The Myth of Immortality: An Analysis of the Maximum Lifespan of US Females

Jan Feifel, Martin Genz, and Markus Pauly

Ulm University

Abstract. Questions related to the existence and specification of a limit to human lifespan lead to heated discussions in several scientific fields such as biology, demography, medicine, or actuarial sciences. In the present paper, we contribute to this discussion from a statistical point of view. To this end, we use combined mortality data of US females obtained from the International Database on Longevity as well as the Human Mortality Database. The use of old-age mortality data typically raises two issues: sparse information on the old ages and censored observations. Up to the present paper, this censoring issue has been ignored in previous investigations on the maximum human lifespan. We address this accordingly by combining sub-sampling and cross-validation techniques with recent results on censored extreme value theory. As the main result, we estimate the maximum lifespan of US females for moving intervals of nine calendar years between 1980 and 2003.

Key words and phrases: Censoring, Extreme value theory, Subsampling, Human Mortality Database, Supercentenarians, Old-age mortality.

1. INTRODUCTION

The dream of immortality is not only subject to legends, science fiction, and fairy tales but also has attracted the attention of researchers from many different fields of science. Who has not yet dreamed about the possibility to extend the individual lifespan to ages that have never been reached before? Following, for example, de Grey (2003) immortality may become reality in near future. On the other hand, there may be a natural limit to human life which probably is just not yet in sight. Or maybe we have already reached this limit. For example Dong, Milholland and Vijg (2016) found *'evidence for a limit to human lifespan'* which is supported by a comprehensive data analysis. However, they base their claims on statistics like the maximum reported age at death which completely ignores people who are still alive at highest ages. Others like, e.g., Fries (1980) found natural constraints to longevity while a few decades later reality exceeded their forecasts. Recently, Rootzén and Zholud (2017) concluded that *'human life*

Jan Feifel is Research Fellow, Institute of Statistics, Ulm University, 89081 Ulm, Germany Martin Genz is Research Fellow, Institute for Finance and Actuarial Sciences and Institute of Insurance Science, Ulm University, 89081 Ulm, Germany Markus Pauly is Professor, Institute of Statistics, Ulm University, 89081 Ulm, Germany

is unbounded but short'. However, their statistical analysis is based on small sample sizes and thus may lack power, see also Section 3 below. In any case, as pointed out by Weon and Je (2009), *'the existence of maximum human lifespan remains a puzzle in aging research'*.

From a statistical point of view, all of these questions are related to the support of the density function of the underlying age distribution at death (the so-called deaths curve) and lead to investigations about its right endpoint. The current paper studies them for US females by means of modern statistical methods from extreme value theory (EVT). Since the deaths curve typically changes over time Börger, Genz and Ruß (2017) its right-endpoint also changes over time Wilmoth et al. (2000); Cohen and Oppenheim (2012). We, therefore, study the evolution of the deaths curve's right tail over moving time periods.

For this analysis, we use data from the International Database on Longevity (IDL) and the Human Mortality Database (HMD). To avoid gender effects and to ensure an acceptable data quality and quantity we focus on data of US females. The HMD is one of the largest publicly available databases on mortality in the world. However, many observations in the HMD are right-censored, particularly at the age of 110. This prohibits a gain of information about the right tail behavior of the deaths curve. However, it is well-known that there have been humans which survived the age of 110 (see e.g. Robine and Allard (1998) for the famous case of Jeanne Calment) and thus the HMD data alone is not suitable for any EVT analysis. In contrast, the IDL is more informative regarding the death counts of the highest ages: It provides the exact age and time of death of so-called supercentenarians, i.e. individuals who at least reached their 110th birthday. The IDL covers death counts of 309 US female supercentenarians in the time span 1980–2003. This allows for a statistical analysis in the classical EVT framework de Haan and Ferreira (2006); Reiss and Thomas (2007); Falk, Hüsler and Reiss (2010); Resnick (2013) as has, e.g., been applied to estimate the maximum attainable age for the Netherlands and Belgium Aarsen and de Haan (1994); Gbari et al. (2017); Einmahl, Einmahl and de Haan (2017) as well as for Canada and Japan Watts, Dupuis and Jones (2006). Since such an analysis has not been carried out for the US alone, we start our investigation with the IDL. Since 309 is a rather small number for quite a large time horizon we increase the inferential accuracy by subsequently combining the datasets of the IDL and the HMD. We then study this combined data set (CDS) over moving time intervals, each of 9 years length. In this way, we evaluate the progression of the lifespan distribution from 1980–2003. As a by-product, working at smaller time intervals is more convincing with respect to the usual iid assumption underlying EVT methods. Moreover, it allows for an adequate treatment of the present truncation by incorporating survivors of each time interval as right-censored observations. To take care of this issues in an adequate way we employ methods from *censored* EVT following Einmahl, Fils-Villetard and Guillou (2008), see also Gomes and Neves (2011); Worms and Worms (2014) or Gomes and Guillou (2015) for other censored EVT approaches.

To our knowledge, this is the first paper on estimating the maximum lifespan in a population that

- employs methods from censored EVT and
- includes old-aged survivors into their analysis
- while adequately dealing with the underlying censoring issue.

The remainder of this paper is organized as follows: Section 2 gives a detailed description of two employed databases and discuss their structure. Moreover, in this section, we describe how we construct a combined data set which exploits the advantages of both databases and we make a quantitative analysis of this data set. Thereafter, in Section 3 we introduce and apply classical EVT methods to analyze the IDL data alone and illustrate our findings. As this leads to results which are neither statistically significant nor rigorous, we analyze the combined data set by means of censored EVT in Section 4. The results are illustrated and discussed in Section 5 and the paper closes with some conclusions and an outlook in

Section 6.

2. DATA DESCRIPTION

In this section we cover different aspects of the data used for our research. Unfortunately, there is no publicly available database which includes complete mortality and survival data for a sufficiently long history while providing the complete age range for all covered populations. Nevertheless, we admit that there are databases giving detailed information on survival as well as the time and age at death. But often these databases are not publicly available and only cover single populations. In particular, this complicates the comparison of results. We therefore decided to analyze publicly available data. In what follows, we first describe some characteristics of the databases underlying our investigations. We then describe how to exploit their advantages, and finally make a quantitative analysis for the estimation of the deaths curve's right endpoint.

2.1 Characteristics of the HMD and IDL Databases

Our research is based on data from two different databases: the International Database on Longevity IDL (2017) and the Human Mortality Database HMD (2015). As outlined in the introduction we focus on US females to have a rather homogeneous population and a sufficiently large sample size.

The **IDL** contains data for 15 industrialized countries across the world. Each of its records relates to one individual dying after its 110th birthday, where the individual's age at death is given exactly in days from their birthday and it is supposed to be free of age ascertainment bias Maier et al. (2010). For the US the IDL provides the age and time of death of 309 females dying at the age of 110 or beyond between 1980 and 2003. The oldest US female recorded in the IDL reached an age of 43560 days (i.e. 119 years and 97 days). These figures illustrate a major issue of the IDL: Its exactness and lack of ascertainment bias inevitably lead to a reduction in the number of observations. Thus, the data may underestimate the count of supercentenarians dying in the years between 1980 and 2003. Maier et al. (2010) explain the process of validation of the IDL data and concludes that the IDL fits the profile of the US female population well in general though it probably underestimates the absolute number of people dying in this age range. A further disadvantage of the IDL is the lack of survival data: We only know the time and age of death of some very old US females but there might be some individuals still alive in similar or even higher ages. Moreover, the IDL stopped collecting data after 2003.

In contrast to the IDL the **HMD** covers mortality and population size (i.e. survival) data of 38 different countries, where death counts are obtained in up to three dimensions: age at death, calendar year of death, and calendar year of birth. The HMD edits the input data to obtain a homogeneous structure over calendar years per population but also across populations. To reduce complexity we pass on the information of the year of birth and only employ the single-year and single-age death counts and population size (survivor) data. Unfortunately the HMD only supplies age and time discrete data, i.e. we only know how many members of a population die during the (integer) year t and in the (integer) age x, but we neither know the exact date of death, nor their exact age at death. We therefore follow Wilmoth et al. (2017) and assume the date of death as well as the date of birth during a calendar year to be uniformly distributed. In this way we can redistribute the death counts per calendar year and thus work with an exact time and age at death. Similarly, we can derive an exact age structure of the population at the end of a calendar year.

For our analysis, we extracted HMD population size data for US females for the time horizon from 1980 – 2003. In sum it includes almost 26 million deaths over all ages whereof 1853 females are deterministically right-censored at 110, i.e. died at the age of 110 or beyond. This means that the HMD only contains single age mortality data up to the age of 109 and a cumulative entry *110+* thereafter. This constricts the

estimation of the maximum attainable age. For this reason, we additionally merge the IDL into the HMD for the analysis. This combination exploits the quantity of the HMD and the quality of the IDL data.

2.2 Combining the Data – the CDS

The input data of the HMD for US females builds on several official population and health statistics of the United States. In this sense, the data is likely to be complete with regard to the US population in general. Thus, it is reasonable to assume the IDL population being a sub-population of the HMD population. With this assumption we combine them as follows: First, as we are interested in old-age mortality, we only include the HMD data (death counts and survivors) beyond the age of 97. The inclusion of survivors thereby adds right-censored observations to our time interval (1980–2003) of interest. Second, for each calendar year, we randomly delete as many deaths counts from the HMD deaths counts' entry 110+ as we observe deaths in this age range in the IDL. Finally, we derive the number of survivors of each calendar year from the data of the IDL. This is possible as we know the exact age and time of death of each individual recorded in the IDL. We randomly delete as many survivors of the HMD as we have survivors implicitly given by the IDL for each calendar year and each age in the age span beyond 97. This merges the IDL into the HMD and avoids double counts in the resulting data set which we denote as **CDS** (combined data set). Later on, we analyze the CDS with censored EVT techniques introduced in Einmahl, Fils-Villetard and Guillou (2008) which require the assumption of iid observations. However, since the structure of mortality typically changes over time Börger, Genz and Ruß (2017), this presumption may not be reasonable for the complete time horizon of the CDS of 24 calendar years. Since evaluating on an annual basis would lead to rather small sample sizes, particularly with respect to the proportion of the IDL data in the CDS per period, we found a compromise: We run our analysis on time windows of nine adjacent calendar years. In each step, these windows are shifted by one-year, i.e. we start with the window between 1980 and 1988, proceed with the window between 1981 and 1989, and so on, ending with the window between 1995 and 2003. We treat the records for fixed windows as realizations from independent and identically distributed random variables that may be subject to independent right-censoring. The assumption on identically distributed random variables on each window appears to be more convincing compared to the entire period 1980–2003, see the Supplement for density plots underpinning this assumption. Moreover, analyzing moving windows enables us to evaluate the evolution of the deaths curve's right endpoint over time.

2.3 Quantitative Analysis of the data bases

To get a picture of the data bases we oppose the values of the IDL and the HMD (for the records beyond 97 years; without double counts) in Table 1. For each window the absolute numbers of death counts and censored observations are shown. For the HMD the latter is split into censored individual surviving the time window (i.e. the subject died at a later time point) and censoring due to HMD-specific missing information for ages beyond 110 (110+ entries) which is responsible for less than 1% of the censored observations. As the IDL collects the explicit age at death, observations from the IDL are only censored due to survival of the specific time window. This quantity decreases in the last four windows since the IDL only covers data up to the year 2003. As a small caveat, we note that this issue may lead to a slight under-estimation of the right endpoint in the last four windows. In addition to the absolute death counts in the IDL also its portion in the number of death counts in the age range beyond 110 from the CDS are shown. This portion ranges between 10% and 20%, which means that for each window at least one out of ten US female supercentenarians dying is an uncensored observation in the CDS. The portion of the IDL data in all deaths of the CDS per window is much smaller (between about 0.03% and 0.05%)

	HMD			IDL		
Window	Observed death counts	Censored observations due to		Observed death counts	*% of all CDS deaths in the range 110+*	Censored due to survival
		110+ entries	survivors			
1980-88	184752	624	85577	72	10.34	14
1981-89	198520	616	90676	79	11.37	14
1982-90	212466	624	95641	87	12.24	18
1983-91	227203	614	101020	84	12.03	32
1984-92	240896	617	105983	102	14.19	24
1985-93	257341	656	108517	118	15.25	23
1986-94	272599	657	110999	129	16.41	22
1987-95	287816	666	112084	135	16.85	28
1988-96	301362	675	115006	145	17.68	24
1989-97	314376	706	118120	152	17.72	30
1990-98	327949	732	119849	157	17.66	36
1991-99	342915	729	121968	168	18.73	45
1992-00	356859	752	123306	181	19.4	27
1993-01	370065	760	124548	173	18.54	11
1994-02	380731	751	125481	159	17.47	3
1995-03	391504	746	127637	142	15.99	0

TABLE 1

Shown are the window-specific sample sizes of the censored and uncensored observations in HMD and IDL together with the portion of the IDL in the higher age entries (older than 110 years) of the CDS.

but this is mainly due to the high number of observations from the HMD in the age range between 97 and 109. Finally, the corresponding CDS death and censoring counts are obtained by summing up the corresponding columns, respectively. It is apparent that these values reveal an increasing trend over the windows. This is in line with the findings of, e.g. Robine and Paccaud (2005), who found that the number of old Swiss people increases over time.

In what follows we first study the IDL data with classical EVT techniques for the whole time interval 1980–2003 as the number of IDL counts within each window are rather small (Table 1). We then use the results as motivation for a more sophisticated analysis of the CDS with censored EVT methods for the moving windows.

3. CLASSICAL EVT ANALYSIS OF THE IDL

Consider the classical EVT framework given by iid random variables $X_i \sim F$ with unknown cumulative distribution function (cdf) F. Here X_i indicates the age at death of subject i and we are particularly interested in the *right endpoint* of the cdf F, i.e. $x_F := \sup\{x : F(x) < 1\} \leq \infty$. Since we do not want to presume the existence of a limit to human lifespan, we should also allow for an infinite lifetime, i.e. for $x_F = \infty$. For estimating x_F there exist two well-established EVT approaches, each employing a different class of distributions: the *generalized extreme value distributions (GEV)* and the *generalized Pareto distributions (GPD)*. In what follows, we briefly describe these two approaches and refer to de Haan and Ferreira (2006) for more details and proofs.

The GEV is based on the Fisher-Tippet Theorem (see Fisher and Tippett (1928)). It describes the class of potential limit distributions G_γ of the standardized maximum $(\max_{1 \leq i \leq n}\{X_i\} - b_n)/a_n$ for suitable

sequences $a_n > 0$ and b_n. The class of GEV $G_\gamma(ax + b)$, is given by

$$(3.1) \qquad G_\gamma(x) = \begin{cases} \exp\left(-(1 + \gamma x)^{-\frac{1}{\gamma}}\right), & 1 + \gamma x > 0 \quad \gamma \neq 0 \\ \exp\left(-\exp(-x)\right), & x \in \mathbb{R} \qquad \gamma = 0, \end{cases}$$

where a and b are scaling and location parameters which can be estimated from the current dataset. While this concept is rather general, a GPD approach seems to be more suitable with respect to the left truncated IDL data since it considers values above a given threshold t Balkema and de Haan (1974), say $t = 110$ in case of the IDL. To this end, we are interested in the distribution $F_t(x) = \mathbb{P}(X_1 - t \leq x | X_1 > t)$. For example de Haan and Ferreira (2006) show that the limit equation $\lim_{t \to x_F} \sup_{x_F - t < x < 0} |F_t(x) - H_{\gamma,\sigma}(x)| = 0$ can only be fulfilled for distributions from the GPD given by

$$(3.2) \qquad H_{\gamma,\sigma}(x) = \begin{cases} 1 - \left(1 + \frac{\gamma x}{\sigma}\right)^{\frac{-1}{\gamma}} & \text{for } \gamma \neq 0 \\ 1 - \exp(-\frac{x}{\sigma}) & \text{for } \gamma = 0, \end{cases}$$

where σ is again a scaling parameter.

In both cases, the *extreme value index (EVI)* γ is the key parameter that characterizes the tail of the limit distribution. In particular, the value of γ separates two different cases: If $\gamma < 0$ holds (case (i)), the considered distribution has a finite right endpoint which corresponds to a finite maximal age. Contrary, if $\gamma > 0$ holds (case (ii)), the right endpoint is infinite and US females may have the potential to live forever. For the decision between the cases (i) and (ii) and for the estimation of x_F we need adequate estimators for γ. The recent literature discusses several different estimators, some of which are valid only for EVIs in certain intervals (see e.g. Pickands (1975) or Dekkers, Einmahl and de Haan (1989) for early work on this topic and de Haan and Ferreira (2006) and Resnick (2013) for reviews). However, we particularly allow for positive values of the EVI. This basically leaves us with two choices of estimators for γ: the *moment estimator (Mom)* for the GEV approach and the *maximum likelihood estimator (MLE)* for the GPD approach. The latter is defined as that value of γ which maximizes the likelihood during a fit of the GPD model to the data while the moment estimator is defined as

$$(3.3) \qquad \hat{\gamma}_n^{Mom}(k) := M_n^{(1)}(k) + 1 - \frac{1}{2}\left(1 - \left(M_n^{(1)}(k)\right)^2 / M_n^{(2)}(k)\right)^{-1},$$

where $M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\log(X_{n-i+1,n}) - \log(X_{n-k,n})\right)^j$ and $X_{i,n}$ denotes the i-th order statistic. Other estimators such as Pickands or Hill estimators have not been considered due to restricted EVI values (Hill) or less efficiency (Pickands), see e.g. the discussion in Section 3 of de Haan and Ferreira (2006). In what follows, $\hat{\gamma}_n^{(\cdot)}(k)$ denotes the EVI estimator, where k is the number of upper order statistics used and $(\cdot)$ indicates the short-form of the estimator (i.e. MLE or Mom, respectively). The choice of k as a function of the number of observations n is crucial for a good estimate. Since both estimators are consistent for γ and even asymptotically normal if $k \to \infty$ while $\frac{k}{n} \to 0$, we should not choose k too small or too large. To this end, e.g. de Haan, Mercadier and Zhou (2016) describe the following observation: the smaller the values of k, the higher the variance of the EVI estimator while for increasing k the bias of the EVI estimator gets larger. Finding a suitable k between these extremes is called the *bias-variance trade-off*. To get a first idea of the choice of k, we perform a graphical inspection: Figure 1 shows both the moment and maximum likelihood estimates for the IDL depending on k and the point-wise 95% confidence intervals (CI) for both estimates (where we employed the methods of Aarsen and
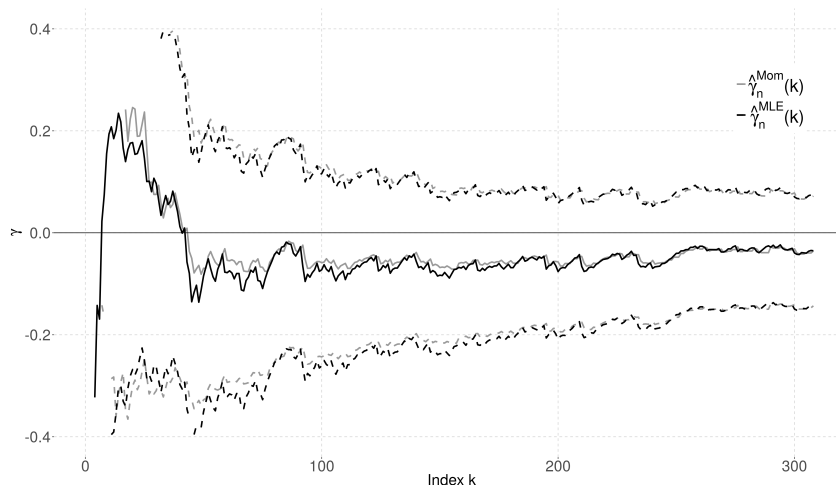
FIG 1. *Point estimates $k \mapsto \hat{\gamma}_n^{(\cdot)}(k)$ of the EVI based on the MLE and Mom approach as a function of k together with point-wise 95% CI for the IDL.*

de Haan (1994) for the construction of the Mom-based CI and carried them over to the MLE-based CI for a better comparison).

From Figure 1 we can see that both EVI estimates are positive for $k \leq 46$ and negative for $k > 46$. Interpreting values of $k \leq 46$ as too small this suggests a negative EVI-estimate corresponding to a finite maximal attainable age. However, the CI or more decisive the one-sided CI (not shown here) portend that the null hypothesis $H_0 : \gamma > 0$ of a positive EVI cannot be rejected at the significance level of 5%. In particular, the asymptotic one-sided z-test for H_0 based on the MLE (Mom) gives an approximate p-value of 0.2349 (0.2183). Thus, the hypothesis of a potentially infinite lifetime cannot be rejected based on data from the IDL and the methods described above. If we ignore this statistical insignificance for now and only take into account that the EVI-estimates are negative for $k > 46$ so we can nevertheless estimate a maximum attainable age. To this end, we have to find a suitable k. A general recommendation, or rule-of-thumb, is to choose k in a region where these estimates stabilize. Incorporating the bias-variance trade-off as well as the asymptotic framework, this points us to a choice of k between 100 and 200. The minimum, maximum and median of the MLE and Mom estimates for γ with k between 100 and 200, the corresponding value of k, and the 95% confidence intervals are given in Table 2.

		$\hat{\gamma}_n^{(\cdot)}(k)$	(95% CI for γ)	k	$\hat{x}_F^{(\cdot)}(k)$	(95% CI for x_F)
min	MLE	-0.0711	(-0.2173, 0.0751)	155	132.30	(91.62, 172.98)
	Mom	-0.0942	(-0.2773, 0.0888)	106	127.49	(99.53, 155.45)
median	MLE	-0.0572	(-0.2205, 0.1061)	128	136.71	(68.37, 205.05)
	Mom	-0.0663	(-0.2247, 0.0921)	142	133.11	(84.14, 182.08)
max	MLE	-0.0385	(-0.1978, 0.1207)	140	147.60	(02.92, 292.28)
	Mom	-0.0418	(-0.2032, 0.1196)	139	144.49	(21.32, 267.66)

TABLE 2

MLE- and Mom-estimates of the EVI (γ) and the right-endpoint (x_F) for the IDL data together with 95% confidence intervals for different choices of k.

As already seen in Figure 1 the CIs are rather wide and each contains the value of 0, i.e. a potential

infinite lifetime cannot be rejected at level 5%. This may be explained by the small sample size and a potentially negative γ that is quite close to zero (i.e. does not lie far in the alternative). However, each of the estimated EVI values corresponds to a different GPD (MLE case) or GEV (Mom case) and can be used to estimate the right-endpoint of the underlying distribution.

For illustration, exemplary QQ-plots for both median values are given in Figure 2. There the empirical quantiles of the IDL data are plotted against simulated quantiles of the respective distribution functions and the solid lines denote the corresponding regression line, respectively. Both plots show an acceptable fit with a considerable advantage for the GPD model (left panel).

This additionally reassures the potential applicability of both EVI approaches. We also investigated whether the underlying distribution function lies in the maximum domain of attraction of G_γ for the above choices of $\gamma < 0$. By means of the test of Dietrich, de Haan and Huesler (2002) with the critical value proposed in Huesler and Li (2006) we could not reject this.
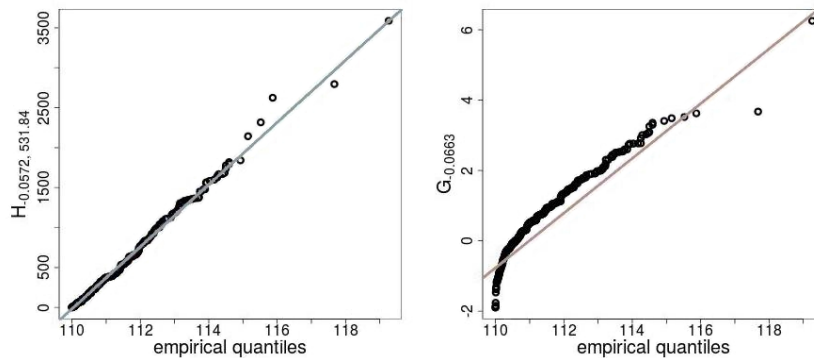

FIG 2. *QQ-plots of empirical quantiles over simulated quantiles; left panel: pareto distribution with $\gamma = -0.0572$; right panel: generalized extreme value distribution with $\gamma = -0.0663$. The solid lines visualize the regression line for the plots.*

In the next step, we determine the right endpoint of the corresponding models. As explained around Equation (3.1) above, we need to estimate the scaling and location parameters a and b, in order to find a GEV. To this end we apply the method described in Section 3 of de Haan and Ferreira (2006) and finally obtain estimates for the maximum lifespan $\hat{x}_F^{Mom}(k)$ (GEV) and $\hat{x}_F^{MLE}(k)$ (GPD), respectively. The estimators again depend on the tuning parameter k and are consistent and asymptotically normal under similar regularity assumptions. With this, we can construct 95% CIs. The last two columns of Table 2 show the results: Given are the minimum, median, and maximum of the MLE and Mom endpoint estimates with the corresponding asymptotic 95% CIs and the value of k referring to the endpoint estimates $\hat{x}_F^{Mom}(k)$ and $\hat{x}_F^{MLE}(k)$, respectively. The obtained estimates for the maximum lifespan fluctuate between 127.49 (GEV with minimum EVI Mom-estimate) and 147.6 years (GPD with maximum EVI MLE-estimate). This suggests, that the sample size is probably not sufficient for more precise results. Moreover, the confidence intervals are again (too) widespread. In particular, for the larger EVI estimates the asymptotic behavior of the underlying variance estimator (which is of order $O(1/\gamma^4)$ for $\gamma \to 0$) cannot be countervailed by the sample size which leads to the given CIs with poor relevance.

4. A CENSORED EVT ANALYSIS OF THE CDS

The results of the previous section are neither statistically significant nor rigorous. As a consequence, we need to significantly increase the sample size of the dataset. Moreover, the neglect of present censor-

ing as well as the underlying iid assumption for the whole time span may lead to implausible estimates. To take care of these issues we analyze the CDS over 16 moving time windows as described in Section 2 by means of censored EVT. To this end, Einmahl, Fils-Villetard and Guillou (2008) provide several theoretical results with regard to the statistical analysis of randomly right-censored extreme value data. In their framework, the age at death of the i-th individual is modeled via iid random variables X_i with distribution function F_X. These random variables may be independently right censored by censoring random variables $C_i \overset{iid}{\sim} F_C$, where F_C denotes their distribution function. Consequently, the observations in the dataset are given by $Z_i = \min(X_i, C_i)$ and a censoring indicator is given by $\delta_i = \mathbf{1}\{Z_i = X_i\}$. Let $\gamma_{(\cdot)}$ and $x_{F_{(\cdot)}}$ denote the EVI and the endpoint of the distribution function of $X, C,$ and Z, respectively.

To answer the target questions on the maximum lifespan we need to determine γ_X and x_{F_X} based on the available data. The classical EVT approaches from the previous subsection, however, would only lead to reasonable estimates for γ_Z and x_{F_Z} (when only using the observations Z_i). On the other hand, a complete case analysis (when deleting all censored observations) would generally result in estimates that are strongly biased. To address censoring adequately, Einmahl, Fils-Villetard and Guillou (2008) introduce estimators for γ_X and x_{F_X} which are valid in the following cases:

$$(4.1) \qquad \begin{cases} \text{case 1: } \gamma_X > 0, \ \gamma_C > 0 \\ \text{case 2: } \gamma_X < 0, \ \gamma_C < 0, \ x_{F_X} = x_{F_C} \\ \text{case 3: } \gamma_X = \gamma_C = 0, \ x_{F_X} = x_{F_C} = \infty. \end{cases}$$

With regard to the estimation of the maximal attainable age at death only the second case corresponds to a finite right-endpoint. For the sake of generality we nevertheless also allow for non-negative EVIs γ_X and γ_C.

In what follows, we consider intervals of nine adjacent calendar years for the time span 1980–2003. Starting with 1980–1988 these windows are shifted by steps of one calendar year, see Table 1. As the CDS records of death and censoring counts for the 16 time windows (see Table 1) do not refuse the first two cases of Assumption (4.1) at first sight we presume that the methods of Einmahl, Fils-Villetard and Guillou (2008) are applicable to estimate the right endpoint of the deaths curve for each window. The windows contain between about 300,000 and 500,000 of (partially censored) observations. Since this figure is still too large for most algorithms with respect to the computing time (even when using modern computing clusters), we follow a sub-sampling approach: Let $\mathcal{Z}_n := \{Z_1, ... Z_n\}$ denote the CDS sample for one representative window. We then randomly draw m times from $\mathcal{Z}_n$ without replacement to construct a sub-sample of size m, where $m < n$. Independently repeating this N times we obtain sub-samples $\mathcal{Z}_m^i := \{Z_1^i, ..., Z_m^i\}$, $1 \leq i \leq N$. We then estimate the EVIs (of the given window) separately for each sub-sample. Following Politis, Romano and Wolf (1999) we should choose m from a theoretical perspective such that $\frac{m}{n} \to 0$ as $m \to \infty$. Moreover, the size of the sub-samples m should be sufficiently large, such that each sub-sample includes IDL data with a sufficiently high probability. Taking these issues as well as the computation time into account we calculated estimates for $m = 7500$ and $N \in \{1000, 5000\}$. For the choice of k we employ a cross-validation type procedure that is related to the Leave-One-Out method. This allows for a more objective choice of the tuning parameter k, at least compared to the rule of thumb approach from Section 3. To this end, we draw one more sub-sample $\mathcal{Z}_m^0$ from $\mathcal{Z}_n$ of length m that serves as a test sample. All other sub-samples $\mathcal{Z}_m^i$, $1 \leq i \leq N$, serve as training samples. For each of these sub-samples $\mathcal{Z}_m^i$, $0 \leq i \leq N$, we estimate the EVI $\hat{\gamma}_{X,i}^{(\cdot)}(k_i)$ and thereby vary the number of upper order statistics $1 \leq k_i \leq m$. To this end, we use the methods described by Einmahl, Fils-Villetard and Guillou (2008).
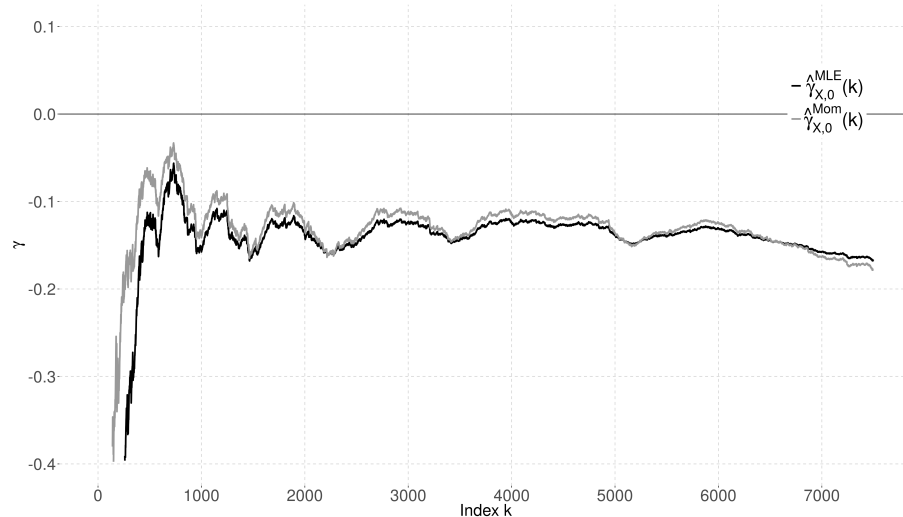
FIG 3. *Different EVI estimates as a function of the number of upper order statistics k_0 in case of the trainings dataset of the first window between 1980 and 1988 from the CDS.*

Figure 3 shows different EVI estimates $k_0 \mapsto \hat{\gamma}_{X,0}^{(\cdot)}(k_0)$ for the test sample of the first interval between 1980 and 1988 for $N = 5000$. Since there are several stable plateaus visible, it is not clear, where the upper order statistic should be chosen to get a good bias-variance tradeoff. We therefore compare the EVI estimates for the test set $k_0 \mapsto \hat{\gamma}_{X,0}^{(\cdot)}(k_0)$ with the corresponding EVI estimates from each trainings set $k_i \mapsto \hat{\gamma}_{X,i}^{(\cdot)}(k_i)$. For each of these comparisons we obtain the estimator

$$(4.2) \qquad \hat{k}_{0,i} := \operatorname*{argmin}_{1 \leq k_0 \leq m} \sum_{k_i=1}^{m} k_i^{5/4} \left(\hat{\gamma}_{X,0}^{(\cdot)}(k_0) - \hat{\gamma}_{X,i}^{(\cdot)}(k_i) \right)^2, \quad 1 \leq i \leq N.$$

This corresponds to the EVI estimate from the test set $\mathcal{Z}_m^0$ that has the smallest penalized quadratic distance from all EVI estimates of the ith trainings set $\mathcal{Z}_m^i$. The choice of the penalty $k_i^{5/4}$ reduces the influence of the bias since it supports intermediate values and penalizes larger values of k (for which the bias is supposed to become too large). Now we can derive an estimator for the EVI by taking the mean of all $\hat{\gamma}_{X,0}^{(\cdot)}(\hat{k}_{0,i})$, $1 \leq i \leq N$, obtained from Equation (4.2):

$$\hat{\gamma}_{X,0}^{(\cdot)} = N^{-1} \sum_{i=1}^{N} \hat{\gamma}_{X,0}^{(\cdot)}(\hat{k}_{0,i}).$$

To additionally state corresponding 95% confidence intervals of the form $(\hat{\gamma}_L, \hat{\gamma}_U)$ we use the asymptotic normality of the EVI estimators given in Einmahl, Fils-Villetard and Guillou (2008). Proceeding in this way for all time windows we obtain the MLE- and Mom-based EVI estimates with corresponding 95%-confidence intervals stated in Table 3 (for $N = 5000$). Moreover, the limiting distribution of the EVI-estimators directly yields estimates for the right endpoint given any estimate for the EVI. However, no mathematically sound confidence intervals for x_F have been developed so far. We therefore seize on the former suggestion that every EVI-estimate yields a right endpoint estimate and proceed as follows: For

window	Mom $\hat{\gamma}_n^{(\cdot)}(k)$	(95% CI for γ)	MLE $\hat{\gamma}_n^{(\cdot)}(k)$	(95% CI for γ)
1980-88	-0.14266	(-0.14307,-0.14226)	-0.13500	(-0.13533,-0.13467)
1981-89	-0.14225	(-0.14266,-0.14183)	-0.13827	(-0.14075,-0.13579)
1982-90	-0.15943	(-0.15973,-0.15913)	-0.14813	(-0.14966,-0.14660)
1983-91	-0.15920	(-0.15954,-0.15886)	-0.13979	(-0.14017,-0.13941)
1984-92	-0.13900	(-0.14761,-0.13038)	-0.12407	(-0.12458,-0.12356)
1985-93	-0.13899	(-0.13956,-0.13843)	-0.12448	(-0.12513,-0.12382)
1986-94	-0.14553	(-0.14596,-0.14511)	-0.12650	(-0.12707,-0.12594)
1987-95	-0.14616	(-0.14663,-0.14569)	-0.12669	(-0.12791,-0.12547)
1988-96	-0.15376	(-0.15430,-0.15322)	-0.15025	(-0.15080,-0.14970)
1989-97	-0.15359	(-0.15585,-0.15134)	-0.15062	(-0.15102,-0.15022)
1990-98	-0.14601	(-0.14819,-0.14383)	-0.14717	(-0.14757,-0.14677)
1991-99	-0.14524	(-0.14564,-0.14484)	-0.14676	(-0.14717,-0.14635)
1992-00	-0.15112	(-0.15162,-0.15061)	-0.13735	(-0.13789,-0.13681)
1993-01	-0.13187	(-0.13607,-0.12768)	-0.13764	(-0.14035,-0.13492)
1994-02	-0.15253	(-0.15318,-0.15188)	-0.15808	(-0.15859,-0.15757)
1995-03	-0.14023	(-0.14231,-0.13815)	-0.16646	(-0.16818,-0.16475)

TABLE 3

MLE- and Mom-estimates of the EVI (γ), upper order statistics k, and 95% confidence intervals for the CDS for $N = 5000$.

the upper and lower limits $\hat{\gamma}_L^{(\cdot)}$ and $\hat{\gamma}_U^{(\cdot)}$ of the 95 % empirical confidence interval of $\gamma_{X,0}$ we calculate corresponding endpoint estimates, say $\hat{x}_{F_X,L}^{(\cdot)}$ and $\hat{x}_{F_X,U}^{(\cdot)}$, to obtain an (at least descriptive) confidence interval $(\hat{x}_{F_X,L}^{(\cdot)}, \hat{x}_{F_X,U}^{(\cdot)})$ for x_F. The results for $N = 5000$ are shown in Table 4 and discussed below.

5. DISCUSSION OF THE CDS RESULTS

In this section, we discuss our results and their plausibility. We start by describing the results on the EVI given in Table 3 (the results for a subsample size of $N = 1000$ were similar and thus not shown here). First, it is apparent that all EVI estimates are below zero for every window. Moreover, even the upper bounds of their 95% confidence intervals are below zero. Since this even holds after multiplicity adjustment (results not shown) we can argue for a significantly negative EVI ($\gamma_{F_X} < 0$). This particularly indicates the existence of a limit to the lifespan of US females in every window; at least in this retrospective view. A similar investigation for the censored observations alone also indicated a significant negative $\gamma_{F_C} < 0$ so that Case 2 of Assumption (4.1) appears to be plausible. Accordingly, we are able to estimate the right endpoint of the deaths curve as described in the former section.

Table 4 shows these estimates together with their confidence intervals for both the maximum likelihood and the moment estimator in case of $N = 5000$ subsamples (the results for $N = 1000$ were again similar and thus not shown). As we can see from this table, both right endpoint estimates range between 120.278 and 126.610 over all time windows. Further, we cannot detect any major trends neither in the results for the maximum likelihood estimator nor in those of the moment estimator.

Moreover, the confidence intervals range between 117.280 and 130.830 with a maximum interval length of 9.283 years (MLE estimation in the window between 1993 and 2001). For the interpretation of these results, we shall recall from Section 2 that the IDL stopped collecting data after 2003. Thus, the later time spans contain less IDL survival data which may lead to a slight underestimation. Anyhow, the observed rather constant behavior of the deaths curve's right endpoint is in line with recent findings of Einmahl, Einmahl and de Haan (2017) who also did not find 'indications of trends in these upper limits over the last 30 years' for Dutch residents based on classical (non-censored) EVT.

window	Mom $\hat{x}_{F_X}$	(95% approx. CI for $\hat{x}_{F_X}$)	MLE $\hat{x}_{F_X}$	(95% approx. CI for $\hat{x}_{F_X}$)
1980-88	122.802	(120.240, 127.740)	125.409	(122.072, 129.602)
1981-89	122.898	(120.304, 127.801)	124.839	(122.642, 129.510)
1982-90	120.781	(117.280, 123.487)	122.880	(121.530, 124.833)
1983-91	120.732	(117.432, 123.566)	124.099	(123.995, 129.278)
1984-92	123.001	(120.793, 124.673)	126.510	(123.700, 130.830)
1985-93	122.871	(120.881, 125.343)	126.610	(123.661, 130.671)
1986-94	121.832	(118.322, 126.409)	126.310	(123.391, 127.733)
1987-95	121.661	(118.465, 126.420)	126.278	(123.533, 127.622)
1988-96	120.562	(119.143, 122.889)	121.897	(119.420, 125.152)
1989-97	120.797	(119.508, 122.043)	122.206	(119.408, 125.067)
1990-98	122.107	(119.987, 123.567)	122.380	(119.214, 124.939)
1991-99	122.186	(118.805, 126.413)	122.354	(119.232, 124.976)
1992-00	121.107	(118.078, 124.314)	123.482	(119.541, 128.454)
1993-01	123.489	(120.877, 125.906)	124.340	(119.267, 128.550)
1994-02	120.278	(118.502, 122.782)	120.704	(117.576, 124.642)
1995-03	122.478	(120.417, 124.407)	120.946	(119.382, 122.930)

TABLE 4

MLE- and Mom-estimates of the right endpoint (x_{F_X}) and 95% approximative confidence intervals for the CDS for
N = 5000.

As explained in Section 2.1, the oldest US female recorded in the IDL reached an age of 119 years and 97 days and she died in the year 1999. Though the lower bound of the confidence intervals for some windows is lower than 119 years, the estimates for the right endpoint are always larger. Since the HMD is regarded to be complete with respect the US population our results seem to be plausible.

Advantages of the CDS approach. We finally indicate the advantages of working with the CDS. First, utilizing the HMD alone did not lead to reliable estimates (results shown in the Supplement). In particular, in the analysis of the HMD alone we obtained EVT estimates of the right endpoint that are below the age of 119 years. However, from the IDL we know at least one proven case of a US female surviving the age of 119 (see Section 2.1). On the other hand, we could infer more information from investigating the CDS than building on the IDL alone: The evaluation of the deaths curve's right endpoints on moving windows was not possible on the IDL data alone due to too small sample sizes per window (see Table 1). We, therefore, performed its EVT analysis for the whole time span 1980–2003 in Section 3 for which the stated EVI confidence intervals even include the value of zero (see Table 2). However, the existence of a finite lifespan could be accepted by means of the significantly larger CDS. Since the latter database is larger (in particular it contains the IDL) and incorporates survival data, the corresponding results from Table 3 are more reliable than the results for the IDL alone.

Setting the numbers into context. Comparing our results to previous findings in recent literature is difficult because the basics (i.e. data and methods) differ considerably. In particular, the present statistical analysis is the first that accounts for censoring of the underlying data accordingly. Nevertheless, to at least associate the numbers with the context of a maximum human lifespan we summarize some of the most recent results: Utilizing HMD data alone, de Beer, Bardoutsos and Janssen (2017) find that the maximum human lifespan may increase to 125 years. Li, Ng and Chan (2011) estimate the maximum lifespan for Australians to 112.2 and to 109.43 for New Zealand females and males, respectively. To this end, they use HMD data. Bravo and Corte-Real (2012) also use HMD data and obtain similarly low estimates at 112.77

and 111.78 for Spanish and Portuguese females, respectively. However, at least for Spain and Australia, we can find IDL records for females dying at higher ages (both after their 114th birthday). Einmahl, Einmahl and de Haan (2017) analyzed about 285,000 death counts of male and female Dutch residents between 1986–2015 (who at least reached an age of 92) for single years (which makes the sample size for each estimation even smaller) and found an average estimated upper endpoint at 115.7 years. Not only in relation to our results, these figures seem to be exceptionally low. For instance, even more than two decades ago, Aarsen and de Haan (1994) infer a 95% confidence interval for the Netherlands between 113 and 124 years (based on about 20,000 combined observations from Dutch men and women). More recently, Weon and Je (2009) obtain maximum age estimates around 126 years for Swedish females and Hanayama and Sibuya (2015) find estimates for the maximum lifespan of Japanese at 123 years. For both analyses, HMD data is used. In contrast, Gbari et al. (2017) use a country-specific database for Belgian citizens which includes 36,616 death counts for females dying in the age range 95+ after 1981. Based on this data they quantify the maximum attainable age of Belgian females between 120.3 and 122.73 years. Our estimates for US females are in the upper range of all these recent results. These disparities in the results may be caused by different sources of data, differences between the populations under study, or the diversity of the employed methods. In particular, the sample size of the data plays an important role in the quality of the results. In this regard, the data we use outperforms the data from many previous researches (see Table 1). Moreover, for example, the case of Jeanne Calment Robine and Allard (1998) indicates that it is possible to survive to ages beyond 120 years. In this spirit we consider our results to be reasonable.

6. CONCLUSION AND OUTLOOK

The present paper investigates the question of quantifying the maximum human lifespan for US females. To this end, we used two different sources of data: the International Database on Longevity (IDL) and the Human Mortality Database (HMD). We discussed the specific characteristics of each database and additionally exploited their advantages by constructing a comprehensive combined data set, the CDS. These data include both death counts and censored survival data. We then started our investigation of inferring the existence of a possibly finite lifespan by means of extreme value theory (EVT). Starting with a classical EVT analysis of the IDL alone, the hypothesis of an infinite lifespan could not be rejected. This is presumably due to its rather small sample size of 309 for the analyzed time span 1980–2003. Moreover, the obtained point estimates for the maximum attainable age fluctuated considerably and are thus not very meaningful. To obtain more reliable results we subsequently analyzed the CDS which we even arranged into intersecting windows to study the evaluation of the lifespan over time. Due to the involved censoring, we employed more convenient censored EVT methods from Einmahl, Fils-Villetard and Guillou (2008) which we equipped with a computationally efficient algorithm based on sub-sampling and cross-validation.

Finally, we illustrated and discussed our results. We found significant evidence for a finite lifespan in the combined data and obtained reliable point estimates for the maximum attainable age of US females for each window. These estimates vary between 120 and 127 years which appears to be reasonable compared to existing results in the literature (particularly, the oldest US female recorded in the IDL reached an age of more than 119 years).

In addition, we like to note, that independent of our analysis other authors like Einmahl, Einmahl and de Haan (2017) and Rootzén and Zholud (2017) have been motivated to apply (uncensored) EVT methods on different datasets after reading the nature paper by Dong, Milholland and Vijg (2016). However, judging from our findings, we recommend the application of censored EVT methods to other popula-

tions for future research; employing the idea of the current CDS: Whenever there is a comprehensive data set which is right censored (as e.g. in the HMD) and a comparatively small but exact dataset (as e.g. given by the IDL) on the same population this method can be used and benefits from much larger and reliable data. In addition, we like to stress that for enhanced analysis additional research is needed from a methodological point of view to construct asymptotically correct confidence limits for the right endpoint, to develop Bayesian EVI estimate versions under censoring and to accordingly treat more complicated censoring issues as, e.g., interval-censored data for the Weibull model.

ACKNOWLEDGEMENTS

REFERENCES

AARSEN, K. and DE HAAN, L. (1994). On the maximal life span of humans. *Mathematical Population Studies* **4** 259-281.

BALKEMA, A. A. and DE HAAN, L. (1974). Residual Life Time at Great Age. *The Annals of Probability* **2** 792-804.

BÖRGER, M., GENZ, M. and RUSS, J. (2017). Extension, Compression, and Beyond - A Unique Classification Framework for Mortality Evolution Patterns. Preprint, ifa-ulm.de.

BRAVO, J. M. and CORTE-REAL, P. (2012). Modeling Longevity Risk using Extreme Value Theory: An Empirical Investigation using Portuguese and Spanish Population Data. Preprint, Evora University.

COHEN, J. E. and OPPENHEIM, J. (2012). Is a limit to the median length of human life imminent? *Genus* **68**.

DE BEER, J., BARDOUTSOS, A. and JANSSEN, F. (2017). Maximum human lifespan may increase to 125 years. *Nature* **546** E16.

DE GREY, A. (2003). The foreseeability of real anti-aging medicine: focusing the debate. *Experimental Gerontology* **38** 927–934.

DE HAAN, L. and FERREIRA, A. (2006). *Extreme value theory: an introduction*. Springer, New York; [Heidelberg].

DE HAAN, L., MERCADIER, C. and ZHOU, C. (2016). Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance and Stochastics* **20** 321-354.

DEKKERS, A. L. M., EINMAHL, J. H. J. and DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics* **17** 1833-1855.

DIETRICH, D., DE HAAN, L. and HUESLER, J. (2002). Testing extreme value conditions. *Extremes* **5** 71-85.

DONG, X., MILHOLLAND, B. and VIJG, J. (2016). Evidence for a limit to human lifespan. *Nature* **538** 257–259.

EINMAHL, J., EINMAHL, J. H. J. and DE HAAN, L. (2017). Limits to human life span through extreme value theory. Preprint, Tilburg University.

EINMAHL, J. H. J., FILS-VILLETARD, A. and GUILLOU, A. (2008). Statistics of extremes under random censoring. *Bernoulli* **14** 207-227.

FALK, M., HÜSLER, J. and REISS, R.-D. (2010). *Laws of small numbers: extremes and rare events*. Springer, New York; [Heidelberg].

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** 180–190.

FRIES, J. F. (1980). Aging, natural death, and the compression of morbidity. *The New England Journal of Medicine* **303** 130-135.

GBARI, S., POULAIN, M., DAL, L. and DENUIT, M. (2017). Extreme Value Analysis of Mortality at the Oldest Ages: A Case Study Based on Individual Ages at Death. *North American Actuarial Journal* **21** 397-416.

GOMES, M. I. and GUILLOU, A. (2015). Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review* **83** 263–292.

GOMES, M. I. and NEVES, M. M. (2011). Estimation of the extreme value index for randomly censored data. *Biometrical Letters* **48** 1–22.

HANAYAMA, N. and SIBUYA, M. (2015). Estimating the Upper Limit of Lifetime Probability Distribution, Based on Data of Japanese Centenarians. *J. Gerontol. A. Biol. Sci. Med. Sci.* **71** 1014-1021.

HMD (2015). Human Mortality Database (http://www.mortality.org).

HUESLER, J. and LI, D. (2006). On testing extreme value conditions. *Extremes* **9** 69-86.

IDL (2017). International Database on Longevity (http://www.supercentenarians.org).

LI, J. S. H., NG, A. C. Y. and CHAN, W. S. (2011). Modeling old-age mortality risk for the populations of Australia and New Zealand: An extreme value approach. *Mathematics and Computers in Simulation* **81** 1325-1333.

MAIER, H., GAMPE, J., JEUNE, B., ROBINE, J.-M. and VAUPEL, J. W., eds. (2010). *Supercentenarians*. Springer, New York; [Heidelberg].

PICKANDS, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics* **3** 119-131.

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer, New York; [Heidelberg].

REISS, R.-D. and THOMAS, M. (2007). *Statistical analysis of extreme values*. Springer, New York; [Heidelberg].

RESNICK, S. I. (2013). *Extreme values, regular variation and point processes*. Springer, New York; [Heidelberg].

ROBINE, J. M. and ALLARD, M. (1998). The oldest human. *Science* **279** 1831-1831.

ROBINE, J. M. and PACCAUD, F. (2005). Nonagenarians and centenarians in Switzerland, 1860-2001: a demographic analysis. *J Epidemiol Community Health* **59** 31-37.

ROOTZÉN, H. and ZHOLUD, D. (2017). Human life is unlimited–but short. *Extremes* **20** 713–728.

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** 162-178.

WEON, B. M. and JE, J. H. (2009). Theoretical estimation of maximum human lifespan. *Biogerontology* **10** 65–71.

WILMOTH, J. R., DEEGAN, L. J., LUNDSTRÖM, H. and HORIUCHI, S. (2000). Increase of Maximum Life-Span in Sweden, 1861-1999. *Science* **289** 2366-2368.

WILMOTH, J. R., ANDREEV, K., JDANOV, D. and GLEI, D. A. (2017). Methods protocol for Human Mortality Database (Version 6).

WORMS, J. and WORMS, R. (2014). New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. *Extremes* **17** 337–358.