CCSO Centre for Economic Research

University of Groningen

CCSO Working Papers **June, 2006**

CCSO Working paper 2006/04

Modelling the development of world records in running

Gerard H. Kuper
Department of Economics, University of Groningen

Elmer Sterken
Department of Economics, University of Groningen

CCSO Working Papers
http://www.eco.rug.nl/ccso

Modelling the Development of World Records in Running

Gerard H. Kuper Elmer Sterken*

June 2006

Abstract

We model the development of world records of metric running events from the 100 meter dash to the marathon for men and women. First, we review methods to fit time-series curves of world records in general. We discuss methods to estimate curves and review candidate functional forms that fit the systematic shape of the progress of world records. Next, we fit the asymmetric Gompertz-curves for 16 events and compute implied limit values. In order to assess the implied limits we use the Francis (1943)-model to relate limit records and distance in a log-log specification. We compare men and women and conclude that there is a fixed difference in record times between the two sexes. Finally, using the log-log relationship between time and distance we calculate the development of the world record of the mile in a robustness check.

JEL-classification: C2.

Key-words: Curve fitting, lower bounds, running.

*University of Groningen, Department of Economics, P.O. Box 800, 9700 AV Groningen, The Netherlands. Phone: +31-50-363-3723, E-mail: e.sterken@rug.nl

1 Introduction

Over the past 50 years multiple studies have made attempts to model the development of world or Olympic records, for instance for running events. Deakin (1967) explored progress in the mile record, Chatterjee and Chatterjee (1982) for the 100, 200, 400, and 800 meter in the Olympic Games, Blest (1996) for running distances of 100 meter to the marathon, and recently Nevill and Whyte (2005) for 800 meter to the marathon by male and female runners. Modelling world records through time attracts attention from different perspectives. First, all studies are inspired by the apparent problems in the analysis of world records. In terms of time used to complete an event for instance we only observe nonpositive changes. These nonpositive changes can be very infrequent: world records sometimes survive for about 20 to 25 years. But at some instances improvements are really substantial, leading to extreme values in the distribution of the first difference of the series. Or technological innovation of sports gear shifts up the human frontier; an example is the klapskate in speed skating (see Kuper and Sterken, 2003). A second element in the analysis is the interest in the ultimate human performance. Given the development of the world record up to now, can we predict the fastest time ever? And thirdly, can we compare contemporaneous performances? How does the world record 10K running for men compare to the 5K for men? Is there a phase difference in the development of records of the various events? And can women outperform men in the far future?

Observing the time series of the world record of a well-developed sports event reveals an inverted S-shape pattern (we will treat all developments of records as monotonic declining time series of the time used to complete an event). In the early phase of the development of running events, competition is not fierce, and amateurism dominates. At the inflection point the rate of progress is large, because more sportsmen get involved, more professional help is available, rewards become more visible, etc. After this rapid development phase there is a phase of saturation. It is hard to improve the record and only at a few instances a highly talented individual is able to break it. For some sports events we do not observe such a shape of the time series, because the development is much faster through cross-fertilization (e.g. the 3000 meter steeple for women). For some events typical observations are available for some pieces of the curve. This sometimes leads to the use of simple piecewise linear techniques to model the development of world records. These linear approximations are computationally attractive, but theoretically poor. Nevill and Whyte (2005) contribute to the debate on whether linear approximations are helpful in describing world records. Tatem et al. (2004) use linear approximations of the development of the best male and female 100 meter sprints at the Olympic Games and conclude that in 2156 women will run as fast as men. Whipp and Ward (1992) also employed a linear approximation of marathon times of men and women and predicted in 1992 that female marathon runners would run as fast as men in 1998. Of course linear approximations cannot be correct: world record times cannot become negative. So the question is to find more biologically sound and statistically robust nonlinear (S-shaped) functions that provide a superior fit (see Nevill and Whyte, 2005).

Besides finding the best fit, most studies are trying to get insight into the upper-limit of speed (of running) or the lower limit of time to be used to complete an event. For some functional forms, a nice limit value can be derived from the properties of the functions.

For instance Kuper and Sterken (2003) give detailed lower limits of time to be used on skating events, while Nevill and Whyte (2005) give predicted peak world records for the men's 800 meter, 1500 meter, mile, 5000 meter, 10000 meter, marathon and women's 800 meter, and 1500 meter. For skating the predictions of future world records are highly dependent on the no-technical progress assumption. Contrary to running, speedskating is a technology-intensive sport (skates, ice rinks, clothing), so that shocks to technological progress are visible in the improvements of world records. Ultimate human performances are so conditional on this typically hard to predict factor.

In this paper we contribute to the literature on modelling world records. We focus on running events, since these events are well-developed, highly competitive, not intensively affected by the problem of hard to predict technical innovations, and have a typical long history. For instance for the one mile record we have official data since 1865 (for men). We first review methods to model world records. There are two approaches. First, one can fit the historical curves of individual events. Secondly, one can compare different records at the same moment of time and detect outliers. We do both. First we compute the approximation of historical lower bounds of time needed to complete running events using the single-event historical data. We apply, after a careful selection and discussion of alternative specifications, the Gompertz-curve. The Gompertz-curve is a relatively simple curve that allows for an asymmetric S-shape. From these specifications we compute the implied infinite lower bounds. These lower bounds are compared in a cross-sectional setting as a robustness check. Finally, we use one event, the one mile run, to compare the forecasting performance of our methodology.

The set-up of the paper is as follows. First, we discuss the existing methodology to model world records. From this analysis we conclude to use the Gompertz-model. In Section 4 we present the results of fitting the Gompertz-curve for the 100, 200, 400, 800, 1500, 5000, 10000 meter and marathon events for men and women. In Section 5 we test for robustness of the methods by relating the limit values implied by the Gompertz-curves and distance in the famous log-log model of distance and time (see Section 2 for a description). We summarize and conclude in Section 6.

2 Modelling World records

In this section we present two approaches to model world records. First we review candidate functional forms to fit the historical development of world records. Next we review the literature on comparing world records at the same moment of time on various distances. Before doing so, we first discuss in short the problems in estimation. How do we cope with the extreme nature of world records? For instance Kuper and Sterken (2003) estimate a so-called Chapman-Richards approach (see hereafter for a precise definition) to the development of the trend using normally distributed residuals and calculate the extreme residual to compute the shift to the extreme frontier. Smith (1988) proposes a maximum likelihood method of fitting a model to a series of record times. Let Y_t be the best performance in a particular event in the tth year:

$$Y_t = X_t + c_t \tag{1}$$

where X_t is an iid-variable and c_t a nonrandom trend. The records in an T-year period are given by:

$$Z_t = \min(Y_1, \ldots, Y_t) \tag{2}$$

for $1 \leq t \leq T$. The sequence Z_t is observable, but the underlying data Y_t not. We concentrate on two parameterizations. First, we can think of the density function $f(x;\theta)$ of X_t. Let $F(x;\theta)$ be the cumulative density function of $f(x;\theta)$. The statistical problem is then a censored data problem in which the value of Y_t, in a nonrecord year, is censored at the record level. Smith (1988) analyses the normal distribution (which we do not describe here), the so-called Gumbel extreme-value distribution:

$$F(x;\mu,\sigma) = 1 - \exp[-\exp((x-\mu)/\sigma)] \tag{3}$$

and the generalized extreme-value (GEV) distribution:

$$F(x;\mu,\sigma,k) = 1 - \exp[-(1+k(x-\mu)/\sigma)^{\frac{1}{k}}] \tag{4}$$

The GEV with $k = 0$ gives the Gumbel distribution. For $k \geq 1$ the GEV-function does not yield a local maximum. Smith indeed finds that these extreme-value functions have fitting problems and proceeds with the normal distribution. Sterken (2005) applies a similar approach to age-dependent running performance data in a stochastic frontier analysis. Stochastic frontiers allow for measurement error in the data, but assume that a large fraction of the deviation of observed records to the frontier is due to inefficiency. The stochastic frontier analysis can so be seen as an alternative approach to trim extreme values.

The main attention in this paper focuses on the modelling of the trend c_t though. We review the ideas on the parametrization of c_t next. After that we review the relationship between c_t-limit values for various distances.

2.1 Fitting the curve

In this section we briefly review candidate functional forms of the nonrandom trend c_t that are or could be used in fitting the progress of world records. After that

2.1.1 Linear trend

The linear trend $c_t = \theta_1 + \theta_2 t$ is one of the first candidates to use (see Whipp and Ward, 1992, and Tatem et al., 2004). Although a linear trend cannot be true in the long run (there is no limit value of e.g. speed), local linear approximations often perform rather well. So if the focus of the fitting is purely descriptive (and not on forecasting or computing limit values) local linear approximations, like e.g.data envelopment analysis, are true candidates.

2.1.2 Exponential

The exponential function $c_t = \theta_1 - \theta_2 \exp(-\theta_3 t)$ is the mostly used asymptotic model (see e.g. Deakin, 1967). For positive values of θ_3 the limit value of the series is θ_1.

Ratkowsky (1990) describes this function as a close to linear model, since the estimates of the parameters encompass the ones of a linear model. The main critique on the model is its monotonic change over time, which is not apparent in standard S-shaped curves.

2.1.3 Modified Weibull equation

The modified Weibull $c_t = \theta_1 - \exp(-\theta_2 t^{\theta_3})$ is function is also in the exponential class and has limit value θ_1 for $\theta_2 > 0$. The Weibull function is a very flexible function and widely used to model growth and yield data. The parameters θ_2 and θ_3 are scale and shape parameters. For $\theta_3 = 1$ we get the exponential form.

2.1.4 Chapman-Richards

This model, which is also known as the Von Bertalanffy equation, $c_t = \theta_1 - \theta_2[1 - \exp(-\theta_3 t)]^{\theta_4}$ contains a number of special cases. If $\theta_4 = 1$ we get the so-called natural growth model. If $\theta_4 = -1$ we get the logistic model (see hereafter). The Chapman-Richards function has been used by Grubb (1998) for running and Kuper and Sterken (2003) for skating. The limit value is $\theta_1 - \theta_2$.

2.1.5 Antisymmetric exponential function

One of the disadvantages of the Chapman-Richards form is that it is a symmetric S-shaped function for certain parameter combinations. Blest (1996) and Grubb (1998) propose the antisymmetric exponential function with a positive limit θ_1:

$$c_t = \theta_1 + \theta_2 \exp[-\theta_3(t - \theta_4)] \text{ if } t \geq \theta_4 \quad (5)$$

$$c_t = \theta_1 + \theta_2[2 - \exp(\theta_3(t - \theta_4)] \text{ if } t < \theta_4 \quad (6)$$

2.1.6 Generalized logistic

The generalized logistic function $c_t = \theta_1/[\theta_2 + \exp(\theta_3 - \theta_4 t)]$ has two horizontal asymptotes. One at $c = 0$ for $t \to \infty$ and at $C = \theta_1/\theta_2$ for $t \to -\infty$. The inflection point is at $t = \theta_3/\theta_4$. This inflection point is interesting in the modelling of world records. According to Nevill and Whyte (2005) this point reveals the period of greatest gain (acceleration) in world record performance. For instance for running events this period is in the years 1940-1960.

2.1.7 Gompertz

Gompertz (1825) proposed $c_t = \theta_1 + \theta_2 \exp[-\exp(\theta_3(t - \theta_4))]$. Similar to the logistic function the Gompertz curve has asymptotes at θ_1 and $\theta_1 + \theta_2$. The inflection point is at $t = \theta_4$. The Gompertz-curve is asymmetric about its point of inflection. The parameters θ_2 and θ_4 control the shape of the function.

2.1.8 Schnute's equation

Schnute (1981) proposed a generalization of the Chapman-Richards, Gompertz and logistic functions: $c_t = \theta_1^{\theta_2} + (\theta_3^{\theta_2} - \theta_1^{\theta_2})[1 - \exp(-\theta_4(t - T_0))]/[1 - \exp(-\theta_4(t - T))]$. Contrary to the Chapman-Richards it does not impose an asymptotic trend. The parameters θ_1 and θ_3 are the values of c at the first and the last year of observation T_0 and T respectively.

2.2 Cross-sectional approach

Apart from analyzing the progress of world records for separate events, one could also consider the cross-sectional evidence. For instance for running Francis (1943), Lietzke (1954), and Grubb (1998) analyzed the log-log relationship between running time and distance. Kennelly (1905) stated the relationship between time t and distance d after a study of race horses and human athletes as

$$\log t = 9/8 \log d - \text{constant} \tag{7}$$

Define velocity $v = d/t$. Francis (1943) considered the relation:

$$(\log d - 1.5)(v - 3.2) = 6.081 \tag{8}$$

and found that the 5000m record in 1943 was a sign of high performance. In this model the exp(1.5) is an asymptote, which gives the distance at which the maximum speed is attained. Rewriting the Francis equation in general terms $v = A/[\log d - B] + C$, Mosteller and Tuckey (1977) proposed the relation:

$$v = A(d - B)^{\lambda} + C \tag{9}$$

where C is the speed at long distances, B the distance at which the maximum speed is obtained and A the decrease in speed with transformed distance.

Lietzke (1954) noted that, starting from the log-log relation between time used and distance:

$$\log d = k \log t + \log a \tag{10}$$

we get $t = (d/a)^{1/k}$. This implies that $v = d/t = a^{1/k} d^{(k-1)/k}$. So:

$$\log v = \frac{k-1}{k} \log d + \frac{1}{k} \log a \tag{11}$$

Lietzke labels the constant $(k - 1)/k$ as the exhaustion constant and estimates different values for running, horse racing, swimming, cycling, walking, and even auto racing.

Blest (1996) estimates a model for the Olympic records t_{ij} for the distance d_j at the time of the i-th Olympiad:

$$t_{ij} = A_i d_j^{\gamma_i} \tag{12}$$

where A_i and γ_i are parameters to be estimated. An alternative model is $t_{ij} = Ad_j^{\gamma_i}$, where A is an average value over n Olympiads:

$$A = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log A_i\right) \tag{13}$$

No matter what model is used, there seems to be a clear log-log relation between speed and distance, or running time and distance. We explore this relation for the limit values of the Gompertz-curves of individual distances that we compute hereafter.

3 Selection of the functional form

Choosing a model for a set of data can be a difficult task, especially when time series are short. In case of modelling world records there is a second complication: some records stand for a very long time which makes it difficult to observe patterns in the improvement.

We first plot the data (in seconds) against time. This provides valuable information on the shape of the development of world records. For long time series we find a non-linear pattern. Also there is prior knowledge that is useful in selection the functional form. First, there are limits to world records, in other words, there is a lower bound which rules out linear models. Second, increased professionalism in sports leads to saturation in the sense that recently improvements are generally smaller. This, however, may not be true in sports where technology plays an important role (like in speedskating). The two observations suggest to apply S-shaped curves. These S-shaped curves can be symmetric or asymmetric around the inflection point.

In this paper we apply the Gompertz model derived in 1825. This model is used in various applications ranging from population and body growth to biomass and forest growth rates. The four-parameter Gompertz model is given by

$$c_t = \theta_1 + \theta_2 \exp\left(-\exp\left(\theta_3\left(t-\theta_4\right)\right)\right) \tag{14}$$

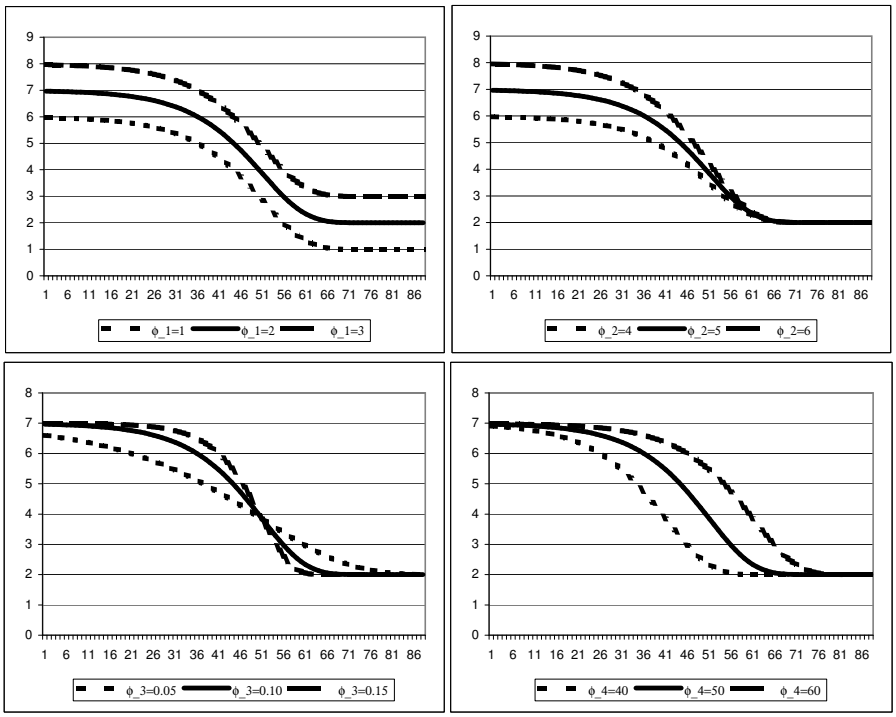
c_t denotes running time in seconds for a certain distance and t is the time index. Four parameters $\theta_1, \theta_2, \theta_3$ and θ_4 are estimated. Parameter θ_1 allows the lower asymptote to be different from zero. Parameters θ_2 and θ_3 control the shape of the curve: For positive values of θ_3 the curve is monotonically decreasing. Parameter θ_3 also determines the smoothness of the curve, small values makes the curve linear. The smoothness decreases if θ_3 increases. Parameters θ_1 and θ_4 shift the curve: θ_1 shifts the curve up and down and θ_4 shifts the curve along the x-axis. Figure 1 illustrates this. The lower and upper limits are identified as:

$$\lim_{t\to\infty} c_t = \theta_1 \tag{15}$$

$$\lim_{t\to-\infty} c_t = \theta_1 + \theta_2 \tag{16}$$

with $\theta_1 > 0$ and $\theta_2 > 0$. The lower limit is calculated from one parameter, whereas the upper limit is the sum of two parameters.

Figure 1: The shape of the Gompertz function for different parameter values keeping the other parameters constant



Unlike the logistic model, the Gompertz model is not symmetric about the inflection point (see Schabenberger et al., 1999). Note that in point of inflection (where $t = \theta_4$):

$$c_t = \theta_1 + \frac{\theta_2}{e} \tag{17}$$

Where e is the base of the natural logarithms, sometimes called the Eulian number and is defined as

$$e = \lim \left(1 + \frac{1}{n}\right)^n \approx 2.718281828 \tag{18}$$

The point of inflection is below the center of the limits (asymmetry) since $e > 2$:

$$\theta_1 + \frac{\theta_2}{e} < \theta_1 + \frac{\theta_2}{2} \tag{19}$$

This implies that the Gompertz model has a shorter period of fast growth.

At point of inflection the slope is:

$$\frac{dc_t}{dt} = -\frac{\theta_3 \theta_2}{e} < 0 \tag{20}$$

since $\theta_2, \theta_3 > 0$.

The Gompertz function is flexible enough to allow for various patterns. Figure 1 shows some examples.

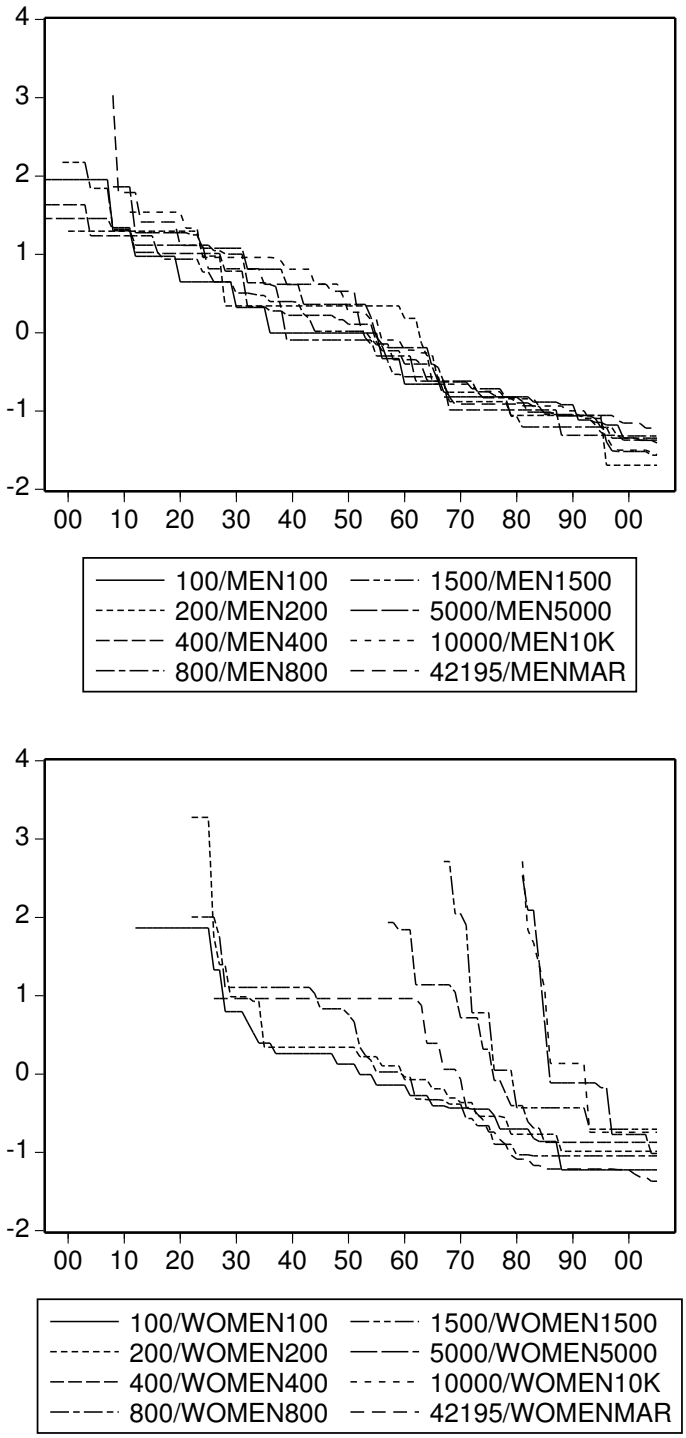
4 Running data

Our main source of information is the official website of the Olympic movement[1] which publishes information on current world records and progress of world records for various events. In this section we show the current record holders and we illustrate the progress of the world records graphically.

For running we analyze distances from the 100 meter sprint to the marathon for both men and women. We transform the raw data of running times to seconds per meter. The normalised time series are shown in Figure 2. This figure show a remarkable improvement in the world record running times. Figure 2 indicates that for some women distances there are relatively few observations. Also many of the women's world records are relatively old (see Table 1). This makes it difficult to estimate the parameters in the model. Examples of old records are the records on the sprint distances from the late Florence Griffith Joyner, who set the current 100 meter and 200 meter world records in the summer of 1988, almost twenty years ago. The oldest world record in running is the 800 meter record held by the Czech athlete Jarmila Kratochvílová. Also Marita Koch's 400 meter world record (from 1985) is yet to be beaten. The men's world records are improved more often. However,

[1] http://www.olympic.org/uk/utilities/reports/.

Figure 2: The development of world records running for men and women (in seconds, normalised)



Note: Series are scaled using zero mean and unit standard deviation.

there examples of world records lasting for a very long time. Jim Hines's 100 meter world record, set in 1968 at the Olympic Games in Mexico was not improved until 1983. Pietro Mennea's 200 meter world record of 1979 lasted until 1996. Recently, Justin Gatlin's 100 meters world record of 9.76 seconds has been corrected to 9.77, equaling the previous mark set by Asafa Powell in 2005.

Table 1: Current running world records (June 2006)

Men	Performance	Athlete	Nat.	Place	Date
100m	9.77	Asafa Powell	JAM	Athens (GRE)	14-06-2005
200m	19.32	Michael Johnson	USA	Atlanta (USA)	01-08-1996
400m	43.18	Michael Johnson	USA	Sevilla (ESP)	26-08-1999
800m	1:41.11	Wilson Kipketer	DEN	Cologne (GER)	24-08-1997
1500m	3:26.00	Hicham El Guerrouj	MAR	Rome (ITA)	14-07-1998
5000m	12:37.35	Kenenisa Bekele	ETH	Hengelo (NED)	31-05-2004
10000m	26:17.53	Kenenisa Bekele	ETH	Brussels (BEL)	26-08-2005
Marathon	2:04:55	Paul Tergat	KEN	Berlin (GER)	28-09-2003
Women	Performance	Athlete	Nat.	Place	Date
100m	10.49	Florence Griffith-Joyner	USA	Indianapolis (USA)	16-07-1988
200m	21.34	Florence Griffith-Joyner	USA	Seoul (KOR)	29-09-1988
400m	47.60	Marita Koch	GDR	Canberra (AUS)	06-10-1985
800m	1:53.28	Jarmila Kratochvílová	TCH	Munich (GER)	26-07-1983
1500m	3:50.46	Yunxia Qu	CHN	Beijing (CHN)	11-09-1993
5000m	14:24.53	Meseret Defar	ETH	New York City (USA)	03-06-2006
10000m	29:31.78	Junxia Wang	CHN	Beijing (CHN)	08-09-1993
Marathon	2:15:25	Paula Radcliffe	GBR	London (GBR)	13-04-2003

5 Results of fitting the Gompertz curves

In this section we model the development of the world records. The world records describe the possibility frontier of athletes. We model the trend in world record times. We try to predict the pattern of world record improvements based on past performance assuming that there will not be major technological improvements that shift out the possibility frontier. With only a few observations it is difficult to estimate the shape of the frontier in a reliable way. Despite these problems we will try to determine the frontier, which we use to derive the limits for eight running events - 200m, 400m, 800m, 1500m, 5k, 10k and the marathon - for both men and women. We use annual data for the world records as they are on December 31 of each year. The last year included in our analysis is 2005 (so Gatlin's 100 meter and Defar's 5000 meter world records are not included in our estimations). To fit the Gompertz model we minimize the sum of squared errors in a grid search in which the parameter θ_2 is fixed and the remaining parameters are estimated using non-linear least squares. In presenting the results we give the optimal value of θ_2. The regression model is

$$c_t = \theta_1 + \theta_2 \exp\left(-\exp\left(\theta_3\left(t - \theta_4\right)\right)\right) + \eta_t \tag{21}$$

where η_t is the error term. To find the limit we add the smallest error to parameter θ_1.

$$\text{limit} = \theta_1 + \min \eta_t \tag{22}$$

The results are in Table 2. The table gives the parameter estimates (with the standard errors in the rows below indicated by (se)). We also indicate the start of the sample, the number of observations, the number of improvements in the sample, indicators of the fit of the model and the implied limit values. We also indicate the predicted times for the year of the next Olympiad in Bejing (2008). Especially for the records for women we find a rather flat limit after 2008.

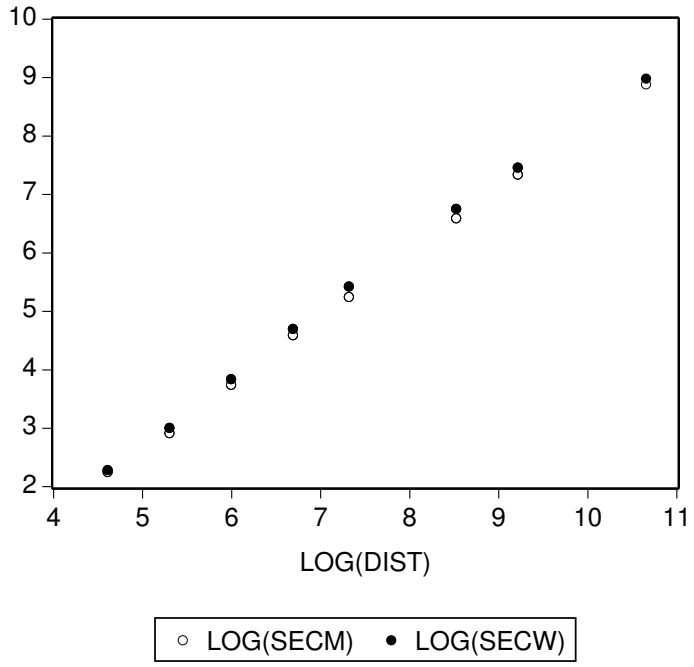
Table 2: Modelling running time (seconds) for the period until 2005

Men	100m	200m	400m	800m	1500m	5000m	10000m	marathon
θ_1	9.6370	18.8528	42.9993	101.1068	195.5526	743.6015	1575.143	7513.437
(se)	(0.0380)	(0.1043)	(0.0749)	(0.1687)	(1.7961)	(3.7731)	(5.2033)	(38.2890)
θ_2	236	3	7	16	7840	237	390	3000
θ_3	0.0029	0.0211	0.0256	0.0290	0.0027	0.0186	0.0264	0.0287
(se)	(0.0002)	(0.0009)	(0.0005)	(0.0006)	(0.0002)	(0.0004)	(0.0005)	(0.0010)
θ_4	574.7330	78.2325	54.0637	53.0789	589.4005	54.8990	61.4153	48.1727
(se)	(32.7003)	(4.9838)	(1.4810)	(1.3574)	(31.7592)	(2.5540)	(1.7174)	(1.6126)
First year	1896	1900	1896	1896	1899	1920	1911	1908
Observations	110	106	110	110	107	86	95	98
Improvements	16	10	16	17	30	25	30	24
R^2	0.9788	0.9472	0.9837	0.9792	0.9854	0.9862	0.9828	0.9587
SSR	0.2172	2.1964	5.5854	43.6060	250.6848	1614.260	13380.38	2026859
Limit (sec)	9.55	18.54	42.42	99.01	190.98	733.60	1544.48	7256.36
Limit (h:m:ss.s)	9.55	18.54	42.42	1:39.01	3:10.98	12:13.60	25:44.48	2:00:56
2008	9.71	18.94	42.51	1:39:08	3:20.64	12:26.76	25:53.15	2:01:11
Women	100m	200m	400m	800m	1500m	5000m	10000m	marathon
θ_1	10.1857	21.5373	47.5564	113.1286	230.9393	868.3973	1767.419	8467.090
(se)	(0.1198)	(0.1570)	(0.0607)	(0.2230)	(0.2929)	(1.5989)	(4.6382)	(46.3125)
θ_2	3140	5960	7	40	348	30400000	33000000	4800
θ_3	0.0030	0.0061	0.1033	0.0472	0.0458	0.0140	0.0136	0.1599
(se)	(0.0003)	(0.0006)	(0.0039)	(0.0013)	(0.0025)	(0.0020)	(0.0015)	(0.0095)
θ_4	-631.6025	-290.9621	77.9019	58.8077	50.2936	-100.4812	-99.3344	76.5443
(se)	(64.8919)	(31.9656)	(0.3921)	(0.4595)	(1.2963)	(27.2444)	(20.9273)	(0.3344)
First year	1920	1922	1957	1940	1967	1981	1981	1962
Observations	86	83	49	66	39	25	25	44
Improvements	19	16	12	21	7	8	7	20
R^2	0.9497	0.8813	0.9846	0.9852	0.9726	0.9183	0.9533	0.9742
SSR	1.7635	28.2718	3.3782	63.2202	56.4728	386.7078	2564.169	2229555
Limit (sec)	9.88	20.30	46.67	110.42	227.87	859.17	1744.23	7982.27
Limit (h:m:ss.s)	9.88	20.30	46.67	1:50.42	3:47.87	14:19.17	29:04.23	2:13:02
2008	10.17	20.34	46.67	1:50.42	3:47.87	14:19.27	29:04.97	2:13:02

6 Limit values of time and distance

In the previous section we have estimated limit running times for each running distance and for both sexes. This information is shown in Figure 3 which indicates that the relationship between the limits of running time and distance is linear in log-log form. This confirms what has been found by Lietzke (1954) and Grubb (1998).

Figure 3: The log-log relationship between the limits of running time for men (SECM) and women (SECW) and distance



Finally we estimate the pooled log-log model to determine parameter k in Equation (10). The model allows for a different intercept for men and women (fixed effects). We assume a cross-section SUR specification to allow for contemporaneous correlation between men and women. We also estimate robust coefficient standard errors.[2] The null for redundant fixed effects (FE) is rejected at 5%, so in Table 3 we report the fixed effects estimates. Note however that the sample is very small.

Our estimate for k is slightly smaller than the one Kennelly reported in 1905. Kennelly's estimate is 9/8, while our estimate is 11/10, which is significantly smaller than 9/8 at 5%. This implies that if distance increase by 10%, the limit time increases by 11%, irrespective of gender. This is what Lietzke refers to as the exhaustion constant or the fatigue rate. As the fixed effects (FE) model is not rejected, the gender differences 2×0.0551 are significant. The estimates imply that the women are 11% slower than men.

Finally, we estimate the limit time for the one-mile run (1609 meters), which is an irregular running distance at main championships. The current world record is held by Hicham El Guerrouj (MOR) who set a world record time of 3:43.13 on July 7, 1999 in Rome (ITA). The current women world record on the mile is 4:12.56, set by Russian Svetlana Masterkova on August 14, 1996 in Zürich (SUI). The development of the one-mile world records are

[2]In Table 3 we applied the White period robust coefficient variance estimator which accommodates arbitrary serial correlation and time-varying variances in the disturbance. Other robust coefficient variance estimators give similar results.

Table 3: Cross-section SUR estimation results for the log-log relationship between limit time t and distance d, with the White period robust coefficient variance estimator

Model: $\log t = k \log d + \log a$		
Parameter	Coefficient	Robust standard error
a	-2.8058	0.0697
k	1.1067	0.0082
Fixed Effects		
a-Men	-0.0551	
a-Women	0.0551	
R^2	0.9998	
s.e. of regression	1.0859	
Durbin-Watson	1.4669	
Redundant FE $F(1, 13)$ (p-value)	48.8978	(0.0000)

shown in Figure 4.

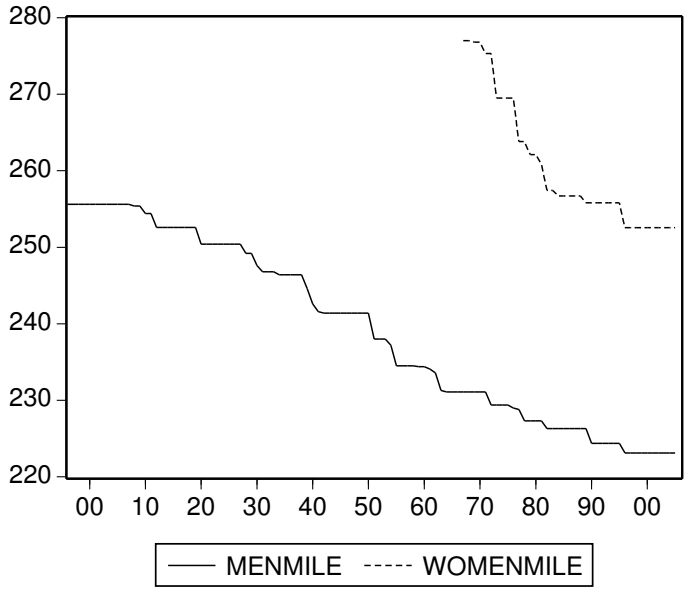
Table 4: Modelling running time for the one-mile run for men (seconds)

	Men		Women	
	Coefficient	Std. error	Coefficient	Std. error
θ_1	223.3653	0.2115	253.2729	0.4123
θ_2	39		57	
θ_3	0.0330	0.0004	0.0672	0.0034
θ_4	91.2639	0.6382	108.6841	0.4200
First year	1896		1967	
Observations	110		39	
Improvements	35		10	
R^2	0.9944		0.9730	
SSR	82.4707		74.043	
Limit (sec)	220.97		250.73	
Limit(m:ss.s)	3:40.97		4:10.73	

The estimates in Table 3 implies a one-mile limit time for men of 3:22.48 (202.48 seconds). The estimated women one-mile limit is well below the 4 minute barrier at 3:46.04 (226.04 seconds). The question is how these one-mile limit times differ from separate estimates of the Gompertz curve for the one-mile run. Two matters complicate this exercise. First, the current world records are rather old. Second, the one-mile run is an irregular running distance. Table 4 shows the results.

The Gompertz model suggests limit world records which are considerably slower than the times implied by Table 3. If the one-mile run would be a regular running event, the limit

Figure 4: The development of one-mile world records for men and women (in seconds)



world record would be 18.5 seconds faster for men and 24.7 seconds for women.

7 Summary and conclusions

In this paper we analyze the development of world records. We focus on running events and fit the historical curves of individual events using an asymmetric S-shape Gompertz curve. We compute the implied infinite lower bounds for each event and for men and women. With only a few observations for some events it is difficult to estimate the shape of the frontier in a reliable way. Despite these problems we have determined this frontier, which we use to derive the limits for eight running events - 200, 400, 800, 1500, 5000, 10000 meter, and the marathon - for both men and women. We use annual data for the world records until 2005.

The lower bounds calculated from the Gompertz curve estimations are compared in a cross-sectional setting to arrive at a relationship between time and distance. Our estimate is slightly different from the one Kennelly reported in 1905. Our results imply that if distance increase by 10%, the limit time increases by 11%, irrespective of gender.

As the fixed effects (FE) model is not rejected, the gender differences are significant. The estimates imply that the woman are 11% slower than men. Finally, we forecast the lower bound of the mile using the log-log specification between time and distance. If the one-mile run would be a regular running event like the other events, the gain in the long run would be 18.5 seconds for men and 24.7 seconds for women.

References

Blest, D.C. (1996), "Lower bounds for athletic performances", *The Statistician*, **45**, 243–253.

Chatterjee, S. and S. Chatterjee (1982), "New lamps for old: an exploratory analysis of running times in Olympic Games", *Applied Statistics*, **31**, 14–22.

Deakin, M.A.B. (1967), "Estimating bounds on athletic preformance", *Mathematics Gazette*, **51**, 100–103.

Francis, A.W. (1943), "Running records", *Science*, **98**, 315–316.

Gompertz, B. (1825), "On the nature of the function expressive of the law of human mortality and on a new mode of determining life contingencies", *Philos. Trans. Roy. Soc. London A*, **115**, 513–585.

Grubb, H.J. (1998), "Models for comparing athletic performances", *The Statistician*, **47**, 509–521.

Kennelly, A.E., editor (1905), *A study of racing animals*, The American Academy of Arts and Sciences.

Kuper, G.H. and E. Sterken (2003), "Endurance in speed skating: the development of world records", *European Journal of Operational Research*, **148**, 293–301.

Lietzke, M.H. (1954), "An analytical study of world and Olympic racing records", *Science*, **119**, 333–336.

Mosteller, F. and J.W. Tuckey, editors (1977), *Data Analysis and Regression*, Addison-Wesley.

Nevill, A.M and G. Whyte (2005), "Are there limits to running world records?", *Medicine and SCience in Sports and Exercise*, **37**, 1785–1788.

Ratkowsky, D.A., editor (1990), *Handbook of Nonlinear Regression Models*, Marcel Dekker.

Smith, R.L. (1988), "Forecasting records by maxiumum likelihood", *Journal of the American Statistical Association*, **83**, 331–388.

Sterken, E. (2005), "A stochastic frontier approach to running performance", *IMA Journal of Management Mathematics*, **16**, 141–149.

Tatem, A.J., C.A. Guerra, P.M. Atkinson, and S.I. Hay (2004), "Momenteous sprint at the 2156 Olympics?", *Nature*, **431**, 525.

Whipp, B.J. and S.A. Ward (1992), "Will women soon outrun men?", *Nature*, **355**, 25.