Modelling running speed in athletic track events

Patrick Royston
London W5,
UK.
email: patrick.royston@ctu.mrc.ac.uk

18 April 1996

Abstract

An analysis of world record and Olympic athletic track data reveals a remarkably simple statistical model for predicting speed from the sex of the athlete and the distance run. The model predicts that a hypothetical race of length between 130 and 136 m would produce the fastest mean running speeds humanly possible. The model is adapted to produce performance indexes with which to compare results between races of different length. The indexes are proposed as the possible basis of systems for ranking track athletes and for comparing the speeds of different running tracks.

1 Introduction

Chatterjee & Chatterjee (1982) described an analysis of running times in the Olympic Games between the years 1900 and 1976. Arguing that a lower limit on running time must exist for each individual event, they fitted exponential models of the form $y_j = \theta_1 \exp(j\theta_2)$, where y_j is the time for year j ($0 \leq j \leq 76$). They elaborated the model to include the effect of distance run (considering only the 100, 200, 400 and 800 m events) by modelling the logarithm of the time in terms of year and distance. The paper was subsequently criticised in letters to the *Journal* by Reid & Sandland (1983) in terms of inaccuracies in the data, and by Wootton & Royston (1983) for lack of evidence in the data for the claimed exponential levelling-off of running times with year of event. Robinson & Tawn (1995) modelled exceptional times by use of the generalized extreme value distribution, applying the technique to the women's 3000 m track event.

In this paper I reanalyse the Olympic data for male athletes (including longer-distance events) in a manner different from either Reid & Sandland (1983) or Robinson & Tawn (1995). I also estimate the differences between male and female performances by analysing world record times as at 1992, obtaining a strikingly simple statistical model which predicts athletic performance in terms of distance run, sex and whether starting blocks are used. The model is used to develop indexes which enable athletic performances to be ranked by in relation to appropriate *distance and sex-adjusted* speed standards, such as that of the current world record, that of the speed of the race in question, or a weighted mean of the two. A similar approach may allow running tracks to be compared according to their best distance-adjusted speeds, averaged over a suitable spectrum of competitive events. Finally, a new athletic event is proposed: a 133 m sprint, which the model predicts will produce the fastest mean running times humanly possible.

2 Analysis of 1992 world records

The data consist of world record times up to and including 1992 for men and women in the following track events (distance in metres; asterisks indicate data available only for men): 100, 200, 400, 800, 1000*, 1500, 1609.3 (i.e. 1 mile), 2000*, 3000, 5000, 10000. The data are given in Table 1.

Since running time depends strongly on event distance and (owing to the 200-fold range of observed times) would be expected to have non-constant variance, speed in m/sec rather than time was analysed. The effect of distance on speed was modelled using fractional polynomials (Royston & Altman (1994)), which provide many more functional forms than do conventional polynomials. A fractional polynomial with powers $(-1, -1)$ was the best fit for each sex, higher order models not providing a statistically significant improvement. The functional form of this model is $\beta_0 + \beta_1 x^{-1} + \beta_2 x^{-1} \ln x$. Another important consideration is that 100, 200 and 400 m races are run from starting-blocks, which must boost performance somewhat. A term representing a speed increment for starting-blocks was therefore included in the model. Note that since the effect of distance is included in the model, the additional effect of starting-blocks on speed may be regarded as modifying distance-adjusted rather than 'raw' speed.

Further investigation showed that the speed/distance relationship appeared to be parallel between the sexes, the males being an estimated 0.735 m/sec faster than the females at all distances. The final fitted model (with SEs in parentheses) was as follows:

$$y = 5.930\,(0.027) - 2100\,(54)\,x_1^{-1} + 540\,(13)\,x_1^{-1} \ln x_1 + 0.410\,(0.054)\,x_2 - 0.735\,(0.020)\,x_3 + e,$$

where y is speed, x_1 is distance, x_2 is 1 if blocks used, 0 otherwise, x_3 is 0 for male, 1 for female, and e is residual error with an estimated SD of 0.045 m/sec. The intercept speed of 5.930 m/sec is attained with $x_1 = \infty$, $x_2 = x_3 = 0$, that is for a male running a hypothetical race of arbitrarily long distance without starting blocks. The observed and fitted values are shown in Figure 1.

** Figure 1 near here **

The fit is surprisingly good.

An interesting feature of the model (see Figure 1) is that a maximum predicted speed occurs for a hypothetical race whose estimated length, x_1^* say, lies between 100 and 200 m. According to the model, $x_1^* = \exp(1 + r) = 133.0$ m, where $r = -b_1/b_2 = 3.89$ is minus the ratio of the estimated regression coefficients for x_1^{-1} and $x_1^{-1} \ln x_1$. Thus the speed of the fastest possible race depends only on r. By using the delta method and after simplifying the algebra, the expression

$$\text{var}(x_1^*) \simeq (x_1^*/b_2)^2 \left[\text{var}(b_1) + 2r\,\text{cov}(b_1, b_2) + r^2\,\text{var}(b_2)\right]$$

was obtained. It gives $\text{SE}(x_1^*) \simeq 1.7$ m and (with the usual assumption of an asymptotic normal distribution) an approximate 95% confidence interval for x_1^* of $(129.6, 136.4)$ m.

3 Analysis of Olympic records

A similar analysis was applied to cumulative record running times for Olympic track events between 1912 and 1992. Only data for male athletes was used as less data is available for females. The aim was to smooth the times over recent history and to indicate the direction of future trends. Preliminary analysis showed that the $(-1, -1)$ fractional polynomial and a starting-blocks effect still described the running speeds well for each individual Olympiad. Furthermore, the coefficients for x_1^{-1}, $x_1^{-1} \ln x_1$ and x_2 (blocks) did not vary significantly between different years, leading to the following model for the speeds in year $1900 + j$ $(12 \leq j \leq 92)$:

$$y_j = b_{0j} - 2283\,(28)\,x_1^{-1} + 585\,(7)\,x_1^{-1} \ln x_1 + 0.31\,(0.03)\,x_2 + e,$$

the SD of e being 0.065 m/sec. The negative ratio (r) of estimated coefficients of the fractional polynomial terms is 3.90, hence the length of the hypothetical fastest race is 134 m, extremely similar to the value of 133 m from the 1992 world record data. The estimated effect of starting-blocks is somewhat smaller for the Olympic data than for the 1992 world record data. The estimated intercepts, b_{0j}, which may be regarded as distance-adjusted speeds, increase steadily from 5.04 m/sec in 1912 to 5.82 m/sec in 1992. They are plotted in Figure 2, together with fitted exponential curves derived from all data between 1912 and 1992, and from postwar data only (1948-1992). (Fractional polynomials were not used to smooth these relationships as an arbitrary choice of time origin would be required. The exponential fits do not depend on the time origin.)

** Figure 2 near here **

The fits are fairly good; due to features unique to each Olympiad and to the step-function nature of the observations, an excellent fit would not be expected. Nevertheless, the gradual improvement in male athletic performance over time is nicely illustrated. The postwar fitted curve indicates a more marked levelling-off of performance than does the overall curve, and may be a more reliable guide to future performance than the overall curve. Figure 3 shows the observed and fitted speeds for each event, using the fitted curves $7.354 - 2.486 \exp(-0.00543j)$ for the intercepts b_{0j}.

** Figure 3 near here **

The pattern of speeds for the 100 and 200 m events differs slightly from the rest, having improved more rapidly in earlier times and less rapidly recently. The results for the marathon are included for information. They indicate slower speeds than would be expected from equivalent track events, not surprising given that marathons are road races with features such as hills and bends which increase effort and reduce overall speed. Nevertheless it is interesting that recent marathon speeds seem to be approaching those of a theoretical track event.

4 Performance indexes

In several sports (such as cricket and tennis) it has proved possible to rank individuals according to indices of their recent performances, and such rankings are routinely reported in the media. Ranking has not been attempted for athletic track events, presumably owing to the difficulty of comparing times over different distances. The analysis given in the previous sections has shown that it is statistically meaningful to adjust world-record running speeds for distance, sex and type of event (sprint from blocks or standing start).

4.1 Definitions and examples

The idea of adjustment suggests the following class of performance indexes (PIs), π:

$$\pi = 100 \frac{y - y_{\text{adj}}}{y_0 - y_{\text{adj}}}$$

where y is the actual race speed (distance/time) for a given athlete, $y_{\rm adj} = -2100x_1^{-1} + 540x_1^{-1}\ln x_1 + 0.410x_2 - 0.735\,x_3$ is a speed adjustment and y_0 is a speed standard for athletes of sex x_3 over a distance of x_1 m. The value of π is scaled to be 100 for a speed of y_0, a value near 100 indicating good performance (relative to y_0). Since an allowance for sex has been included in $y_{\rm adj}$, the index in theory permits comparison between male and female performances, though in practice one is unlikely to want to do so.

Two types of choice for y_0 seem appropriate, though many others are possible. If (world) record speeds are used, an *absolute* PI, $\pi_{\rm abs}$, is obtained. If the speed of the winning athlete is used, a *relative* PI, $\pi_{\rm rel}$, results. The former favours fast speed irrespective of finishing position in the race, whereas the latter favours high placing irrespective of speed. Essentially, an athlete in a definitive race (i.e. not a heat or a qualifying event) always runs to obtain the highest possible placement and may try to break a record, such as a world, championship, national or personal best. A weighted mean PI of the form $\pi_\alpha = (1-\alpha)\pi_{\rm abs} + \alpha\pi_{\rm rel}$, with $0 \le \alpha \le 1$, may therefore be appropriate for general purposes. Choice of a suitable value of α could be based on analysis of previous results and on a consensual judgment by experts. Here I shall take $\alpha = 0.5$ for illustration.

Calculation of world rankings between athletes in a given season would involve averaging an individual's values of π_α obtained in particular events, or perhaps his/her k seasonal best π_αs for a suitable value of k, and ordering the results. By including appropriate speed standards via y_0, other track events, such as the 100, 110 and 400 m hurdles, steeplechase, relay events and the marathon, could be accommodated if desired.

For example, the current world record time for the 100 m men's sprint is 9.86 sec. The winning time in the 1995 world championship final was 9.97 sec (Donovan Bailey), whereas a previous world champion (Linford Christie) had a hamstring injury and only managed 10.12 sec. The value of $y_{\rm adj}$ for the male 100 m race is 4.278 m/sec. For Bailey, $\pi_{\rm abs} = 98.1, \pi_{\rm rel} = 100$ and $\pi_{0.5} = 99.1$, whereas for Christie (who had previously run the race in under 10 sec more than once) the corresponding values are $95.6, 97.4$ and 96.5, indicating the extent to which Christie's performance fell short on that occasion.

The performance index $\pi_{\rm abs}$ and/or distance-adjusted speeds may also be useful for assessing the (relative) speeds of different running tracks. Certain tracks are anecdotally believed by commentators and others to be 'fast', but no one has proposed a synthesis of performance times which would measure the overall speed objectively. Values of distance-adjusted speeds or of $\pi_{\rm abs}$ averaged over the best performances for selected events on the given track may be suitable for the purpose. Care would be needed to select the correct events and performances to arrive at fair and representative results.

4.2 Application to Olympic records

To illustrate the historical improvements in PI values that have taken place, I have calculated π_{abs} for the male Olympic data. The results are plotted in Figure 4.

** Figure 4 near here **

The patterns are somewhat different for the 100 and 200 m dashes compared with longer races. Performance in the dashes has tended to exceed that in other races until about the last two decades. This is likely to be an artifact of the hand-held timing methods used before 1972: race times measured by stopwatch may be as much as 0.2 sec less than those measured electronically (Reid & Sandland (1983)). The new record in the 400 m set in the 1968 Mexico City meeting was a dramatic improvement. The gradual increases in distance-adjusted speed (Figure 2) are reflected in the values of π_{abs} in the individual events, but with some differences between events as just noted. The evidence for a levelling-off of π_{abs} with time is not strong. We may therefore anticipate a continuation of the trend towards faster times in future, but perhaps with a reduction in the rate of increase in π_{abs}.

5 Discussion

It is rather remarkable that the fractional polynomial with powers $(-1, -1)$ is so consistently successful in describing the relation between speed and distance in the datasets considered here. Certainly no conventional polynomial model could compete with it; even one of seventh order was inadequate. This finding confirms success in using fractional polynomials to model essentially noise-free relationships, described by Royston & Altman (1997). The speed adjustment made for starting from blocks is crude, but more complicated approaches were tried and did not seem to improve the fit. There was no discernible interaction between sex and either the term representing starting-blocks or the fractional polynomial terms.

Using a detailed modelling approach, Robinson & Tawn (1995) adduced some evidence that a record of 486.11 sec in the women's 3000m, set in 1993 by the Chinese athlete Wang Jungxia, was out of line with the then current best times for that event. According to the present analysis of 1992 data, the expected speed for the race is 5.936 m/sec, compared with Junxia's speed of 6.171 m/sec. Her performance is therefore more than 5 standard deviations better than the record in 1992. Certainly by conventional statistical criteria it is atypical compared with the best that could reasonably be expected in 1993.

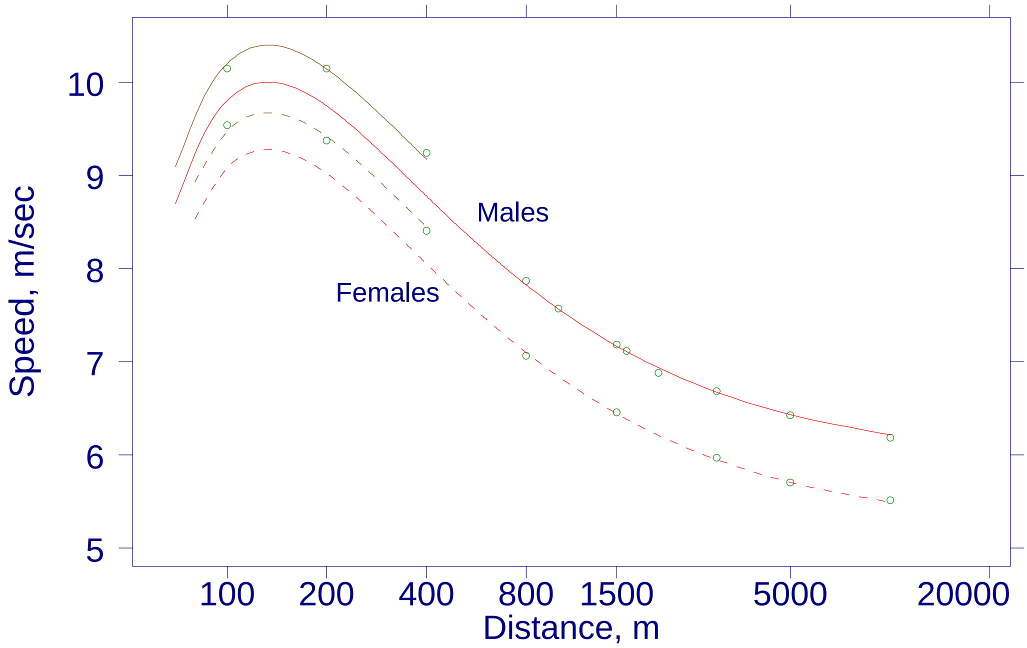
If the proposed indexes are to be of practical value for ranking athletic performances, it is desirable that their means and variances for high-quality athletes are approximately constant between track races of different length. If this were not so, athletes competing in events for which the variance of the PI is low might tend to bunch together in the overall rankings. There is however a certain circularity here, in that a low variance could be said to indicate consistent performance standards, which should not be penalised. One way to investigate properties of (say) the index π_α would be to compute the values of π_α from (say) the fastest 50 athletes in each relevant track event over a given season, and compare the means and variances of π_α between events using standard methods.

Clearly considerable work must be done by statisticians, athletics experts and athletics official bodies before (possibly refined versions of) the performance indexes are accepted as the basis of a fair and sensible way of ranking athletes. It is beyond the scope and intent of the present paper to define such a programme of work.

Finally, the analyses of the 1992 world records and of the Olympic records suggest that a race of between 130–136 m would yield the fastest mean speeds humanly possible. According to the model, by 1992 these speeds were 10.41 m/sec for men and 9.67 m/sec for women, with corresponding times over 133 m of 12.78 sec and 13.75 sec. There may be some interest in experimenting with a new event, a 133 m sprint, in which the predictions could be tested. The idea of a 'fastest possible race' certainly has some appeal.

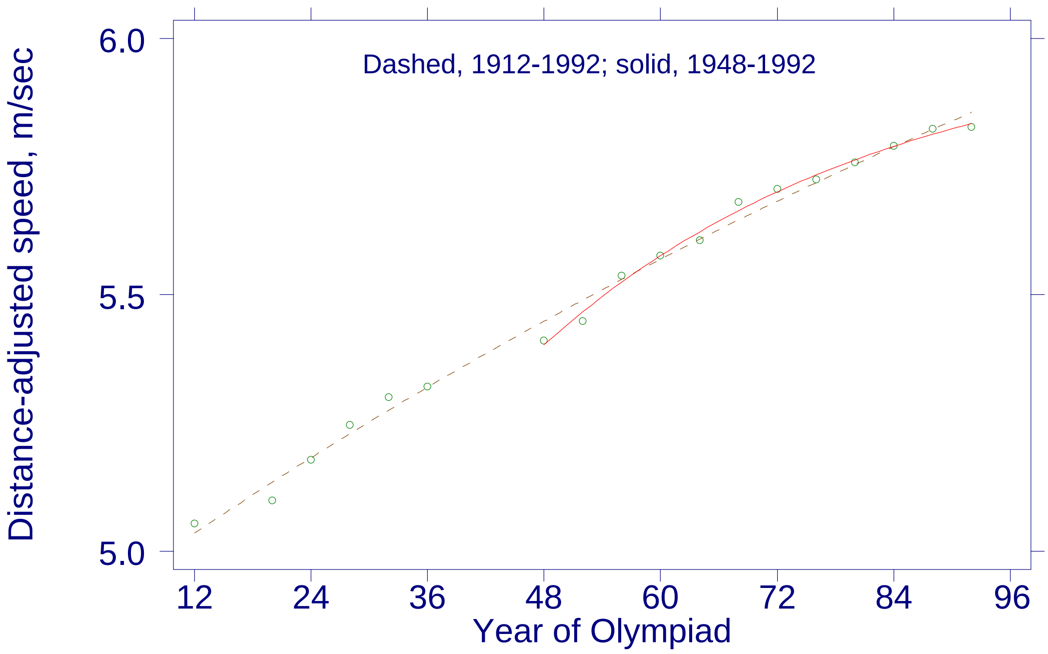
Legend

Figure 1.



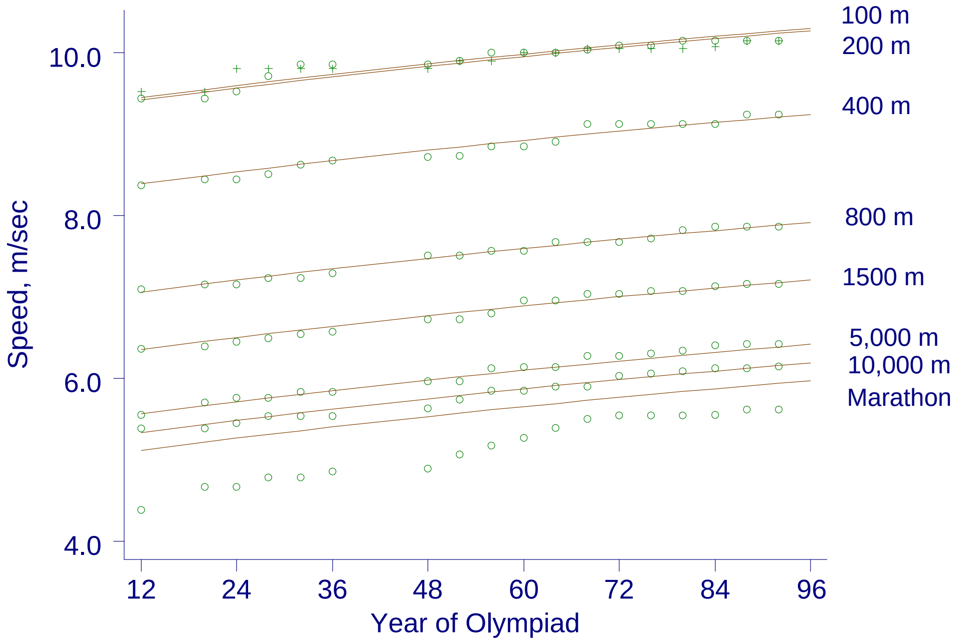
Observed and fitted 1992 world record speeds for several track events. The values for non-sprint races have been extrapolated backwards to show estimated equivalent speeds if starting blocks had not been employed. Males, solid lines; females, dashed lines.

Figure 2.



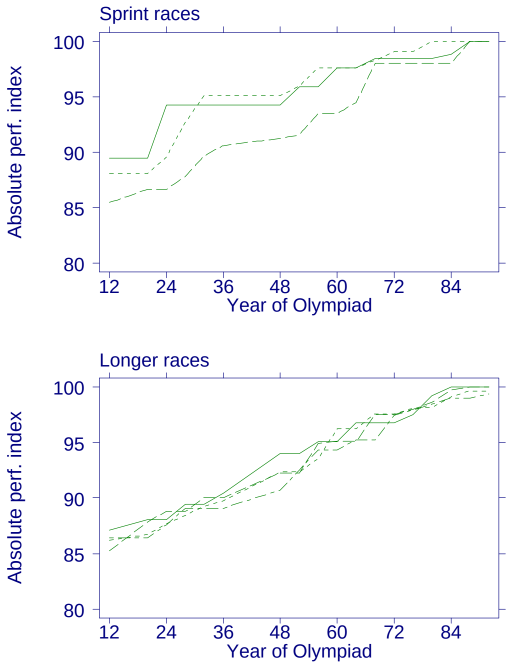
Distance-adjusted speeds for Olympic records, 1912-1992. Solid line, fitted exponential regression line based on 1912-1992 data; dashed line, same for 1948-1992 data.

Figure 3.



Observed and fitted Olympic speeds for several track events. For 100 and 200 m events, + indicates 100 m, o indicates 200 m.

Figure 4.



Performance indices (based on 1992 world records) for Olympic results, 1912-1992. Upper panel: sprint races (100, 200, 400 m); lower panel: longer races (800, 1500, 5000 and 10000 m).

References

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

Reid, D. D. & Sandland, R. L. (1983), 'New lamps for old? [letter]', *Applied Statistics* **32**, 86–87.

Robinson, M. E. & Tawn, J. A. (1995), 'Statistics for exceptional athletics records', *Applied Statistics* **44**, 499–511.

Royston, P. & Altman, D. G. (1994), 'Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling', *Applied Statistics* **43**, 429–467.

Royston, P. & Altman, D. G. (1997), 'Approximating statistical functions by using fractional polynomial regression', *The Statistician* **46**, 1–12.

Wootton, R. & Royston, J. P. (1983), 'New lamps for old [letter]', *Applied Statistics* **32**, 88–89.

Distance (m)	**Men**		**Women**	
	min	sec	min	sec
100		9.86		10.49
200		19.72		21.34
400		43.29		47.60
800	1	41.73	1	53.28
1000	2	12.18	–	–
1500	3	28.86	3	52.47
1609.3 (mile)	3	46.32	4	15.61
2000	4	50.81	–	–
3000	7	28.96	8	22.62
5000	12	58.39	14	37.33
10000	26	58.38	30	13.74

Table 1: World record times for athletic track events, up to and including 1992. Source: Microsoft Encarta, 1994 edition.