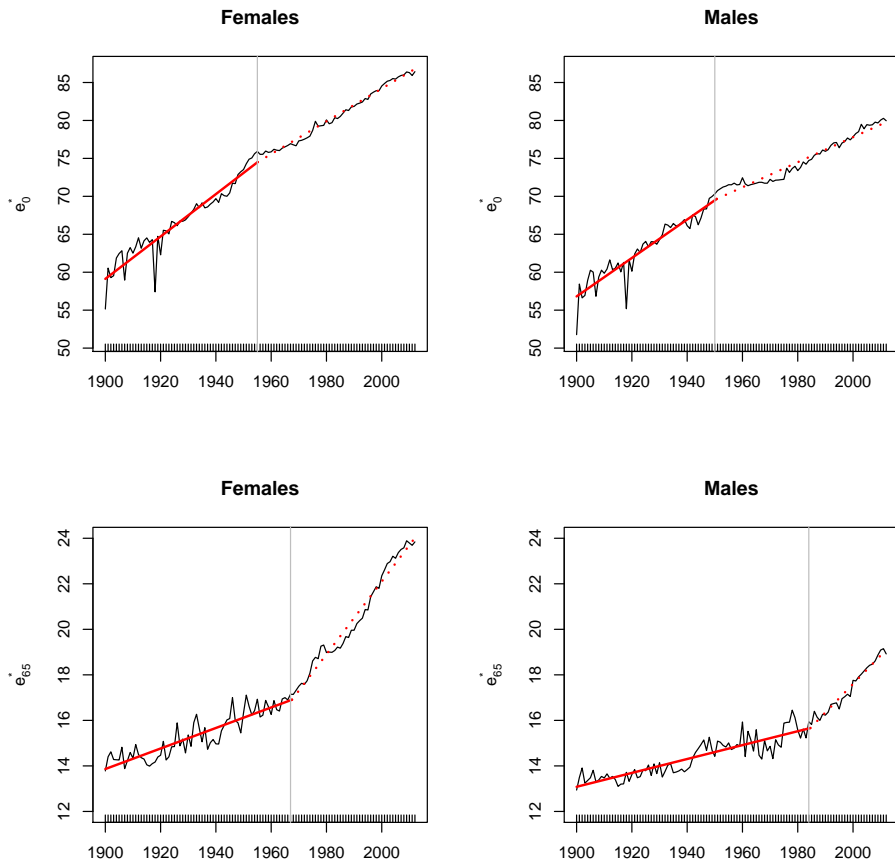
3. Methods

3.1 Motivation

Extreme value statistics facilitate the investigation of stochastic processes at very high or very low levels, and extreme value distributions arise as the limiting distributions of the maxima or minima of a set of random variables. In this paper we are interested in studying the evolution of the maximum life expectancies among all countries.

Figure 3 presents the BPLE at birth for females. It is well known that life expectancy has been increasing over time at all ages, and this is evident in a strong upward trend in the time series of the BPLE. The left panel presents this data from 1955 along with another series representing the same data but with the linear least-squares regression trend removed in order to produce a stationary series. The right panel shows a nonparametric estimate of this stationary data (we use the kernel density) and a fitted GEV distribution. The data has also been scaled by subtracting the intercept term of the fitted regression. Graphically, the fitted distribution appears to provide a very reasonable fit to the annual maximum data. Based on this evidence, one can conclude that extreme value theory may be useful in analyzing the BPLE.

Figure 3:   **Left panel: Raw and detrended data. Right panel: Kernel density and fitted GEV distribution**
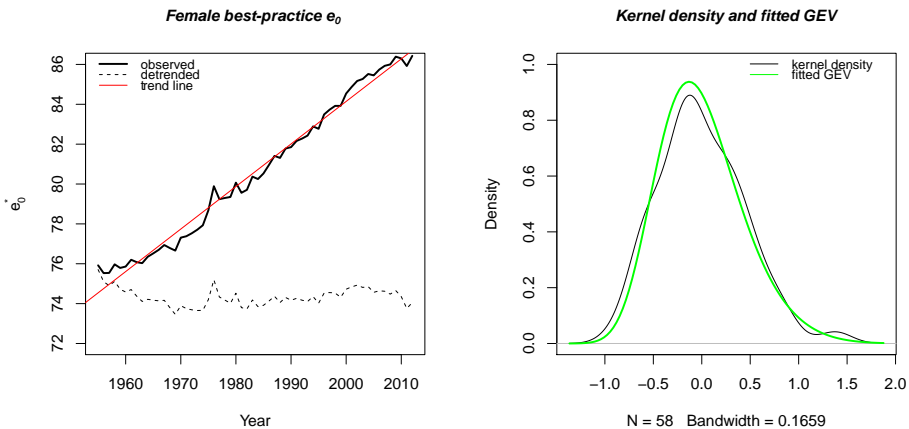
3.2 Classical extreme value theory: Basics

In this section we present some basics of the general classical extreme value theory. The specifics for our particular model will be put forward in subsequent sections. The limiting distributions of extremes give rise to the extreme value distributions. In this article, the parameterizations 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, though, is not particularly useful as the distribution of $F(x)$ is unknown. However, it is possible to find the distribution of M_n (for large n), say G, without any reference to F.

The distribution of M_n is degenerate because, as n tends to infinity, $F^n(z) \to 0$ for any z less than the upper end point of the support of F. 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 exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$, such that as $n \to \infty$,

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

where $G(z)$ is a nondegenerate distribution function, then G must be a member of the generalized extreme value (GEV) family of distributions. This is a striking 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

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

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. This leads to three types of distributions: When $\xi < 0$, the distribution has finite support and is short-tailed; leading to the Weibull distribution. When $\xi > 0$, there is polynomial tail decay, leading to heavy tails; the GEV is of the Fréchet type. The case where $\xi = 0$ is taken to be the limit of Eq. 2 as $\xi \to 0$ and there is exponential tail decay leading to light tails, and the GEV is of the Gumbel type with distribution function

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

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

Briefly, 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 $\hat{\xi}$. (See Section 3.3 for details on our approach.) For inference, estimates of extreme quantiles of the maxima are obtained by solving for z_p in equation 2:

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

where the distribution function of the GEV, $G(z_p) = 1 - p$ and p is the tail probability or the probability of realizing a value at least as large as z_p.

In extreme value terminology, the quantiles of the distribution, z_p, are sometimes called return levels and are associated with the so-called return period $1/p$. If we consider 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 and corresponds to the $1/(1 - 0.01) = 100$-year return period. It is the amount which one expects to see once every 100 years, on average.

3.3 The setup

In applications the X_i usually represent values of a process taken at regular intervals, perhaps daily mean temperatures or hourly measurements of wind speed, so that M_n represents the maximum of the process over n time units of observation. Our application of the theory is somewhat different.

For us, because BPLE is an already defined metric and partly because annual life expectancies are more common and useful in practice, we choose block lengths of one year. Each block represents a sample of size $n = 241$ period life expectancies, $X_1, X_2, \ldots,$ X_{241} from each country represented in our dataset. From this sample, we select the maximum, Z_{241}. This is done for each year $t = 1 \ldots t_{max}$, where t_{max} is the last year of the data, resulting in a series of block maxima $Z_{241,t}; t = 1 \ldots t_{max}$ to which a GEV model is fitted.

The linear trend over time in e_x^* is accounted for by allowing the location parameter of the GEV model to vary linearly with time such that, instead of a fixed location parameter μ, a more flexible parameter is adopted. Thus, time is introduced as a covariate into the parametrization of the GEV distribution, so we assume a location parameter of the form, $\mu_t = \mu_0 + \mu_1 t$, where t represents calendar time. More specifically, t is an index commencing at 1 in the first year of the data (e.g., 1950 for $e_{0,f}^*$) and increases by one unit per consecutive year. The parameter μ_1 could be interpreted as roughly the annual rate of increase in life expectancy. The parameter μ_0 is the fitted initial level of the location parameter for the first year of data.

Since e_x^* has been trending upward linearly over time, μ, the location parameter of the GEV, is the most obvious parameter of the GEV to capture this feature, but we point out that time dependence could also be introduced into the other parameters, provided that the additional complexity is justified and strongly supported by the data. In general, however, the shape parameter, ξ, is not usually altered as it can be difficult to estimate in practice (Coles 2001). For our model we assume that all the time dependence in e_x^* is sufficiently captured by the time-varying location parameter, μ_t.

Although several methods are available, maximum likelihood estimation (MLE) is often a sensible choice because of its relative flexibility and ability to easily incorporate covariate information (Coles 2001; Coles and Dixon 1999). We propose MLE for finding the parameters of the GEV distribution μ_0, μ_1, σ, and ξ. Denote by $z_1, \ldots, z_{t_{max}}$ the sequence of the maxima taken over each data block. Assuming independent maxima, the log likelihood function is defined as

$$\ell(\mu_0, \mu_1, \sigma, \xi) = - \sum_{t=1}^{t_{max}} \left(\log \sigma + (1 + 1/\xi) \log \left[1 + \xi \left(\frac{z_t - (\mu_0 + \mu_1 t)}{\sigma} \right) \right] + \right.$$

$$\left. \left[1 + \xi \left(\frac{z_t - (\mu_0 + \mu_1 t)}{\sigma} \right) \right]^{\frac{-1}{\xi}} \right),$$

provided that $1 + \xi(z - \mu)/\sigma > 0$ for $t = 1, \ldots, t_{max}$. If any $\xi = 0$, an alternative log-likelihood function, derived from the Gumbel distribution, is used. There is no analytical

solution; therefore, numerical procedures are used to jointly estimate the value of the parameters $\hat{\mu}_0, \hat{\mu}_1, \hat{\sigma}$, and $\hat{\xi}$, which maximizes the log-likelihood function. Numerical optimization is done using the R package 'extRemes'.

3.4 A note on assumptions

We now briefly discuss the key assumptions of the classical GEV model and contrast them with our setup and data realities. It is important to bear in mind that the GEV is based on asymptotics. The classical theory is based on maxima taken from blocks consisting of independent, identically distributed variables. Our model does not assume a stationary distribution of extremes as a separate GEV is assumed for each year, as elaborated previously. Therefore, our discussion is mainly on the issue of dependence between e_x^*. First, a comment on within-block dependence is in order. Although there may be some common underlying trends in closely connected countries, we consider the overall dependence in life expectancy between the 241 countries within each block to be sufficiently independent as to not compromise the use of the GEV distribution. Life expectancy in, say, Angola would not under normal circumstances be related to that in Sweden, for example.

On the other hand, between blocks – at extremal levels – there exists dependence which we, for expository purposes, categorize loosely as temporal and nontemporal. This is not a rigid distinction, and the effects may not be easy to disentangle in practice. From a temporal perspective, human mortality trends evolve very stably and, barring catastrophes such as war or pandemics, a high period life expectancy should be followed by the same. We also observe that countries tend to achieve the BPLE in clusters; this pattern is most evident amongst females at birth where only eight countries have been the leader since 1840 (Oeppen and Vaupel 2002).

From a nontemporal perspective, the world's populations are becoming more intimately connected via transportation, communication, trade, and technology. Wilson (2011) documented the convergence of global mortality levels over recent decades. The low-mortality countries which have held the BPLE are becoming more similar in their lifestyles, and globalization is also acting on practices affecting mortality, contributing to converging mortality patterns (Wilson 2011; White 2002; Edwards and Tuljapurkar 2005). As a consequence, there has been a significant degree of convergence in e_0 within our group of low-mortality countries, but less so in e_{65}.

We note that this same convergence mechanism enhances the assumption of an identical distribution: Although countries are of different sizes, ethnic mix, and socioeconomic status, the empirical evidence – particularly among the low-mortality countries – indicates convergence towards similar mortality regimes. On the one hand, the conditions which stimulate convergence act to make the mortality regimes of the low-mortality

countries in our study more similar in distribution, enhancing the arguments for using the GEV. On the other hand, these same factors affect mortality across all countries, leading to greater dependence amongst the mortality patterns and trajectories.

We concede that mortality convergence may be a feature of our group of low-mortality countries, resulting in dependence between consecutive BPLEs and likely introducing some bias into the parameter estimates, thus resulting in underestimation of standard errors. For modeling purposes we make no distinction between whether any dependence is temporal or driven by some other underlying mechanism. What we can say with certainty is that the effects of dependence manifest themselves through time. As a result, we assume that any dependence in life expectancies is due to time effects.

To estimate the magnitude of the impact of any dependence between $(e^*_{x,t}, e^*_{x,t+1})$ we conduct a simple sensitivity analysis. The parameter estimates for the GEV models are recalculated, but rather than using the full dataset we exclude some annual maxima by using data spaced further apart in time. If temporal dependence is strong, then fitting the model using data spaced further apart should substantially impact the GEV parameter estimates. In other words, if dependence is not substantial, then refitting the model using data spaced further apart should not excessively alter the parameter point estimates, and the change should be within a reasonable order of magnitude. The results of this (see Appendix B) suggest that temporal dependence is not an impediment for fitting the GEV model.

4. Results

4.1 Fitted models

Torri and Vaupel (2012) used data from 1900 to fit BPLE. They argued that this period was used because it is nearly linear, while the longer time series available from 1840 varies in a piecewise linear fashion (Vallin and Meslé 2009). However, as demonstrated in Figure 2, there have also been structural breaks during the 20th century; therefore, we do not fit the BPLE arbitrarily from 1900 but from the point of the most recent structural break, which varies depending on the underlying population. By doing this, we ensure that the correct pace of life expectancy increase is attributed to the correct time period and population. Hence, the fitting periods are from 1950 for $e^*_{0,m}$ (representing males at birth), from 1955 for $e^*_{0,f}$ (females at birth), from 1967 for $e^*_{65,f}$ (females at age 65), and from 1984 for $e^*_{65,m}$ (males at age 65).

In summary, we fit time-dependent GEV models:

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

where the location parameter $\mu_t = \mu_0 + \mu_1 t$, 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 over which the model is fitted.

Appendix A presents fitted model diagnostics and shows that the models fit the data well. Using the parameter estimates from Table 1 results in the following fitted GEV models for extreme life expectancy. Note that the shape parameter of the GEV for $e_{0,f}^*$ is zero, indicating a Gumbel distribution. The estimated value for this parameter was about –0.05, but a likelihood ratio test of significance indicated that this value was not significantly different from zero. The negative shape parameters for the others indicate Weibull distributions.

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**

	Neg. likelihood	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\sigma}$	$\hat{\xi}$
Female e_0	33.0	74.0 (0.11)	0.22 (0.003)	0.37 (0.030)	0
Male e_0	65.0	69.4 (0.15)	0.16 (0.003)	0.75 (0.082)	-0.46 (0.101)
Female e_{65}	14.7	16.6 (0.11)	0.16 (0.004)	0.36 (0.048)	-0.43 (0.129)
Male e_{65}	-4.7	15.5 (0.10)	0.12 (0.005)	0.21 (0.028)	-0.29 (0.111)

For female life expectancy at birth:	$GEV(74.0 + 0.22t,\ 0.37)$
For male life expectancy at birth:	$GEV(69.4 + 0.16t,\ 0.75, -0.46)$
For female life expectancy at 65:	$GEV(16.6 + 0.16t,\ 0.36, -0.43)$
For male life expectancy at 65:	$GEV(15.5 + 0.12t,\ 0.21, -0.29)$

Because EVT is concerned with rare events, it is convenient to make inferences on the extreme quantiles (return levels) of the fitted model. Using Equation 3 and the fitted parameter estimates from Table 1, our time-varying return levels are easily calculated. If we use male life expectancy at birth data, in year t, the 98th percentile (50-year return level) estimates the highest life expectancy we might expect to encounter on average every 50 years and is given by

$$z_{0.02}(t) = (69.4 + 0.16t) + 1.62\left[1 - \{-log(1 - \frac{1}{50})\}^{0.46}\right].$$

The quantiles for the other populations are derived in a similar way. Figure 4 presents graphically the 20-year and 100-year return levels, the upper 95% and 99% quantiles of the fitted GEV distributions. Also presented are the corresponding lower quantiles. The quantiles of the different populations are at different distances from each other, reflecting the different tail behavior of the fitted GEVs. The Gumbel of the $e_{0,f}^*$ has a heavier right tail and lighter left tail than the Weibull of the other three populations, so that the

lines representing the upper quantiles are more spaced out and are closer together for the lower quantiles. The reverse is true for the Weibull. The return levels can also be used in an objective way to identify periods of unusually high or low life expectancy because realizations of high or low life expectancy do not necessarily imply extremities. Suppose that the criteria for extreme life expectancy is a realization higher than the 99th percentile. Then one can immediately identify from Figure 4 the data of interest. Indeed, if the model-fitting period is at least as long as the return period of interest, then it would be possible to compare the theoretical tail probability with the observed. For example, the 98th quantile implies a 2% tail probability or a probability that the quantile would be realized once every 50 years. If the 98th percentile is exceeded regularly, one can assume that mortality experience has been unusually light, leading to the high life expectancy, and investigate further. If we consider the upper 95th percentile from Figure 4 we can make these types of comparisons. Over the period 1955–2012 for $e^*_{0,f}$ we would expect to observe a life expectancy in the upper 5% tail about three times over that 60-year period. In reality we observe four. For an assessment of the predictive performance of the model refer to Appendix D.

4.2 Applications

4.2.1 Projections

The return levels from Figure 4, similar to the data, evolve in a linear way. If we assume that the current rate of increase in e^*_x continues over a given forecast horizon, then it is possible to project return levels. Table A–1 presents the 100-year, 20-year, and 2-year (median) return levels in 2040 and 2050 (prediction intervals not shown). For example, in 2040, the median $e^*_{65,m}$ at age 65 is 22.4 years, but the 98th percentile is 22.9 years. Appendix B provides plots of the quantiles for the years 2013 to 2050. A key assumption is that the linear trend for μ continues beyond the range of data we have observed.

Table 2: **Projected extreme life expectancy return levels in 2040 and 2050**

	Group	Median ($z_{0.5}$)	20 Yr Level ($z_{0.05}$)	50 Yr level ($z_{0.02}$)
2040	Female e_0	92.9	93.9	94.2
	Male e_0	84.2	85.2	85.3
	Female e_{65}	28.4	28.9	29.0
	Male e_{65}	22.4	22.8	22.9
2050	Female e_0	95.1	96.1	96.4
	Male e_0	85.9	86.8	87.0
	Female e_{65}	30.0	30.5	30.6
	Male e_{65}	23.7	24.0	24.1

Figure 4: **Best-practice life expectancies and the fitted GEV median, 20-year and 100-year return levels**
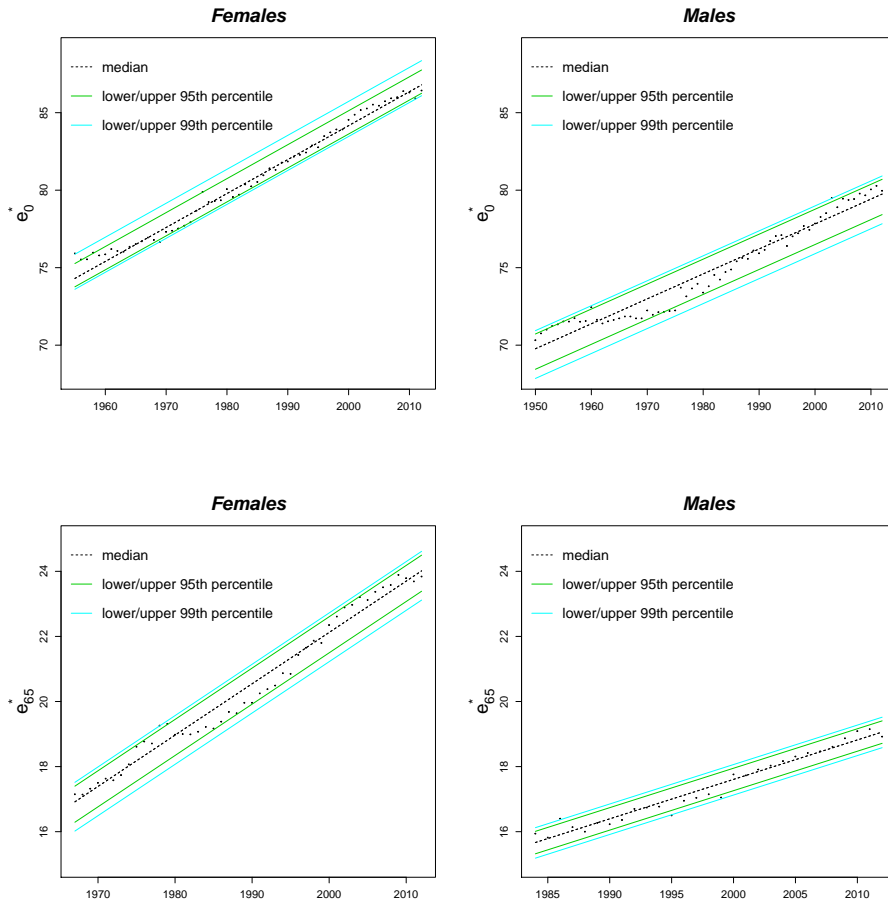
4.2.2 Probabilities

In addition to quantiles, another way of making inferences for GEV models is in the calculation of probabilities (see Table 3). Using Equation 2 and the estimated parameters, one can, given the quantiles, calculate the probabilities. For example, the probability that the maximum female life expectancy at birth will exceed 90 years in 2025 is approximately 59%; by 2050 the probability climbs to over 99.9%. Similarly, there is about

a one-third chance that by 2025 the maximum observed life expectancy of 65-year-old females will reach 27 years.

Table 3: **Probability of the maximum female life expectancies at birth, $e_{0,f}^*$, exceeds certain levels for the years 2020 and 2050**

Year	$P(e_{0,f}^* > 90)$	$P(e_{0,m}^* > 85)$	$P(e_{65,f}^* > 26)$	$P(e_{65,m}^* > 24)$
2025	0.23	< 0.001	0.57	< 0.001
2050	> 0.999	0.86	> 0.999	0.05

4.2.3 Other inference

Conversely, if we have the life expectancy quantile and assume some level of probability, we can estimate the future time when the given life expectancy level will be observed. For example, if we adopt a significance level of 95%, then maximum female life expectancy at birth observed in the world would reach 100 years around 2070; with 95% probability the country with the highest male life expectancy at age 65 would reach 27 years around 2080.

The foregoing analysis is based on worldwide maxima. The blocks to which the block maxima method was applied contained all countries in the world. However, provided that there is sufficient high-quality life expectancy data available and the underlying assumptions are still valid, the methods of the previous section can be applied to some subregion.

5. Discussion

Since e_x^* is the maximum annual life expectancy at age x from among the different nations, we have shown that the theory of extreme value statistics can be applied. We used data beginning in 1900, found structural breakpoints, and fit GEV distributions from the most recent breakpoint. This was done in order to ensure that the correct pace of life expectancy improvement was captured for each specific subpopulation.

We showed that the underlying trend in e_x^* can be used in combination with a fitted GEV to project quantiles of extreme life expectancy. Our concern here is inference about extreme or very high life expectancy, so by definition it is unlikely that the estimated life expectancy return levels would be realized regularly in practice, but the methodology serves as a useful objective tool in quantifying the possibility of observing such levels. We used a simple approach for projection, but more sophisticated methods can be employed. For example, the time-varying parameters of the fitted GEV could be projected stochastically using time series methods. See, e.g., Huerta and Sansó (2007).

Projection of the BPLE can be an important component in forecasting life expectancy. According to the work on this method introduced by Oeppen and Vaupel (2002) and further elaborated by Torri and Vaupel (2012), forecasts of country-specific life expectancy can be obtained by first projecting the BPLE and then modeling and forecasting the gap between the BPLE and the life expectancy in any given population. Torri and Vaupel (2012) used classical univariate time series ARIMA models to forecast $e_{0,f}^*$ and $e_{0,m}^*$. While this is an obvious and straightforward approach, there are some advantages to our methodology. First, there is the strong theoretical justification for using EVT to fit maxima data, and second, with our approach it is possible to not only project e_x^* but also to straightforwardly obtain probabilities about future values of e_x^*.

We calculated the median of $e_{0,f}^*$ in 2050 to be 95.1 years, which is lower than the 96.59 years found by Torri and Vaupel (2012). For $e_{0,m}^*$ it is 85.9 years versus 88.38 years in Torri and Vaupel (2012). This is unsurprising, since, as was shown in Section 2, the pace of life expectancy increase for females has slowed to about 0.21 years per year from around 1955, so that the constant drift term of about 0.24 years per year in the ARIMA model overestimates the future value of BPLE. Similarly, for males at birth the pace has slowed to about 0.17 years per year from 1950 versus a constant ARIMA drift of about 0.20 years per year. On the other hand, using the UN projections of life expectancy for all countries, we obtain median $e_{0,f}^*$ and $e_{0,m}^*$ in 2050 of 91.35 years (Hong Kong) and 86.2 years (Israel) respectively. Our projection for females is higher. This is because the UN median projection model assumes a much slower pace of increase than ours of around 0.1 years per year. For males, however, our model and the UN's result in very similar projections of BPLE.

6. Concluding remarks

In this analysis we used extreme value theory to analyze the time series of annual maximum life expectancy, which to our knowledge has not been previously attempted. We were able to use the fitted model to project future values of BPLE and make probability statements about future values of maximum life expectancy.

This approach could be useful in a number of contexts: first, as an input into the computation of population-specific life expectancies, where the best-practice level and the lag between country and best-practice level are modeled and forecast separately. Perhaps exact country-specific life expectancies are not required, but interest is in the possibly of realizing high (or low) extreme levels. This may be of interest to insurers or pension funds that might require an estimate of the probability of extreme life expectancy exposures which may trigger loss on some product or contract. Of course, EVT is useful not only for extreme maxima but also for extreme minima. For example, one might be interested in ascertaining the likelihood of realizing some low level of life expectancy, which in

turn could be driven by war, disease, or some other catastrophic event that affects life expectancy in an adverse way.

Future work will include the evaluation of robust prediction intervals for any projections. We can also investigate other EVT approaches, such as using more order statistics, where one can model the five highest life expectancies, for example. Models making use of covariates other than time and applied to parameters other than the location parameter could also be explored. Finally, more sophisticated approaches to forecasting the GEV could be explored.

7. Acknowledgements

References

Aarssen, K. and De Haan, L. (1994). On the maximal life span of humans. *Mathematical Population Studies* 4(4): 259–281. doi:10.1080/08898489409525379.

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

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

Coles, S. (2001). *An Introduction to statistical modeling of extreme values*. New York: Springer. doi:10.1007/978-1-4471-3675-0.

Coles, S.G. and Dixon, M.J. (1999). Likelihood-based inference for extreme value models. *Extremes* 2(1): 5–23. doi:10.1023/A:1009905222644.

Davies, R. (2002). Hypothesis testing when a nuisance parameter is present only under the alternative: Linear model case. *Biometrika* 89(2): 484–489. doi:10.1093/biomet/89.2.484.

Edwards, R.D. and Tuljapurkar, S. (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review* 31(4): 645–674. doi:10.1111/j.1728-4457.2005.00092.x.

Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* 24(2): 180–190. doi:10.1017/S0305004100015681.

Galambos, J. and Macri, N. (2000). The life length of humans does not have a limit. *Journal of Applied Statistical Science* 9(4): 253–264.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* 44(3): 423–453. doi:10.2307/1968974.

Gumbel, E. (1958). *Statistics of extremes*. New York: Columbia University Press.

Gumbel, E.J. (1937). *La durée extrême de la vie humaine*. Paris: Hermann et cie.

Han, Z.J. (2005). Living to 100 and beyond: An extreme value study. In: Johansen, R.J. (ed.). *Living to 100 and Beyond Symposium Monograph*.

Hanayama, N. and Sibuya, M. (2015). Estimating the upper limit of lifetime probability distribution, based on data of Japanese centenarians. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences:* glv113.

Huerta, G. and Sansó, B. (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics* 14(3): 285–299. doi:10.1007/s10651-007-0014-3.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). (Available at http://www.mortality.org.). Data downloaded on feb. 2, 2015.

Lee, R. (2006). Perspectives on mortality forecasting. III. The linear rise in life expectancy: History and prospects. Stockholm: Swedish Social Insurance Agency: 19–44.

Li, J.S.H., Hardy, M.R., and Tan, K.S. (2008). Threshold life tables and their applications. *North American Actuarial Journal* 12(2): 99–115. doi:10.1080/10920277.2008.10597505.

Oeppen, J. and Vaupel, J.W. (2002). Broken limits to life expectancy. *Science* 296(5570): 1029–1031. doi:10.1126/science.1069675.

Ševčíková, H., Li, N., Kantorová, V., Gerland, P., and Raftery, A.E. (2016). Age-specific mortality and fertility rates for probabilistic population projections. In: Schoen, R. (ed.). *Dynamic Demographic Analysis*. New York: Springer: 285–310. doi:10.1007/978-3-319-26603-9_15.

Shkolnikov, V.M., Jdanov, D.A., Andreev, E.M., and Vaupel, J.W. (2011). Steep increase in best-practice cohort life expectancy. *Population and Development Review* 37(3): 419–434. doi:10.1111/j.1728-4457.2011.00428.x.

Thatcher, A.R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 162(1): 5–43. doi:10.1111/1467-985X.00119.

Torri, T. and Vaupel, J.W. (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting* 28(2): 519–531. doi:10.1016/j.ijforecast.2011.01.009.

United Nations, Department of Economic and Social Affairs, Population Division (2015). *World Population Prospects: The 2015 Revision*. Tech. rep., United Nations, Department of Economic and Social Affairs, Population Division. https://esa.un.org/unpd/wpp/Download/Standard/Population.

United Nations Statistics Division (2013). Countries or areas, codes and abbreviations [electronic resource]. http://unstats.un.org/unsd/methods/m49/m49alpha.htm.

Vallin, J. and Meslé, F. (2009). The segmented trend line of highest life expectancies. *Population and Development Review* 35(1): 159–187. doi:10.1111/j.1728-4457.2009.00264.x.

Vaupel, J.W. (2012). *How Long Will We Live? A Demographer's Reflexions on Longevity*. Tech. Rep. 20, Scor Global Risk Centre.

Watts, K.A., Dupuis, D.J., and Jones, B.L. (2006). An extreme value analysis of advanced age mortality data. *North American Actuarial Journal* 10(4): 162–178. doi:10.1080/10920277.2006.10597419.

White, K.M. (2002). Longevity advances in high-income countries, 1955–96. *Population and Development Review* 28(1): 59–76. doi:10.1111/j.1728-4457.2002.00059.x.

Wilmoth, J.R. (1998). Is the pace of Japanese mortality decline converging toward international trends? *Population and Development Review* 24(3): 593–600. doi:10.2307/2808156.

Wilson, C. (2011). Understanding global demographic convergence since 1950. *Population and Development Review* 37(2): 375–388. doi:10.1111/j.1728-4457.2011.00415.x.
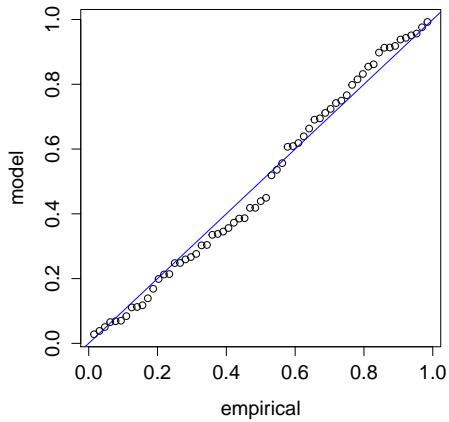
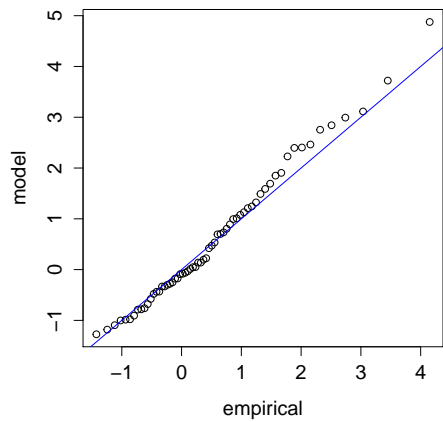# Appendix A: Diagnostic tests

Residual probability plot
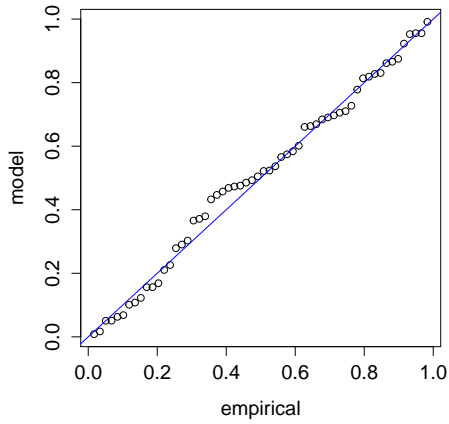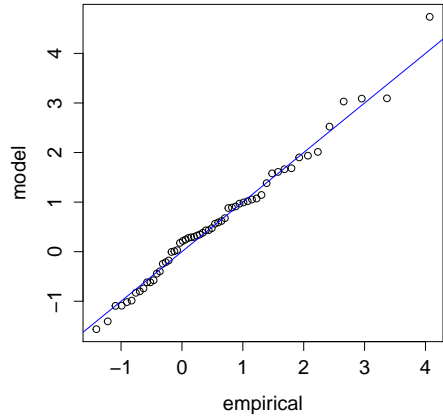
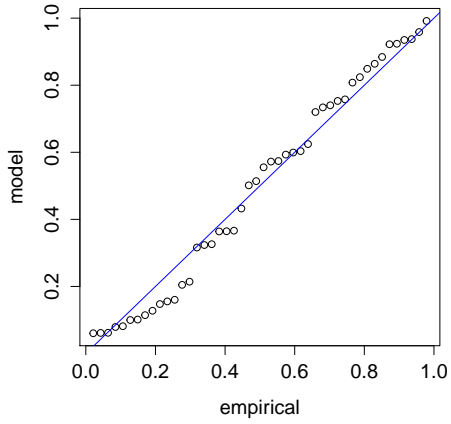### Residual quantile plot (Gumbel scale)

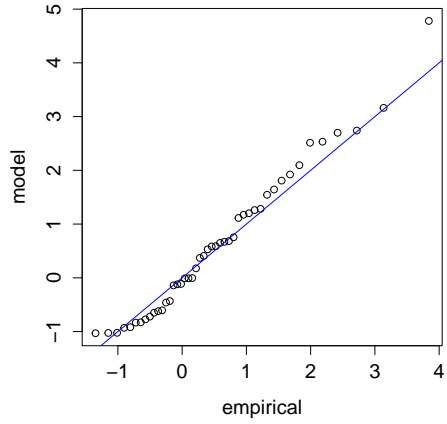### Residual probability plot

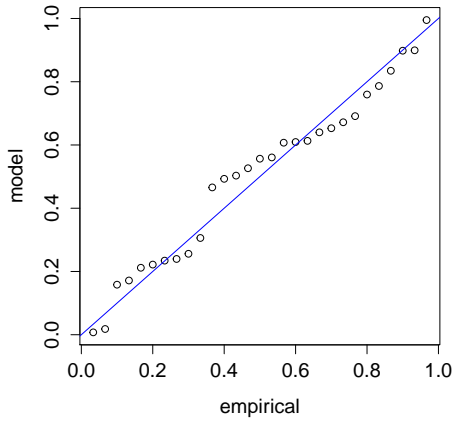### Residual quantile plot (Gumbel scale)
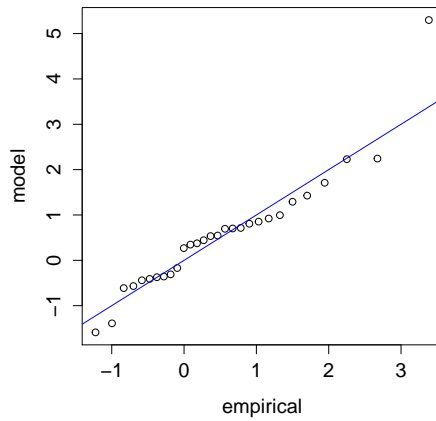
Residual probability plot

Residual quantile plot (Gumbel scale)

Residual probability plot

## Residual quantile plot (Gumbel scale)

Appendix B: Parameter sensitivity tests for temporal dependence

Table A–1: **Sensitivity tests of GEV parameter estimates using data spaced further apart in time. Estimates of μ_0 are rounded to one decimal place; all other parameter estimates are rounded to two decimal places**

	Maxima	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\sigma}$	$\hat{\xi}$
	Annual	74.0	0.22	0.37	0
Female e_0	Every 2nd year	74.0	0.22	0.37	0
	Every 3rd year	74.1	0.22	0.41	0
	Annual	69.4	0.16	0.75	-0.46
Male e_0	Every 2nd year	69.7	0.15	0.78	-0.73
	Every 3rd year	69.7	0.15	0.95	-0.99
	Annual	16.6	0.16	0.36	-0.43
Female e_{65}	Every 2nd year	16.7	0.16	0.40	-0.67
	Every 3rd year	16.9	0.15	0.46	-1.04
	Annual	15.5	0.12	0.21	-0.29
Male e_{65}	Every 2nd year	15.4	0.12	0.14	0.13
	Every 3rd year	15.6	0.12	0.21	-1.05

The estimated parameter values for the different populations in general are not greatly affected by using data spaced farther apart. The parameter $\hat{\xi}$ is the key parameter of interest in this sensitivity analysis as it determines the shape of the distribution and, hence, the parametric family to which the fitted model belongs. By using data further apart, we look for two things: first, that the change in parameter values due to the re-estimation is not disproportionately large and, second, that there be no change in the distributional family on account of the re-estimated shape parameter. These criteria are met in all cases except for the male e_{65} population. We attribute this to unreliable parameter estimates due to the scarcity of data, where using data every second year results in 15 data points and using data every third year results in 10 data points.

Appendix C: Projected quantiles
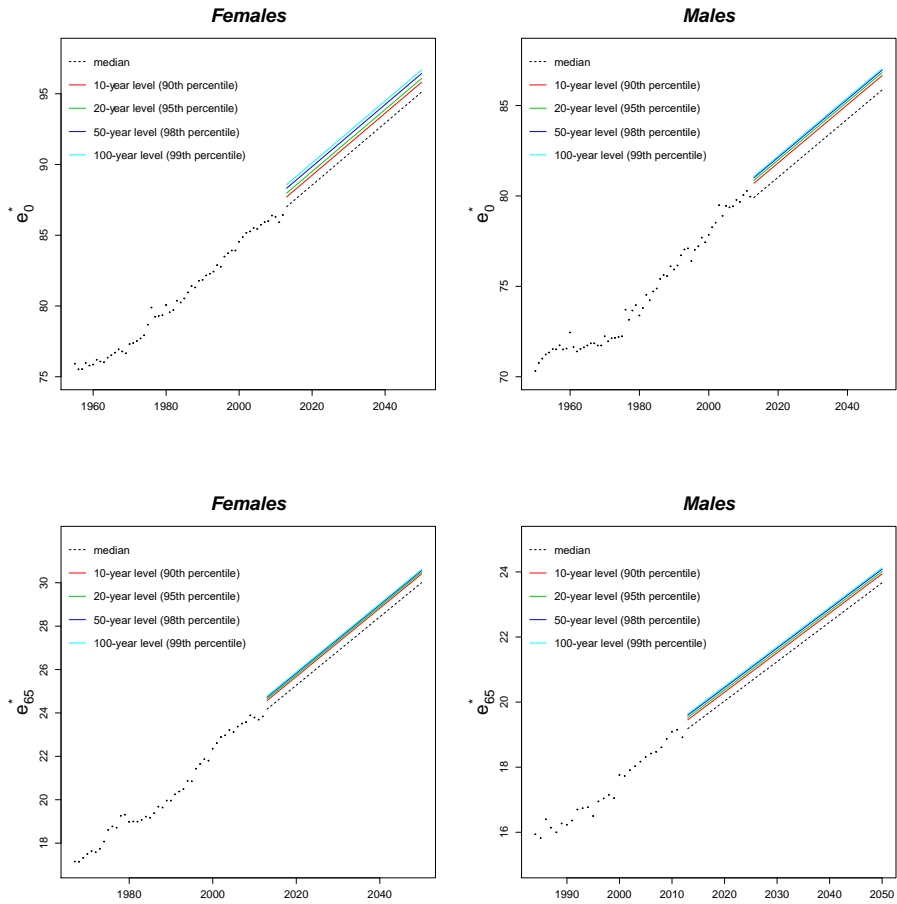
Appendix D: Predictive performance

We initially considered comparing the quantiles from the fitted model to sample quantiles but decided against it as the model quantiles would generally be higher than actuals, plus many of the life expectancies calculated for less developed regions are themselves crude approximations.

As a rough check we compare $\hat{z}_{0.1}(t_i)$ to $\max[e_x(t_i + 1), \ldots, e_x(t_i + 10)]$ for $t = 1985, \ldots, 2002$ and $x = 0, 65$. We fit the model using data up to 1984 (the most recent structural breakpoint of the four subpopulations). Because the 10-year return level is attained, on average, once every 10 years, we compare return levels calculated each year with the rolling maxima observed over successive forward-looking 10-year periods. Only data up to 2002 could be used due to the 10-year forecast, which precluded any longer return periods being used.

The evaluation is done using the standard forecast accuracy measures: mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). The forecasting performance appears reasonable, with an average error between about 2% for female e_0 and 5.84% for female e_{65}.

Table A–2:   **Predictive assessment of fitted models**

	MAE	RMSE	MAPE
Female e_0	1.68	1.70	1.98 %
Male e_0	1.28	1.33	1.61 %
Female e_{65}	1.34	1.36	5.84 %
Male e_{65}	0.53	0.55	2.97 %