Records in Athletics Through Extreme-Value Theory

John H. J. EINMAHL and Jan R. MAGNUS

We are interested in two questions on extremes relating to world records in athletics. The first question is: What is the ultimate world record in a specific athletic event (such as the 100-m race for men or the high jump for women), given today's state of the art? Our second question is: How "good" is a current athletic world record? An answer to the second question also enables us to compare the quality of world records in different athletic events. We consider these questions for each of 28 events (14 for both men and women).

We approach the two questions with the probability theory of extreme values and the corresponding statistical techniques. The statistical model is of a nonparametric nature; only some "weak regularity" of the tail of the distribution function is assumed. We derive the limiting distribution of the estimated quality of a world record.

While almost all attempts to predict an ultimate world record are based on the development of world records over time, this is not our method. Instead, we use *all* top performances. Our estimated ultimate world record tells us what, in principle, is possible in the near future, given the present knowledge, material (shoes, suits, equipment), and drug laws.

KEY WORDS: Endpoint estimation; Exceedance probability; Ranking; Statistics of extremes; World record.

1. INTRODUCTION

What is the total (insured) loss of the first hurricane topping Katrina? Is a specific runway at JFK airport long enough for a safe landing? How high should the dikes in Holland be in order to protect the Dutch against high water levels? How long can we live? How fast can we run? How far can we throw? How high (far) can we jump?

Each of these issues relates to extremes. In this article we are interested in the last three of these issues, relating to world records in athletics. More specifically, we wish to answer two questions. The first question is: What is the ultimate world record in a specific athletic event (such as the 100-m race for men or the high jump for women), given today's state of the art? Our second question is: How "good" is a current athletic world record, that is, how difficult is it to improve? An answer to the second question enables us to compare the quality of world records in different athletic events.

We approach these two extremes-related questions with the probability theory of extreme values and the corresponding statistical techniques. The statistical model is of a nonparametric nature; only some "weak regularity" of the tail of the distribution function is assumed. Somewhat related work on records in sports is given in Barão and Tawn (1999) and Robinson and Tawn (1995), who considered the annual best times in the women's 3,000-m event and drug-related questions for the same event, respectively. Smith (1988) proposed a maximum likelihood method of fitting models to a series of records and applied his method to athletic records for the mile and the marathon.

Almost all attempts to predict an ultimate world record are based on the development of world records over time. This is *not* our method, and we do not try to predict the world record in the year 2525. Instead, we use *all* top performances (see Table 1). Our estimated ultimate record tells us what, in principle, is possible in the near future, given the present knowledge, material (shoes, suits, equipment), and drug laws.

Our selection of athletic events is based on the Olympic Games. While at the first of the modern Olympic Games in 1896, only a few hundred male athletes competed in 10 events, at the 2008 Beijing Olympics, male athletes competed in 24 events: running (100 m, 200 m, 400 m, 800 m, 1,500 m, 5,000 m, 10,000 m, marathon, 110-m hurdles, 400-m hurdles, 20-km walk, 50-km walk, steeplechase); throwing (shot put, javelin throw, discus throw, hammer throw); jumping (long jump, high jump, pole vault, triple jump); relay events (4×100 m and 4×400 m); and the decathlon.

Women first competed at the 1924 Paris Olympics in 5 events, but at the 2008 Beijing Olympics, they competed in 23 events—only the 50-km walk is still excluded. Furthermore, women run 100-m hurdles (instead of 110-m) and compete in a heptathlon (instead of a decathlon).

For the purposes of our study we select 14 events: 8 running events, 3 throwing events, and 3 jumping events, as follows:

> *Running:* 100 m (D), 200 m (H), 400 m (D), 800 m (H), 1,500 m (D), 10,000 m, marathon, 110/100-m hurdles (DH)
> *Throwing:* shot put (DH), javelin throw (DH), discus throw (D)
> *Jumping:* long jump (DH), high jump (DH), pole vault (D).

Our selection thus includes all events that make up the decathlon (D) and heptathlon (H), supplemented by the 10,000 m and the marathon.

This article is organized as follows. In the following section we describe the data and how they were collected. In Section 3 we develop the required extreme-value theory and present the limiting distribution of the estimated quality of a world record (Thm. 1). In Section 4 we apply the theory to the data and answer our two questions. Sensitivity issues are discussed in Section 5, and Section 6 offers some conclusions. The Appendix contains the proof of Theorem 1.

2. THE DATA

For each of the 28 events (14 for both men and women), we collected data on the personal best of as many of the top athletes in each event as we could, taking care that no "holes" occur in the list. Thus, if an athlete appears on our list, then *all* athletes with a better personal best also appear on our list. We emphasize that we are interested in personal bests and not in the development of the world record over time. As a consequence,

John H. J. Einmahl is Professor of Statistics (E-mail: *j.h.j.einmahl@uvt.nl*) and Jan R. Magnus is Professor of Econometrics (E-mail: *magnus@uvt.nl*), Department of Econometrics & OR and CentER, Tilburg University, 5000 LE Tilburg, The Netherlands. The authors are grateful to the co-editor, associate editor, two referees, and Franc Klaassen for their thoughtful and constructive comments; to Andrea Krajina for help with the simulations; and to John Kleppe for exceptional and extensive research assistance.

Journal of the American Statistical Association
December 2008, Vol. 103, No. 484, Applications and Case Studies
DOI 10.1198/016214508000000698

Table 1. Data summary

Event	Men			Women		
	Depth	Worst	Best	Depth	Worst	Best
Running						
100 m	970	10.30	9.78	578	11.38	10.49
110/100-m hurdles	805	13.83	12.91	432	13.20	12.21
200 m	780	20.66	19.32	561	23.14	21.34
400 m	658	45.74	43.18	538	52.02	47.60
800 m	722	1:46.61	1:41.11	537	2:01.05	1:53.28
1,500 m	781	3:38.74	3:26.00	531	4:09.03	3:50.46
10,000 m	1,239	28:30.03	26:20.31	876	33:04.00	29:31.78
Marathon	1,546	2:13:36	2:04:55	1,024	2:36:06	2:15:25
Throwing						
Shot put	392	19.80	23.12	223	18.42	22.63
Javelin throw	422	77.00	98.48	279	54.08	71.54
Discus throw	335	62.84	74.08	222	62.52	76.80
Jumping						
Long jump	629	8.00	8.95	434	6.61	7.52
High jump	436	2.26	2.45	392	1.90	2.09
Pole vault	512	5.50	6.14	407	4.00	4.92

each athlete appears only *once* in our list, namely with his or her top result, even if he or she has broken the world record several times. Our observation period ends on April 30, 2005.

The data were obtained from two websites, namely, a Swedish website compiled by Hans-Erik Pettersson (*web.telia.com/~u19603668/athletics_all-time_best.htm#statistik*) for the period up to mid-2001, and the official website of the International Association of Athletics Federations (IAAF) (*www.iaaf.org/statistics/toplists/index.html*) for each year from 2001 onward. These two sites provide a list of the top athletes (and their results) per event. The Swedish website provides additional information under the headings "Doubtful Wind Reading," "Doubtful Timing," and "Subsequent to Drug Disqualification." These concern records not recognized by the IAAF and, consequently, are not included in our lists. The same applies to information under the heading "Hand Timing," times clocked by hand in a period when electronic timing was available. These records are also not recognized by the IAAF and are not included in our lists. Times clocked by hand from the period when electronic timing was *not* available are recorded with an accuracy of .1 seconds (rather than .01 seconds) and have been interpreted to be exact to two decimal places. For example, a hand-clocked time of 9.9 seconds is recorded by us as 9.90.

The raw data, thus, consist of six lists per event: one for the period up to mid-2001 from the first website and five lists for 2001, 2002, 2003, 2004, and 2005 (the last list runs only to April 30) from the second website. In combining the six lists two further actions are required. First, we considered the worst performance in each of the six lists and then took the best of these worst performances as the "lower" bound in the combined list for each event. This guarantees that there are no holes in the combined list. Second, we removed all multiple entries of the same athlete, so that each athlete appears only once with his or her personal best. (This is not as easy is it might appear, because names are sometimes misspelled and athletes sometimes change their name, typically women after marriage.) The end result is a list per event of top athletes with their personal bests. Table 1 gives an overview of the number of athletes (the "depth") and the worst and best results for each event in the sample. The data consist of about 10,000 observations for the men and 7,000 observations for the women. On average, we have about twice as many observations for the running events as for jumping and throwing. In particular, the number of observations for the throwing events (on average 383 for the men and 241 for the women) is on the low side.

All distances in the jumping and throwing events are measured in meters, and the more meters, the better. All times in the running events are measured in seconds, and the *fewer* seconds, the better. This discrepancy is somewhat inconvenient, and we, therefore, transformed running times to speeds, so that the higher the speed, the better. Thus, 10.00 seconds in the 100 m is transformed to a speed of 36.00 km/h.

Some data occur in clusters, especially in the shorter distances such as 100 m. These clusters occur not because the actual times are the same, but because the timing is imperfect. Because clusters can cause problems in the estimation, we "smoothed" these data. For example, suppose m athletes run a personal best of $d = 10.05$ seconds in the 100 m. Then we smooth these m results over the interval (10.045, 10.055) by

$$d_j = 10.045 + .01\frac{2j-1}{2m}, \qquad j = 1, \dots, m.$$

We could have randomized the smoothing, or we could even have introduced a more general smoothing scheme applied to all the data, but this proved to have little effect on the results.

3. EXTREME–VALUE THEORY

Consider one athletic event, say the 100-m race for men, and let $X_1, X_2, \dots, X_n$ denote the personal bests of all n male 100-m athletes in the world. The precise definition of "athlete" is left vague, and, therefore, the definition and possible measurement of n are difficult. Clearly, n is much larger than the "depth" in Table 1, which refers only to the top athletes (in this case, 970). Fortunately, the value of n turns out to be unimportant.

We consider these n personal bests as iid observations from some distribution function F. Let $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ be the associated order statistics, so that $X_{n,n}$ denotes the world record. (Recall that we transformed running times to speeds, so that the higher the jump, the farther the throw, and the higher the speed, the better.) As an analog to the central limit theorem for averages, we know that if the maximum $X_{n,n}$, suitably centered and scaled, converges to a nondegenerate random variable, then sequences $\{a_n\}$ $(a_n > 0)$ and $\{b_n\}$ exist such that

$$\lim_{n\to\infty} \Pr\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = G_\gamma(x), \tag{1}$$

where

$$G_\gamma(x) := \exp\left(-(1+\gamma x)^{-1/\gamma}\right)$$

for some $\gamma \in \mathbb{R}$, with x such that $1 + \gamma x > 0$. [By convention, $(1+\gamma x)^{-1/\gamma} = e^{-x}$ for $\gamma = 0$.] If condition (1) holds, then we say that F is in the max-domain of attraction of G_γ and γ is called the extreme-value index. This will be the main regularity condition on the right tail of F. Note that (1) implies (by taking logarithms) that

$$\lim_{t\to\infty} t(1 - F(a_t x + b_t)) = -\log G_\gamma(x) = (1+\gamma x)^{-1/\gamma}, \qquad G_\gamma(x) > 0, \tag{2}$$

where t now runs through $\mathbb{R}^+$ and a_t and b_t are defined by interpolation. We may take $b_t = U(t)$ with

$$U(t) := \left(\frac{1}{1-F}\right)^{-1}(t) = F^{-1}\left(1 - \frac{1}{t}\right), \qquad t > 1,$$

where -1 denotes the left-continuous inverse.

We need to estimate γ, a_t, and b_t. Let, for $1 \leq k < n$,

$$M_n^{(r)} := \frac{1}{k}\sum_{i=0}^{k-1}(\log X_{n-i,n} - \log X_{n-k,n})^r, \qquad r = 1, 2.$$

We consider two estimators for $\gamma \in \mathbb{R}$. The first is the moment estimator

$$\hat{\gamma}_1 := M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1};$$

see Dekkers, Einmahl, and de Haan (1989). The second, $\hat{\gamma}_2$, is the so-called maximum likelihood estimator; see Smith (1987). Next, we define the following estimators for $a_{n/k}$ and $b_{n/k}$:

$$\hat{a}_j := \hat{a}_{j,n/k} := \begin{cases} X_{n-k,n} M_n^{(1)}(1 - \hat{\gamma}_j) & \text{if } \hat{\gamma}_j < 0 \\ X_{n-k,n} M_n^{(1)} & \text{otherwise,} \end{cases}$$

for $j = 1, 2$, and

$$\hat{b} := \hat{b}_{n/k} := X_{n-k,n}.$$

Observe that $b_{n/k} = U(n/k)$ and that $\hat{b}$ is just its empirical analog.

This article has two purposes. The first purpose is to estimate the right endpoint

$$x^* := \sup\{x \mid F(x) < 1\}$$

of the distribution function F, that is, the ultimate world record. When estimating the endpoint we assume that $\gamma < 0$; note that $x^* = \infty$ when $\gamma > 0$. It can be shown that condition (1) is equivalent to

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \qquad x > 0. \tag{3}$$

For large t we can write heuristically

$$U(tx) \approx U(t) + a(t)\frac{x^\gamma - 1}{\gamma}.$$

Because $\gamma < 0$ this yields, for large x and setting $t = n/k$,

$$x^* \approx U\left(\frac{n}{k}\right) - a\left(\frac{n}{k}\right)\frac{1}{\gamma}.$$

We, therefore, estimate x^* with

$$\hat{x}_j^* := \hat{b} - \frac{\hat{a}_j}{\hat{\gamma}_j}, \qquad j = 1, 2, \tag{4}$$

when $\hat{\gamma}_j < 0$, and $\hat{x}_j^* := \infty$ otherwise.

Under appropriate conditions, including (1) and $k \to \infty$, $k/n \to 0$ as $n \to \infty$ ($\hat{x}_2^*$ also requires $\gamma > -1/2$), both $\hat{x}_1^*$ and $\hat{x}_2^*$ are consistent and asymptotically normal estimators of x^*. In particular, for $\hat{x}_1^*$ we have under certain conditions

$$\frac{\sqrt{k}(\hat{x}_1^* - x^*)}{\hat{a}_1} \xrightarrow{d} \mathrm{N}\left(0, \frac{(1-\gamma)^2(1-3\gamma+4\gamma^2)}{\gamma^4(1-2\gamma)(1-3\gamma)(1-4\gamma)}\right);$$

see Dekkers et al. (1989, p. 1851). The estimation of extreme quantiles and endpoints has been thoroughly studied; see de Haan and Ferreira (2006, chap. 4) for a detailed account.

The second purpose of this article is to assess the quality of the world record. We measure this quality by $n(1 - F(X_{n,n}))$, which is the expected number of exceedances of the current world record $X_{n,n}$ (conditional on this world record), if we take n iid random variables from F that are independent of the X_i. The lower this number, the more difficult it is to improve the world record and, hence, the better it is. It might seem more natural to measure the quality of the world record based on $x^* - X_{n,n}$. This quantity can, however, be infinite. More important, it does not take into account the tail behavior of F. Observe that our measure of quality is equal to $n(F(x^*) - F(X_{n,n}))$. We shall discuss our quality measure again at the end of Section 4.3.

From (2), with $a_t x + b_t = X_{n,n}$ and $t = n/k$, we have heuristically

$$n(1 - F(X_{n,n})) \approx k\left(1 + \gamma\frac{X_{n,n} - b_{n/k}}{a_{n/k}}\right)^{-1/\gamma}.$$

Hence, we "estimate" $n(1 - F(X_{n,n}))$ by

$$Q_j := k\left[\max\left(0, 1 + \hat{\gamma}_j\frac{X_{n,n} - \hat{b}}{\hat{a}_j}\right)\right]^{-1/\hat{\gamma}_j}, \qquad j = 1, 2;$$

see Dijk and de Haan (1992) or de Haan and Ferreira (2006, chap. 4). It is important to observe that Q_j can be computed without knowing n.

We will need a second-order refinement of the so-called domain of attraction condition (1), phrased in terms of U as in (3).

We assume that there exists a function $A(\cdot)$ of constant sign satisfying $\lim_{t\to\infty} A(t) = 0$, such that, for $x > 0$,

$$\lim_{t\to\infty}\left(\frac{U(tx)-U(t)}{a(t)} - \frac{x^{\gamma}-1}{\gamma}\right)\Big/A(t) = \frac{1}{\rho}\left(\frac{x^{\gamma+\rho}-1}{\gamma+\rho} - \frac{x^{\gamma}-1}{\gamma}\right), \quad (5)$$

with $\rho \le 0$, where we interpret $(x^0 - 1)/0$ as $\log x$. We now present the limiting distribution of Q_j, the estimated quality of the world record. A proof of Theorem 1 is presented in the Appendix.

Theorem 1. Let $\gamma > -1/2$. Let F be continuous and assume that U satisfies the second-order condition (5) with $\rho < 0$. Assume further that $k \to \infty$, $k/n \to 0$, and $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$ as $n \to \infty$. Finally, assume that

$$\sqrt{k}\left(\frac{\hat{a}_{j,n/k}}{a_{n/k}} - 1\right), \qquad \sqrt{k}\left(\frac{\hat{b}_{n/k} - b_{n/k}}{a_{n/k}}\right), \qquad \sqrt{k}(\hat{\gamma}_j - \gamma)$$

are all $O_p(1)$ for $j = 1, 2$. [In fact, any estimators of $a_{n/k}$, $b_{n/k}$, and γ can be used for which these $O_p(1)$ requirements are fulfilled.] Then

$$Q_j \xrightarrow{d} \mathrm{Exp}(1), \qquad j = 1, 2,$$

as $n \to \infty$.

We will see in the proof that $Q_j/[n(1 - F(X_{n,n}))] \xrightarrow{p} 1$. Hence, all the asymptotic randomness of Q_j comes from $X_{n,n}$ and not from the estimation of F.

4. WORLD RECORDS

We now apply the estimators of the previous section to the data discussed in Section 2 in order to answer our questions: (1) What are the ultimate world records? (2) How good are the current world records?

4.1 Estimation of the Extreme-Value Index

Our first goal is to estimate γ, the extreme-value index, for the 14 selected athletic events, men and women separately. To estimate γ, we must first know whether it exists, that is, whether condition (1) holds for some $\gamma \in \mathbb{R}$. We have tested the existence, using Dietrich, de Haan, and Hüsler (2002) and Drees, de Haan, and Li (2006). The test results indicate that only the distribution function of two events, namely, the pole vault for both men and women, fails to satisfy condition (1). Hence, we drop the pole vault from our analysis and continue with 13×2 athletic events for which we want to estimate γ.

In general, for estimation problems in extreme-value theory, the estimator is plotted as a function of k (the number of upper order statistics used for estimation minus 1). It is a difficult practical problem to find a good value for k on which to base the estimator. Typically, for small k the estimator has a high variance, and the plot is unstable; for large k the estimator has a bias. This is illustrated in Figure 1, where we plot $\hat{\gamma}_1$ and $\hat{\gamma}_2$ as a function of k in the 100-m event for men. We see that both estimators behave roughly the same and that γ is clearly negative. Drawing such plots for all events confirms that $\gamma < 0$ for the large majority.

It is not immediately obvious from Figure 1 (and similar figures for the other 25 events) what our estimate for γ should be. We, therefore, also consider two additional estimators that have

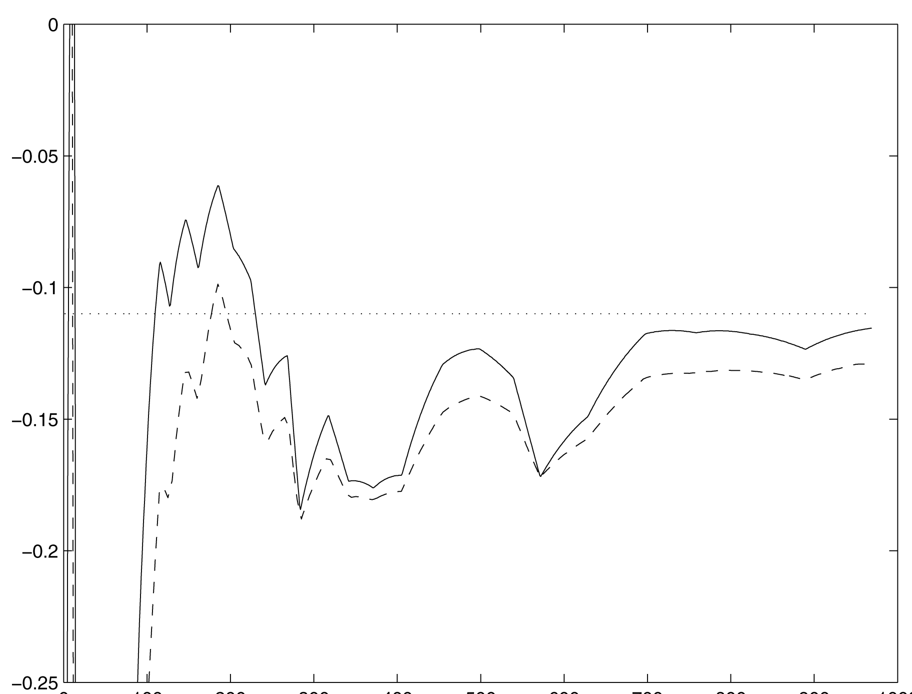

Figure 1. Moment estimator (solid line) and maximum likelihood estimator (dashed line) versus k for the men's 100 m. The selected estimate is the dotted horizontal line.

good properties when $\gamma < 0$. The first additional estimator

$$\hat{\gamma}_3 := 1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1}$$

is simply the "second part" of the moment estimator, because it is well known that $M_n^{(1)}$ (the Hill estimator) is a good estimator of $\gamma > 0$. The second additional estimator has a similar structure:

$$\hat{\gamma}_4 := 1 - \frac{1}{2}\left(1 - \frac{(N_n^{(1)})^2}{N_n^{(2)}}\right)^{-1},$$

where $N_n^{(r)} := (1/k)\sum_{i=0}^{k-1}(X_{n-i,n} - X_{n-k,n})^r$ for $r = 1, 2$; see, for example, Ferreira, de Haan, and Peng (2003).

For every event we looked at the plots of these four estimators and tried to find the first stable region in k of the estimates. More specifically, paying particular attention to values of k between 50 and 250, we tried to identify a set of consecutive values of k of length at least 50 where the estimated values do not fluctuate much, so that the procedure is insensitive to the choice of k in such a region. For example, for the moment estimator in the men's 100 m such a stable region runs from about $k = 100$ to $k = 200$. Then we took averages over the region and over the different estimators. For estimates close to 0 or positive we mainly used $\hat{\gamma}_1$ and $\hat{\gamma}_2$. This procedure led to the results in Table 2. We see that indeed all our estimates of γ are negative, except the one for the men's long jump.

The heuristic treatment of choosing k requires further explanation and justification. An important question is how sensitive this estimation procedure is with respect to a wrong choice of k. We return to this issue in Section 5.

4.2 The Ultimate World Records

We now address our first question, namely, the estimation of the right endpoint of the probability distribution, that is, the ultimate world record. We could proceed as for the estimation of γ, by plugging the four estimators of γ in the definition of $\hat{x}_j^*$ (and in $\hat{a}_j$), $j = 1, 2, 3, 4$; see (4). For $j = 1, 2$ these estimators are shown for the men's 100 m in Figure 2. A much more stable plot, however, is obtained when we replace $\hat{\gamma}_j = \hat{\gamma}_j(k)$ by our (fixed) selected estimate of γ in Table 2. So we still plot our endpoint estimator (4) versus k, but the dependence on k is now only through $X_{n-k,n}$ and $M_n^{(1)} = M_n^{(1)}(k)$; see the dashed–dotted line in Figure 2. We estimate x^* on the basis of the latter plot. This leads to the results in Table 3, where we have transformed the speeds of the running events back to times. In this table we also present the current world records for comparison. Note that the data collection ended on April 30, 2005, but here and in the rest of this article we use the current world records as of October 25, 2007. (In fact, 6 of the 28 world records have been sharpened in the $2\frac{1}{2}$-year period between April 2005 and October 2007: the 100 m, 110-m hurdles, 10,000 m, and marathon for men and the javelin throw and pole vault for women.) Table 3 also presents a rough estimate of the standard error of $\hat{x}^*$ based on the asymptotic normality of $\hat{x}_1^*$ (see Sec. 3).

Because we have assumed that $\gamma < 0$, we do not present an estimate for x^* when the estimate of γ is positive or so close to 0 that it is not clear if indeed $\gamma < 0$. This happens in five events: the men's 400 m and the 10,000 m and long jump for both men and women. The relatively high values of $\hat{\gamma}$ indicate that a substantial improvement in the current world record is possible for these five events.

It appears that very little progress is still possible in the men's marathon (only 20 seconds), but much more in the women's marathon (almost 9 minutes). In contrast, the javelin throw for women appears to be close to its frontier (80 cm), while for the men an improvement of 8 meters is possible.

Table 2. Estimates of γ

Event	Men	Women
Running		
100 m	−.11	−.14
110/100-m hurdles	−.16	−.25
200 m	−.11	−.18
400 m	−.07	−.15
800 m	−.20	−.26
1,500 m	−.20	−.29
10,000 m	−.04	−.08
Marathon	−.27	−.11
Throwing		
Shot put	−.18	−.30
Javelin throw	−.15	−.30
Discus throw	−.23	−.16
Jumping		
Long jump	.06	−.07
High jump	−.20	−.22
Pole vault	—	—

4.3 Quality of the Current World Records

Our second question relates to the ordering of world record holders by means of the estimated quality Q of the world record. Essentially, this quality is measured by transforming all 26 different distributions to the (same) uniform $(0, n)$ distribution. For finding Q we use a similar procedure as for estimating x^*. Again for the men's 100 m, Q_1 and Q_2 are shown in Figure 3, as well as a version of Q with $\hat{\gamma}$ fixed. We use mainly the latter plot to find Q. Based on the asymptotic theory of Theorem 1, we present e^{-Q} (rather than Q itself), which has in the limit a uniform (0, 1) distribution, thus providing not only a relative but also an absolute criterion of quality. Because this transformation is decreasing, a higher value of e^{-Q} means that the record is better. In Table 4 the values of e^{-Q} are presented for the 26 world records and the corresponding world record holders. Although far from perfect, some "uniformity on (0, 1)" of the 26 values can be observed. (The deviation from uniformity is partially caused by using the current world records, instead of those on the day that data collection ended.) Observe that the various events as well as the gender are well mixed. (For the women's javelin event our data collection only began on April 1, 1999, because a new rule was put into effect then. This might lead to a *small* overvaluation of Menéndez's world record.)

Table 4 demonstrates that a world record can have a high quality while it can still be much improved (as with the marathon for women), but that it can also be close to its limit while of relatively low quality (as with the 100-m hurdles for

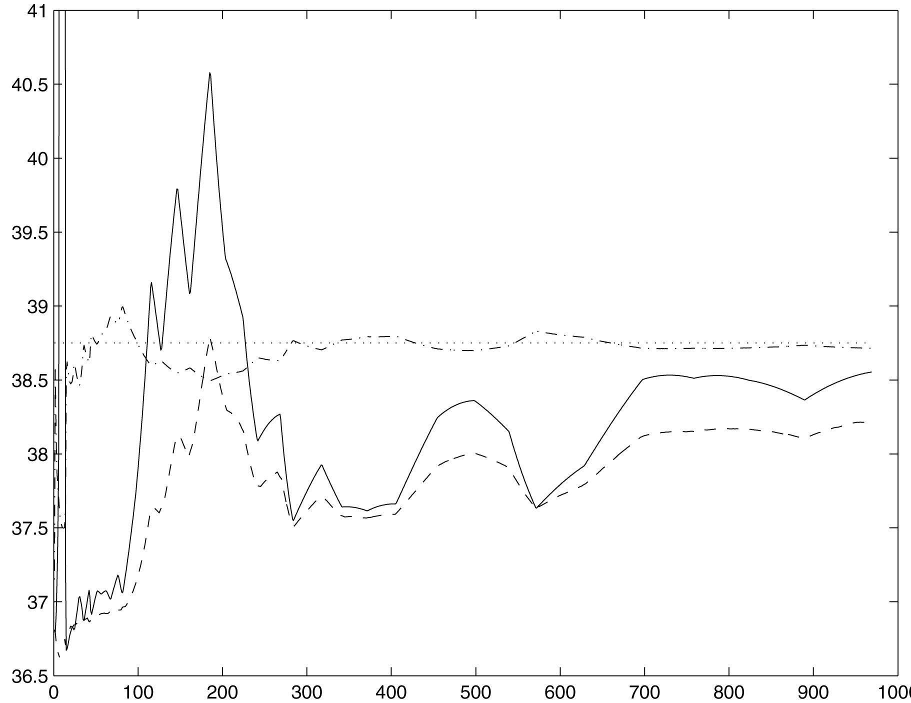

Figure 2. Endpoint estimators (in km/h) versus k for the men's 100 m: moment (solid line), maximum likelihood (dashed line), fixed $\hat{\gamma}$ (dashed–dotted line). The selected estimate is the dotted horizontal line.

women). This is due, in part, to the difference in $\hat{\gamma}$: In the women's marathon $\hat{\gamma} = -.11$, much higher than the $\hat{\gamma} = -.25$ of the women's 100-m hurdles.

We have mentioned before that Q is an estimate of $n(1 - F(X_{n,n}))$ and that Q can be computed without knowing n, the number of all athletes in the world in a specific event. If we did know n (or if a credible estimate were available), we could estimate $1 - F(X_{n,n})$, the conditional probability of achieving a new world record. This would provide an alternative measure for the quality of the current world record. Because we cannot estimate n satisfactorily, this alternative measure cannot be computed. If, however, we are willing to assume that n is the same for all events (which is not unreasonable in this context), then the order in Table 4 would not be affected if we consider $1 - F(X_{n,n})$ instead of $n(1 - F(X_{n,n}))$.

5. SENSITIVITY ANALYSIS

We briefly comment on the (lack of) sensitivity of our procedure. Initially, each of our estimates $\hat{\gamma}$, $\hat{x}^*$, and Q depends on the sample fraction k, as discussed in Section 4. This produces graphs $\hat{\gamma}(k)$, $\hat{x}^*(k)$, and $Q(k)$ as displayed in Figures 1–3. We

Table 3. Ultimate world records ("endpoint") with standard errors and the current world records

	Men			Women		
Event	Endpoint	Standard error	World record	Endpoint	Standard error	World record
Running						
100 m	9.29	.39	9.74	10.11	.40	10.49
110/100-m hurdles	12.38	.35	12.88	11.98	.19	12.21
200 m	18.63	.88	19.32	20.75	.57	21.34
400 m	—	—	43.18	45.79	1.83	47.60
800 m	1:39.65	1.44	1:41.11	1:52.28	1.39	1:53.28
1,500 m	3:22.63	3.31	3:26.00	3:48.33	2.78	3:50.46
10,000 m	—	—	26:17.53	—	—	29:31.78
Marathon	2:04:06	57	2:04:26	2:06:35	10:05	2:15:25
Throwing						
Shot put	24.80	1.25	23.12	23.70	.86	22.63
Javelin throw	106.50	10.30	98.48	72.50	2.99	71.70
Discus throw	77.00	2.85	74.08	85.00	8.10	76.80
Jumping						
Long jump	—	—	8.95	—	—	7.52
High jump	2.50	.05	2.45	2.15	.05	2.09

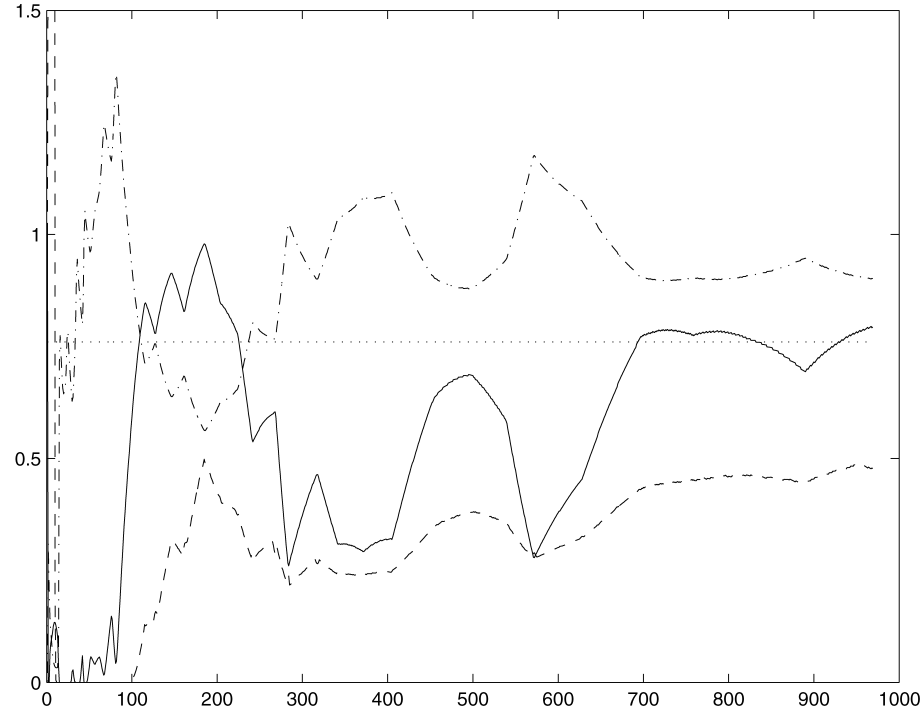

Figure 3. Q versus k for the men's 100 m: moment (solid line), maximum likelihood (dashed line), fixed $\hat{\gamma}$ (dashed–dotted line). The selected estimate is the dotted horizontal line.

then choose our estimates heuristically from the graphs by finding the appropriate (first) stable region, as described in Section 4.1. Because this method is a common procedure both in the literature and in practice, some evidence that the procedure is not sensitive to the choice of k is important, not only for the current study but also for other studies using similar graphical methods.

Table 4. Quality of world records and ordering of world record holders

Athlete	Event	Record	Year	e^{-Q}
Osleidys Menéndez	Javelin (W)	71.70	2005	.98
Haile Gebrselassie	Marathon (M)	2:04:26	2007	.95
Jan Zelezný	Javelin (M)	98.48	1996	.93
Michael Johnson	200 m (M)	19.32	1996	.92
Javier Sotomayor	High jump (M)	2.45	1993	.86
Florence Griffith-Joyner	100 m (W)	10.49	1988	.86
Yunxia Qu	1,500 m (W)	3:50.46	1993	.86
Paula Radcliffe	Marathon (W)	2:15:25	2003	.86
Marita Koch	400 m (W)	47.60	1985	.78
Jarmila Kratochvílová	800 m (W)	1:53.28	1983	.78
Wilson Kipketer	800 m (M)	1:41.11	1997	.74
Hicham El Guerrouj	1,500 m (M)	3:26.00	1998	.74
Jürgen Schult	discus (M)	74.08	1986	.74
Florence Griffith-Joyner	200 m (W)	21.34	1988	.74
Michael Johnson	400 m (M)	43.18	1999	.67
Stefka Kostadinova	High jump (W)	2.09	1987	.64
Gabriele Reinsch	Discus (W)	76.80	1988	.55
Junxia Wang	10,000 m (W)	29:31.78	1993	.50
Natalya Lisovskaya	Shot put (W)	22.63	1987	.50
Asafa Powell	100 m (M)	9.74	2007	.47
Randy Barnes	Shot put (M)	23.12	1990	.45
Kenenisa Bekele	10,000 m (M)	26:17.53	2005	.33
Yordanka Donkova	100-m hurdles (W)	12.21	1988	.33
Galina Chistyakova	Long jump (W)	7.52	1988	.30
Mike Powell	Long jump (M)	8.95	1991	.27
Xiang Liu	110-m hurdles (M)	12.88	2006	.20

Clearly, it is desirable to have extreme-value-type estimators that do not depend too much on (small) changes in the choice of k. The value of k itself (the horizontal axis in the graphs) is not of interest; it is the *vertical* level that is important. If we can choose a credible stable region, where the estimates do not fluctuate much as a function of k, then the actual value of k does not matter much, and the estimates are not sensitive.

We report on two simulation experiments, one for $\gamma > 0$ (the "heavy-tailed" situation most common in applications) and one for $\gamma < 0$ (our application). Our conclusion is that the estimates are reasonably well determined by the heuristic procedure. Our procedure being graphical and nonautomatic, we cannot perform hundreds of replications. Hence, we take 10 samples of size 1,000 each.

In our first experiment we take samples from the absolute value of the standard Cauchy distribution with density

$$f(x) = \frac{2}{\pi(1+x^2)}, \qquad x \geq 0.$$

This distribution is in the max-domain of attraction of G_γ, with extreme-value index $\gamma = 1$. We use the moment estimator to estimate γ. As we see from Figure 4, the heuristic approach with this sample leads to a good estimator, namely, $\hat{\gamma} = .95$. This could, however, be a coincidence. Hence, we perform the estimation for each of our 10 samples and obtain a mean estimate of .998 with a standard deviation of .101.

In our second experiment we consider a distribution with a finite right endpoint, that is, where $\gamma \leq 0$. In fact, we choose

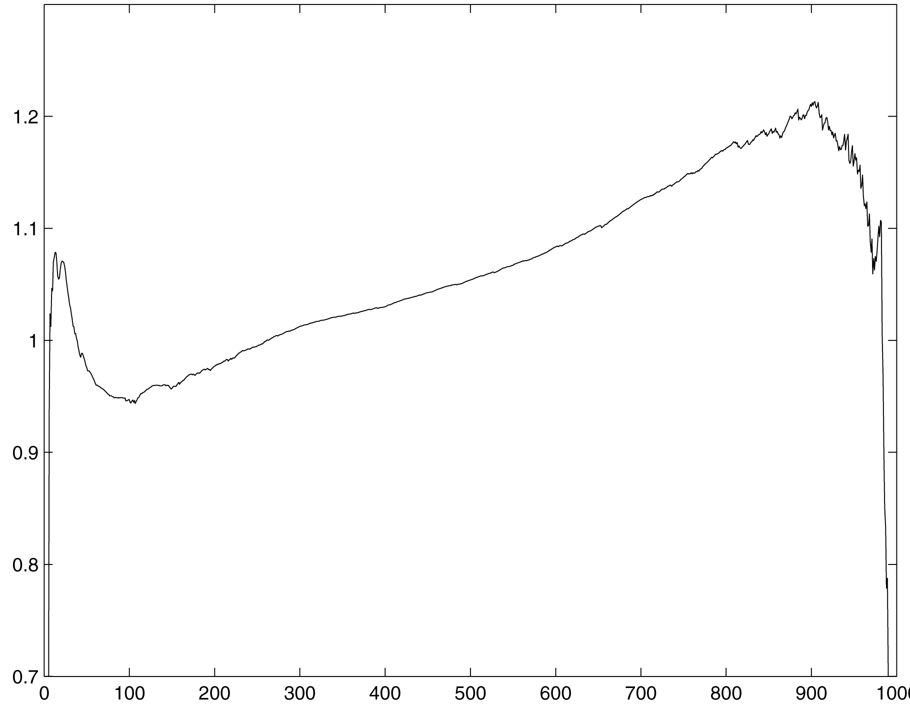

Figure 4. Moment estimator versus k for the absolute value of Cauchy random variables ($n = 1{,}000$, $\gamma = 1$).

an extreme-value index that corresponds to the tail thickness in our own application, namely, $\gamma = -.20$. We simulate from the reversed Burr distribution

$$F(x) = 1 - \left(\frac{\beta}{\beta + (x^* - x)^{-\tau}}\right)^{\lambda}, \qquad x < x^*,$$

with parameters $\beta > 0$, $\tau > 0$, $\lambda > 0$, and endpoint $x^* > 0$. For the simulation we choose $\beta = 1$, $\tau = 10$, $\lambda = 1/2$, and $x^* = 10$, so that $\gamma = -1/(\tau\lambda) = -.20$.

We estimate γ with the moment estimator. Then we estimate x^* in two ways, first by using the estimate of γ depending on k and then by using a fixed estimate of γ, chosen from the $\hat{\gamma}$ versus k plot (as in Sec. 4). Again we take 10 samples of size 1,000. The results of one sample are graphed in Figure 5. Repeating this exercise 10 times, we obtain a mean of $\hat{\gamma}$ of $-.216$ with a standard deviation of .049. For the estimates of x^* based on the fixed estimate of γ, we find a mean of 9.985 with a standard deviation of .136. The estimates of x^* without fixing $\hat{\gamma}$ have a mean of 10.025 with a standard deviation of .132. We see that the estimates are close to the true value $x^* = 10$ and that the two procedures are comparable in terms of statistical performance. However, as can be verified from Figure 5, the problem of finding a stable region in the plot of $\hat{x}^*$ is made easier by using the fixed-$\hat{\gamma}$ procedure.

Based on these simulations and many other experiments not discussed here, we conclude that the graphical heuristic procedure is sufficiently insensitive with respect to the choice of k. Some caution and use of common sense are, however, still required.

6. SUMMARY AND CONCLUSION

Because a record is an "extreme value," it seems reasonable to apply extreme-value theory to world records in athletics. In particular, we attempted to answer two questions: (1) What is the ultimate world record in a specific athletic event given today's state of the art? (2) How "good" is the current world record (i.e., how difficult is it to improve)? We considered 14 events for both men and women: 8 running events, 3 throwing events, and 3 jumping events, such that all events that make up the decathlon and the heptathlon are included, supplemented by the 10,000 m and the marathon.

We do not predict the ultimate world record from the development of world records over time (as is the common procedure). In fact, we use more information than the sequential world records, namely, as many personal bests of top athletes as we could obtain. Combining various publicly available lists gives us, on average, 730 male top athletes and 502 female top athletes per event. We consider our data as the upper order statistics of a random sample and use semiparametric statistical techniques based on extreme-value theory to answer our questions.

Our estimated ultimate world records (Table 3) show that substantial improvements over the current record are still possible, for men especially in the shot put (7.3%) and the javelin (8.1%) and for women especially in the marathon (7.0%) and the discus (10.7%). To achieve these ultimate records is not easy, because the estimated distribution function is quite flat near the right endpoint due to the relatively high values of $\hat{\gamma}$. On the other hand, very little improvement is possible in the 800 m and 1,500 m for women (both .9%) and particularly in the marathon for men (.3%), all having a relatively low $\hat{\gamma}$.

We measure the quality of the current world records by

$$n\big(F(x^*) - F(X_{n,n})\big) = n(1 - F(X_{n,n})),$$

the expected number of exceedances of the world record. High-quality world records are difficult to improve, whereas records

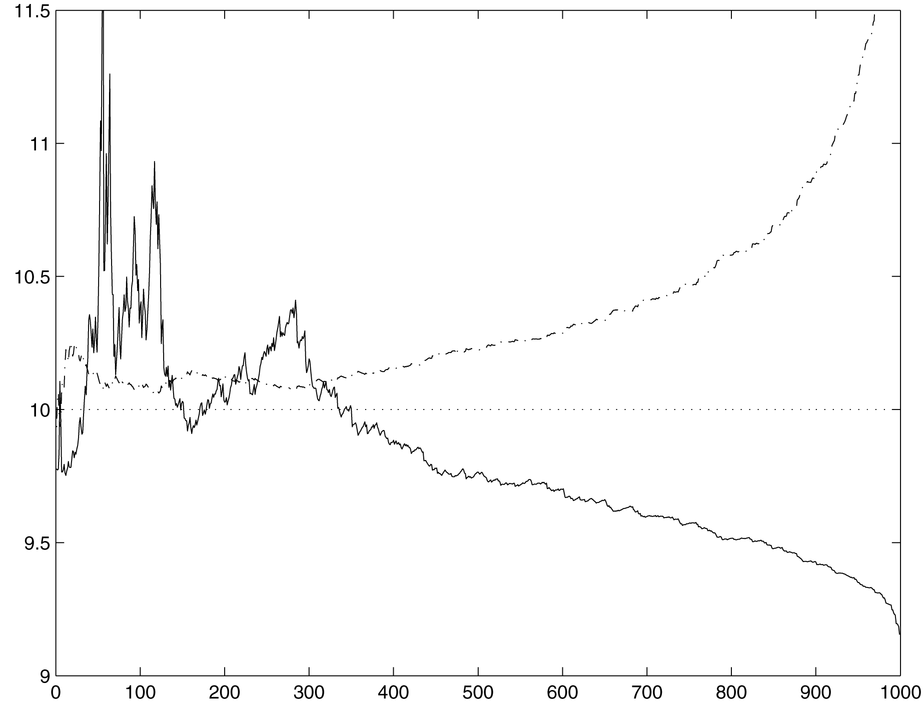

Figure 5. Endpoint estimators versus k for 1,000 reversed Burr random variables: moment (solid line), fixed $\hat{\gamma}$ (dashed–dotted line), true $x^* = 10$ (dotted horizontal line).

with low quality are likely to be improved in the near future. The javelin throw for both men and women and the marathon and 200 m for men have records with the highest estimated quality. In the near future, improvements in the 110/100-m hurdles and the long jump records for both men and women and the 10,000 m for men are likely.

The main theoretical result of our article is Theorem 1, where we obtain the limiting distribution of an "estimator" of $n(1 - F(X_{n,n}))$. It turns out that the limiting distribution is completely determined by $X_{n,n}$ and that it is not affected by the estimation of F.

Our method for measuring the quality of world records could be a basis for an improved method of combining performances in multi-event sports like the decathlon and speed skating. Outside sports statistics, endpoint estimation can be applied in many other scientific disciplines. For example, de Haan (1981) estimated the minimum of a function, and Aarssen and de Haan (1994) considered the maximal life span of humans.

APPENDIX: PROOF OF THEOREM 1

We observe first that the continuity of F in conjunction with the probability integral transform implies that the three random variables

$$n(1 - F(X_{n,n})), \qquad n(1 - U_{n,n}), \qquad nU_{1,n}$$

have the same distribution, where $U_{n,n}$ and $U_{1,n}$ denote the maximum and the minimum, respectively, of a random sample of size n from the uniform $(0, 1)$ distribution. It is easy to see that $nU_{1,n} \xrightarrow{d} \text{Exp}(1)$. Therefore, the same is true for $n(1 - F(X_{n,n}))$. Thus, it suffices to show that

$$\frac{Q_j}{n(1 - F(X_{n,n}))} \xrightarrow{p} 1,$$

which, in turn, is implied if we can show that

$$R := \frac{\widetilde{Q}_j}{n(1 - F(X_{n,n}))} \xrightarrow{p} 1,$$

where

$$\widetilde{Q}_j := k\left(1 + \hat{\gamma}_j \frac{X_{n,n} - \hat{b}}{\hat{a}_j}\right)^{-1/\hat{\gamma}_j}, \qquad j = 1, 2.$$

We set $\hat{p} := 1 - F(X_{n,n})$ and $\hat{d}_n := k/(n\hat{p})$, we write $a := a_{n/k}$ and $b := b_{n/k}$, and we suppress j in the remainder of the proof ($j = 1, 2$). Now define

$$A_n := \sqrt{k}\left(\frac{\hat{a}}{a} - 1\right),$$

$$B_n := \sqrt{k}\left(\frac{X_{n-k,n} - b}{a}\right),$$

$$\Gamma_n := \sqrt{k}(\hat{\gamma} - \gamma)$$

and notice from the conditions of Theorem 1 that all three are $O_p(1)$.

We first consider the case where $\gamma \neq 0$. Rewrite R as

$$R = \hat{d}_n\left[1 + \frac{\hat{\gamma}}{\gamma}\frac{a}{\hat{a}}(\hat{d}_n^{\gamma} - 1) \times \left(\frac{X_{n,n} - b}{a} \cdot \frac{\gamma}{\hat{d}_n^{\gamma} - 1} - \frac{\hat{b} - b}{a} \cdot \frac{\gamma}{\hat{d}_n^{\gamma} - 1}\right)\right]^{-1/\hat{\gamma}}.$$

Using the facts that $X_{n,n} = U(1/\hat{p})$ and $b = U(n/k)$ and defining

$$S := \left(\frac{U(1/\hat{p}) - U(n/k)}{a}\frac{\gamma}{\hat{d}_n^{\gamma} - 1} - 1\right)\Big/ A\left(\frac{n}{k}\right),$$

$$Y_n := \frac{\hat{\gamma}}{\gamma}\frac{a}{\hat{a}} = \left(1 + \frac{\Gamma_n}{\gamma\sqrt{k}}\right)\Big/\left(1 + \frac{A_n}{\sqrt{k}}\right),$$

we obtain (see also prop. 8.2.9 in de Haan and Ferreira 2006)

$$
\begin{aligned}
R &= \hat{d}_n\left[1 + Y_n(\hat{d}_n^{\gamma} - 1)\left(1 + A\left(\frac{n}{k}\right)S - \frac{\gamma}{\hat{d}_n^{\gamma} - 1}\frac{B_n}{\sqrt{k}}\right)\right]^{-1/\hat{\gamma}} \\
&= \hat{d}_n\left[1 + Y_n(\hat{d}_n^{\gamma} - 1)\left(1 + A\left(\frac{n}{k}\right)S\right) - \gamma Y_n\frac{B_n}{\sqrt{k}}\right]^{-1/\hat{\gamma}} \\
&= \left[\hat{d}_n^{-\hat{\gamma}} + \hat{d}_n^{-\hat{\gamma}}(\hat{d}_n^{\gamma} - 1)Y_n\left(1 + A\left(\frac{n}{k}\right)S\right)\right. \\
&\qquad \left. - \hat{d}_n^{-\hat{\gamma}}\gamma Y_n\frac{B_n}{\sqrt{k}}\right]^{-1/\hat{\gamma}} \\
&= \left[\hat{d}_n^{\gamma-\hat{\gamma}}\left(\hat{d}_n^{-\gamma}\left[1 - Y_n\left(1 + A\left(\frac{n}{k}\right)S\right) - \gamma Y_n\frac{B_n}{\sqrt{k}}\right]\right.\right. \\
&\qquad \left.\left. + Y_n\left(1 + A\left(\frac{n}{k}\right)S\right)\right)\right]^{-1/\hat{\gamma}} \\
&=: [T_1(T_2 + T_3)]^{-1/\hat{\gamma}}.
\end{aligned}
$$

We have

$$T_1 := \hat{d}_n^{\gamma-\hat{\gamma}} = \hat{d}_n^{-\Gamma_n/\sqrt{k}} = \exp\left(-\frac{\Gamma_n}{\sqrt{k}}\log\frac{k}{n\hat{p}}\right) \xrightarrow{p} 1.$$

From (5) with $\rho < 0$, it follows that $S \xrightarrow{p} -1/(\rho + \min(\gamma, 0))$; see de Haan and Ferreira (2006, lemma 4.3.5). Hence, $A(n/k)S \xrightarrow{p} 0$ and $T_3 \xrightarrow{p} 1$. Further, because

$$\sqrt{k}\left(1 - Y_n\left(1 + A\left(\frac{n}{k}\right)S\right) - \gamma Y_n\frac{B_n}{\sqrt{k}}\right) = O_P(1)$$

and

$$\frac{1}{\sqrt{k}\hat{d}_n^{\gamma}} = \frac{(n\hat{p})^{\gamma}}{\sqrt{k}k^{\gamma}} \xrightarrow{p} 0,$$

we see that $T_2 \xrightarrow{p} 0$. Hence, $R \xrightarrow{p} 1$ for $\gamma \neq 0$.

Next, we consider the case $\gamma = 0$. By convention, $(\hat{d}_n^0 - 1)/0$ is interpreted as $\log\hat{d}_n$. Then

$$R = \hat{d}_n\left[1 + \hat{\gamma}\log\hat{d}_n\frac{a}{\hat{a}}\left(1 + A\left(\frac{n}{k}\right)S - \frac{1}{\log\hat{d}_n}\frac{B_n}{\sqrt{k}}\right)\right]^{-1/\hat{\gamma}},$$

so that

$$\log R = \log\hat{d}_n - \frac{1}{\hat{\gamma}}\log\left(1 + \hat{\gamma}\log\hat{d}_n\left[1 + O_P\left(\frac{1}{\sqrt{k}}\right)\right]\right).$$

It follows that

$$|\log R| = O_P\left(\frac{1}{\sqrt{k}}\right)\log\hat{d}_n + O_P(|\hat{\gamma}|\log^2\hat{d}_n) \xrightarrow{p} 0$$

and, hence, that $R \xrightarrow{p} 1$ for $\gamma = 0$. This completes the proof.

[Received September 2006. Revised October 2007.]

REFERENCES

Aarssen, K., and de Haan, L. (1994), "On the Maximal Life Span of Humans," *Mathematical Population Studies*, 4, 259–281.

Barão, M. I., and Tawn, J. A. (1999), "Extremal Analysis of Short Series With Outliers: Sea-Levels and Athletics Records," *Applied Statistics*, 48, 469–487.

de Haan, L. (1981), "Estimation of the Minimum of a Function Using Order Statistics," *Journal of the American Statistical Association*, 76, 467–469.

de Haan, L., and Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, New York: Springer.

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 Hüsler, J. (2002), "Testing Extreme Value Conditions," *Extremes*, 5, 71–85.

Dijk, V., and de Haan, L. (1992), "On the Estimation of the Exceedance Probability of a High Level," in *Order Statistics and Nonparametrics: Theory and Applications*, eds. P. K. Sen and I. A. Salama, Amsterdam: North-Holland, pp. 79–92.

Drees, H., de Haan, L., and Li, D. (2006), "Approximations to the Tail Empirical Distribution Function With Application to Testing Extreme Value Conditions," *Journal of Statistical Planning and Inference*, 136, 3498–3538.

Ferreira, A., de Haan, L., and Peng, L. (2003), "On Optimising the Estimation of High Quantiles of a Probability Distribution," *Statistics*, 37, 401–434.

Robinson, M. E., and Tawn, J. A. (1995), "Statistics for Exceptional Athletics Records," *Applied Statistics*, 44, 499–511.

Smith, R. L. (1987), "Estimating Tails of Probability Distributions," *The Annals of Statistics*, 15, 1174–1207.

——— (1988), "Forecasting Records by Maximum Likelihood," *Journal of the American Statistical Association*, 83, 331–338.