RECORDS IN ATHLETICS THROUGH EXTREME –VALUE THEORY

By John H.J. Einmahl, Jan R. Magnus

September 2006

TILBURG ◆ UNIVERSITY

Records in athletics
through extreme-value theory[*]

John H.J. Einmahl
Tilburg University

Jan R. Magnus
Tilburg University

September 20, 2006

Abstract: In this paper we shall be 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 athletics event (such as the 100m for men or the high jump for women), given today's state of the art? Our second question is: how 'good' is a current athletics world record? An answer to the second question will also enable us to compare the quality of world records in different athletics events. We shall consider these questions for each of twenty-eight events (fourteen 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 nonparametric nature, but some 'weak regularity' of the tail of the distribution function will be assumed. We will 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 top performances over time, this will not be our method. Instead, we shall only use the top performances themselves. Our estimated ultimate world record tells us what, in principle, is possible now, given today's knowledge, material (shoes, suits, equipment), and drugs laws.

JEL codes: L83, C13, C14.
Key words: Endpoint estimation, exceedance probability, ranking, statistics of extremes, world record.

---

1   Introduction

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

These questions all relate to extremes. In this paper we shall be 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 athletics event (such as the 100m for men or the high jump for women), given today's state of the art? Our second question is: how 'good' is a current athletics world record? An answer to the second question will enable us to compare the quality of world records in different athletics events.

We shall approach these two extremes-related questions with the probability theory of extreme values and the corresponding statistical techniques. The statistical model is of nonparametric nature, but some 'weak regularity' of the tail of the distribution function will be 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 3000m event and drugs-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 athletics records for the mile and the marathon.

Almost all attempts to predict an ultimate world record are based on the development of top performances over time. This will *not* be our method and we are not trying to predict the world record in the year 2525. Instead, we shall only use the top performances themselves (see Table 1). Our estimated ultimate record tells us what, in principle, is possible *now*, given today's knowledge, material (shoes, suits, equipment), and drugs laws.

Our selection of athletics 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 ten events, at the 2004 Athens Olympics, male athletes competed in twenty-four events:

Running: 100m, 200m, 400m, 800m, 1500m, 5000m, 10,000m, marathon, 110m hurdles, 400m hurdles, 20km walk, 50km walk, steeplechase;

Throwing: shot put, javelin throw, discus throw, hammer throw;

Jumping: long jump, high jump, pole vault, triple jump;

two relay events (4×100m and 4×400m), and the decathlon.

Women first competed at the 1924 Olympics in five events, but in Athens they competed in twenty-two events (not in the steeplechase and the 50km walk). Furthermore, women run 100m hurdles (instead of 110m) and compete in a heptathlon (instead of a decathlon).

For the purposes of our study we selected fourteen events: eight running events, three throwing events, and three jumping events, as follows:

Running: 100m (D), 200m (H), 400m (D), 800m (H), 1500m (D), 10,000m, marathon, 110/100m hurdles (DH);

Throwing: shot put (DH), javelin throw (DH), discus throw (D);

Jumping: long jump (DH), high jump (DH), pole vault (D).

The selection was made such that all events that make up the decathlon (D) and heptathlon (H) are included, supplemented by the 10,000m and the marathon.

The paper 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 (Theorem 1). In Section 4 we apply the theory to the data and answer our two questions. An appendix contains the proof of Theorem 1.

2 The data

For each of the twenty-eight events (fourteen for both men and women) we collected data of the personal best of as many of the top athletes in each event as we could. We emphasize that we are primarily interested in personal bests and not in the development of the world record. As a consequence, 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 are 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 onwards. These two sites provide a list of the top athletes (and their results) per event. The lists of both sites have been merged and multiple entries of the same athlete have been removed. The Swedish website provides additional information under the headings 'doubtful wind reading,' 'doubtful timing,' and 'subsequent to drugs disqualification.' These concern records not recognized by the IAAF, and consequently 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 not included by us. Times clocked by hand from the period when electronic timing was *not* available are recorded with a precision of 0.1 seconds (rather than 0.01 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 lists for 2001–2004 and 2005 (until April 30) from the second website. We considered the worst performance in each of the six lists. The best of these worst performances was taken as the lower bound for each event. This guarantees that there are no 'holes' in the combined list. Next we removed all multiple entries of the same athlete, so that each athlete appears only once with his or her personal best. 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 per event and the depth covered. The data consist of about 10,000 observations for the men and 7000 observations for the women. On average, we have about twice as many observations for the running events than for jumping and throwing. Especially 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 thus transformed running times to speeds, so that the higher the speed, the better. Thus, 10.00 seconds on the 100m is transformed to a speed of 36.00 km/h.

Some data occur in clusters, especially in the shorter distances such as 100m. These clusters occur not because the actual times are the same, but because the timing is imperfect. Since these 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 on the 100m. Then we smooth

	Men			Women		
Event	depth	worst	best	depth	worst	best
100m	970	10.30	9.77	578	11.38	10.49
110/100m hurdles	805	13.83	12.91	432	13.20	12.21
200m	780	20.66	19.32	561	23.14	21.34
400m	658	45.74	43.18	538	52.02	47.60
800m	722	1:46.61	1:41.11	537	2:01.05	1:53.28
1500m	781	3:38.74	3:26.00	531	4:09.03	3:50.46
10,000m	1239	28:30.03	26:17.53	876	33:04.00	29:31.78
marathon	1546	2:13:36	2:04:55	1024	2:36:06	2:15:25
shot put	392	19.80	23.12	223	18.42	22.63
javelin throw	422	77.00	98.48	279	54.08	71.70
discus throw	335	62.84	74.08	222	62.52	76.80
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	5.01

Table 1: Data summary.

these m results over the interval (10.045, 10.055) by

$$d_j = 10.045 + 0.01\frac{2j-1}{2m} \qquad (j = 1, \ldots, m).$$

3 Extreme-value theory

Consider one athletics event, say the 100m for men, and let $X_1, X_2, \ldots, X_n$ denote the personal bests of all n male 100m athletes in the world. The precise definition of 'athlete' is left vague, and therefore the definition and possible measurement of n is 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 i.i.d. 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 transform running times to speeds, so that the higher the jump, the farther the throw, and the higher the speed, the better.) As an analogue 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

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

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 (1) holds, then we say that F is in the max-domain of attraction of G_γ; 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

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

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' stands for the left-continuous inverse.

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

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

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 analogue.

This paper 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

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

For large t we can write heuristically

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

Since $\gamma < 0$, this yields, for very 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

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

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 $\gamma > -1/2$), these estimators 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, Einmahl, and de Haan (1989, p. 1851). The estimation of extreme quantiles and endpoints has been thoroughly studied; see de Haan and Ferreira (2006, Chapter 4) for a detailed account.

The second purpose of this paper 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 i.i.d. random variables from F that are independent of the X_i. The lower this number, the better is the world record. It might seem more natural to measure the quality of the world record by

$x^* - X_{n,n}$. This quantity can be infinite, however. Moreover, and more importantly, it does not take into account the tail behavior of F. Observe that our measure of quality is equal to $n(F(x^*) - F(X_{n,n}))$.

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, Chapter 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 of constant sign satisfying $\lim_{t \to \infty} A(t) = 0$, such that for $x > 0$:

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

with $\rho \leq 0$, 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 the three random variables

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

are all $O_p(1)$ for $j = 1, 2$. 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. We also note that any estimators of $a_{n/k}, b_{n/k}$ and γ can be used as long as the three $O_p(1)$ requirements are fulfilled.

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: (a) what are the ultimate world records, and (b) 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 fourteen selected athletics events, for men and women respectively. In order to estimate γ we must first know whether it exists, that is, that relation (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 relation (1). Hence we drop the pole vault from our analysis and continue with 13×2 athletics 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 one). It is a difficult practical problem to find a good value for k to base the estimator on. Typically, for small k the estimator has a high variance and the plot is unstable; for large k the estimator has a bias. Our strategy has been to find the first stable region in the plot of the estimator versus k and to use the corresponding vertical level as our estimate.

This is illustrated in Figure 1, where we plot $\hat{\gamma}_1$ and $\hat{\gamma}_2$ as a function of
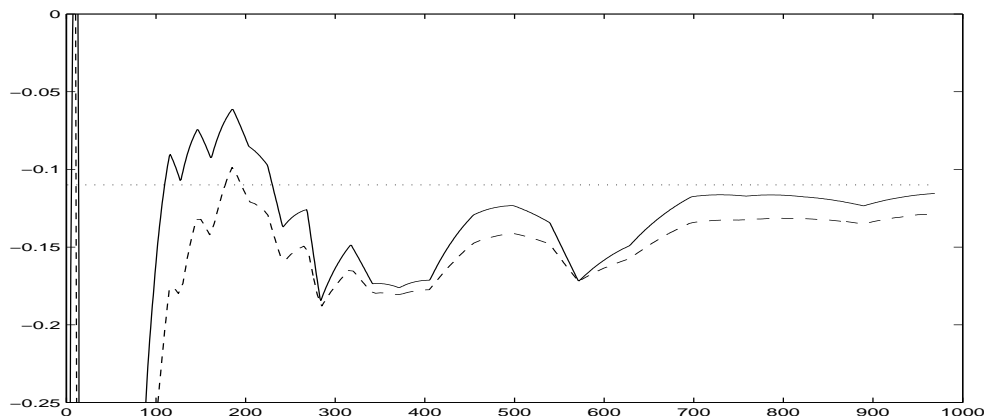

Figure 1: Moment estimator (solid line) and maximum likelihood estimator (dashed line) versus k for the men's 100m. The selected estimate is the dotted horizontal line.

k in the 100m 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 twenty-five events) what our estimate for γ should be. We therefore also consider two additional estimators that have 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, since it is well-known that $M_n^{(1)}$ (the Hill estimator) is a good estimator when $\gamma > 0$. 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 $N_n^{(r)} := \frac{1}{k}\sum_{i=0}^{k-1}(X_{n-i,n} - X_{n-k,n})^r$ 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. For example, for the

Event	Men	Women
100m	-0.11	-0.14
110/100m hurdles	-0.16	-0.25
200m	-0.11	-0.18
400m	-0.07	-0.15
800m	-0.20	-0.26
1500m	-0.20	-0.29
10,000m	-0.04	-0.08
marathon	-0.27	-0.11
shot put	-0.18	-0.30
javelin throw	-0.15	-0.30
discus throw	-0.23	-0.16
long jump	0.06	-0.07
high jump	-0.20	-0.22
pole vault	—	—

Table 2: Estimates of γ.

moment estimator in the men's 100m 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 zero 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.

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 100m in Figure 2. A much more stable
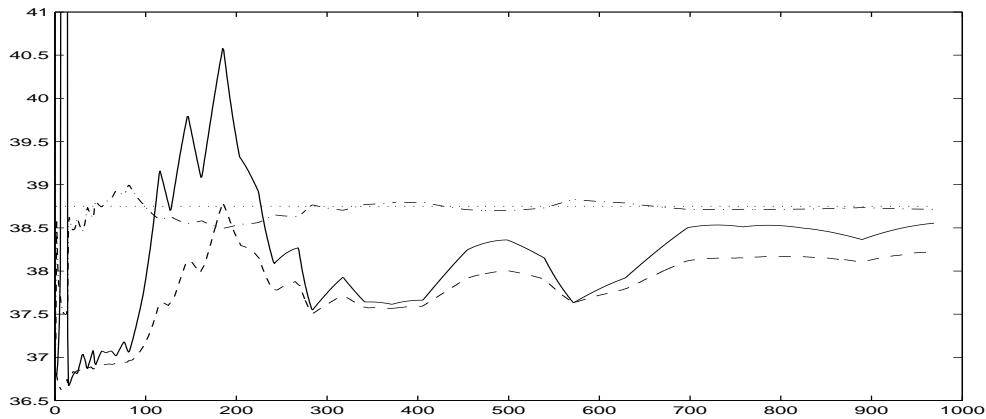

Figure 2: Endpoint estimators (in km/h) versus k for the men's 100m: moment (solid line), maximum likelihood (dashed line), fixed $\hat{\gamma}$ (dashed-dotted line). The selected estimate is the dotted horizontal line.

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 sequel the world records are updated until the date on this manuscript. Also in Table 3 a rough estimate of the standard error of $\hat{x}^*$, based on the asymptotic normality of $\hat{x}_1^*$ (see Section 3), is presented.

Since we have assumed that $\gamma < 0$, we do not present an estimate for

Event	Men			Women		
	endpoint	prec.	world rec.	endpoint	prec.	world rec.
100m	9.29	0.39	9.77	10.11	0.40	10.49
110/100m h.	12.38	0.35	12.88	11.98	0.19	12.21
200m	18.63	0.88	19.32	20.75	0.57	21.34
400m	—	—	43.18	45.79	1.83	47.60
800m	1:39.65	1.44	1:41.11	1:52.28	1.39	1:53.28
1500m	3:22.63	3.31	3:26.00	3:48.33	2.78	3:50.46
10,000m	—	—	26:17.53	—	—	29:31.78
marathon	2:04:06	57	2:04:55	2:06:35	10:05	2:15:25
shot put	24.80	1.25	23.12	23.70	0.86	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
long jump	—	—	8.95	—	—	7.52
high jump	2.50	0.05	2.45	2.15	0.05	2.09

Table 3: Ultimate world records, their precisions, and the current world records.

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

It appears that not much progress is possible in the men's marathon (only 49 seconds), but much more in the women's marathon (almost nine minutes). In contrast, the javelin throw for women appears to be close to its frontier (80cm), while for the men an improvement of eight meters is possible.

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 twenty-six 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 100m, 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. Since this transformation is decreasing, a higher value of e^{-Q} means that the
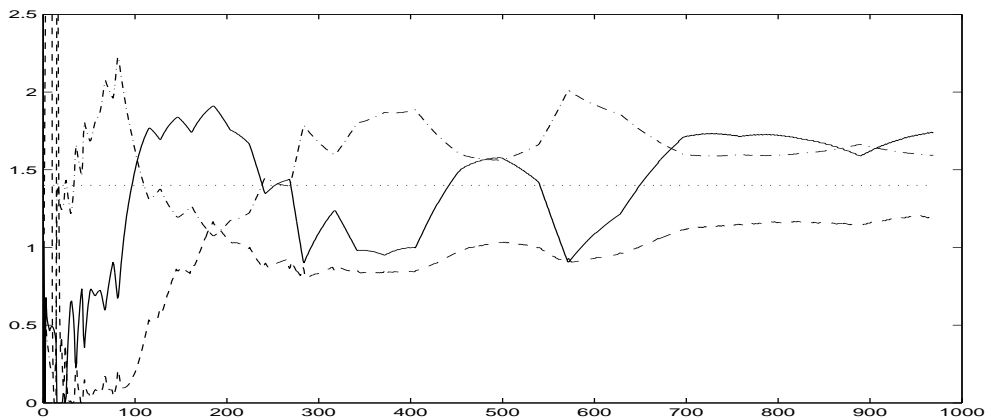
Figure 3: Q versus k for the men's 100m: moment (solid line), maximum likelihood (dashed line), fixed $\hat{\gamma}$ (dashed-dotted line). The selected estimate is the dotted horizontal line.

record is better. In Table 4, the values of e^{-Q} are presented for the twenty-six world records and the corresponding world record holders. Although far from perfect, some 'uniformity on $(0,1)$' of the twenty-six values can be observed. Note that the list is led by the javelin throwers, but in general the various events as well as the gender are well mixed.

The table demonstrates that a world record can have a high quality while it can still be much improved (like the marathon for women), but that it can also be close to its limit while of relatively low quality (like the 100m hurdles for women). This is due, in part, to the fact that the $\hat{\gamma} = -0.11$ of the women's marathon is much higher than the -0.25 of the women's 100m 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. Since we cannot estimate n, 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 relative qualities reported in Table 4 would not change.

athlete	event	record	year	e^{-Q}
Osleidys Menéndez	javelin (W)	71.70	2005	0.98
Jan Zelezný	javelin (M)	98.48	1996	0.93
Michael Johnson	200m (M)	19.32	1996	0.92
Javier Sotomayor	high jump (M)	2.45	1993	0.86
Florence Griffith-Joyner	100m (W)	10.49	1988	0.86
Yunxia Qu	1500m (W)	3:50.46	1993	0.86
Paula Radcliffe	marathon (W)	2:15:25	2003	0.86
Marita Koch	400m (W)	47.60	1985	0.78
Jarmila Kratochvílová	800m (W)	1:53.28	1983	0.78
Wilson Kipketer	800m (M)	1:41.11	1997	0.74
Hicham El Guerrouj	1500m (M)	3:26.00	1998	0.74
Jürgen Schult	discus (M)	74.08	1986	0.74
Florence Griffith-Joyner	200m (W)	21.34	1988	0.74
Michael Johnson	400m (M)	43.18	1999	0.67
Stefka Kostadinova	high jump (W)	2.09	1987	0.64
Paul Tergat	marathon (M)	2:04:55	2003	0.61
Gabriele Reinsch	discus (W)	76.80	1988	0.55
Junxia Wang	10,000m (W)	29:31.78	1993	0.50
Natalya Lisovskaya	shot put (W)	22.63	1987	0.50
Randy Barnes	shot put (M)	23.12	1990	0.45
Kenenisa Bekele	10,000m (M)	26:17.53	2005	0.33
Yordanka Donkova	100m h. (W)	12.21	1988	0.33
Galina Chistyakova	long jump (W)	7.52	1988	0.30
Mike Powell	long jump (M)	8.95	1991	0.27
Asafa Powell	100m (M)	9.77	2005	0.25
Xiang Liu	110m h. (M)	12.88	2006	0.20

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

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

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

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

$$A_n := \sqrt{k} \left(\frac{\hat{a}}{a} - 1 \right), \quad B_n := \sqrt{k} \left(\frac{X_{n-k,n} - b}{a} \right), \quad \Gamma_n := \sqrt{k}(\hat{\gamma} - \gamma),$$

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

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

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

$$S := \frac{\frac{U(1/\hat{p}) - U(n/k)}{a} \frac{\gamma}{\hat{d}_n^\gamma - 1} - 1}{A \left(\frac{n}{k} \right)}, \quad Y_n := \frac{\hat{\gamma}}{\gamma} \frac{a}{\hat{a}} = \frac{1 + \frac{\Gamma_n}{\gamma \sqrt{k}}}{1 + \frac{A_n}{\sqrt{k}}},$$

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

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

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

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

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

$$=: [T_1(T_2 + T_3)]^{-1/\hat{\gamma}}.$$

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 $T_3 \xrightarrow{p} 1$. Finally we note that

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

and that

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

This implies that $T_2 \xrightarrow{p} 0$ and completes the proof 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,

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

and hence

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

It follows that

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

thus concluding the proof.

References

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.

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, 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.

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

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.

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

Department of Econometrics & OR and CentER
Tilburg University
P.O. Box 90153
5000 LE Tilburg
The Netherlands

j.h.j.einmahl@uvt.nl
magnus@uvt.nl