How good is Tiger Woods?

Baker, RD and McHale, IG

How good is Tiger Woods?

Rose D. Baker, C. Math, FIMA and Ian G. McHale
Centre for Operational Research and Applied Statistics, University of Salford

Abstract

A major objective of professional sport is to find out which player or team is the best. Unfortunately the structure of some sports means that this is often a difficult question to answer. For example, there may be too many competitors to run a round-robin league, whilst knock-out tournaments do not compare every player with every other player. The problem gets worse when one has to compare players whose performance varies over time. Fortunately mathematical modelling can help and in this article, we use the Plackett-Luce model to estimate time-varying player strengths of golfers. We use the model to investigate how good golf's current biggest attraction, Tiger Woods, really is.

Introduction

'Who is the greatest golfer?' is an often discussed topic at the 19th hole of any golf course. In recent times, it seems likely that Tiger Woods is. We know Tiger is good, in fact we can be pretty certain he is the best amongst his contemporaries, but just how good is he? How much better than his competitors is he, and given his recent drop in form, can we expect Tiger to get back to his best? It is difficult, if not impossible to give definitive answers to these questions. However, as is often the case in life, mathematical modelling and statistical analysis can shed light on the situation. We use a ratings or ranking model, and fit it to tournament data by the method of maximum likelihood.

Ratings models have typically been used to compare candidates at elections, or compare the popularity of a set of choices made by consumers. Given the main objective of professional sport is to find the best player or team, ratings models are naturally applied to sport. Further, data on sport are often widely accessible making it an ideal playground for a statistician. Here we use a ratings model to compare the ability of golfers. The model adopted is the Plackett-Luce model which we use to estimate competitor 'strengths'.

The strengths estimated from the standard Plackett-Luce model are static in that they do not vary with time. When used to estimate player strength in the context of golf, it seems unreasonable to assume each and every player has a constant strength throughout their playing career. It is more likely that a player's strength will vary during his or her career. As such, we need a model that can allow a player's strength to vary with time. Our solution is to estimate the strength parameters at regularly-spaced time points, with a smooth interpolation between them. Splines are very often used in statistical modelling (e.g. Harrell, 2001) but it is even simpler and arguably better to use the newer but little-known method of barycentric interpolation. This was mentioned in a recent article in this magazine (Trefethen, 2011), and we believe we are the first workers to use it in a statistical analysis. Adopting this methodology means that, unlike previous work, we model the evolution of player strength deterministically rather than stochastically.

Other authors, when modelling player or team strengths in other sports, have most commonly adopted stochastic rather than deterministic evolution of player strength. This is sometimes justifiable. In team sports, for example, there is a non-neglible stochastic component of strength evolution: as the team buys and sells top players, its strength forms part of a random walk. In individual sports however, there is a strong systematic component in strength evolution. For example, a player may gain experience in the early part of his career, reach a peak and then decline in strength until he retires. Indeed, as we find, this would appear to describe the career path of a typical golf professional.

A Plackett-Luce model with time-varying strengths

The Bradley-Terry model (Bradley and Terry, 1952) is a ratings model used for contests between two players. Such a model is appropriate for obtaining strength estimates of tennis players for example, where players compete in head-to-head matches, but many players are observed playing many matches. In golf, the situation is more complicated since many players compete in the same 'match', and we observe many of these matches. Each match is of course a golf tournament, and at the end of each tournament we have a rankings list of the players who competed. But when we have many such tournaments, with different sets of competitors, how can we combine the results? The answer comes from the generalisation of the B-T model, to be used when n players are competing in each 'match', named the Plackett-Luce model, introduced by Luce (1959), and Plackett (1975). There are more complex models (e.g. Stern, 1990) which however require computation of multivariate integrals.

Here the probability that the ith player wins is

$$P_i = \frac{\pi_i}{\sum_{j=1}^{n} \pi_j},$$

where the jth player has strength π_j. After a player has been ranked first, another player ranks second by winning out of the remaining players, and so on down to the last, who 'wins' with probability one.

Time-varying player strengths

In order to allow for time-dependent player strengths, π, we take as model variables for player i the strengths y_{ik} tabulated at m_i equally-spaced epochs t_{ik} from the first year of tournament play to the last. Next, we interpolate these tabulated strengths, using the barycentric interpolation formula

$$\pi_i(t) = \frac{\sum_{k=1}^{m_i} w_{ik} y_{ik}/(t - t_{ik})}{\sum_{k=1}^{m_i} w_{ik}/(t - t_{ik})},$$

where $w_{ik} = (-1)^k$, with first and last weights being halved. This method is described in Berrut and Trefethen (2004), and is claimed to give smaller errors than spline interpolation. It is trivial to obtain the differential $\partial\pi_i(t)/\partial y_{ik}$ which will be needed in computing the likelihood function.

To obtain a more precise estimate of a player's maximum strength, after a preliminary likelihood maximisation, a new set of interpolation epochs was chosen so that the time of maximum strength fell at a tabulated time. This necessitated the use of unequally-spaced points, when the weight formula is

$$w_{ik} = (-1)^k \{ \frac{1}{t_{ik} - t_{i,k-1}} + \frac{1}{t_{i,k+1} - t_{ik}} \},$$

with terms with out-of-range epochs omitted.

The likelihood function

Hunter (2004) describes the maximisation of Bradley-Terry likelihood functions, including the Plackett-Luce extension. He writes the log-likelihood function (for constant strengths) as

$$\ell(\pi) = \sum_{j=1}^{N} \sum_{i=1}^{m_j-1} \{\ln \pi_{a(j,i)} - \ln \sum_{s=i}^{m_j} \pi_{a(j,s)}\},$$

where there are N tournaments, with m_j players competing in the jth, and $a(j,i)$ denotes the identity of the player who came ith in the jth tournament.

The sum is up to $m_j - 1$ only because the probability for the last player is unity. With ties, however, the formula must be modified to

$$\ell(\pi) = \sum_{j=1}^{N} \sum_{i=1}^{m_j} \{\ln \pi_{a(j,i)} - \ln \sum_{s=p(j,i)}^{m_j} \pi_{a(j,s)}\},$$

where $p(j,i)$ is the position (placement) of the ith player in the jth match; several players can have the same position. This approximation for ties is that made by Cox and Oakes (1984) in the context of the proportional hazards model; the analogy with ranking models was pointed out in a short paper by Su and Zhou (2006).

The meaning of the model parameters

A player's strength $\pi(t)$ is a function of model parameters, but does not directly have any observable meaning. However, it is easy to find such meanings. Thus, the player with greater maximum strength would have been able to beat the other player more often than not. For lifetime achievement we looked at the total strength, i.e. the integral of strength over player's playing lifetime. This can be given a simple meaning in terms of observables. Imagine that a player of strength π_1 regularly competes against a player of very high ability π_2. The proportion of games won is $\pi_1/(\pi_1 + \pi_2) \simeq \pi_1/\pi_2$. Over the player's playing lifetime, the number of games won is $(\rho/\pi_2) \int \pi(t)\, \mathrm{d}t$, where ρ is the rate at which these competitions occur. Hence $\int \pi(t)\, \mathrm{d}t$ is proportional to the total number of games that the player would win against a very strong player. This is a sensible measure of lifetime achievement.

Estimation

When maximising the likelihood function for the tabulated 'strengths', we cannot use the slow but sure method of likelihood maximisation recommended by Hunter (2004).Solving for the tabulated 'strengths' y_{ik} rather than directly for the π means that we cannot manipulate the likelihood maximisation condition so that y_{ik} is the subject. This is the basis of the 'fixed point' method. A further complicating factor here is that when fitting the model to data from golf tournaments we need to maximise a likelihood function with respect to hundreds, or even thousands, of parameters. This is no easy feat, but the tractability of the likelihood function means that first and second derivatives can be found analytically, which helps. An *ad hoc* method was developed which worked well enough. The difficulty is that one does not want to try to invert the huge Hessian matrix. After some experimentation, a Newton-type method was evolved, whereby only the diagonal term of the Hessian was computed. This of course throws away the attractive property of second-order convergence, but function maximisation still proceeded rapidly until near the maximum.

Backtracking was used if a step decreased the likelihood function. In addition, single-variable minimisations were interspersed, because these can be done without using the full Hessian, and the routine switched to 'single variable mode' if backtracking failed. It tried switching out of this mode again if enough successful steps had been made using the single-variable iteration. The single variable to be iterated was chosen as the one promising the largest increase in log-likelihood, based on the first and second differentials. Finally, several random restarts were made.

A small problem is that only the ratio of the strengths π is determinable from tournament data. Hence the strength of one player at one time point was fixed at unity, e.g. $y_{11} = 1$. The y_{ik} were rescaled after each iteration to preserve this property.

The lack of a Hessian matrix again makes it harder to do statistical inference, i.e. to find errors on model parameter estimates, and to do significance tests. Here the famous bootstrap method comes to the rescue, giving us a complete methodology for parameter estimation and inference.

Implementing this methodology requires some heavy computing, even with the relatively simple Plackett-Luce model, and one of the authors was obliged to upgrade her computer! Then using fortran95 from the Numerical Algorithms Group (NAG) and use of their library of numerical subroutines made rapid computation possible. There was a wide variety of complications to be sorted out, such as the already mentioned existence of ties. Another obvious problem is that if a player performs very well, their strength parameter may tend to infinity, causing numerical problems and greatly slowing down likelihood maximisation. A small 'prior' term serves to constrain the parameters, removing the problem and greatly speeding up likelihood maximisation.

Results

We obtained data from all four of men's golf's major tournaments (the Masters, the US Open, the British Open and the USPGA) from 1996 to 2011. For each tournament, we had the finishing position of all players. In all there were 63 tournaments and 228 players. Table 1 shows the top 10 players according to total 'lifetime' strength.

INSERT TABLE 1

As expected Tiger Woods is the best player of his generation. More interesting perhaps is the identity of the other players in the list. Jim Furyk is rated as the fourth best player since 1996 yet is not typically regarded so highly in golfing circles. Furyk splits the big five of recent times, namely, Woods, Mickelson, Els, Goosen and Singh. The dominance of the US-based players is evident. Only Lee Westwood, and to some extent Sergio Garcia are players based in Europe. The rise of European golf in the last few years (at the time of writing in early 2012, European players have won seven of the last eight major titles) may mean that this list will be dominated with European golfers in future years.

Although of some interest, the results shown in Table 1 do not display the time-varying strengths of the players. Figure 1 shows these for the top 4 players.

INSERT FIGURE 1

Ernie Els and Tiger Woods' careers seem to be following a familiar pattern - an increasing strength in the early part of their careers with a decline in later years. This is a very common shape of such strength trajectories. Tiger's peak is near 2001 - a year in which Tiger dominated the world of golf. Jim Furyk and Phil Mickelson are following a more peculiar career path. Jim Furyk has two clear peaks in his ability, whilst Phil Mickelson looks like he may be entering another peak having peaked once in 2003 (although recent form suggests this may not be long-lived). Examining the trajectories of all players in our analysis shows that the inverted U-shape is very much the norm career path for modern golfers.

Tiger Woods' closest rival has been Phil Mickelson. Our parameter estimates suggest that, at their respective peak strengths (14.99 and 10.08), a match between the two would be won by Tiger with probability $14.99/(14.99 + 10.08) = 0.598$.

Conclusions

In answer to the questions set at the start of the paper, Tiger Woods does indeed seem to be the best golfer of his generation, but given the patterns observed in players' strengths, it seems unlikely that he will once again reach the heights of his powers observed around 2001. Golfers typically increase in strength, and having reached a peak, face a slow decline in playing ability until they retire. Of course, part of Tiger's recent decline has been down to injury (the few tournaments Tiger has played recently have seen him perform somewhat below the standards he set in the early part of his career), which, as Tiger gets older is more and more likely to happen, suggesting his decline will continue. However, his decline in form was also partly due to his personal problems. As such, if these are overcome, he may find himself entirely focussed on his golfing legacy once more, and if anyone can have a 'twin peak career', Tiger can.

Future work in this area will be to answer the ultimate question for any golf fan: 'who is the greatest golfer ever?'. To do so, one would need data from a much longer time frame (golf's major competitions began in 1860 with the British Open), and even more heroic computing, but getting answers to questions such as 'who would win a match between Tiger Woods and Jack Nicklaus' would make the effort worth it.

Bibliography

[1] Berrut, J. P. and Trefethen, L. N. (2004). Barycentric Lagrange interpolation, SIAM Review, 46, 501-517.

[2] Bradley, R. A., and Terry, M. E. (1952). Rank analysis of incomplete block designs I: the method of paired comparisons. Biometrika, 39, 324345.

[3] Cox, D. R. and Oakes, D. (1984). Analysis of Survival Data, Chapman and Hall, New York.

[4] Harrell, Jr., F.E. (2001). Regression modeling strategies. Springer series in statistics, New York.

[5] Hunter, D. R. (2004). MM Algorithms for generalized Bradley-Terry models, The Annals of Statistics, 32, 384-406.

[6] Luce, R. D. (1959). Individual Choice Behavior: A Theoretical Analysis. Wiley, New York.

[7] Plackett, R. L. (1975). The analysis of permutations, Applied Statistics, 24, 193-202.

[8] Stern, H. (1990). Models for Distributions on Permutations, Journal of the American Statistical Association, 85, 410, 558-564.

[9] Su, Y. and Zhou, M. (2006). On a connection between the Bradley-Terry model and the Cox proportional hazards model, Statistics and Probability Letters, 76, 698-702.

[10] Trefethen, L. N. (2011). Six myths of polynomial interpolation and quadrature, Mathematics Today, August.

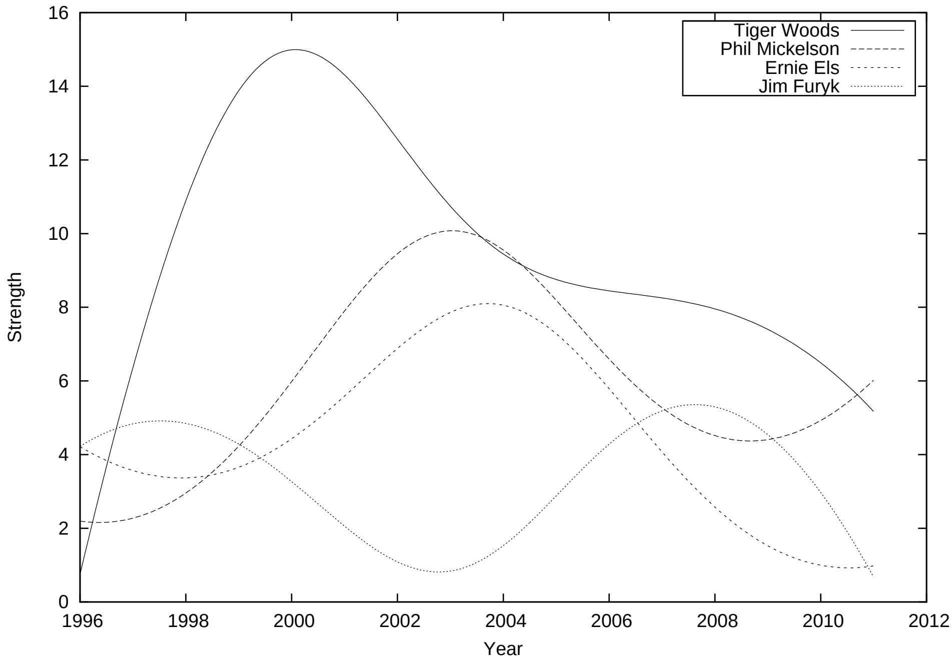

Figure 0.1: Lifetime strengths of the four players with highest lifetime strength.

Figures and Tables

Player	total lifetime strength (cv)
Tiger Woods	12.64 (0.17)
Phil Mickelson	8.16 (0.19)
Ernie Els	6.12 (0.18)
Jim Furyk	4.56 (0.17)
Vijay Singh	4.55 (0.16)
Retief Goosen	4.11 (0.21)
Lee Westwood	3.3 (0.15)
Mike Weir	2.84 (0.16)
Sergio Garcia	2.81 (0.15)
Stewart Cink	2.66 (0.11)

Table 0.1: Top 10 players by total ‘lifetime’ strength with coefficient of variation.