Who wants to live forever?
An analysis of the maximum lifespan in the US

Jan Feifel*, Martin Genz† and Markus Pauly*

February 17, 2017

Abstract

Questions related to the existence and specification of a maximum 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 mortality data of US females obtained from the International Database on Longevity as well as the Human Mortality Database. The employment of data on old age mortality typically raises two mature issues: sparse information on the old ages and censored observations. To tackle these challenges we use enhanced methods from extreme value theory and an adequate combination of both databases. 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 subsampling and cross-validation techniques with recent results on censored extreme value theory. Finally, we obtain estimates for the maximum human lifespan of US females in the age range between 125.8 and 128.7 years.

Keywords: Censoring, Extreme value theory, Subsampling, Human Mortality Database, Supercentenarians

* University of Ulm, Institute of Statistics, Germany
† Institute for Finance and Actuarial Sciences and University of Ulm, Institute of Insurance Science, Germany
email: jan.feifel@uni-ulm.de

1 Introduction

Who among us has not yet thought about the possibility to extend the individual lifespan to ages that have never been reached before? May it even be possible to live forever as maintained by some authors (e.g. de Grey [2003] or Kurzweil and Grossman [2005])? Or is there any natural limit to human life which probably is just not yet in sight? Did we actually reach this limit already? The latter claim was just recently supported by explorative data analyses in a Nature paper by Dong et al. [2016]. Therein the authors found *'evidence for a limit to human lifespan'* leading to remarkable press review all over the world and heated discussion among scientists. They explain their claim with *'natural constraints'* which has been asserted by many authors before (e.g. Fries [1980]). 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, however, all of these questions are related to the support of the distribution function of the underlying age distribution at death (the so-called deaths curve) and lead to investigations about its right endpoint. The current paper answers these questions with modern statistical methods, especially from *censored* extreme value theory (EVT). Since it is ad hoc not clear whether the support is finite or not, we also infer this presumption.
For our analysis, we use data from the International Database on Longevity (IDL) and the Human Mortality Database (HMD). To avoid gender effects while guaranteeing an acceptable data quality and quantity we focus on mortality data of US females. The HMD is probably one of the largest publicly available databases in the world. However, it has the particular disadvantage that a lot of observations are right-censored; particularly at age 110. This prohibits a gain of information about the right tail behavior of the deaths curve. However, it is common knowledge that there have been humans which survived the age of 110 (e.g. the famous case of Jeanne Calment, see Robine and Allard [1998]) and thus the structure of the HMD data is not particularly suitable for our purposes. In contrast, the IDL is more informative regarding the death counts of the highest ages: It provides the exact lifespan of so-called supercentenarians, i.e. individuals who at least achieved their 110^{th} birthday. The IDL covers 309 US female supercentenarians which allow for a statistical analysis in the typical extreme value framework as, e.g. presented in the monographs by de Haan and Ferreira [2006], Reiss and Thomas [2007], Falk et al. [2010] and Resnick [2013]. These rather classical EVT methods have already been applied to estimate the maximum attainable age in different populations, see Aarsen and de Haan [1994] and Gbari et al. [2016] for the Netherlands and Belgium as well as Watts et al. [2006] for Canada and Japan, to name only few of them. To our knowledge, there is no similar analysis for the US. The underlying inference procedures, however, are deduced from asymptotic considerations which may lead to (slightly) imprecise estimates when applied to the IDL. To increase the inferential accuracy we combine the two datasets and analyze them with existing methods from censored EVT following Einmahl et al. [2008] (see also Gomes and Neves [2011], Worms and Worms [2014] or Gomes and Guillou [2015] for other censored EVT approaches).

In statistics, it is well known that accounting for censoring is important. In particular, disregarding the censoring issue may lead to biased estimates. Nevertheless, we could not find a single paper on estimating the maximum lifespan that dealt with censoring accordingly. This

gap is closed by the present article.

The remainder of this paper is organized as follows. Section 2 gives a detailed description of the employed datasets and discusses some issues with the structure of the databases. Thereafter, classical EVT is applied to analyze the IDL in Section 3, whereas Section 4 deals with the analysis of the joined dataset by modern censored EVT approaches. The paper closes with a discussion and outlook in Section 5.

2 Data Description

In this section, we cover different aspects of the databases under consideration. Unfortunately, to our knowledge there is no publicly available database which includes complete mortality data for a sufficiently long history while providing the complete age range for all covered populations. However, with some restrictions there are at least two different databases which can be employed for our purpose: the International Database on Longevity (see IDL [2015]) and the Human Mortality Database (see HMD [2015]). As outlined in the introduction we thereby focus on female US citizens for a reliable data quality and quantity that permits a reasonable statistical analysis. In the remainder of this section, we briefly describe both the IDL and the HMD and discuss some aspects of their structure with the focus on US females. Furthermore, we compare these databases and their characteristics directly and derive ramifications for subsequently applied inference methods.

The IDL contains data for 15 industrialized countries across the world. Each of its records relates to one individual dying after its 110^{th} birthday, where the individual's age at death is given exactly in days from their birthdays. This exactness is a special feature of the IDL. As a further advantage, it is supposed to be free of age ascertainment bias (see Maier et al. [2010]). For the United States, the IDL contains 395 persons dying at the age of 110 or beyond between 1980 and 2003, whereof 309 are females. 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 even underestimate the count of supercentenarians dying in the years between 1980 and 2003. This also fits the assessment of Kestenbaum and Fergusen (see Maier et al. [2010]), who guess that there were more than 400 supercentenarians in the US (i.e. females and males) during the timespan covered by the IDL. However, they further explain the process of validation of the IDL data and suggest the conclusion, that the IDL fits the profile of the US female population well in general. To give a statistical framework for the subsequent inference given in Section 3 we treat the records of the IDL as realizations from independent and identically distributed random variables that are left-truncated at age 110.

In contrast to the IDL, the HMD covers mortality data and population sizes 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. A special feature of the HMD is the availability of edited data and unedited input data (at least not edited by the HMD). For US females, the unedited input data covers the highest age ranges (even beyond the age 110) only for some calendar

years between 1953 and 1988. Moreover, in parts, the number of deaths is even given per Lexis triangle. However, for other calendar years, the input data is only available up to an age where some people are still alive. This makes comparisons across different calendar years difficult, particularly for higher age ranges. Hence, we decided to use the edited HMD data, which has a homogeneous structure. In order to reduce complexity we additionally employ the 1×1 number of deaths period data, i.e. we pass on the information of the year of birth. Note however, that this data is manipulated in different ways, particularly by extrapolating models in higher age ranges up to the age 109 (the right-censoring point), see the explanations of the HMD method sheet given in Wilmoth et al. [2007] for details. Therefore, we may be confronted with certain misestimations of the number of deaths in advanced age ranges and these manipulations may have an impact on the results of the joined analysis of IDL and HMD in Section 4. Due to the lack of not manipulated data we cannot quantify this impact and thus assume the data to sufficiently represent the reality.

For US females the HMD covers the years between 1933 and 2014. In sum it includes 73 million deaths over all ages, whereof almost 4000 females are right-censored at 110, i.e. died at the age of 110 or beyond. The sources for the input data set are 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. The HMD supplies only age discrete data (i.e. in completed calendar years), which requires some adjustments to the data. We assume the deaths to be uniformly distributed over each calendar year, which enables us to redistribute the deaths uniformly during each year of age.

In summary, the HMD is in terms of quantity by far better than the IDL. In contrast, the data of the IDL has a significantly higher quality. This disparity cannot be resolved in any way. As mentioned above, the HMD is likely to represent the US population in general, i.e. we act on the assumption that every US citizen is recorded in this database. Thus, we can assume, that each individual recorded in the IDL has also a representation in the HMD as an individual dying at the age of 110 or beyond. We use this assumption in Section 4 where we merge the data of both databases in order to achieve a more informative combined data set.

Another disparity between the HMD and the IDL is the age range covered by the databases. While the HMD only contains single age mortality data up to the age of 109 and a cumulative entry at 110 and beyond, the IDL covers the age range beyond 109. These circumstances imply different strategies for the statistical analysis of these databases as outlined in the two subsequent sections.

3 Classical EVT analysis of the IDL

In this section, we 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. we search for $x_F := \sup\{x : F(x) < 1\} \leq \infty$. 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 $x_F = \infty$. A consistent

estimator for x_F is then given by $M_n = \max_{i\leq n}\{X_i\}$ with cdf F^n. To infer this estimator, there are two well-established EVT approaches, each employing a different class of distributions: the generalized extreme value distributions (GED) 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 GED 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 $(M_n - b_n)/a_n$ for suitable sequences $a_n > 0$ and b_n. The class of GED $G_\gamma(ax + b)$, is given by

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

where a and b are scaling and location parameters which can be calculated from the current dataset.

While this concept is rather general, the second approach based on the class of GPDs seems to be more suitable with respect to the left truncated IDL data. In the GPD approach we are not only considering the maxima of the realizations of random variables, but their values above a given threshold t, say $t = 110$ in our case (see also Balkema and de Haan [1974]). 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

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

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 limiting 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 the 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 et al. [1989] for early work on this topic and de Haan and Ferreira [2006] and Resnick [2008] 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 GED 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 semi-parametric GPD model to the data while the moment estimator is defined as

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

where $M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}(\log(X_{n-i+1,n}) - \log(X_{n-k,n}))^j$ 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 normally distributed 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 et al. [2016] describe the following observation: the smaller the values of k are, the higher is the variance of the EVI estimator while for increasing k the bias of the EVI estimator gets larger. Finding a suitable k between these extremities is called the *bias-variance tradeoff*. To get a first idea of the choice of k, we make 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 de Haan [1994] for the construction of the Mom-based CI and carried them over to the MLE-based CI for a better comparison).

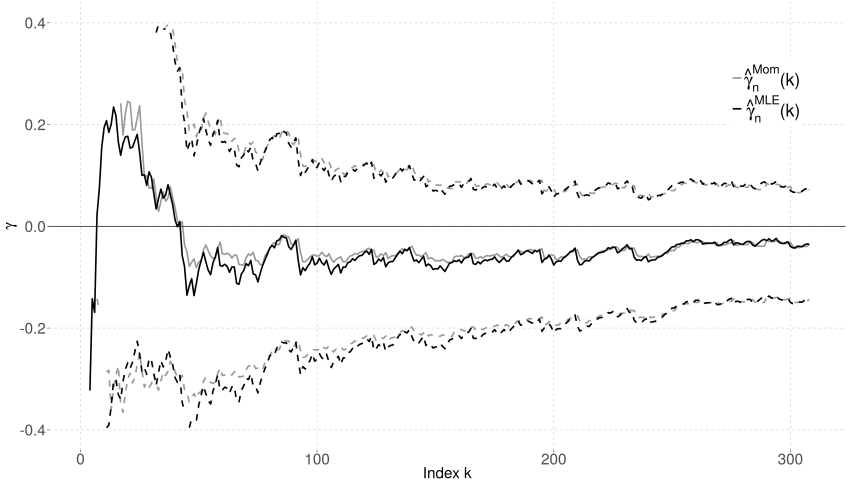

Figure 1: Point estimates $k \mapsto \hat{\gamma}_n^{(\cdot)}(k)$ of the EVI based on the MLE and Mom approach as a function of k together with point-wise 95% CI for the IDL.

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 0.34 (0.32). 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 $k > 46$ 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. Additionally incorporating the bias-variance tradeoff 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 k between 100 and 200, together with the corresponding value of k and the 95% confidence intervals are given in Table 1.

	$\hat{\gamma}_n^{MLE}(k)$	k	95% CI	$\hat{\gamma}_n^{Mom}(k)$	k	95% CI
min	-0.0711	155	$\pm$0.1462	-0.0942	106	$\pm$0.1830
median	-0.0572	128	$\pm$0.1633	-0.0663	142	$\pm$0.1584
max	-0.0385	140	$\pm$0.1592	-0.0418	139	$\pm$0.1614

Table 1: MLE- and Mom-estimates of the EVI for the IDL data together with 95% confidence intervals for different choices of k

As already seen in Figure 1 the CIs are extensive and each inherits the value of 0 which 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 GED (Mom case) and can be used to estimate the right-endpoint of the underlying distribution.

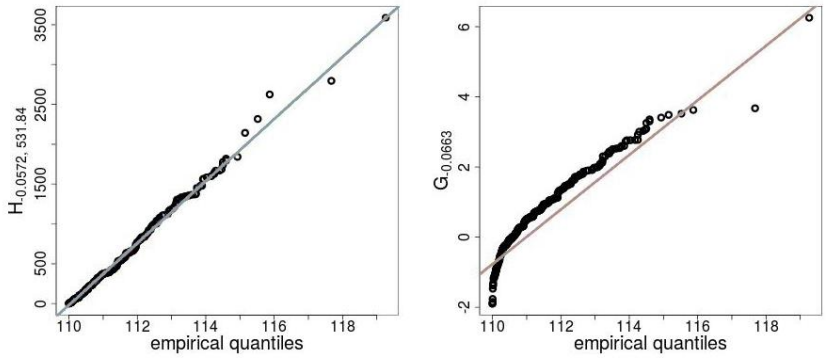

Figure 2: QQ-plots of empirical quantiles over simulated quantiles; left panel: pareto distribution with $\gamma = -0.0572$; right panel:generalized extreme value distribution with $\gamma = -0.0663$. The solid lines visualize the regression line for the plots.

For illustrational purposes, 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. 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 GED. This can be done by applying the method described

in Section 3 of de Haan and Ferreira [2006] and we finally obtain estimates for the maximum lifespan $\hat{x}_F^{Mom}(k)$ (GED) 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. From the latter, we can construct 95% CIs. Table 2 shows the results: It gives the minimum, median, and maximum of the MLE and Mom endpoint estimates with the corresponding asymptotic 95% CIs and the value of k which refer to the values of $\hat{x}_F^{Mom}(k)$ and $\hat{x}_F^{MLE}(k)$, respectively. The obtained estimates for the maximum lifespan vary between 127.49 (GED with minimum EVI Mom-estimate from Table 1) and 147.6 years (GPD with maximum EVI MLE-estimate from Table 1). 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.

	$\hat{x}_F^{MLE}(k)$	k	95% CI	$\hat{x}_F^{Mom}(k)$	k	95% CI
min	132.30	155	$\pm$40.68	127.49	106	$\pm$27.96
median	136.71	128	$\pm$68.34	133.11	142	$\pm$48.97
max	147.60	140	$\pm$144.68	144.49	139	$\pm$123.17

Table 2: MLE- and Mom-estimates of the right-endpoint for the IDL data together with 95% confidence intervals for different choices of k

We finally note, that 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 et al. [2002] with the critical value proposed in Huesler and Li [2006] we could not reject this. Some statisticians may interpret this as a slight underpinning of the results given in Table 2.

4 A joined statistical analysis of the HMD and IDL

The results of the previous section are neither precise nor statistically significant. As a consequence we need to significantly increase the sample size of the data set. The HMD data purges this challenge with a sample size of more than 73 million deaths over all ages whereof almost 4000 females die at the age 110 or beyond (see Section 2). Unfortunately, this dataset is right censored at age 110 and thus we cannot apply the classical EVT method we used in Section 3. However, Einmahl et al. [2008] provide several theoretical results with regard to the statistical analysis of right censored extreme value data.
In their framework, the age at death of the i-th individual is modeled via independent and identically distributed 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 is the distribution function of the censoring distribution. 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}. The classical EVT approaches from the previous sections, however, would only lead to reasonable estimates for γ_Z and x_{F_Z} (when only working with the observations Z_i). On the other hand, a complete case analysis (when deleting all censored observations) would generally lead to estimates that are strongly biased. To address censoring adequately, Einmahl et al. [2008] introduce estimators for γ_X and x_{F_X} which are valid in the following cases:

$$\begin{cases} \text{case 1: } \gamma_X > 0,\ \gamma_C > 0 \\ \text{case 2: } \gamma_X < 0,\ \gamma_C < 0,\ x_{F_X} = x_{F_C} \\ \text{case 3: } \gamma_X = \gamma_C = 0,\ x_{F_X} = x_{F_C} = \infty. \end{cases} \tag{4.4}$$

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.
Because of the right censoring of the HMD data at the age 110, the model described above can be simplified by setting $C_i \equiv 110$, i.e. considering deterministic censoring variables with $F_C(x) = \mathbb{1}(110 \leq x)$. However, a more general setting would allow us to incorporate the information from the IDL. Thus, in what follows, we prefer the following model: $C_i \overset{iid}{\sim} (1-\varepsilon)\mathbb{1}(110 \leq x) + \varepsilon D$, where $\varepsilon << 1$ and D is a cdf with support $(110, \infty)$, both unknown. This model allows for observed values after age 110 (as in the IDL) as well as for

Combining the data. In order to build a combined dataset (CDS) on the basis of the HMD and the IDL we have to rearrange the structure of the HMD data slightly (also recall the description of the databases in Section 2). First, we have to trim the HMD data in the dimension of calendar years: The IDL covers the time span between 1980 and 2003, while the HMD has much more extensive calendar years coverage for US females. Thus, we cut off those calendar years of the HMD data, where we have no data from the IDL. Secondly, we must trim the HMD data in the age dimension: The applied EVT methods require a threshold which avoids moving too far to the left (i.e. employing too much data from younger age ranges, without loosing too much necessary information). To this end, we cut off the ages below age 90 from the HMD data. The choice of the age 90 seems to be reasonable since for US females this age is greater than the modal value of the deaths curve, i.e. $90 > \operatorname{argmax}_{0 \leq x \leq \infty}\ f_Z(x) := \operatorname{argmax}_{0 \leq x \leq \infty}\ \frac{\partial F_Z(x)}{\partial x}$. Thus the part of the deaths curve which lies beyond that age can be regarded to be the right tail of the deaths curve (see also Drees et al. [2006]). Thirdly, we must thin out the HMD data: As we already discussed in Section 2, we can reasonably assume that each person recorded in the IDL is also recorded in the HMD. Therefore, we delete 309 randomly selected HMD records "dying at 110 or beyond" for avoiding double counts in the CDS. All these incorporating restrictions reduce the sample size n in the CDS to about 1.4 million.

Since this figure is still too large for most algorithms with respect to their computing times (even when using modern computing clusters) we apply the following subsampling approach:

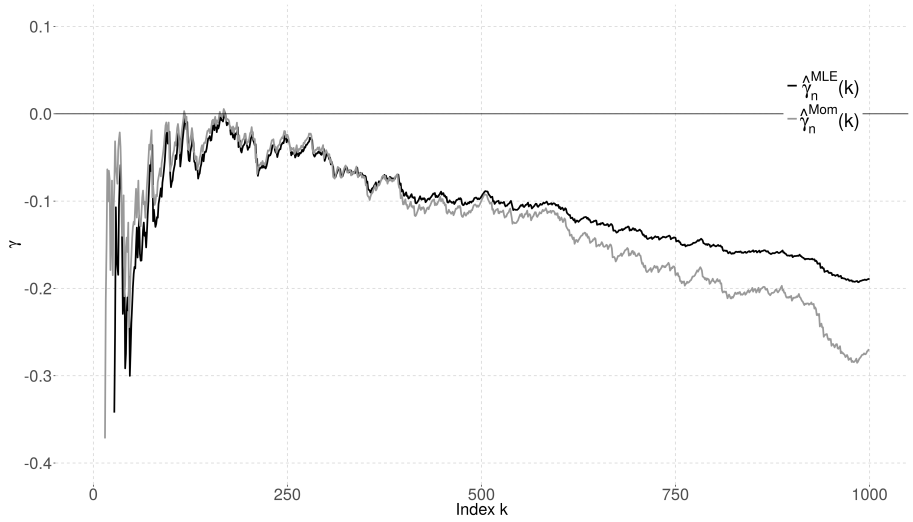

Figure 3: Different EVI estimates as a function of the number of upper order statistics k_0 in case of the trainings dataset from the CBS with $m = 1000$.

Denote the complete CDS sample as $\mathcal{Z}_n := \{Z_1, ...Z_n\}$ and randomly draw m times from $\mathcal{Z}_n$ without replacement to construct one subsample of size m, $m < n$. We independently repeat this N times and thus obtain N subsamples $\mathcal{Z}_m^i := \{Z_1^i, ..., Z_m^i\}, 1 \leq i \leq N$. We then estimate the EVIs separately for each subsample.

Following Politis et al. [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 subsamples m should be sufficiently large, such that each subsample includes (uncensored) IDL data with a sufficiently high probability. Taking these issues as well as the computation time into account we calculated estimates for different choices of $m \in \{1000, 5000\}$ 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; leading to 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 subsample $\mathcal{Z}_m^0$ from $\mathcal{Z}_n$ of length m that serves as a test sample. All other subsamples $\mathcal{Z}_m^i, 1 \leq i \leq N$, will serve as training samples. For each of these subsamples $\mathcal{Z}_m^i, 0 \leq i \leq N$, we then calculate EVI estimates $\hat{\gamma}_{X,i}^{(\cdot)}(k_i)$ for varying numbers $1 \leq k_i \leq m$ of upper order statistics. To this end we use the methods described by Einmahl et al. [2008]. Figure 3 shows the plots of different EVI estimates $k_0 \mapsto \hat{\gamma}_{X,0}^{(\cdot)}(k_0)$ for the test sample for $m = 1000$. For $m = 5000$ we found similar results (not shown here).

It can be readily seen from Figure 3 that the censored MLE (solid line) and Mom estimates (dashed line) behave similarly up to $k_0 \approx 600$ and diverge thereafter. A stable plateau is reached between 400 and 590 and the rule of thumb choice would recommend some k from this interval; corresponding to a (of course rather naive) EVI estimate of

$$\hat{\gamma}_X^{(\cdot)} \approx -0.11.$$

However, as mentioned above we proceed more objective and compare the EVI estimates for

the test set $k_0 \mapsto \hat{\gamma}^{(\cdot)}_{X,0}(k_0)$ with the corresponding EVI estimates from each trainings set $k_i \mapsto \hat{\gamma}^{(\cdot)}_{X,i}(k_i)$. For each of these comparisons we obtain the estimator

$$\hat{k}_{0,i} := \underset{1\le k_0\le m}{\operatorname{argmin}} \sum_{k_i=1}^{m} k_i^{5/4} \left(\hat{\gamma}^{(\cdot)}_{X,0}(k_0) - \hat{\gamma}^{(\cdot)}_{X,i}(k_i) \right)^2, \quad 1 \le i \le N. \tag{4.5}$$

This corresponds to the EVI estimate from the test set $\mathcal{Z}^0_m$ that has the smallest penalized quadratic distance from all EVI estimates of the i^{th} trainigs set $\mathcal{Z}^i_m$. 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 $\hat{\gamma}^{(\cdot)}_{X,0}(\hat{k}_{0,i}), 1 \le i \le N$ obtained from (4.5):

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

Table 3 shows the results for both estimators (i.e. MLE and Mom), two subsample sizes $m \in \{1000, 5000\}$ and a fixed number of replications $N = 5000$. The results for $N = 1000$ were similar (not shown here). We also give the corresponding 95% confidence intervals, utilizing the asymptotic normally distributed estimators for the EVI given in Einmahl et al. [2008]. We

m	1000		5000	
	$\hat{\gamma}^{(\cdot)}_{X,0}$	95%-CI	$\hat{\gamma}^{(\cdot)}_{X,0}$	95%-CI
(c, Mom)	-0.1146	$\pm$ 0.03630	-0.1112	$\pm$0.01658
(c, MLE)	-0.1132	$\pm$ 0.03116	-0.1091	$\pm$0.01444

Table 3: EVI estimates for γ_X obtained from the cross-validation type method for different choices of $m \in \{1000, 5000\}$ together with 95% confidence intervals. All calculations are based on $N = 5000$ training sets and one test set drawn from the CDS.

see from Table 3 that all EVI estimates lie between -0.115 and -0.109 which corresponds to the (intuitive) choice -0.11 from the rule of thumb method. In particular, they are much smaller than the EVI estimates obtained from the IDL dataset alone, see Section 3. Moreover, it seems that the choice of m mainly affects the width of the confidence intervals (which is clear), whereas the EVI point estimates are only slightly affected. However, for each combination of m and estimator (MLE or Mom) we obtain significantly negative estimates for the EVI γ_X. This corresponds to a finite right endpoint. The resulting censored MLE and Mom endpoint estimates for $N = 5000$ and $m \in \{1000, 5000\}$ are shown in Table 4. The resulting point estimates for the maximum lifespan now vary between 125.82 (censored Mom-estimate based on subsamples of size $m = 5000$) and 128.73 years (censored MLE based on subsamples of size $m = 5000$). Compared to the results from Section 3, where we analyzed the IDL data exclusively, these estimates show significantly smaller values. Running a similar analysis based on the classical uncensored EVT approach exclusively on the HMD data, with $m = 5000$ we

m	1000	5000
	$\hat{x}^{(\cdot)}_{F_X}$	$\hat{x}^{(\cdot)}_{F_X}$
(c, Mom)	125.90	125.82
(c, MLE)	128.49	128.73

Table 4: Censored MLE and Mom-estimates of the right-endpoint obtained from the results of the cross validation type method for different choices of $m \in \{1000, 5000\}$. All calculations are based on $N = 5000$ trainings set and one test set drawn from the CDS.

get $\hat{x}^{MLE}_{F_Z} = 130.62$ and $\hat{x}^{Mom}_{F_Z} = 128.06$, respectively. Of course, this would not only ignore censoring in the HMD data but also is an incorrect application of the estimators. However, these considerably higher estimates illustrate two facts: the importance of respecting the involved censoring in a thorough statistical analysis and the surplus of combining different data sources for the quality of the results.

5 Conclusion and Discussion

We have investigated the question of quantifying the maximum human lifespan for US women. In particular, based on extreme value theory (EVT) methods we inferred the existence of a possibly finite lifespan for two data sets: The International Database on Longevity (IDL) and a combination of the IDL with the human mortality database (HMD). The IDL contains very accurate data from 309 female US supercentenarians dying between 1980 and 2003. Presumably, due to its small sample size, the hypothesis of an infinite lifespan could not be rejected for the IDL. Moreover, the obtained point estimates for the maximum attainable age fluctuated considerably and were thus not very reliable. We therefore additionally considered observations from the more extensive HMD and combined them with the IDL leading to a considerably larger data set. Since observations from the HMD are right-censored we additionally had to employ censored EVT methods. In particular, based on theoretical findings from Einmahl et al. [2008], we found significant evidence for a finite lifespan in the combined data set and obtained reliable point estimates for the maximum attainable age of US females. These estimates vary between 125.82 and 128.73 years which seems reasonable since the oldest US female recorded in the IDL reached an age of more than 119 years.

These values cannot be compared directly to previous findings in recent literature since the basics (i.e. data and methods) differ considerably. In particular, to our knowledge, the present statistical analysis is the first that accordingly accounts for censoring of the underlying data. Nevertheless, to at least associate the numbers with the context of a maximum human lifespan we summarize some of the most recent results: Weon and Je [2009] obtain maximum age estimates around 126 years for Swedish females, while Bravo and Corte-Real [2012] find estimates of only 112.77 and 111.78 years for Spanish and Portuguese females, respectively. Li et al. [2011] estimate a similarly low maximum lifespan of 112.20 years for Australian and even only 109.43 years for New Zealand citizens. Aarsen and de Haan [1994] compute 95% confidence

estimates for the Netherlands between 113 and 124 years and Hanayama and Sibuya [2015] estimate the maximum human lifespan for Japanese to 123 years. Most recently, Gbari et al. [2016] quantify the maximum attainable age of Belgian females between 120.3 and 122.73 years. The disparity in the results may be an artifact caused by different sources of data, differences between the populations under study (e.g. socioeconomic conditions, etc.) as well as the diversity of the employed methods.

For future research, we recommend the application of censored EVT methods to other cohorts. Moreover, an extension of the analysis by a calendar years dimension would be desirable. In particular, we believe that even more reliable results may be obtained by considering the evolution of the maximal attainable age at death *over time*. However, due to the lack of existing theory for extremal processes that are subject to right censored data, we could not yet follow this approach. The same fact holds for censored extreme value regression models that incorporate covariables.

References

K. Aarsen and L. de Haan. On the maximal life span of humans. *Mathematical Population Studies*, 4(4):259–281, 1994.

A. A. Balkema and L. de Haan. Residual life time at great age. *The Annals of Probability*, 2(5): 792–804, 1974.

J. M. Bravo and P. Corte-Real. Modeling longevity risk using extreme value theory: An empirical investigation using portuguese and spanish population data. Technical report, Evora University, Portugal, 2012.

A. de Grey. The foreseeability of real anti-aging medicine: focusing the debate. *Experimental Gerontology*, 38(9):927–934, 2003.

L. de Haan and A. Ferreira. *Extreme value theory: an introduction*. Springer, New York; [Heidelberg], 2006.

L. de Haan, C. Mercadier, and C. Zhou. Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance and Stochastics*, 20(2):321–354, 2016.

A. L. M. Dekkers, J. H. J. Einmahl, and L. de Haan. A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, 17(4):1833–1855, 1989.

D. Dietrich, L. de Haan, and J. Huesler. Testing extreme value conditions. *Extremes*, 5(1): 71–85, 2002.

X. Dong, B. Milholland, and J. Vijg. Evidence for a limit to human lifespan. *Nature*, 538(7624): 257–259, 2016.

H. Drees, L. de Haan, and D. Li. Approximations to the tail empirical distribution function with application to testing extreme value conditions. *Journal of Statistical Planning and Inference*, 136:3498–3538, 2006.

J. H. Einmahl, A. Fils-Villetard, and A. Guillou. Statistics of extremes under random censoring. *Bernoulli*, 14(1):207–227, 2008.

M. Falk, J. Hüsler, and R.-D. Reiss. *Laws of small numbers: extremes and rare events*. Springer Science & Business Media, 2010.

R. A. Fisher and L. H. C. Tippett. 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, 4 1928. ISSN 1469-8064.

J. F. Fries. Aging, natural death, an the compression of morbidity. *The New England Journal of Medicine*, 303(3):130–135, 1980.

K. Y. A. S. Gbari, M. Poulain, L. Dal, M. Denuit, et al. Extreme value analysis of mortality at the oldest ages: a case study based on individual ages at death. Technical report, UCL, 2016.

M. I. Gomes and A. Guillou. Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review*, 83(2):263–292, 2015.

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

N. Hanayama and M. Sibuya. Estimating the upper limit of lifetime probability distribution, based on data of japanese centenarians. *J. Gerontol. A. Biol. Sci. Med. Sci.*, 71(8):1014–1021, 2015.

HMD. Human mortality database, 2015. URL `http://www.mortality.org`.

J. Huesler and D. Li. On testing extreme value conditions. *Extremes*, 9(1):69–86, 2006.

IDL. International database on longevity, 2015. URL `http://www.supercentenarians.org/`.

R. Kurzweil and T. Grossman. *Fantastic voyage: live long enough to live forever*. Rodale, 2005.

J. Li, A. Ng, and W. Chan. Modeling old-age mortality risk for the populations of australia and new zealand: An extreme value approach. *Mathematics and Computers in Simulation*, 81(7): 1325–1333, 2011.

H. Maier, J. Gampe, B. Jeune, J.-M. Robine, and J. W. Vaupel, editors. *Supercentenarians*. Springer, New York, 2010.

J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistic*, 3(1): 119–131, 1975.

D. N. Politis, J. P. Romano, and M. Wolf. *Subsampling*. Springer Series in Statistics. Springer New York, 1999.

R.-D. Reiss and M. Thomas. *Statistical analysis of extreme values*. Springer, 2007.

S. I. Resnick. *Extreme values, regular variation and point processes*. Springer, New York, 2008.

S. I. Resnick. *Extreme values, regular variation and point processes*. Springer, 2013.

J.-M. Robine and M. Allard. The oldest human. *Science*, 279:1831.1831, 1998.

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

B. M. Weon and J. H. Je. Theoretical estimation of maximum human lifespan. *Biogerontology*, 10(1):65–71, 2009.

J. Wilmoth, k. Andreev, D. Jdanov, and D. Glei. Metods protocol for human mortality database, 2007. URL `http://www.mortality.org/Public/Docs/MethodsProtocol.pdf`.

J. Worms and R. Worms. New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. *Extremes*, 17(2):337–358, 2014.