**Research Article**

Rose D. Baker and Ian G. McHale*

# Estimating age-dependent performance in paired comparisons competitions: application to snooker

https://doi.org/10.1515/jqas-2023-0082
Received September 5, 2023; accepted December 22, 2023;
published online February 1, 2024

**Abstract:** We first present a model for the outcome of snooker matches in which player strengths are allowed to vary deterministically with time. The results allow us to identify the greatest players of all time, and to examine the relationship between age and performance. Second, we present a random effects model which uses the estimated strengths from our first model, to forecast player performance, and to assess the extent to which early promise has been maintained. Ronnie O'Sullivan and Stephen Hendry are the two candidates for the title of the greatest of all time. We find that peak performance occurs between the ages of 25 and 30, younger than would be expected when compared to findings in other sports. Outside sport, these findings contribute to the general literature on variation of performance with age.

**Keywords:** paired comparisons; snooker; age-dependence; barycentric interpolation; random-effects model

## 1 Introduction

This paper presents a model for the outcome of snooker matches. Player strengths are estimated using a time-varying paired comparisons model that allows these strengths to vary with time and age throughout the careers of players. We use this model to investigate two main issues. First, we identify the maximum strength achieved by players, and which player had the highest strength of all, identifying the greatest snooker player of all-time. Second, we use the results in a second model to consider the general relationship between age and performance, and model how performance evolves with age. The first question is of interest to sports fans (who is the greatest snooker player of all-time?), whilst the second question is of general interest to society (how are age and performance related?).

Snooker is our sport of choice for two reasons: first, there are data on the results of each match from the last 100 years, and there are accompanying details of the players involved in matches including their ages. Second, snooker is a sport with a wide fan base that is continuing to grow in popularity. This has been driven by a rise in prominence in Asia, specifically China. Despite the rise in popularity in terms of both participation and spectating (World Snooker estimates that there are 350 million snooker fans in China alone (Gunia 2022)), the scientific community has not caught up, and in contrast to many other sports, there is very little analysis of snooker in the scientific literature.

Snooker is a cue sport, like pool, but played on a larger 12 foot by 6 foot table covered in a green cloth called a baize. The table has six 'pockets' into which balls can be potted. A frame begins with 15 red balls (each worth 1 point), a yellow, a green, a brown, a blue, a pink, and a black ball (worth 2, 3, 4, 5, 6 and 7 points respectively). Players alternate turns (shots) attempting to pot a red ball into one of the pockets (or prevent the opponent from potting a red ball on his shot) by hitting a white ball onto a red ball. If a player is successful in potting a red ball, he must play a subsequent shot in which he can choose to pot one of the non-red balls, or 'play safe', i.e. attempt to prevent the opponent from potting a red on the next turn. The player with the most points at the end of the frame wins that frame.

The paper is structured as follows. In the next section we provide a review of the literature on the two topics of relevance: time-varying ratings models, and the relationship between age and performance. Section 3 presents the data we use and some descriptive statistics. Our model is presented in Section 4. The results of our analysis for identifying the greatest player of all-time, and the relationship between age and performance are given in Section 5. Section 6 provides modelling and results for forecasting future performance, and considers which players have

***Corresponding author: Ian G. McHale**, University of Liverpool Management School, Liverpool, L69 7ZH, UK, E-mail: ian.mchale@liverpool.ac.uk. https://orcid.org/0000-0002-7686-3879
**Rose D. Baker**, Salford Business School, University of Salford, Salford, M5 4WT, UK

fulfilled their potential. We conclude with some closing remarks in Section 7.

# 2 The literature

## 2.1 Time varying ratings models

Snooker is a form of paired comparison competition in that players compete in head-to-head matches. The two competitors in any one match are selected from a larger population of players. There are several models for paired comparisons which are formulated such that each competitor is associated with a strength, and the probability that one competitor beats another is a function of the two strengths. Note that strengths are taken as positive, and sometimes logged strengths are used.

The Bradley–Terry model (Bradley and Terry 1952) is perhaps the best-known paired comparisons model and assumes that the probability of winning is given by the distribution function of the logistic distribution of the difference in the two logged strengths. The Thurstone–Mosteller model (Thurstone 1927) on the other hand uses a normal distribution of logged strengths. Stern (1990) provides a more general paired comparisons model and shows how his model provides a continuum of paired comparisons models that includes the Bradley–Terry and Thurstone–Mosteller models as special cases.

However, these models are time-invariant in that competitor strengths are assumed to be constant. To allow for the dynamics of competitor strength, one solution is to re-estimate the model as new data are recorded. Indeed, this is the approach taken by, for example, McHale and Morton (2011) who allow for matches to be exponentially weighted such that results in the more recent past carry greater importance when estimating each player's strengths.

An alternative approach is to allow for the stochastic nature of strengths and explicitly model the time-varying changes in competitor strengths. This is the approach adopted by Glickman (1999) and Knorr-Held (2000).

Another alternative is to model the evolving competitor strengths deterministically. For example, Baker and McHale (2014a) presented a methodology to estimate time-varying ratings for paired comparisons, and used it to rank men's tennis players, before adapting the methodology to use empirical Bayes to rank women's tennis players in Baker and McHale (2017). Here we use this methodology in snooker but adapt the analysis to additionally look at how the estimated strengths of players vary with age.

## 2.2 The relationship between age and performance

It is perhaps an obvious statement that performance in almost all fields of human endeavour is related to age. However, understanding when humans are at their 'peak' has long fascinated scientists, and estimating when the peak is, and the rate of decay from this peak, is a question that is of general interest to society.

In a series of articles starting in the 1940s, Lehman charted the age of success in many fields including chemistry and physics (Lehman 1945), and fourteen sports including boxing, golf, tennis, corn-husking, and pistol shooting (Lehman 1951). The somewhat consistent finding is that success is more likely to occur in the late 20s to early 30s.

Since Lehman's descriptive work many authors have continued to use descriptive statistics to examine the relationship between age and performance. For example, Schulz and Curnow (1988) chart the ages of gold medalists in Olympic disciplines, and the age of success in several other sports including tennis, golf and baseball. The insight gained in their work is that peak age (early 20s) in sports requiring power and speed (for example, weightlifting and the 100 m) occurs earlier than in sports (e.g. long-distance running and tennis) requiring stamina and guile (late 20s to early 30s).

These studies are all descriptive in nature. Further, they do not explicitly model the relationship between age and performance at the level of the individual. Instead they chart the age of winners of competitions. This means that they only describe the ageing characteristics of the elite, who may have different ageing characteristics from the other competitors. Hence it is more informative to directly model the evolution of performance metrics over the career of individuals.

Roring and Charness (2007) present a mixed effects model for the evolution of Elo ratings in chess. They find that peak age is in the early 40s, and that better chess players decay at a slower rate then less able players. Baker and McHale (2023) model the evolution of performance metrics in golf. They find that the different skill-sets required for golf (power and speed for driving the golf ball a long distance, and touch and finesse for putting and chipping) do indeed peak at different ages.

To be able to model the relationship between age and performance, one needs a measure of performance. In both Roring and Charness (2007) and Baker and McHale (2023) performance metrics already existed. In the present paper, for snooker, no such off-the-shelf metric exists and we must use a paired comparisons model to estimate player strengths that we employ as the performance metric with which we will examine the relationship between age and performance.

## 3 Data

We collected data from www.cuetracker.net, a website which holds the results of every match in every professional tournament since 1908. Two datasets were available, one giving the results of matches played, the other giving biographical data such as date of birth. Despite being recorded in the database as a victory for one player over another, some matches were not actually played for various reasons. For example, one player was ill. These 'walkovers' were ignored in our analysis, leaving 109,733 matches with 1642 players. There was a total of 789,763 frames played. Of these players, 71.7 % had birth dates supplied, these being the better-known players.

In identifying the greatest snooker player of all time, most fans would simply look at the number of World Championship titles won. There is a complication in this since in its early days, snooker was not a sport played by many people and only a handful of competitors entered the World Championship. Further, as the sport boomed in popularity, there was major reform in how snooker operated in 1977 when the World Championship moved to the Crucible Theatre in Sheffield, UK. Prior to that year, the tournament was organised very differently. For example, in several years, the title holder from the previous year automatically qualified for the final. Further, the number of entries was as low as five. The interested reader can discover more about the format of the World Championships prior to 1977 on the wikipedia page.[1] Most fans nowadays consider 1977 to be the year that the World Championship started. This is the year that the tournament became a knockout format with 32 players entering the final tournament stage (with several pre-qualifying rounds taking place to make it into the final 32).

Discarding the first half-century of results is somewhat unsatisfactory however as it means that the records of some greats are simply deleted from history. For example, Joe Davis won the World Championship 15 times between 1927 and 1946 (in every year it was held). His brother, Fred Davis, then won 8 world championships between 1948 and 1956, and John Pulman won the next 8 titles. Our model can be used to include these in the debate on who is the greatest.

For reference, the players who top the list based on number of World Championships won since 1977 are: Ronnie O'Sullivan and Stephen Hendry (on 7 world titles), Steve Davis (on 6 world titles), and John Higgins and Mark Selby (on 4 world titles).

Our analysis is not constrained to consider only the World Championship. We have results of all matches in all tournaments. Table 1 shows the highest performing players according to match win percentage and frame win percentage, conditional on having played in a minimum number of 100 matches in professional tournaments.

Table 1 demonstrates the problems that often arise when using simple descriptive statistics to rank players. First, in order to avoid 'silly' entries, we have imposed the arbitrary condition that players must have played in at least 100 matches. Second, even with this condition, there are unexpected entries in the list. Together with the biggest names in snooker history (e.g. Ronnie O'Sullivan, Joe Davis, John Higgins, Steve Davis), there are some players who are unknown to all but the most ardent of fans. Roger Garrett and Matt Wilson, for example, would not feature on any list of the top snooker players of all time.

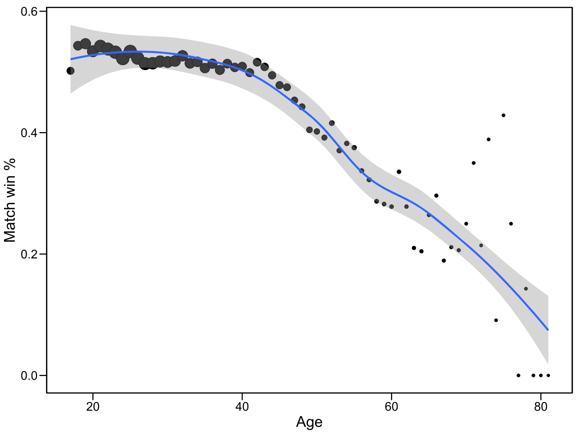
Such lists have these 'errors' for two reasons. First, no account is taken of the strength of opponents faced. Even for the truly top players, there may be complications in that there may be a cluster of super-strong players competing for trophies in the same era. They may have shared the spoils on offer, when in another era any one of them might have been dominant. The rivalry between Rafael Nadal, Roger Federer and Novak Djokovic in men's tennis comes to mind. Second, no account is taken of the amount of evidence there is for how strong (or weak) the player is. Only by using a model of strength can these problems be addressed.

Turning to the second focus of our paper, we now consider some descriptive statistics for examining the relationship between age and performance. Figure 1 shows (a) match win percentage, and (b) frame win percentage as functions of age. Each dot represents the average for all players of that age. The size of the dot corresponds to the number of players of that age over which the average was calculated. The pattern appears to show that performance deteriorates from the early 20s. However, it should be noted that only the very best snooker players turn professional at a young age and so the early part of the curve is biased upwards. The later part may also be biased, as many players whose performance has deteriorated may have retired from the game.

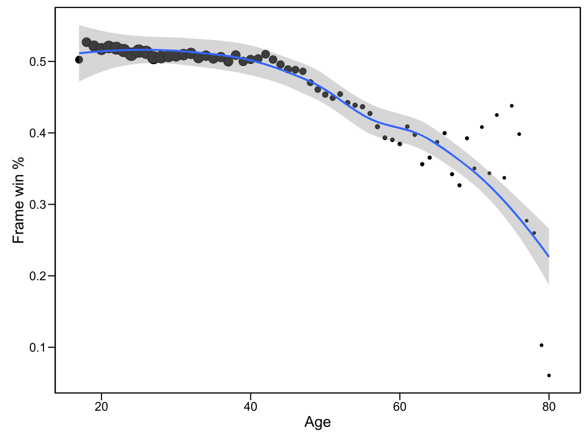
1 https://en.wikipedia.org/wiki/World_Snooker_Championship.

**Table 1:** Naive rankings lists of snooker players based on their match win percentage (left) and frame win percentage (right).

| Rank | Name | Matches | Matches won | Match win % | First year | Last year | Name | Frames played | Frames won | Frame win % | First year | Last year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Ronnie O'Sullivan | 1492 | 1120 | 75.06702 | 1992 | 2023 | Ronnie O'Sullivan | 12,554 | 7664 | 61.04827 | 1992 | 2023 |
| 2 | Joe Davis | 134 | 100 | 74.62687 | 1927 | 1958 | Roger Garrett | 859 | 523 | 60.88475 | 1992 | 1995 |
| 3 | Roger Garrett | 124 | 89 | 71.77419 | 1992 | 1995 | Judd Trump | 9075 | 5460 | 60.16529 | 2005 | 2023 |
| 4 | Judd Trump | 1341 | 952 | 70.9918 | 2005 | 2023 | John Higgins | 13,740 | 8098 | 58.93741 | 1992 | 2023 |
| 5 | John Higgins | 1766 | 1246 | 70.55493 | 1992 | 2023 | Stephen Hendry | 11,844 | 6938 | 58.57818 | 1985 | 2023 |
| 6 | Stephen Hendry | 1294 | 904 | 69.8609 | 1985 | 2023 | Mark Selby | 10,404 | 6062 | 58.26605 | 1998 | 2023 |
| 7 | Mark Selby | 1481 | 1012 | 68.33221 | 1998 | 2023 | Neil Robertson | 8822 | 5098 | 57.78735 | 1998 | 2023 |
| 8 | Alan Burnett | 155 | 105 | 67.74194 | 1995 | 2002 | Steve Davis | 12,530 | 7197 | 57.43815 | 1978 | 2016 |
| 9 | Ding Junhui | 938 | 626 | 66.73774 | 2002 | 2023 | Mark Williams | 12,597 | 7231 | 57.40256 | 1992 | 2023 |
| 10 | Neil Robertson | 1202 | 802 | 66.72213 | 1998 | 2023 | Ding Junhui | 7092 | 4056 | 57.1912 | 2002 | 2023 |
| 11 | Nick Marsh | 123 | 81 | 65.85366 | 1992 | 1996 | Chris Scanlon | 1589 | 905 | 56.95406 | 1992 | 2000 |
| 12 | Mark Allen | 1007 | 663 | 65.83913 | 2003 | 2023 | Mark Allen | 6967 | 3960 | 56.83939 | 2003 | 2023 |
| 13 | Shaun Murphy | 1290 | 846 | 65.5814 | 1997 | 2023 | Stuart Bingham | 9904 | 5626 | 56.80533 | 1995 | 2023 |
| 14 | Mark Williams | 1743 | 1142 | 65.51922 | 1992 | 2023 | Shaun Murphy | 9250 | 5252 | 56.77838 | 1997 | 2023 |
| 15 | Paul Hunter | 342 | 224 | 65.49708 | 1995 | 2006 | Matt Wilson | 1699 | 964 | 56.73926 | 1993 | 2001 |
| 16 | Steve Davis | 1469 | 959 | 65.28251 | 1978 | 2016 | Yan Bingtao | 2479 | 1406 | 56.71642 | 2013 | 2023 |
| 17 | Matt Wilson | 230 | 150 | 65.21739 | 1993 | 2001 | Alan Burnett | 1145 | 649 | 56.68122 | 1995 | 2002 |
| 18 | Scott MacFarlane | 155 | 101 | 65.16129 | 1989 | 2016 | Scott MacFarlane | 1064 | 603 | 56.67293 | 1989 | 2016 |
| 19 | Stephen Maguire | 1158 | 748 | 64.59413 | 1997 | 2023 | Daniel Ward | 688 | 388 | 56.39535 | 2002 | 2020 |
| 20 | Yan Bingtao | 360 | 232 | 64.44444 | 2013 | 2023 | Eddie Lott | 774 | 436 | 56.33075 | 1992 | 1996 |




(a) Match win percentage and age




(b) Frame win percentage and age

**Figure 1:** Relationship between match win percentage and age (left), and frame win percentage and age (right). Each dot represents the percentage as calculated for a single year of age. The size of the dots are proportional to the number of players over which the average was calculated. The smoothed loess curve is also shown on both plots.

In order to understand the relationship between age and performance, we need to study age-related performance at the level of the individual.

# 4 Methodology

## 4.1 A paired comparisons model for snooker

Our starting point for modelling the outcome of snooker matches is the methodology of Baker and McHale (2017) who presented a time-varying paired comparisons model for women's tennis. As in tennis, snooker has 'matches' within matches. In tennis, a player needs to win games to win sets, and sets to win a match. In snooker, players compete in frames and matches are typically 'best-of $x$ frames', where $x$ ranges from 5 to 35. Baker and McHale (2017) choose the set as the unit of victory to be modelled in their analysis of women's tennis. Here we choose the frame. The simple justification of this is that less information is lost when using frame results. For example, a player who wins a snooker match by a score of 10-1 in frames has likely performed better than a player beating the same opponent by a score of 10-9.

The Baker and McHale (2017) model is based on Stern's (Stern 1990) generalization of the Bradley–Terry model which we use to model the likelihood of the observed numbers of frames scored by each player in a match. This uses the ratio of player 'strengths'. The probability that player $i$ beats player $j$ is given by

$$p_{ij} = B(\beta, \beta)^{-1} \int_0^{\alpha_i/(\alpha_i+\alpha_j)} y^{\beta-1}(1-y)^{\beta-1}\,dy, \tag{1}$$

where player $i$'s strength is $\alpha_i$, $\beta$ is a parameter to be estimated and B denotes the beta function. For $\beta = 1$, the model reduces to the familiar Bradley–Terry model $p_{ij} = \alpha_i/(\alpha_i + \alpha_j)$, whilst as $\beta \to \infty$ the Thurstone–Mosteller (TM) model is obtained, as shown by Stern (1990). Note that this statement needs some elaboration, because the model tends to $p_{ij} = \Phi\{\sqrt{\beta/2}(\ln\alpha_i - \ln\alpha_j)\}$, where $\Phi$ is the standard normal distribution function. Thus the difference of logged strengths is given a variance of $2/\beta$, which must be small if $\beta$ is large. We would probably instead want to fit the model $p_{ij} = \Phi\{(\ln\alpha_i - \ln\alpha_j)/\sigma\}$, where $\sigma \simeq 1$. Thus although the Stern model for given $\alpha_i, \alpha_j$ does indeed tend to the TM model, it is not the TM model that one would fit. This is however not a problem, because when the $\alpha_i$ are estimated from data, if $\alpha_i \to \alpha_i^{\beta^{-1/2}}$, we regain the TM model with $\sigma = 1$. Thus with strengths estimated from data, for large $\beta$ we do indeed attain the general TM model.

Stern (1992) stated that with currently available sample sizes, the choice of paired-comparison model did not matter. With our much larger sample size, we can clearly see a better fit to data (larger log-likelihood) with the Stern model. Because residual error is the yardstick against which effects are measured, a better-fitting model will in general yield more accurate predictions.

## 4.2 A time-varying paired comparisons model

As in Baker and McHale (2017), strengths are allowed to vary deterministically with time, and were modelled using barycentric interpolation (Baker and Jackson 2014; Berrut and Trefethen 2004). This is analogous to spline interpolation, but arguably simpler and more accurate. Nodes replace the knots of the spline method, and the model parameters are the strengths at each node for each player. One of these strengths has to be fixed, as only strength ratios are used.

The strength of player $i$ at time $t$ is given by

$$\alpha_i(t) = \frac{\sum_{k=1}^{n_i} w_{ik}\lambda_{ik}/(t - t_{ik})}{\sum_{k=1}^{n_i} w_{ik}/(t - t_{ik})} \tag{2}$$

where $\lambda_{ik}$ is the $k$th strength of player $i$, i.e. the strength at time $t_{ik}$, so that $\alpha_i(t_{ik}) = \lambda_{ik}$. To differentiate between $\alpha_i(t)$ and $\lambda_{ik}$, we call the latter the *tabulated strength*. At time $t_{ik}$, the two are the same. There are $n_i$ such nodes for player $i$. The $w_{ik}$ are weights and we use weights of order zero such that $w_{ik} = (-1)^k$. The tabulated strengths, $\lambda_{ik}$, are the model parameters.

Let there be $n_p$ players in total in the dataset. Let the $l$th player have $n_l$ tabulated strengths $\lambda_{l1} \dots \lambda_{ln_l}$, with log-mean $y_l = \ln\{\sum_{k=1}^{n_l} \lambda_{lk}/n_l\}$. Let there be $n_m$ matches in total in the dataset, and let the winner of the $j$th match be player number $i_{j1}$, the loser number $i_{j2}$. The numbers of frames won are respectively $f_{i_{j1}}, f_{i_{j2}}$, where $f_{i_{j1}} < f_{i_{j2}}$. Let $P_j(\lambda_{i_{j1}}, \lambda_{i_{j2}})$ be the probability that player $i_{j1}$ wins, where $\lambda$ denotes the vector of tabulated strengths. Of course, $P$ is simply a function of the two player strengths as interpolated from the $\lambda$ values at the time of the $j$th match. The log-likelihood is then

$$\ell_0 = \sum_{j=1}^{n_m} \left\{ f_{i_{j1}} \ln P_j + f_{i_{j2}} \ln(1 - P_j) \right\},$$

which we maximize for all parameters including the β parameter from Stern's continuum of paired comparisons models. Players with few games played can present a problem, because their performance may appear spuriously good. The Empirical Bayes method used in Baker and McHale (2017) was applied, and this shrinks the scores of players with few games towards the average and so removes the problem.

The way to shrink scores is to assume that the log-means $y_l$ are normally and independently distributed $Y_L \sim N[\mu, \phi^2]$. The posterior likelihood is then obtained by multiplying the likelihood by the product of the prior distributions for each player's mean log-strength (the normal probability densities). The logarithm of the posterior pdf is then

$$\ell = \ell_0 - \frac{1}{2}\sum_{l=1}^{n_p} \left\{ \frac{(y_l - \mu)^2}{\phi^2} \right\} - \frac{n_p}{2} \ln \phi^2 - \frac{n_p}{2} \ln(2\pi). \tag{3}$$

Our procedure is empirical Bayes because we estimate the 'prior' parameters $\mu$ and $\phi$ from the data.

The log-likelihood for the tabulated strengths must be maximized, and a procedure for doing this is described in Baker and McHale (2014b). Although in this case there were 3500 nodes, with therefore 3499 parameters, the log-likelihood can be quickly maximized by analytically calculating its first and second derivatives for each variable, and ignoring cross-terms. The computation and storage space required to retain the cross-terms would be immense.

Zeroing cross-terms destroys the second-order convergence of the Newton-Raphson method, but iteration nevertheless proceeds quickly to the same function maximum. Random restarts were used to ensure that the global maximum was found and not a local hilltop.

Snooker matches take the format of 'best-of-x-frames' and end once a player reaches the number of frames whereby the other player can no longer win more frames even if all remaining frames were played. This stopping rule (that could be included in the log-likelihood) is not a function of model parameters, and so can be ignored in likelihood-based inference. The only problem is that our goodness of fit tests require the predicted probability of winning, which is different from the tennis case.

If the probability of winning a frame is $p$, and the game stops when either player wins $n$ frames, the probability of winning is the probability $P$ that a player wins at least $n$ out of $2n - 1$ games. This is $1 - F(2n - 1, p)$, where $F$ is the distribution function of the binomial distribution, i.e.

$$P = \sum_{i=n}^{2n-1} \binom{2n-1}{i} p^i (1-p)^{2n-1-i}.$$

Play stops when a player wins $n$ games, but the probability of a win is the same as that calculated assuming that all $2n - 1$ games are played.

In the Stern (1990) model, the probability that a player wins can be written as the distribution function of a symmetric beta-distribution with parameter $\beta$. Here the MLE was $\hat{\beta} = 1.79$, standard error 0.062. This is close to the Bradley–Terry model, but significantly different. To see what the model looks like, with $r = \alpha_1/(\alpha_1 + \alpha_2)$, for the closely related case of $\beta = 2$ the Stern model gives the probability $p$ that player 1 wins a frame as $p = r^2\{1 + 2(1 - r)\}$.

Goodness of model fit was assessed as per Baker and McHale (2017), by comparing the observed and predicted numbers of wins over 10 years around peak strength for the top 50 players. The top 50 players are identified by locating every player's peak strength, then summing the strengths in the 5 years either side of this moment. The top 50 are then the players with the highest 10-year total strengths. This gave a goodness of fit chi-squared of 24.60 on 50 degrees of freedom. One can conjecture that the contribution to this of each player is roughly independent of the others, because a player competes against so many other players. However, it must be accepted that all methods of assessing goodness of fit of complex models are questionable.

A little discussed issue in fitting paired comparisons models is that of 'connectivity'. For the model to be estimable, every player has to be connected through play to every other player, and there must not be two or more disjoint sets of players. For example, players A and C can be connected even though player A never played player C, but they did both play against player B, and so on. This issue is especially important to consider when estimating a paired comparisons model on a dataset spanning more than a century of results. The check of connectivity presented in Baker and McHale (2017) was carried out here and this showed that all players were connected, i.e. there were not two or more disjoint sets of players.

### 4.3 Weighting of different match types

Fitting strength parameters by maximum-likelihood estimation implicitly gives equal weight to all levels of match. However, championship matches will have more effect on the strength estimates, because more frames are played. It may be that skilled players put less effort into playing in lower-level competitions, which would mean that more important matches should be given a higher weight. This idea is of particular relevance in snooker as the World Championship carries so much weight, relative to other tournaments, both in terms of prizes, and history.

To test whether players 'try harder' in certain tournaments, 20 % of matches in which both players ultimately played 15 or more matches were randomly omitted from the likelihood maximization. The outcome of games that had been omitted was then predicted using the model fitted to the remaining data, so that this is an out-of-sample fit. The probability of successfully predicting the winner was then 0.608, Brier score 0.203.

Weighting by level of game can be achieved by weighting by the number of frames $x$ required to win. Several 2-parameter weighting schemes were tried. For example, the weight $w(x)$ given by

$$w(x) = (x + \sqrt{1 + (ax)^2}/a)^\gamma,$$

where $x$ is the number of frames to be attained and $a > 0$. This function is a constant if $\gamma = 0$, and increases if $\gamma > 0$, being convex if $\gamma > 1$, else concave. For $\gamma < 0$, it decreases in a convex or concave way. This is thus a flexible function that can capture the different kinds of weighting that might be needed. However, varying $a$ and $\gamma$ could never improve on

**Table 2:** Rankings of snooker players according to maximum one-year strength, maximum average five-year strength, and maximum average ten-year strength.

| Rank | One-year strength | | | | Five-year strength | | | | Ten-year strength | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Player | Strength | Year | 95 % CI | Player | Strength | Year | 95 % CI | Player | Strength | Year | 95 % CI |
| 1 | Ronnie O'Sullivan | 2.547(0.019) | 2004 | (1, 3) | Ronnie O'Sullivan | 2.452(0.064) | 2010–2015 | (1, 1) | Ronnie O'Sullivan | 2.397(0.100) | 2010–2019 | (1, 2) |
| 2 | Stephen Hendry | 2.547(0.019) | 1993 | (1, 5) | Stephen Hendry | 2.403(0.107) | 1991–1996 | (2, 3) | Stephen Hendry | 2.222(0.202) | 1990–2000 | (1, 3) |
| 3 | Steve Davis | 2.490(0.031) | 1993 | (1, 7) | Judd Trump | 2.173(0.101) | 2016–2021 | (1, 6) | Steve Davis | 2.136(0.221) | 1984–1994 | (2, 5) |
| 4 | Mark Williams | 2.404(0.090) | 2003 | (1, 12) | Steve Davis | 2.169(0.244) | 1990–1995 | (3, 6) | John Higgins | 2.126(0.120) | 1997–2007 | (3, 6) |
| 5 | John Higgins | 2.297(0.018) | 1999 | (2, 9) | John Higgins | 2.151(0.099) | 1997–2002 | (3, 7) | Judd Trump | 2.088(0.146) | 2011–2021 | (4, 8) |
| 6 | Judd Trump | 2.275(0.021) | 2019 | (2, 14) | Mark Williams | 2.128(0.187) | 1999–2003 | (2, 12) | Mark Selby | 2.043(0.112) | 2007–2017 | (4, 9) |
| 7 | Stephen Maguire | 2.229(0.045) | 2004 | (1, 20) | Mark Selby | 2.052(0.121) | 2007–2012 | (5, 11) | Neil Robertson | 1.957(0.081) | 2012–2022 | (4, 9) |
| 8 | Mark Selby | 2.207(0.037) | 2016 | (3, 15) | Stephen Maguire | 1.986(0.228) | 2004–2008 | (1, 21) | Mark Williams | 1.959(0.306) | 1996–2006 | (3, 17) |
| 9 | Paul Hunter | 2.155(0.028) | 2003 | (1, 24) | Neil Robertson | 1.974(0.037) | 2012–2016 | (4, 16) | Ding Junhui | 1.895(0.072) | 2005–2014 | (6, 16) |
| 10 | James Wattana | 2.154(0.015) | 1993 | (3, 21) | Shaun Murphy | 1.941(0.128) | 2004–2009 | (3, 20) | Ken Doherty | 1.896(0.118) | 1992–2002 | (7, 16) |
| 11 | Alan McManus | 2.138(0.026) | 1994 | (3, 23) | Stephen Lee | 1.935(0.092) | 1998–2003 | (2, 24) | Stephen Maguire | 1.900(0.187) | 2003–2013 | (3, 22) |
| 12 | Ken Doherty | 2.116(0.016) | 1993 | (5, 27) | Ken Doherty | 1.913(0.030) | 1998–2002 | (8, 19) | Shaun Murphy | 1.870(0.144) | 2005–2014 | (7, 21) |
| 13 | Shaun Murphy | 2.102(0.045) | 2008 | (5, 28) | James Wattana | 1.907(0.206) | 1990–1995 | (5, 22) | Mark Allen | 1.851(0.075) | 2009–2019 | (7, 21) |
| 14 | Neil Robertson | 2.082(0.018) | 2019 | (4, 23) | Ding Junhui | 1.900(0.082) | 2009–2014 | (7, 22) | Matthew Stevens | 1.837(0.101) | 1998–2007 | (4, 21) |
| 15 | Stephen Lee | 2.041(0.020) | 2002 | (4, 28) | John Parrott | 1.894(0.114) | 1990–1995 | (9, 24) | Stuart Bingham | 1.805(0.098) | 2010–2020 | (7, 22) |
| 16 | John Parrott | 2.027(0.005) | 1993 | (4, 26) | Mark Allen | 1.882(0.073) | 2009–2014 | (8, 25) | Stephen Lee | 1.800(0.174) | 1997–2007 | (7, 24) |
| 17 | Peter Ebdon | 2.015(0.013) | 1995 | (7, 27) | Matthew Stevens | 1.870(0.083) | 1999–2003 | (8, 25) | John Parrott | 1.792(0.138) | 1989–1999 | (11, 25) |
| 18 | Stuart Bingham | 2.012(0.040) | 2006 | (5, 31) | Peter Ebdon | 1.851(0.135) | 1992–1997 | (11, 26) | Peter Ebdon | 1.791(0.140) | 1994–2003 | (11, 26) |
| 19 | Graeme Dott | 1.998(0.016) | 2005 | (4, 31) | Stuart Bingham | 1.838(0.054) | 2014–2018 | (11, 26) | Ali Carter | 1.779(0.042) | 2005–2014 | (11, 26) |
| 20 | Mark Allen | 1.996(0.010) | 2013 | (8, 30) | Jimmy White | 1.833(0.115) | 1990–1995 | (14, 28) | James Wattana | 1.780(0.207) | 1986–1996 | (10, 31) |

the Brier score obtained with $\gamma = 0$. Hence it seems that the results of lower-level matches are not more or less noisy than others, and that the standard of play either does not vary with the level of competition, or else both players change in the same way, so that the ratio of their strengths remains constant. This is perhaps good news for sports integrity in that we find no evidence for varying strengths by player across different levels of matches.

# 5 Results

## 5.1 Player rankings

Table 2 shows player rankings based on average strength over 1, 5 and 10 year periods. It is a *who's who* of snooker. Ronnie O'Sullivan comes top according to the rankings based on each of the three time periods. His best single year is estimated as 2004, but more sustained periods of supremacy start in 2010. Stephen Hendry is a close runner up on all three rankings. If the two snooker giants were to play each other at their respective peak strengths, the probability that Ronnie O'Sullivan wins is 50.8 %. At their respective peaks the two players are almost inseparable.

There are a few surprise entries, and omissions. For example, no player prior to 1993 appears in the one-year list, and no player prior to 1984 (Steve Davis, 1984–1994) appears on the ten-year list. For the 1-year rankings, Fred Davis comes in at rank 24, with his brother Joe as number 35. In general, top players from earlier years do not score highly in the rankings. This is true even with a crude ranking (see Table 1, based on the percentage of frames played that were won). It is not obvious why this should be for anything other than the simple sporting reason – players are better nowadays than the players who dominated in their day. There is certainly a much larger pool of talent now, so perhaps it should be expected that standards are much higher.

The confidence intervals on ranks displayed in Table 2 are derived by bootstrapping the data 100 times using the parametric bootstrap (i.e. having simulated results from the model). Their large size shows that for shorter periods, it could have turned out that several other players would have come top. However, for the 5-year ranking Ronnie O'Sullivan always comes top. For the 10-year ranking, Stephen Hendry performs nearly as well, and given different stochastic outcomes, could have come top, again demonstrating just how closely matched these two players are.

Figure 2 shows the evolution of strengths for three of snooker's greatest players: Ronnie O'Sullivan, Stephen Hendry and Steve Davis. The confidence intervals show how close the three are at their peaks, but perhaps the most remarkable observation is the longevity of Ronnie O'Sullivan's career. Unlike Hendry and Davis, he has yet to decline from his peak strength towards retirement. Just how remarkable this is will be considered in the next section.

## 5.2 Age-dependence of performance

In order to estimate the relationship between age and performance we take averages of the strength curves such as the ones displayed in Figure 2. The result is shown in Figure 3. This is a (fairly) nonparametric estimate of the relationship between snooker performance and age. To obtain this 'average' curve we looked only at curves of players with a minimum of 10 years of play. Each individual's strength curve had to be normalized and this was done by dividing by that individual's highest average of performances at 3 consecutive nodes. The standardized performances have been

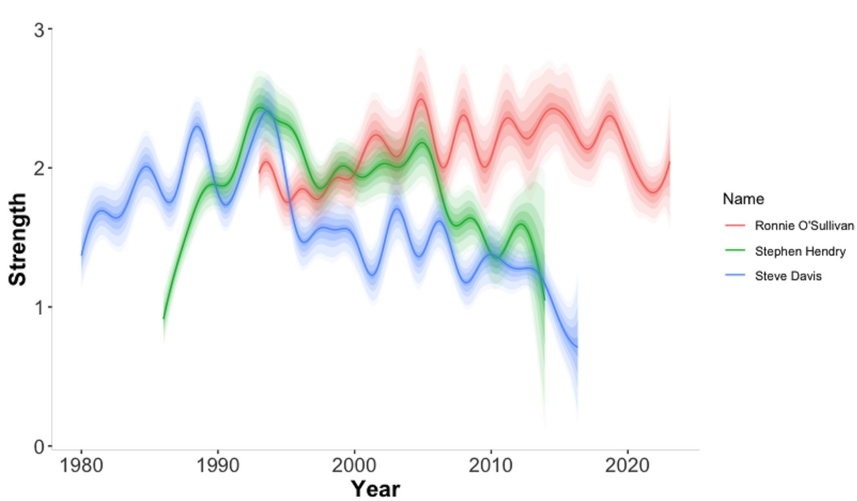



**Figure 2:** Evolution of player strengths for three of snooker's greatest players: Ronnie O'Sullivan, Stephen Hendry and Steve Davis. The shaded areas around the curves represent the 20 %, 50 %, 70 %, 90 % and 95 % confidence intervals.

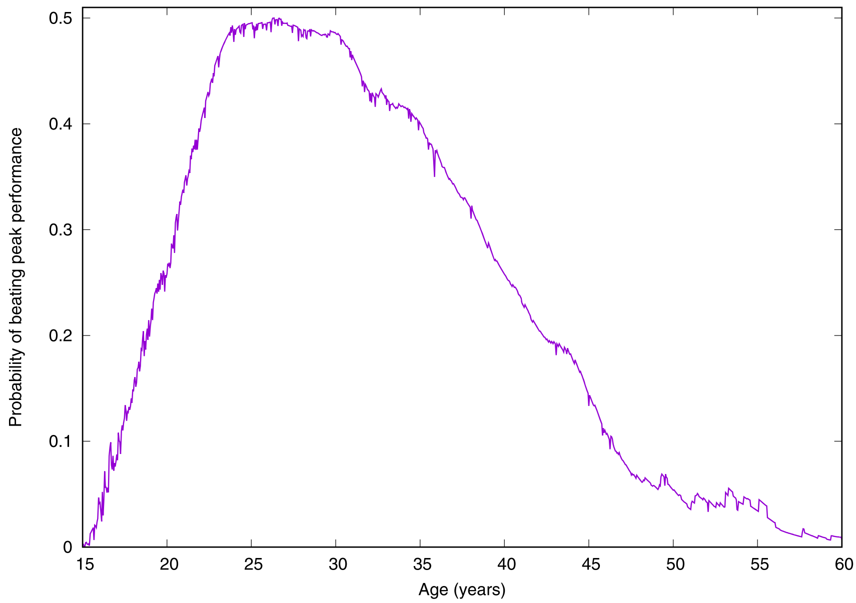



**Figure 3:** Nonparametric estimate of the variation of snooker performance with age.

evaluated at 1000 ages between 15 and 60, and binned. The average performance in each bin has finally been converted to the probability of beating peak performance, in a contest where the player has to be first to make 18 frames. The standard error on the lines is less than 0.02.

This shows how ability quickly rises from age 15 to a peak at 25–30, before declining with age. This is, without doubt, a younger peak age than one would expect, given findings in other sports. Sports requiring power and speed tend to be associated with younger peaks in performance (early to mid-20s). For example, Schulz and Curnow (1988) found peak performance in 100 m sprint of around 24 years. On the other hand, Baker and McHale (2023) found peak performance in golf of 35 years, and Roring and Charness (2007) found the peak age of performance in chess was 43.8 years. These are sports that rely less on speed and power and relatively more on skill and knowledge. That snooker breaks this trend is intriguing as it requires almost no physical prowess other than perhaps mental stamina (for long matches in the World Championships).

Figure 4 shows the same data as in Figure 3, but divided into players with lower and higher than average peak performance. It can be seen that more able players reach their peak performance older than weaker players, at age 30 instead of age 25. This finding replicates that found by Roring and Charness (2007) – the initially more able peak later.

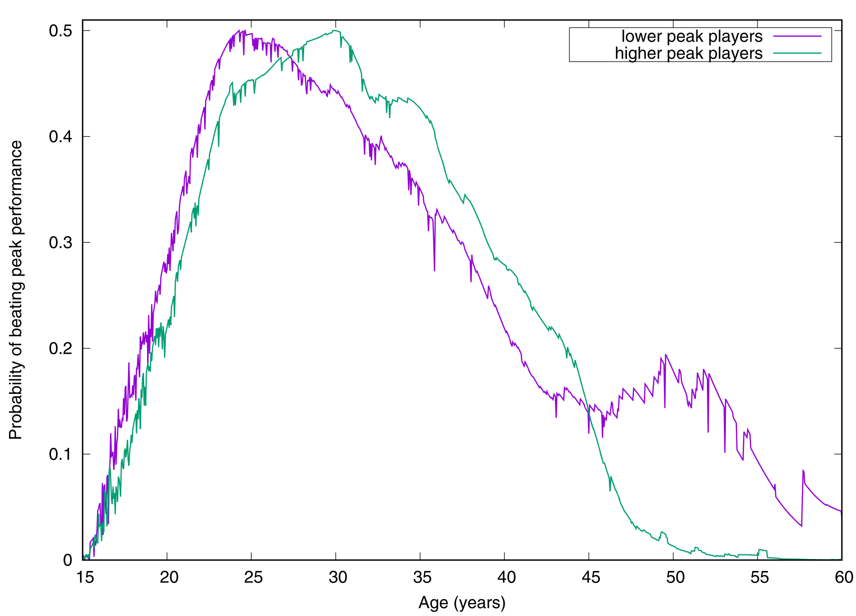



**Figure 4:** Nonparametric estimate of the variation of snooker performance with age, for players with lower and higher than average peak performance.

However, the two curves do not replicate the finding in Roring and Charness (2007) that suggests the performance of the initially more able decays more slowly. Here we see two almost identical rates of decay.

Other modelling choices could have been made; for example one could fit a random-effects model to the original data as did Baker and McHale (2023) for golf. Their model however did not allow for the effect of age on performance to vary with overall ability. Alternatively, one could allow the tabulated strengths, λ, to themselves be a function of age. However, a nonparametric approach, as we have used, requires fewer assumptions.

# 6 Forecasting performance

The descriptive method used for studying age-dependence cannot be used for forecasting the future performance of individual players. This is because it relies on knowing the peak of performance, which cannot be assumed known, especially for young players. However, the random-effects model from Baker and McHale (2023) can be adapted for use here.

We model $y_{it}$, the estimated strength (performance) of the $i$th player at the $t$th node. Before estimation, this was called $\lambda_{it}$ in (2), so $y_{it} = \hat{\lambda}_{it}$. This is modelled as the sum of a barycentric interpolation of performance to the $i$th player's age at the $t$th node, a random error for the player's performance, and a random residual error. A barycentric interpolation coefficient $\mu_j + \epsilon_{ij}$ appears at each of $m$ ages, $a_j$, that are the nodes of the interpolant, and this coefficient is a normally-distributed random variate. The addition of the $\epsilon_{ij}$ allows the performance of individual players to diverge from the mean age-trajectory. We take age $a$ ranging from a minimum of 12.5 to a maximum of 60 with $m = 20$ regularly spaced nodes at 2.5 year intervals; the model fit was fairly insensitive to the number of nodes and their placement. Suppressing the player suffix $i$, with the player's age at the $t$th match $a_t$, the model for the $t$th score $y_t$ is

$$y_t = \sum_{j=1}^{m} r_{tj}\{\mu_j + \epsilon_j\} + \eta + \epsilon_t, \tag{4}$$

where

$$r_{tj} = \frac{w_j/(a_t - a_j)}{\sum\limits_{j=1}^{m} w_j/(a_t - a_j)}. \tag{5}$$

The $w_j$ are the barycentric weights as before.

Writing

$$h_t = \sum_{j=1}^{m} r_{tj}\mu_j,$$

the deterministic part of the right-hand side, simply $y_t = h_t + t + \sum_{j=1}^{m}\epsilon_j + \eta + \epsilon_t$.

The errors are $\epsilon_t \sim N[0, \sigma^2]$, $\eta \sim N[0, \lambda^2]$, and $\epsilon_j \sim N[0, \phi^2]$ and all errors are independent. The parameters are $\boldsymbol{\mu}$, and $\sigma, \lambda, \phi$. If data are not available for a player after a young age, the expected random errors are very small for higher ages, outside the range of the data, as they are shrunk towards zero. Details of the likelihood function and model fitting are given in Appendix.

Likelihood maximisation was done using a function minimiser that does not require derivatives: this requires only computation of the log-likelihood and is quick; fitting takes less only a few seconds. Note that this computational scheme means that we are not restricted to a linear model, although we have used one here.

The vector $\boldsymbol{\delta}$ is defined in Appendix, and is the vector of realized random effects, the expected departure of a player's performance from the mean performance at an age node.

Forecasting was done by setting $\boldsymbol{\delta} = \mathbf{0}$ and extrapolating the curve using (4), adjusting it so that it was continuous with the interpolated curve at the final age node. Forecasts appear reasonable for older players: Figure 5 shows the curve for Ronnie O'Sullivan. Truncating the data and forecasting gives curves that do not reproduce the full peak for top players, but rather decay smoothly. Hence this forecast is most useful for older players.

## 6.1 Characterizing whether players have fulfilled their potential

The parameters $\boldsymbol{\delta}$ from the random-effects model are the posterior expectations of the random effects $\epsilon_j$. They give the performance of a player at a particular age, relative to their overall expected performance $\eta$. The correlation $\rho$ between $\boldsymbol{\delta}$ and age at the nodes characterizes how performance varies with time. A player who maintains good performance over a long timespan must have a random effect that increases or at least does not decrease with age, to compensate for the deterioration in performance caused by age. Conversely, a player whose performance deteriorates quickly after peaking will have a negative correlation. The correlation is independent of the strength of the player, and merely measures how performance changes with age. Thus a weak player could be held to have fulfilled their potential as much as a strong player, if their performance did not quickly fall off with age.

A second measure of declining performance is the skewness of the (squared) strengths. A positive skewness means that performance has declined substantially with

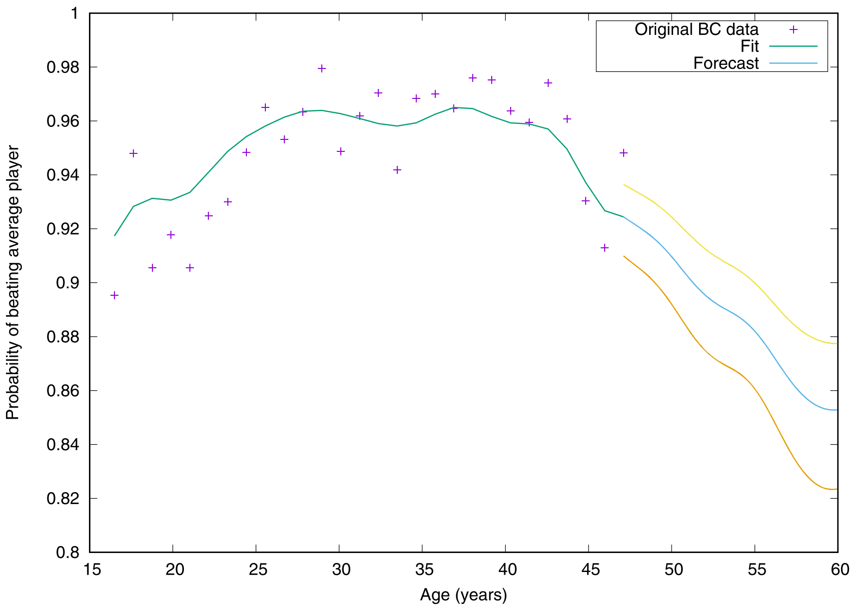



**Figure 5:** Forecast snooker performance with age for Ronnie O'Sullivan, showing the performance at the original barycentric (BC) nodes (purple plus signs), the model fit (green line), and forecast to age 60 (blue line) with standard errror (yellow and orange lines). Performance is the probability of beating an average player with winner to reach 5 frames.

age. This measure has the advantage that it does not require use of the random-effects model, being calculated directly from the fitted strengths.

Table 3 shows results for the top 20 players. Recall that most of these players have not yet retired, so the correlations and skewnesses could change. We see that players known for a long and successful career such as Ronnie O'Sullivan have large positive correlations and negative skewness, whereas some, such as Stephen Hendry, have large negative correlations and positive skewness. Although a successful player, in later years, Hendry's performance declined markedly from his once great levels.

**Table 3:** Inference on fulfilment of potential: correlations $\rho$ and skewness for the 20 top snooker players from 10-year strengths in Table 2.

| Rank | Player | Correlation $\rho$ | Skewness |
|---|---|---|---|
| 1 | Ronnie O'Sullivan | 0.559 | −0.128 |
| 2 | Stephen Hendry | −0.441 | 0.376 |
| 3 | Steve Davis | −0.100 | −0.462 |
| 4 | John Higgins | −0.034 | 0.089 |
| 5 | Judd Trump | 0.368 | −0.246 |
| 6 | Mark Selby | 0.495 | −0.192 |
| 7 | Neil Robertson | 0.613 | −0.195 |
| 8 | Mark Williams | 0.087 | 0.081 |
| 9 | Ding Junhui | 0.106 | 0.053 |
| 10 | Ken Doherty | −0.167 | 0.409 |
| 11 | Stephen Maguire | −0.095 | 0.104 |
| 12 | Shaun Murphy | −0.276 | −0.043 |
| 13 | Mark Allen | 0.452 | −0.138 |
| 14 | Matthew Stevens | −0.465 | 0.310 |
| 15 | Stuart Bingham | 0.489 | −0.225 |
| 16 | Stephen Lee | −0.142 | 0.097 |
| 17 | John Parrott | −0.483 | 0.533 |
| 18 | Peter Ebdon | −0.118 | 0.206 |
| 19 | Ali Carter | 0.486 | −0.024 |
| 20 | James Wattana | −0.500 | 0.477 |

## 7 Conclusions

The paper presents a model for the outcome of snooker matches that estimates the strengths of players, and allows for these strengths to vary deterministically with time. We use the model for two main purposes. First, we identify the greatest ever snooker players. Two candidates stand out: Ronnie O'Sullivan and Stephen Hendry. Intriguingly, there are no players present in the top 20 from before snooker's boom era of the 1980s. Second, we use the estimated strength curves in a random effects model to estimate the relationship between performance and age. We find a peak age of betweeen 25 and 30: earlier than might be expected when compared to similar sports. Further, we find that more able players peak later in their careers than less able players.

The methodology we use extends the model of Baker and McHale (2017) to enable forecasting of future performance and further consider the issue of identifying over-achieving players. Our method uses the correlation between the expected performance and age. We find that there are indeed players that over-achieve (Ronnie O'Sullivan performs above expectations relative to his age), and players that under-achieve (Stephen Hendry dropped in performance earlier than was expected). This type of analysis could be used to help coaches identify players that can be improved and moved back towards their expected performance.

There are additional findings revealed in the paper. For example, we find that weighting of matches from different tournaments, in order to identify behaviour whereby player's try harder in certain tournaments, is not needed.

Future work could be to study other similar sports, or performance outside of sport. The analysis relies only on a performance metric being available, or that a performance metric can be estimated using a model. Further, developing more families of suitable models besides the Stern (1990) model would also be useful for sensitivity analysis.

**Research ethics:** Not applicable.
**Author contributions:** The authors have accepted responsibility for the entire content of this manuscript and approved its submission.
**Competing interests:** The authors state no conflict of interest.
**Research funding:** None declared.
**Data availability:** The raw data can be obtained on request from the corresponding author.

# Appendix

## Details of the ageing model

This appendix gives details of model fitting for the individual player ageing model.

## The likelihood function

Using equation (4), the likelihood for a player is

$$\mathcal{L} = \frac{F}{(2\pi)^{(m+1)/2}} \int_{-\infty}^{\infty} \exp(-\psi/2)\, \mathrm{d}\epsilon\, \mathrm{d}\eta, \tag{6}$$

where

$$F = (2\pi\sigma^2)^{-n/2}\lambda^{-1}(\phi)^{-m}. \tag{7}$$

With $\Delta_t = y_t - h_t$,

$$\psi = \frac{\sum_{t=1}^{n}\left(\Delta_t - \sum_{j=1}^{m} r_{tj}\epsilon_j - \eta\right)^2}{\sigma^2} + \frac{\sum_{j=1}^{m}\epsilon_j^2}{\phi^2} + \frac{\eta^2}{\lambda^2}. \tag{8}$$

Writing $\mathbf{v} = (\epsilon_1 \dots \epsilon_m, \eta)^T$, we have

$$\psi = A - 2\mathbf{B}^T\mathbf{v} + \mathbf{v}^T\mathbf{M}\mathbf{v}, \tag{9}$$

where $\mathbf{M}$ is symmetric, $A = \sum_{t=1}^{n}\Delta_t^2/\sigma^2$ and

$$B_i = \begin{cases} \sum_{t=1}^{n} \Delta_t r_{ti}/\sigma^2 & \text{if } i \le m \\ \sum_{t=1}^{n} \Delta_t/\sigma^2 & \text{if } i = m+1 \end{cases} \tag{10}$$

Also,

$$M_{ij} = \begin{cases} \sum_{t=1}^{n} r_{ti}r_{tj}/\sigma^2 + \delta_{ij}/\phi^2 & \text{if } i,j \le m \\ \sum_{t=1}^{n} r_{ti}/\sigma^2 & \text{if } i \le m, j = m+1 \\ n/\sigma^2 + 1/\lambda^2 & \text{if } i = j = m+1 \end{cases} \tag{11}$$

## Model fitting

To fit the model to data by likelihood-based methods, the likelihood function must be integrated over the $m+1$ normal random variates $\epsilon, \eta$.

We can evaluate the integral, allowing for random effects at each node, by completing the square in the exponent. Then

$$\psi = A - 2\mathbf{B}^T\mathbf{v} + \mathbf{v}^T\mathbf{M}\mathbf{v} = C + (\mathbf{v} - \boldsymbol{\delta})^T\mathbf{M}(\mathbf{v} - \boldsymbol{\delta}), \tag{12}$$

from which we read off $\mathbf{M}\boldsymbol{\delta} = \mathbf{B}$, $C = A - \boldsymbol{\delta}^T\mathbf{B}$. The vector $\boldsymbol{\delta}$ is found by solving the $m+1$ linear equations $\mathbf{M}\boldsymbol{\delta} = \mathbf{B}$. The distribution of $\mathbf{v}$ given $y_1 \dots y_n$ is multivariate normal with mean $\boldsymbol{\delta}$ and covariance matrix $\mathbf{M}^{-1}$.

Maximum likelihood estimators are known to underestimate scale parameters, such as $\sigma$, $\phi$ and $\lambda$, but because of the large sample size, this bias will be negligible.

Note that one can think of this random-effects model in Bayesian terms; the normal distribution of the errors would be the prior pdf, and our likelihood would then become the posterior probability. This approach would then be empirical Bayes, based on maximum posterior probability.

The vector of realized random effects, $\boldsymbol{\delta}$, is found by solving the $m+1$ linear equations. The method of computation used here is efficient for this example, and differs from methods commonly used, such as iterative generalized least squares. This is often used to estimate the model parameters, using the EM algorithm, but the approach described here is more direct. Doing a Cholesky decomposition $\mathbf{M} = \mathbf{L}\mathbf{L}^T$, where $\mathbf{L}$ is lower-diagonal, the $m+1$ linear equations for $\boldsymbol{\delta}$ can be solved, and the determinant $|\mathbf{M}| = \prod_{i=1}^{m+1} L_{ii}^2$ calculated. There is no need to invert the matrix $\mathbf{M}$. Thus finally

$$\mathcal{L} = F\ \exp(-A/2 + \boldsymbol{\delta}^T\mathbf{B}/2)/|\mathbf{M}|^{1/2}. \tag{13}$$

# References


Baker, R. D. and Jackson, D. (2014). Statistical application of barycentric rational interpolants: an alternative to splines. *Comput. Stat.* 29: 1065−1081.

Baker, R. D. and McHale, I.G. (2014a). A dynamic paired comparisons model: who is the greatest tennis player? *Eur. J. Oper. Res.* 236: 677−684.

Baker, R. D. and McHale, I.G. (2014b). Time varying ratings in association football: the all-time greatest team is. *J. Roy. Stat. Soc. A* 178: 481−492.

Baker, R. D. and McHale, I.G. (2017). An empirical bayes model for time-varying paired comparisons ratings: who is the greatest women's tennis player? *Eur. J. Oper. Res.* 258: 328−333.

Baker, R. D. and McHale, I.G. (2023, in press). A flexible mixed model for age-dependent performance: application to golf. *J. Roy. Stat. Soc. Appl. Stat.* 72: 1260−1275.

Berrut, J. P. and Trefethen, L.N. (2004). Barycentric lagrange interpolation. *SIAM Rev.* 46: 501−517.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs I: the method of paired comparisons. *Biometrika* 39: 324−345.

Glickman, M. E. (1999). Estimation in large dynamic paired comparison experiments. *J. Roy. Stat. Soc. C Appl. Stat.* 48: 377−394.

Gunia, A. (2022). How Asia revived the dying sport of snooker. Available at: <https://time.com/6220526/snooker-hong-kong-masters/> (Accessed 27 November 2023).

Knorr-Held, L. (2000). Dynamic rating of sports teams. *J. Roy. Stat. Soc. D* 49: 261−276.

Lehman, H.C. (1945). Man's most creative years quality versus quantity of output. *Sci. Mon.* 61: 127−137.

Lehman, H. C. (1951). Chronological age vs. proficiency in physical skills. *Am. J. Psychol.* 64: 161−187.

McHale, I. G. and Morton, A. (2011). A bradley−terry type model for forecasting tennis match results. *Int. J. Forecast.* 27: 619−630.

Roring, R.W. and Charness, N. (2007). A multilevel model analysis of expertise in chess across the life span. *Psychol. Aging* 22: 291−299.

Schulz, R. and Curnow, C. (1988). Peak performance and age among superathletes: track and field, swimming, baseball, tennis and golf. *J. Gerontol.: Psychol. Sci.* 43: 113−120.

Stern, H. (1990). Models for distributions on permutations. *J. Am. Stat. Assoc.* 85: 558−564.

Stern, H. (1992). Are all linear paired comparison models empirically equivalent? *Math. Soc. Sci.* 23: 103−117.

Thurstone, L. L. (1927). The method of paired comparisons for social values. *J. Abnorm. Psychol.* 21: 384−400.