Probabilistic Modelling in Multi-Competitor Games

Christopher J. Farmer

Master of Science
Division of Informatics
University of Edinburgh
2003

Abstract

Previous studies in to predictive modelling of sports have largely concentrated on those which belong to the group of sports and games which can be termed as *one-on-one*: those where teams or players face each other in pairs e.g. football, basketball and tennis. Here the case of *multi-competitor* games or sports such golf, horse racing and motor sport, which are defined as ones in which three or more players or teams compete against each other at the same time are considered. Models, which given data in the form of previous performance of players, can make probabilistic predictions on the outcome of future tournaments are investigated. The generalised Bradley-Terry model for repeated multiple rankings of individuals known as the Plackett-Luce model is reviewed. A two-way anova model which uses scores rather than rankings is proposed as an alternative more suited for specific application to sports data. A dynamic extension and Kalman filter approach to fitting data to the two-way anova model is described.

The models are firstly developed and tested on artificially generated 'toy' data then are applied to real world data in the form of seven years worth of US PGA tour records. Findings from this study are presented in the form of a comparison of the static models with the Official Golf World Rankings and a case study of the dynamic analysis of two players. A quantitative evaluation of the predictive performance of the models is presented in the form of results from their application in a gambling scenario. The dynamic two-way anova model is found to be competitive in its predictions when compared to those of a human expert.

Table of Contents

1 Introduction **1**

2 Background and Theory **3**

2.1 Ranking systems in sport……………………………………………..3

2.2 Basics of PGA Tour golf tournaments…………………………………..4

2.3 Bradley-Terry Model……………………………………………………5

2.4 Plackett-Luce Model…………………………………………………….5

2.5 MM algorithm…………………………………………………………...6

2.6 Two-way anova model…………………………………………………..8

3 Models applied to toy data in static case **11**

3.1 Plackett-Luce model applied to toy data………………………………..12

3.2 Two-way anova model applied to toy data……………………………..16

3.3 Comparison of Plackett-Luce and two-way anova models……………..25

4 Dynamic model **29**

4.1 Generating dynamic data…………………………………………...…30

4.2 Estimation of α in dynamic case……………………………………..…30

4.3 Dynamic estimation of β…………………………………………….....33

4.4 Comparison of β estimation in dynamic and static models……………..34

4.5 Estimation of score standard deviation in dynamic case……………….…37

5 Application to real world data **40**

5.1 Data Issues……………………………………………………………40

5.2 Calculated α values for tournaments……………………………………42
5.3 Static model analysis of real world data………………………………...45
5.4 Comparison of static models with Official World Rankings…………....47
5.5 Dynamic two-way anova model analysis of real world data……………51
5.6 Evaluation of models by application to gambling scenario……………..55

6 Discussion and Conclusions **60**
6.1 Scores vs Rankings……………………………………………………...60
6.2 Application to other sports………………………………………………61
6.3 Ranking systems………………………………………………………...62
6.4 Future work and scope for improvements………………………………63
6.5 Conclusion………………………………………………………………66

Acknowledgements **67**

Bibliography **68**

A1 Key for Player and Tournament IDs **71**
A1.1 Player IDs……………………………………………………………..71
A1.2 Tournament IDs…………………………………………………….…73

1

1.0 Introduction

Previous studies in to predictive modelling of sports have largely concentrated on those which belong to the group of sports and games which can be termed as *one-on-one*: those where teams or players face each other in pairs. Examples of such include football, basketball, tennis and chess. This project considers the case of *multi-competitor* games or sports, which are defined as ones in which three or more players or teams compete against each other at the same time. Examples of such sports are horse racing, motor sports and many athletic events. The main focus of this project is golf, so although the theory still applies to other sports, terminology applicable to golf will be used. For example the golfing parlance of *players* who compete in *tournaments* could equally well apply to horses that compete in races.

The central objective of the project is to obtain a model, which given data in the form of previous performance of players can make probabilistic predictions on the outcome of future tournaments. There are several problems to overcome in doing this. Initially it requires the formulation of a system for rating the players, i.e. inferring their 'ability' quantitatively somehow from data available on past performance. The model must also be able to incorporate player 'form', i.e. changes in the ability of players over time. A further issue arises in the nature of sports such as golf due to the fact that not every player plays in every tournament, so the relative ability of players must be assessed even although they are not always performing in direct competition with each other.

In section 2 the background to the problem is discussed and the existing work on repeated rankings of groups of individuals – the Bradley-Terry model and its multi-competitor extension the Plackett-Luce model – is reviewed. A two-way anova model which uses scores rather than rankings is proposed as an alternative more suited to the specific application to golf data.

In section 3 artificially generated ‘toy’ data is used to investigate the performance of the Plackett-Luce and two-way anova models in the static case. Techniques for fitting the two-way anova model to data are detailed. Finally a comparison of the relative performance of the two models is presented.

Section 4 describes a dynamic extension to the two-way anova model and a Kalman filter approach for the fitting of data to it. This dynamic model is compared with the static two-way anova model in terms of their relative performance on toy data.

The ultimate intention when developing any model is to apply it in a real world scenario. With this in mind, a significant element of the objectives of this project are to test the developed models in such a scenario. In section 5 all the models discussed in the project are applied to real world data in the form of seven years worth of US PGA tour records. Data collection and pre-processing issues are discussed. Findings from this real world application are presented in the form of a comparison of the static models with the Official Golf World Rankings and a case study of the dynamic analysis of two players. As an attempt to quantitatively evaluate the predictive performance of the models, the results of a study applying them in a gambling scenario are presented.

In section 6 the results and findings from this project are discussed in general. The models studied are compared and contrasted and their potential for application to other sports assessed. Areas of improvement to the work carried out in this project are suggested and possible relevant future work outlined.

2

2.0 Background and Theory

In section 2.1 ranking systems used in sport are discussed in general. 2.2 gives a basic outline of the nature of PGA Tour golf tournaments, for which this project is geared towards modelling. Sections 2.3 and 2.4 cover the theoretical background to the Bradley-Terry model for repeated rankings of a group of individuals and its extension to the multi-competitor case, the Plackett-Luce model. In section 2.5 the MM algorithm method for fitting data to the Plackett-Luce model is outlined. In section 2.6 a different approach, based on the two-way anova model, which uses player scores rather than rankings is proposed.

2.1 Ranking systems in sport

Modelling of the performance of players or teams in sports essentially boils down to the creation of a ranking or rating system. Here a distinction must be made between two separate forms of ratings systems, distinguished by the contrasting motivations behind them. Smith[1] uses the terms 'static' and 'dynamic' to distinguish the two, however here the nomenclature of Sorensen[2] is adopted and the terms 'earned' and 'predictive' used to better describe the underlying motivations behind the systems.

Earned systems are used to determine who has been the best performer over a specified period of time. Examples of such systems include an end of season football league table or the 'Order of Merit' in sports such as golf and tennis, which are based on player's monetary earnings over the course of a season.

Predictive systems are used when the ratings are desired to be tuned for predictive power. The basic difference is that the model now requires a temporal aspect as more recent performances will be of greater importance in the context of player 'form'. This is in direct contrast to an earned model, where it is important to treat all tournaments to be of equal importance regardless of when they occurred within the specified period.

The creation of a 'fair' earned rating system is in many cases an interesting problem in its own right and has lead to much controversy and debate, particularly with reference to the ranking of NCAA college football teams: Wilson's bibliography[3] contains over 50 references to papers on the topic. The main focus of this project, however, is the predictive case, although static models can serve as a basis for comparison of the performance of dynamic models.

2.2 Basics of PGA Tour golf tournaments

The methods described in this project are applied to data from the US PGA golf tour over the period 1996 – 2003. This is the leading golf tour in terms of prize money and is, in general, where the world's top golfers play most of their competitive golf. The tour runs from January to November each year and for the most part consists of one tournament played each week. The format of the tournaments is usually one round of 18 holes played by each player each day from Thursday to Sunday, giving a total of 4 rounds (72 holes) played by each player. Each player attempts to take as few strokes as possible to complete each hole and the winner of the tournament is declared as the player who has the lowest aggregate total of strokes after 72 holes. In most tournaments, after the second round, there is a 'cut' whereby only a certain number of players are allowed to proceed to the final two rounds. This number is normally determined by either a predefined number of best scoring players the tournament organisers wish to retain, or by the number of players within a preset margin of the leader's score.

There are some exceptions to this general format, including tournaments played over a fewer or greater number of rounds, or with different policies for the cut, or no cut at all. In some weeks two tournaments are held simultaneously at different courses

The approach taken to dealing with factors such as the cut and these differences in tournament format in pre-processing the data is detailed in section 5.1.

2.3 Bradley-Terry Model

There is considerable existing work in the statistics literature concerning the repeated ranking of members of a group of individuals. The most fundamental model in this field is generally attributed to Bradley and Terry[4], although the same model had been referred to in work[5] pre-dating their 1952 paper. The Bradley-Terry model deals with the area of paired comparisons, where ranking takes place between members drawn from a group two at a time. In the model each member is assigned a real-valued positive attribute γ. Thus for a group of m individuals, with $\boldsymbol{\gamma} = (\gamma_1,\ldots, \gamma_i,\ldots, \gamma_m)^T$, where γ_i is associated with individual i, the probability of individual i being preferred to individual j is given by

$$p_{ij} = \frac{\gamma_i}{\gamma_i + \gamma_j}. \qquad (1)$$

To put this model in a sporting context, consider the individuals to be sports teams. Here γ_i can be thought of as representing the *ability* of team i and p_{ij} as the probability of team i beating team j in a match between the two. Clearly this model lends itself to the case of one-on-one sports. Indeed, it has been used, in a more generalised form to allow for ties and 'home court advantage', for the purposes of rating sports teams in the cases of basketball and soccer[6] and, in a yet more generalised form to incorporate 'score difference' in American football[7].

The Bradley-Terry model has been widely studied and has many generalisations and applications in a broad range of areas. In-depth explanations of which can be found in many sources[8,9].

2.4 Plackett-Luce Model

The particular generalisation of the Bradley-Terry model which is of specific interest in this project is that known as the Plackett-Luce model. The extension of the Bradley-Terry model to comparisons of three individuals is given by Pendergrass and Bradley[10]. The further generalisation of this to comparisons of any number of individuals is detailed by Marden[11]. The model draws on independent work by Plackett[12] and Luce[13], which gives rise to its name as termed by Marden, and by which it will be referred to here. Note that it is also referred to as the rank-ordered logit model in the economics literature, and as the exploded logit model in the marketing literature.

As in the Bradley-Terry model each individual in the group is assigned a real valued positive parameter, γ. For a group of m individuals, with $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_i,\ldots,\gamma_m)'^T$ the probability of a ranking of the individuals from best to worst, w, is given by

$$P[W = w \mid \gamma] = \prod_{i=1}^{m} \frac{\gamma_{w_i}}{\gamma_{w_i} + \gamma_{w_{i+1}} + \ldots + \gamma_{w_k}} \qquad (2)$$

where w_i is the i th player in the ranking.

This model is probably best understood in terms of a real life example. Considering a golf tournament, where the individuals are players, and γ_i represents the ability of player i, the probability of player i winning the tournament is given by

$$P(player_i\, wins) = \frac{\gamma_i}{\sum_m \gamma_m} \qquad (3)$$

Given that player i has won, i.e. been ranked first, the second placed player is then chosen as simply the winner of a new tournament between the remaining players, with player i removed. This process then continues until all k players have been ranked. Note that the normalisation of the vector $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_i,\ldots,\gamma_m)^T$ results in the value γ_i being equal to the probability of player i winning a tournament between all m players.

2.5 MM algorithm

In order to fit the Plackett-Luce model of eqn (2) using maximum likelihood the MM algorithm of Hunter[14] is used. The algorithm proceeds iteratively by first constructing a minorizing function of the log likelihood, then in turn maximising this function. Hunter shows that this algorithm is guaranteed to converge to the unique maximum likelihood estimator of $\boldsymbol{\gamma}$, provided the assumption that no player either wins or finishes last in all the tournaments they compete in holds. Note that there is no requirement for all players to compete in every tournament imposed by either the model or the MM algorithm, thus no special considerations need be made for this case, which will be common place in the data studied.

Consider the case where there is a total of t players. If the data consists of N rankings, where the j th ranking includes m_j players, let the order of players in the j th ranking be indexed by $a(j,1),\ldots a(j,m_j)$, i.e. from first to last in ranking order. Assuming i.i.d data, Hunter shows that the algorithm takes the form of the update equation: for t=1,…,m

$$\gamma_t^{(k+1)} = \frac{w_t}{\sum_{j=1}^{N} \sum_{i=1}^{m_j-1} \delta_{jit} \sum_{s=i}^{m_j-1} \gamma_{a(j,s)}^{k}} \qquad (4)$$

where w_t is the number of rankings in which the t th player is ranked higher than last and δ_{jit} is the indicator of the event that player t receives a rank no better than i in the j th ranking.

2.6 Two-way anova model

The Plackett-Luce model considers only the rankings of players in tournament results. There is a possibility that this could result in a loss of information. A model based on player scores rather than just rankings should result in more information being retained. With this in mind, a different model which has its foundation in the two-way anova model is proposed here. This standard statistical model for the two-way analysis of variance[15] has the form

$$y_{ij} = \mu + \beta_i + \alpha_j + \varepsilon_{ij} \qquad (5)$$

where the outcome of an event, y_{ij}, is dependent upon a mean effect, μ, two independent effects, β_i and α_j, and a noise term, ε_{ij}. Both the sum of β over all i and the sum of α over all j are equal to zero.

For the purposes of applying this model to the sport of golf, consider the case where there is a total of t players, each of whom has an associated parameter, β, who compete in N tournaments, each of which has a parameter, α. Thus, in vector form: $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_i,\ldots,\beta_t)^T$ and $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_j,\ldots,\alpha_N)^T$. The mean term, μ, of eqn (5) is incorporated in to the β term to form

$$y_{ij} = \beta_i + \alpha_j + \varepsilon_{ij}\,. \qquad (6)$$

The outcome y_{ij} now represents the score of the i th player in the j th tournament. Whereas the Plackett-Luce model dealt with rankings of players, the two-way anova model uses the score – a real observable number – which in the case of golf is the number of strokes a player takes in a tournament, with the winner being the player who takes the fewest strokes. Using scores instead of rankings has the potential advantage of providing more information as to the relative performance of players. For example, a player may win a tournament by 10 strokes from the second placed player, while the third placed player is a further one stroke behind the second placed player. By just using rankings, the difference between the first placed player and the second placed player will be equivalent to the difference between the second placed player and the third placed player, namely one place. By using the scores of the players, however, further information is retained about the performance of each player.

β can be thought of as a measure of player 'ability', similar in this respect to γ in the Plackett-Luce model, although the exact meaning of β and the values it takes are very different to those of γ. Note that by incorporating the mean term in to the β term, the constraint that the sum of β over all i is zero no longer applies, although the equivalent constraint, that the sum of α over all j equals zero, remains. The second parameter, α, termed the 'course effect' is a measure of the 'difficulty' of the tournament it is associated with. A difficult tournament is considered to be one played on a particularly hard course, or perhaps one affected by unfavourable weather conditions. The net result of a difficult tournament is that all participating player's scores are affected and will be greater as a result (remembering that the aim of golf is to get as low a score as possible). Reciprocally an easy tournament will result in all players getting lower scores. Note that the assumption that α affects all players equally for each tournament is made here, which may not be the case in reality, however effects of this nature are difficult to quantify and thus this simplifying assumption is necessary. α can be viewed as effectively a 'noise' parameter, shifting the mean of the scores up or down for each tournament. Thus in order to obtain estimates of β the α values must be obtained and 'stripped out' of the results. A procedure for this is described in section 3.2, where it is also seen that this effectively solves one of the key problems in modelling multi-competitor sports; that of the non-participation of some players in some tournaments as was discussed previously in section 1.

The model can be used to generate scores by sampling from the normal distribution where the mean is formed by the sum of β and α, and the standard deviation given by a parameter σ associated with each player:

$$y_{ij} \sim N\left(\beta_i + \alpha_j, \sigma_i^{\,2}\right) \tag{7}$$

Using the previous definition of α, β is defined as the expected score of a player in an averagely difficult tournament, i.e. one for which $\alpha = 0$, leaving only β as the mean of the score.

This model can be easily extended to the dynamic case by modelling the evolution of abilities over time as random Gaussian walks:

$$\beta_i^{k+1} \sim N\left(\beta_i^k, \sigma_{evol}^2\right) \tag{8}$$

This allows the abilities of players to vary over time, a crucial aspect of a predictive model. The estimation of $\boldsymbol{\beta}$ in this dynamic case is a task that clearly lends itself to time series methods. In section 4.3 a Kalman filter approach to this is detailed.

3

3.0 Models applied to toy data in static case

In this section the performance of the models described in section 2 is investigated in the static case. For the purposes of this artificially generated ‘toy’ data is used. The advantages of this approach as opposed to using ‘real world’ data are that it allows much greater control over possible complexities in the data. No considerations have to be made for unquantifiable influences or noise in the data and particular effects such as increased numbers of players or tournaments can be examined in isolation. This approach of starting at the most simplistic case and gradually building in more levels of complexity allows potential hurdles to be dealt with one at a time rather than attempting to solve all the problems at once, where interacting effects may obscure any beneficial adjustments.

Section 3.1 covers how data can be generated from the Plackett-Luce model and investigates the performance of the model in recovering initial parameters from this generated data. Section 3.2 describes how data is generated using the two-way anova model and the techniques used to analyse this data, firstly in the simplified case where all players play in all tournaments. The additional complexities of non-participation of some players in tournaments and non-equal score standard deviations are then incorporated. In section 3.3 a means of comparison of the Plackett-Luce and two-way anova models is described and the results of such a comparison presented.

3.1 Plackett-Luce model applied to toy data

Toy data was generated from the Plackett-Luce model using the algorithm:

```
Initialise: Number of Players, Number of Tournaments, Average Proportion of
players to play in each tournament

γ* = normalise (numbers randomly picked from the uniform distribution [0,1])

for Number of Tournaments

        initialise group of competing players
        initialise place = 1

        while size of group of completing players > 0

                calculate probability of winning for each player
                select winner of tournament between all remaining players
                assign place to winning player
                remove winning player from group of competing players
                place = place + 1
        end
end
```

As the models are ultimately to be applied to PGA tour golf data, the values of parameters such as the number of players and the number of tournaments are where appropriate chosen to reflect this. Considering the available real world data, a reasonably 'realistic' scenario is having 440 players playing in 325 tournaments, where the probability of any given player playing in any given tournament is 0.4, i.e. on average there are 176 players competing in each tournament.

Figure 1 shows results from using the MM algorithm[16] to estimate γ in such a scenario.

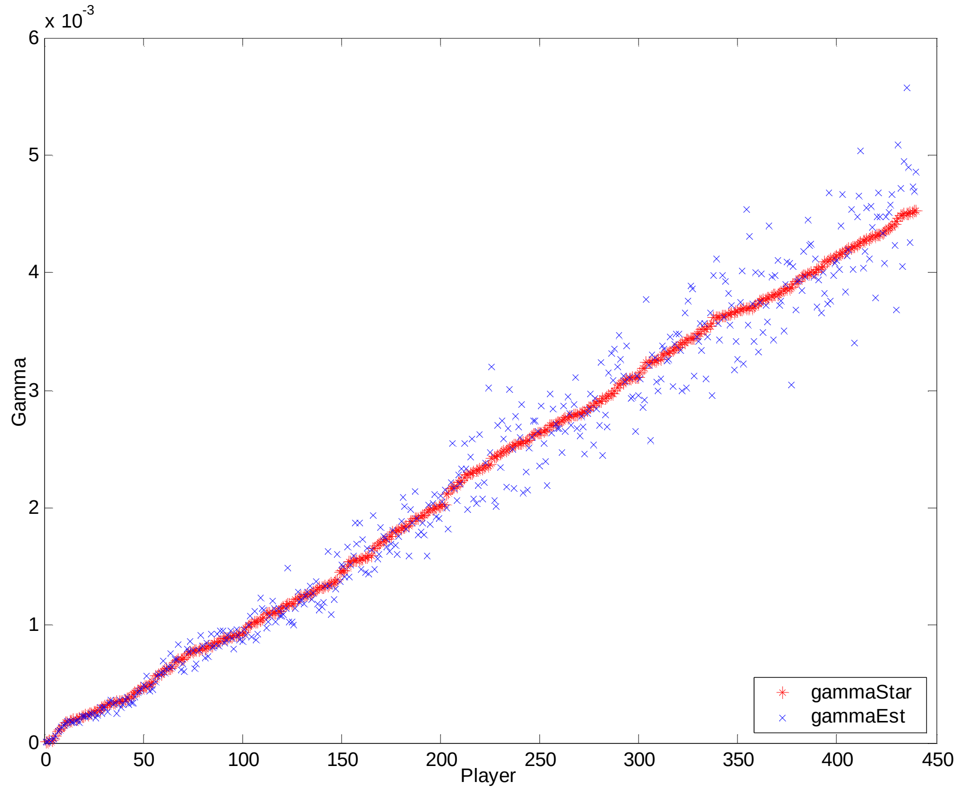

Fig 1. Plot showing the results from the Plackett-Luce model in a realistic scenario. The red points are the true γ values, $\boldsymbol{\gamma^*}$**,** which have been sorted by increasing value of $\boldsymbol{\gamma^*}$. The blue points represent the corresponding estimate of γ, $\boldsymbol{\gamma^{est}}$, for each player obtained using the MM algorithm.

Due to the nature of the Plackett-Luce model, one tournament of 100 players can be thought of as 100 separate 'quasi-tournaments' with the number of players decreasing by one each time a 'winner' is selected and removed. It is thus conceivable to think that perhaps more information is retained about players who are involved in more of these quasi-tournaments. When considering the case of more than one tournament, it is evident that players with lower values of γ will not be selected as quickly and thus will on average compete in more quasi-tournaments. Thus it is feasible that the estimates of the lower values of γ could be more accurate relative to the higher values. From the plot there is indeed a clear increase in the absolute error of the estimates for increasing values of γ (i.e. the blue points aren't as close to the red points in the top right of the plot compared to the bottom left). However when considered in terms of relative error, given by $| \gamma^*_i - \gamma^{est}_i | / \gamma^*_i$, this is found to be on average uniform across all the players. So in terms of relative error there is no such bias towards relatively better estimations of lower γ players.

It is obvious that if the available data is increased – by increasing the number of tournaments – the estimates of γ obtained from the MM algorithm will be better. This effect can be quantified by experiment by playing a large number of tournaments and measuring the average difference between γ^* and γ^{est} for different numbers of tournaments to produce 'learning curves'. Such experiments were carried out and the results are presented in fig 2.

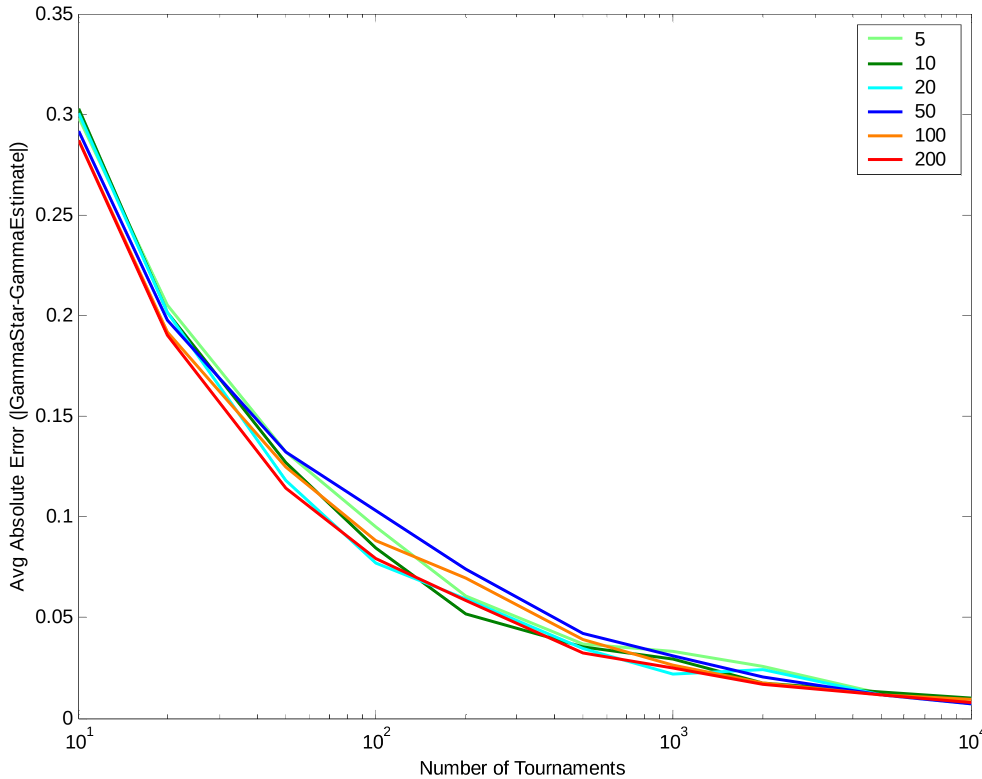

Fig 2. Learning curves for the estimation of γ using the MM algorithm. These show that the summed absolute error between the normalised γ* and γ^{est} asymptotes to zero as the number of tournaments increases. The different colours represent the situations with different numbers of players playing in each tournament. The errors are averages based on many trials: a total of 10000 tournaments were generated for each of the six different player number cases. The MM algorithm was then used on sequential blocks of these results, thus the error values for 10 tournaments are averages over 1000 trials and the values for 100 tournaments are averages over 100 trials etc. There are no values for less than 10 tournaments because of one of the conditions of the Plackett-Luce model: that no player can finish last, or tied last, in all tournaments they compete in. Thus the algorithm often does not converge properly with very small sample sizes where this condition may be violated.

From fig 2 it is apparent that the error in the estimate of $\boldsymbol{\gamma}$ is more or less the same in all of the cases of differing numbers of players for any given number of tournaments. This can be explained by considering once again the basics of the Plackett-Luce model: that one tournament of x players is essentially the same as playing x different tournaments, where the 'winners' of previous tournaments do not compete in the subsequent tournaments. So although the numbers of players are increasing, the information contained in each tournament ranking increases concordantly thus it can be concluded that the error in gamma estimation is dependent on the number of tournament rankings available and not the number of players.

3.2 Two-way anova model applied to toy data

In section 2.6 A model based on the two-way anova model was described. The performance of this model is examined here. Firstly, the case in which all players play in all tournaments.

Data is generated by sampling from eqn (7) in section 2.6 to produce a matrix, $\boldsymbol{S}$, of the scores where the (i,j) th element is the score of player i in tournament j. The term σ_i in eqn (7), which is termed here as the 'score standard deviation', is for the time being assumed to be equal for all players.

The goal is, given the score matrix $\boldsymbol{S}$, to infer the values of β for each player, which in the static case boils down to the task of 'stripping out' the influence of α on the results. This provides a problem as these α terms are not directly measurable. Although they have been previously termed as a 'course effect' this is not a full description as in fact α will not be a constant for one particular course. α should be viewed as a measure of how difficult it was to play in a particular tournament, which is dependent on more than just the course that the tournament is played on: weather conditions for example can have an effect. The 'difficulty' of a tournament must be determined using only the available information: the scores of the players who compete in it. In the simplistic case where all players play in all tournaments this is trivial: the estimate of $\boldsymbol{\alpha}$ is obtained by calculating the mean score for each tournament, then subtracting the mean of these mean scores. Mathematically, define the vector $\boldsymbol{s}^{\boldsymbol{avg}}$ to be the column-wise mean of the matrix of scores $\boldsymbol{S}$, i.e. the j th element, $s^{avg}{}_j$, is the mean of column j of $\boldsymbol{S}$: the mean score in the tournament j. The estimate of $\boldsymbol{\alpha}$ is then given by

$$\boldsymbol{\alpha}^{\boldsymbol{est}} = \boldsymbol{s}^{\boldsymbol{avg}} - \overline{s^{avg}} \, . \qquad (9)$$

This estimate is then used to form the adjusted scores matrix, $\boldsymbol{A}$, where

$$\boldsymbol{A} = \boldsymbol{S} - \boldsymbol{\alpha}^{est\,T} \, . \qquad (10)$$

Note that the adjustment can be either to increase or decrease the scores dependent upon the sign of the estimate of α. Estimates of α for 'easy' tournaments will be negative, and those for 'difficult' tournaments positive, thus subtracting these from the scores has the effect of increasing scores for easy tournaments and decreasing scores for difficult tournaments. This makes sense intuitively by considering the example of two players, player x and player y of equal ability (equal values of β). Player x performs very well to win a difficult tournament, tournament i, scoring say 280. In a different, easier tournament, tournament j, which is won with a score of 265, player y plays poorly and scores 279. A direct comparison of the two scores, without taking α in to account would suggest that player y had performed better than player x. However, by calculating α, which in this case would be greater for tournament i than tournament j, and adjusting the scores accordingly, the adjusted scores, which form the matrix $\boldsymbol{A}$, can now be considered as scores where there is no course or weather effect present between different tournaments, i.e. the score of player x in tournament i can be directly compared with the score of player y in tournament j. It is then clear that $\boldsymbol{\beta}^{\boldsymbol{est}}$, is given simply by the means of the rows of the matrix $\boldsymbol{A}$: i.e. the i th element of $\boldsymbol{\beta}^{\boldsymbol{est}}$ is the mean of the i th row of $\boldsymbol{A}$.

Figs 3 and 4 show results using this approach on data generated using 440 players who all play in 325 tournaments. The values of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ were generated by sampling from normal distributions with means of 288 and 0 respectively. The score standard deviation was set at 10 for all players.

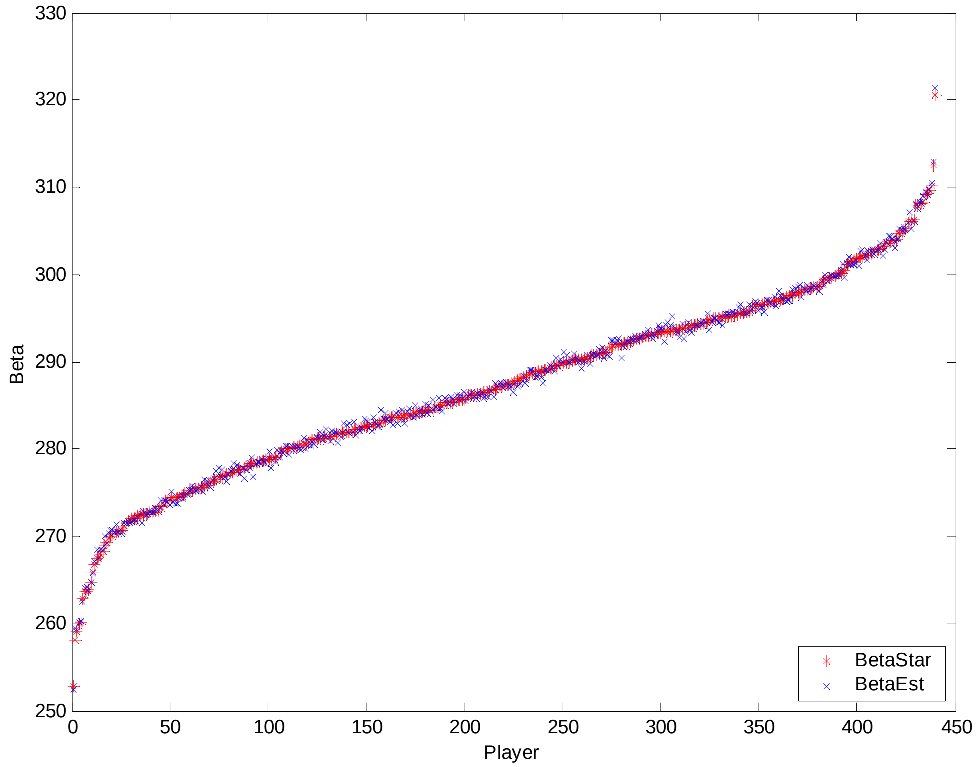

Fig 3. β estimation in the two-way anova model. The estimate of $\boldsymbol{\beta}$ obtained is shown here in blue, with the corresponding true values, $\boldsymbol{\beta^*}$, plotted in red. Note that the players have been sorted here in to ascending $\boldsymbol{\beta^*}$.

The shape of the curve generated in Fig 3 is due to the values of $\boldsymbol{\beta^*}$ being drawn from a normal distribution. A normal distribution was chosen as this produces the situation whereby most players are clumped around the average value, with a few really good players, and a few really bad players, which is probably a more realistic scenario than that produced by drawing from a uniform distribution. This is also the case for the $\boldsymbol{\alpha}$, where again this was done to produce a more realistic situation: it can be reasonably assumed that most tournaments will be of similar difficulty, while there may be a few particularly easy ones and a few particularly hard ones. The results from the estimation of which are shown in fig 4.

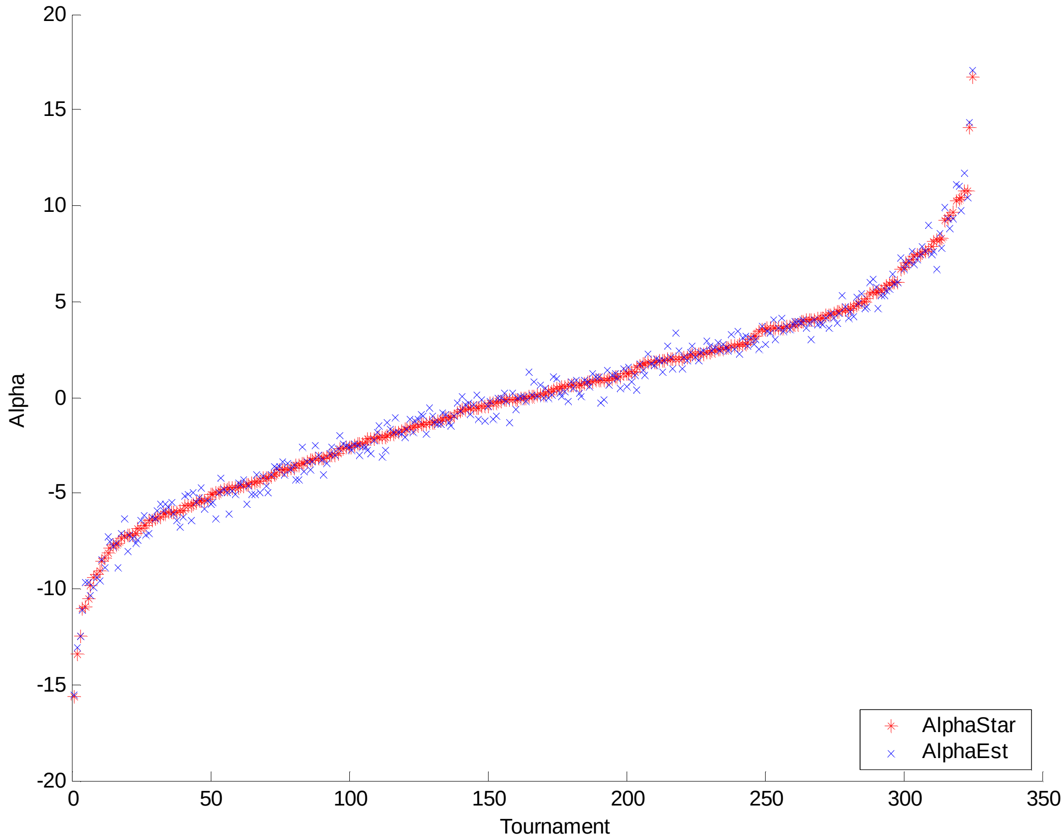

Fig 4. α estimation in the two-way anova model for the same run as in Fig 3. The estimate of $\boldsymbol{\alpha}$ is shown here in blue, with the corresponding true values, $\boldsymbol{\alpha^*}$, plotted in red. Note that here the tournaments have been sorted in to ascending $\boldsymbol{\alpha^*}$.

Adding in an additional level of complexity - that not all players play in all tournaments - requires a new approach to the task of estimating $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. As stated earlier the only information available for the purposes of estimating $\boldsymbol{\alpha}$ is the scores obtained by players playing in the tournament. When all players do not play in all tournaments the approach detailed earlier of taking the mean score in a tournament to get a measure of its difficulty will not work. Consider two tournaments, *i* and *j*. Tournament *i* is very prestigious tournament with lots of prize money on offer and hence attracts a field including all the best players. Tournament j, however, is a relatively small tournament with low prize money and hence many of the top players don't bother playing in it. A player, player x, who is a good player plays in the big tournament, i, and wins it scoring say 280, with the mean score in the tournament being 290. A different player, player y, who is of poorer ability than x plays in the lesser tournament, j, and also wins it scoring 280 with the mean score in the tournament also being 290. The model needs to be able to take in to account that the performance of player x in tournament i was better than that of player y in tournament j. In order to do this the role of $\boldsymbol{\alpha}$ needs to be increased beyond just representing a course/weather conditions effect and also account for differences in field strength between tournaments. A measure of the field strength can be obtained from the β values of the participating players, however $\boldsymbol{\beta}$ can not be properly estimated without knowing, or having a good estimate of, $\boldsymbol{\alpha}$. In this sense $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are reciprocally defining. In order to break this circularity a two-step iterative algorithm can be used. The first step calculates $\boldsymbol{\beta}$ based on the current estimate of $\boldsymbol{\alpha}$ and the second step calculates $\boldsymbol{\alpha}$ based on the new estimate of $\boldsymbol{\beta}$ obtained from the first step. These steps are repeated until convergence. The algorithm in detail is:

For a matrix of Scores, $\boldsymbol{S}$, where the the (i,j) th element is the score of player i in tournament j and the non-participation of a player in a tournament is indicated by a score of zero.

Initialise
Number of players, $numP$ = number of rows in $\boldsymbol{S}$
Number of tournaments, $numT$ = number of columns in $\boldsymbol{S}$
$\boldsymbol{\alpha}^{est}$ = vector of zeros, with number of elements equal to $numT$
Adjusted scores, $\boldsymbol{A} = \boldsymbol{S}$

Until convergence of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$
  Step 1: Calculate $\boldsymbol{\beta}^{est}$ based on current $\boldsymbol{\alpha}^{est}$
  For $i = 1{:}numP$
  $\boldsymbol{\beta}_i$ = mean of the non zero elements of the i th row of $\boldsymbol{A}$
  End
  Step 2: Calculate $\boldsymbol{\alpha}^{est}$ based on current $\boldsymbol{\beta}^{est}$
  Form matrix $\boldsymbol{B}$: the scores adjusted by $\boldsymbol{\beta}^{est}$
  For I =1:$numP$
    For j =$numT$
      $B(i,j) = S(i,j) - \boldsymbol{\beta}^{est}{}_i$
    End
  End
  Elements that were zero, representing non-participation, may now not be zero, so to correct this: For all elements of $\boldsymbol{S}$ that are zero, make corresponding element of $\boldsymbol{B}$ zero.
  Form vector, $\boldsymbol{m}$, the mean β-adjusted score for each tournament
  For j =1:$numT$
    $m(j)$ = mean of the non zero elements of the j th column of B
  end
  Constrain elements of $\boldsymbol{m}$, to form the zero mean vector, $\boldsymbol{\alpha}^{upd}$
  $\boldsymbol{\alpha}^{upd} = \boldsymbol{m}$ – mean of $\boldsymbol{m}$
  Update estimate of $\boldsymbol{\alpha}$
  $\boldsymbol{\alpha}^{est} = \boldsymbol{\alpha}^{est} + \boldsymbol{\alpha}^{upd}$

  Adjust Scores by subtracting $\boldsymbol{\alpha}^{upd}$
  For I =1:$numP$
    For j =$numT$
      $\boldsymbol{A}(i,j) = \boldsymbol{A}(i,j) - \boldsymbol{\alpha}^{upd}(j)$
    End
  End
  For all elements of $\boldsymbol{S}$ that are zero, make corresponding element of $\boldsymbol{A}$ zero.
End

Results using the two step algorithm for the case when not all players play in all tournaments are shown in Figs 5 and 6. These are obtained from data generated again for the 'realistic' scenario of 440 players who play in 325 tournaments where the probability of each player playing in any given tournament is 0.4, i.e. on average there are 176 players competing in each tournament. The score standard deviation was set at 10 for all players. Fig 5 shows β estimation, while fig 6 shows α estimation.

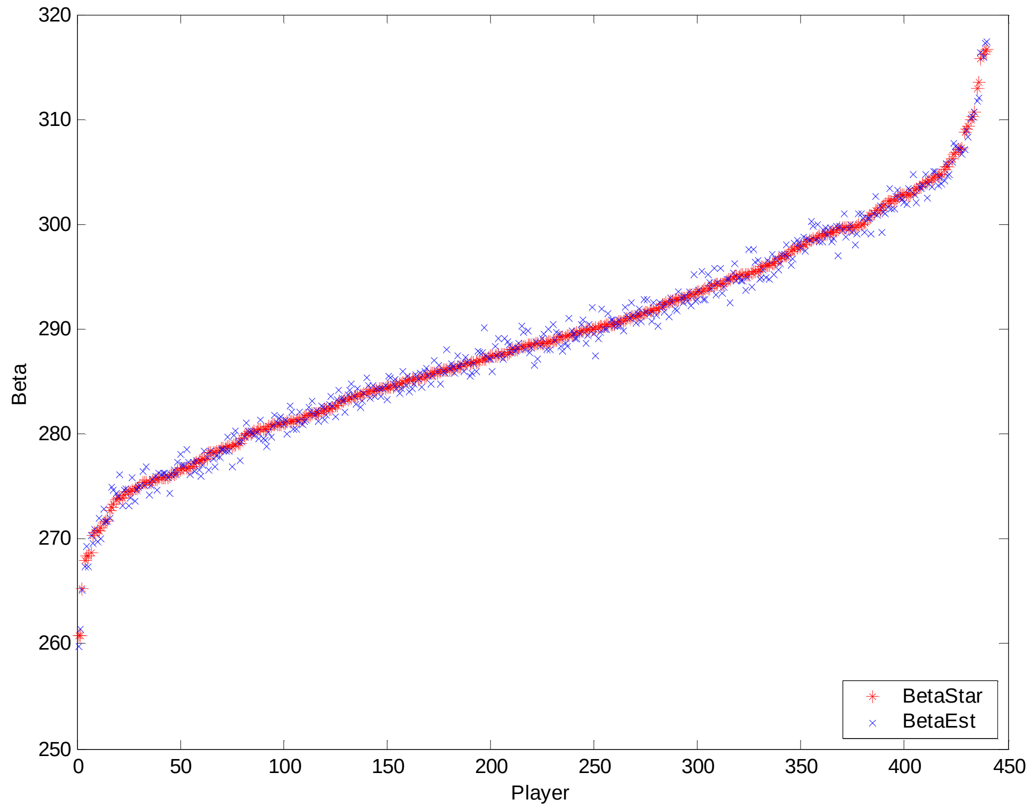

Fig 5. β estimation in the two-way anova model The estimate of $\boldsymbol{\beta}$ obtained by performing the two-step iterative algorithm described above is shown here in blue, with the corresponding true values, $\boldsymbol{\beta}^*$, plotted in red. Note that the players have been sorted here in to ascending $\boldsymbol{\beta}^*$.

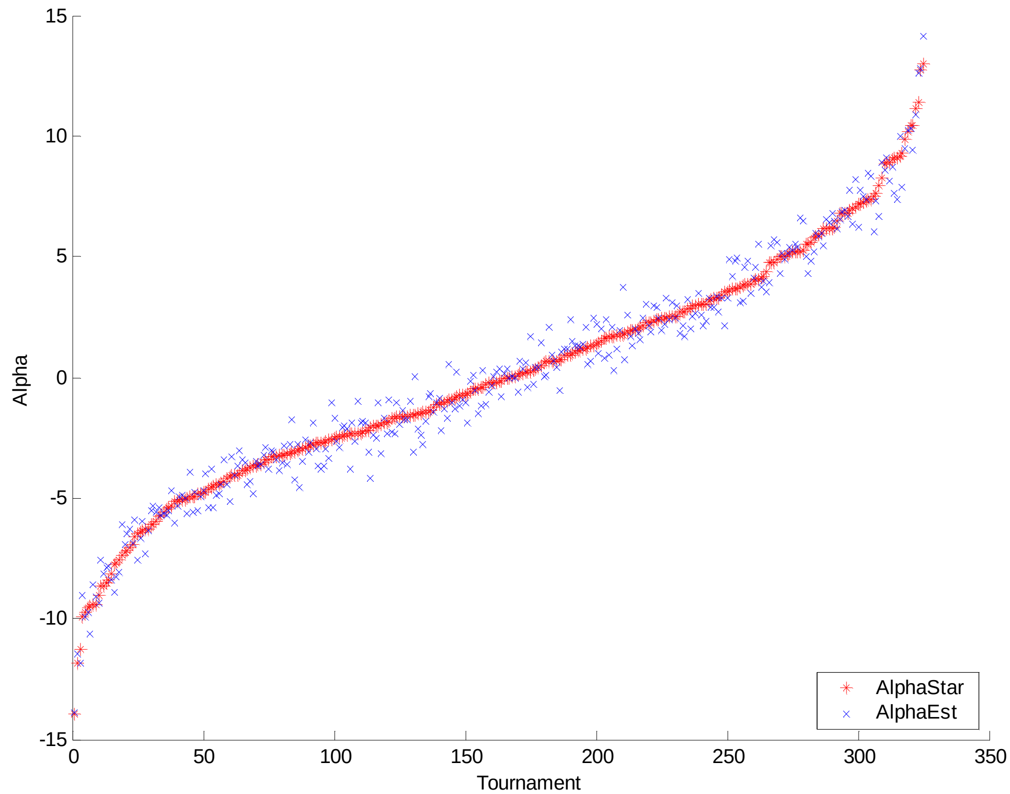

Fig 6. α estimation in the two-way anova model when not all players play in all tournaments. The estimate of $\boldsymbol{\alpha}$ is shown here in blue, with the corresponding true values, $\boldsymbol{\alpha^*}$, plotted in red. Note that here the tournaments have been sorted in to ascending $\boldsymbol{\alpha^*}$.

A further level of complexity can be incorporated by allowing players to have different score standard deviations. This can be thought of in a real-world terms as some players being more 'erratic' than others. While a 'steady' player is likely to regularly perform to close to their ability, an erratic player is more likely to perform well above or below their ability in any given tournament. This is represented in terms of the model by steady players and erratic players having lower and higher score standard deviations respectively.

Estimating the score standard deviation of a player is a relatively simple task in the static β case. The score standard deviation estimate for player i is given by calculating the standard deviation of the non-zero terms of the i th row of the adjusted scores matrix $\boldsymbol{A}$. Fig 7 shows results of the estimation of the score standard deviation of players, again using the 'realistic' number of players and tournaments with the score standard deviations of players randomly assigned from a uniform distribution with upper and lower limits of 15 and 5 respectively.

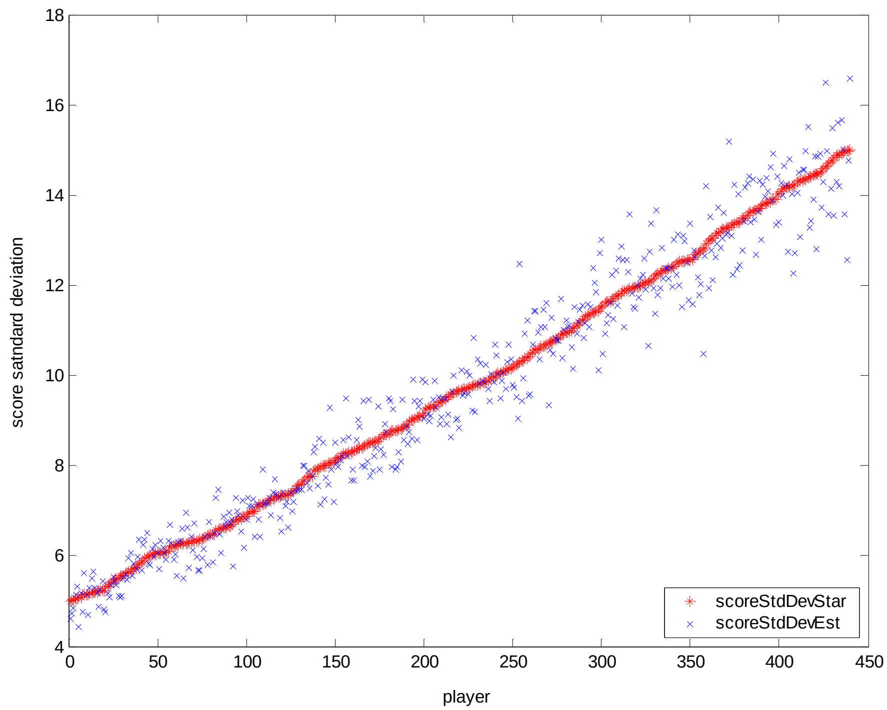

Fig 7. Score standard deviation estimation in the two-way anova model. The estimate of score standard deviation is shown here in blue, with the corresponding true values plotted in red. Note that here the players have been sorted in to ascending order of true score standard deviation.

3.3 Comparison of Plackett-Luce and two-way anova models

The task in fitting the Plackett-Luce model to data is the estimation of $\boldsymbol{\gamma}$, while for the two-way anova model it is $\boldsymbol{\beta}$, thus there is no direct way to compare the relative performance of the models. In order to do this both models must be used to estimate the same parameter. There is no way of obtaining an estimate of $\boldsymbol{\beta}$ using the Plackett-Luce model, however by using Monte Carlo sampling of the two-way anova model estimates of $\boldsymbol{\gamma}$, as defined by the Plackett-Luce model, can be produced.

The procedure used is to generate scores for a given number of tournaments using the two-way anova model. These scores are then easily converted to rankings by giving the player with the best score in a tournament the ranking of 1, the second best the ranking 2 and so on. These rankings are then used in the MM algorithm to find the Plackett-Luce model estimate of $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}^{PL}$. The original scores are then used to obtain an estimate of $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{est}$, in the usual way as previously described. This estimate of $\boldsymbol{\beta}$ is then used to generate a large number of sample tournaments, from which an estimate of $\boldsymbol{\gamma}$ can be obtained. Recalling that $\boldsymbol{\gamma}$, when normalized, can be defined as the vector whose elements are equal to the probability of the corresponding player winning a tournament between all the players, from the large number of generated tournaments an estimate of this probability is obtained by simply dividing the number of tournaments a player wins by the total number of tournaments. Thus an estimate of $\boldsymbol{\gamma}$ from the two-way anova model, $\boldsymbol{\gamma}^{TWA}$, is obtained.

Finally the 'true' value of $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}^*$, is required. An estimate of this is obtained by using the same method as for $\boldsymbol{\gamma}^{TWA}$, but using the true value of $\boldsymbol{\beta}$, $\boldsymbol{\beta}^*$, which was used to generate the initial set of tournament scores.

There is the possibility that if too few tournaments are simulated in the Monte Carlo sampling that the value of $\boldsymbol{\gamma}^{TWA}$ obtained could be affected. The estimates of $\boldsymbol{\gamma}$ were found empirically to be stable after 10^8 sample tournaments.

Fig 8 shows the results using this method of comparison from repeated trials of the 'realistic' scenario as used previously.

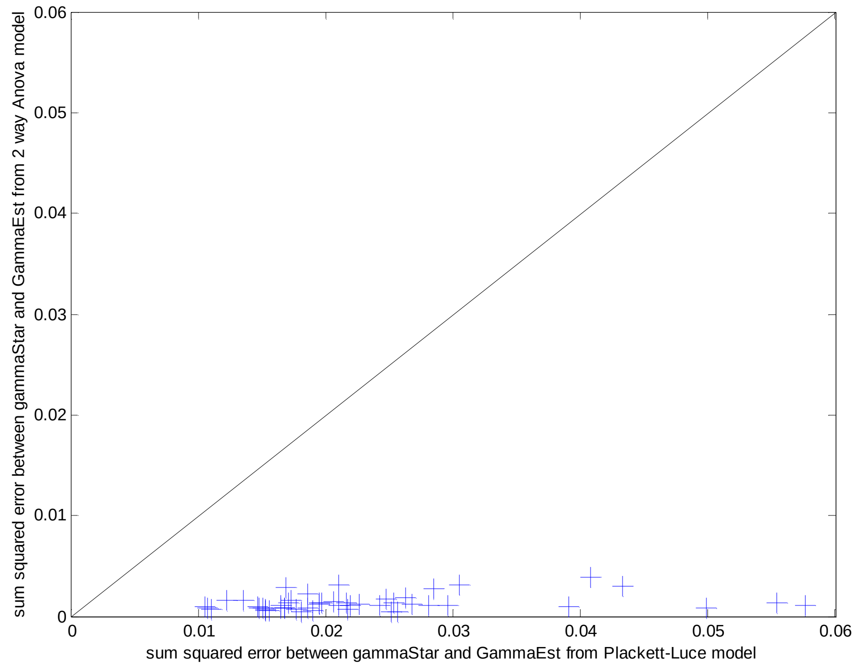

Fig 8. Comparison of accuracy of γ estimation in Plackett-Luce and two-way anova models. Each blue cross, of which there are 50, represents a different trial whereby γ^*, γ^{PL} and γ^{TWA} were obtained using the procedure outlined previously. If the quality of the estimates of the two models were similar then the crosses should be dispersed equally on either side of the central diagonal line. A bias to either side indicates that there is a difference between the two. From the plot, all the crosses are towards the lower right. This strong bias shows that the error in estimation of γ from the Plackett-Luce model is very likely to be greater, and thus the two-way anova model is producing better quality estimations.

From Fig 8 it is clear that the two-way anova model is producing better results. This perhaps isn't really surprising, as there is information lost in the conversion of the scores to the rankings which the Plackett-Luce model uses, thus the γ estimate obtained from the two-way anova model where this information is retained should be superior. As was discussed in section 2 the form of this 'information' is the further distinguishing of player performances. The distance between the 1st and 2nd player could be 10 shots, or could only be 1 shot but in terms of rankings these situations are not distinguished. It follows that when there is limited data available the Plackett-Luce model could be expected to 'squash together' the γ values: i.e. give the best players too low values of γ, and the worst players too high values. To see if this is the case, fig 9 shows the results for one of the 50 trials from fig 8.

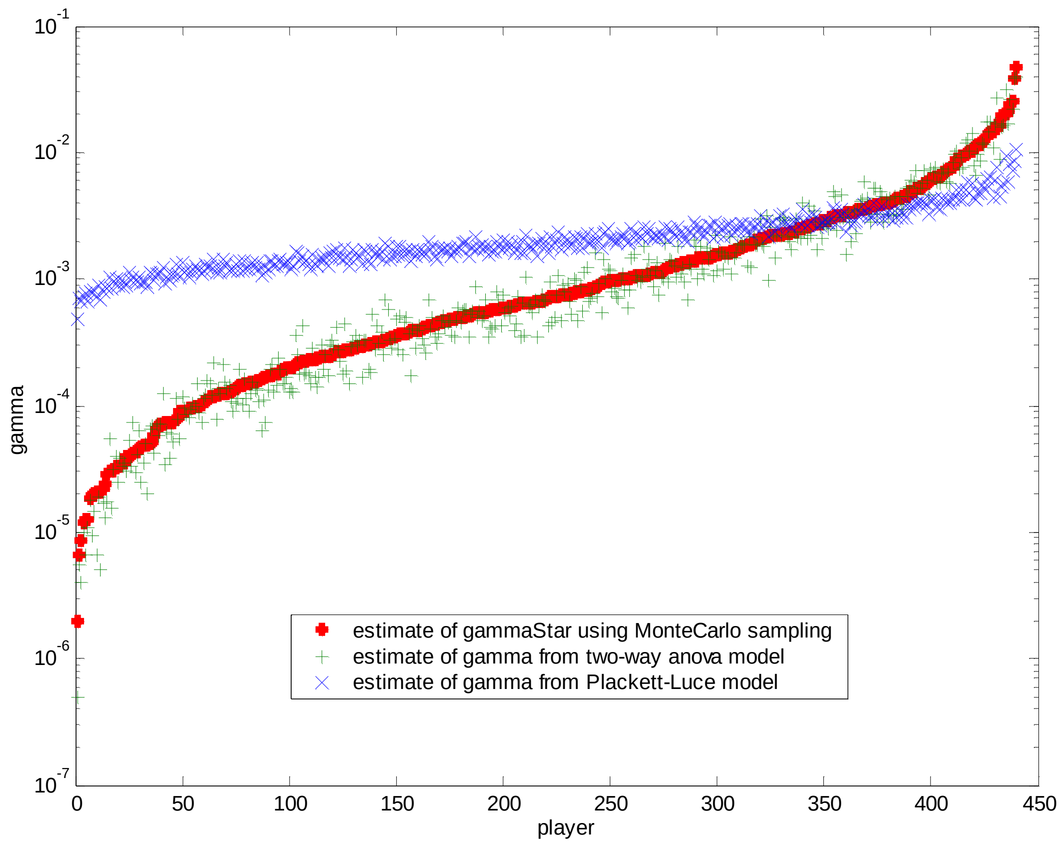

Fig 9. Plot of γ^* (red), γ^{PL} (blue) and γ^{TWA} (green) for a typical trial. The players have been sorted by increasing γ^*. The gamma axis is shown on a log scale for visibility.

From the plot in fig 9 the expected 'squashing' effect predicted for the Plackett-Luce model can be seen. For the players with low gamma (towards the left) the blue Plackett-

Luce prediction is too high and for the higher gamma players - towards the right - it is too low.

4

4.0 Dynamic model

In any real world situation, it is likely that individual players can improve or decline in ability over time. The models discussed in section 3 make no allowance for this: the players are assumed to have constant ability and the results from a tournament played several years ago are deemed to be as important as the results from one played last week. It seems probable that a dynamical model, which is able to differentiate between tournaments, dependent on how long ago they were played, and essentially ‘weight’ them for importance should be able to provide a more accurate estimate of the current ability of a player.

Section 4.1 describes how the two-way anova model is extended to generate dynamic data. In 4.2 the estimation of $\boldsymbol{\alpha}$ given such dynamic data is discussed. Section 4.3 details a Kalman filter approach to the dynamic estimation of $\boldsymbol{\beta}$, the performance of which is compared with static estimation in section 4.4. In section 4.5 a Kalman smoother method for score standard deviation estimation is described.

4.1 Generating dynamic data

The model for dynamic $\boldsymbol{\beta}$ is defined by eqn (8) in section 2.6. In order to generate dynamic toy data the same procedure as for the static case is used, but now the $\boldsymbol{\beta}$ vector is updated following every tournament so that each element takes a random Gaussian walk around its previous value.

Each tournament can be considered to represent one time step, so after initialisation $\boldsymbol{\beta}$'s evolution over time is modelled by, after each tournament, performing the update

For i = 1:number of players

$$\beta_i^{t+1} = \beta_i^{t} + N\left(0, \sigma_{evol}^{2}\right)$$

End.

4.2 Estimation of $\boldsymbol{\alpha}$ in dynamic case

As in the static case, the $\boldsymbol{\alpha}$ values for each tournament must be obtained before scores from different tournaments can be directly compared. Given that players β values can go up as well as down and that there are sufficient players competing in each tournament, the variances in $\boldsymbol{\beta}$ could be assumed to cancel out for the purposes of estimating $\boldsymbol{\alpha}$ and the same procedure as for the static case used. This assumption was tested experimentally by performing repeated trials on data generated using the same 'realistic' number of players and tournaments as used in section 3. $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and score standard deviation vectors were generated as usual. Two separate sets of tournament result data were then generated: one with static $\boldsymbol{\beta}$, and one with dynamic $\boldsymbol{\beta}$ in which the $\boldsymbol{\beta}$ vector was updated after each tournament with a σ_{evol}^{2} value of 1. The two-step algorithm as detailed in section 3, which assumes static data, was then used to estimate $\boldsymbol{\alpha}$ from both data sets. The results from one trial are shown in Fig 10

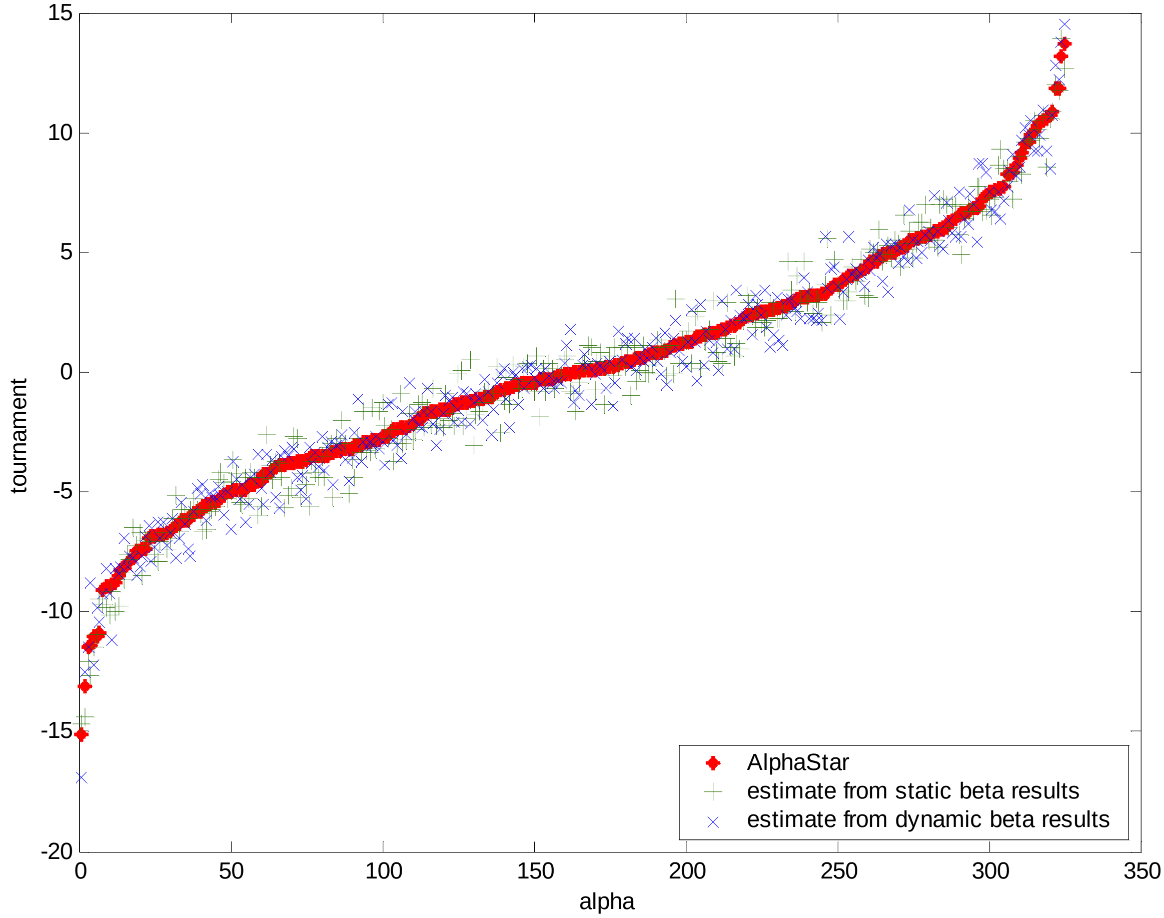

Fig 10. Plot showing α estimation from both static and dynamic data for one trial. The true value of $\boldsymbol{\alpha}$ is plotted in red, while the $\boldsymbol{\alpha}$ estimates from the static and dynamic data are plotted in green and blue respectively.

From fig 10 there is no obvious difference in the accuracy of estimate obtained by the two step algorithm when using either static or dynamic data. A total of 50 such trials were performed. In each case the summed absolute error between the true $\boldsymbol{\alpha}$ and those obtained from the static and dynamic data was calculated. These are plotted in fig 11.

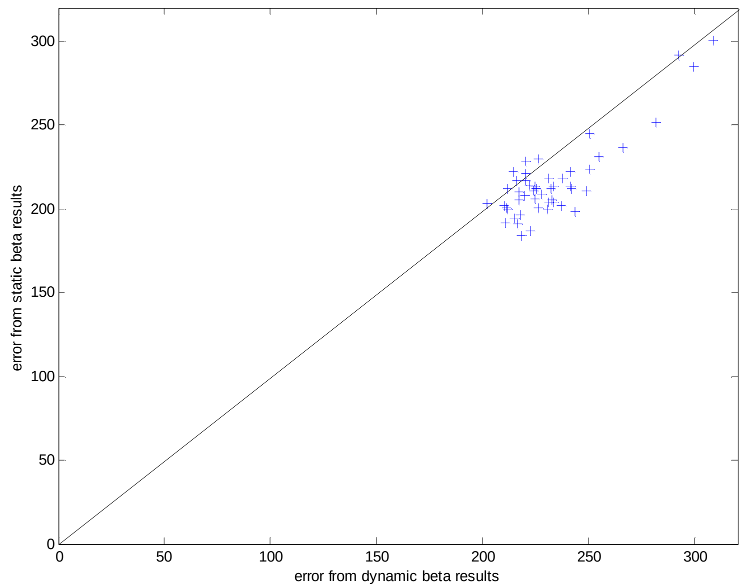

Fig. 11. Comparison of the accuracy of estimates of $\boldsymbol{\alpha}$ obtained from the static and dynamic $\boldsymbol{\beta}$ cases. Each point in the plot represents one of the 50 trials.

From this plot it is clear that there is a slight bias towards the static data producing a more accurate estimate of $\boldsymbol{\alpha}$. This is not unexpected, as the two-step algorithm assumes static data, thus when the data is not in fact static, there is likely to be a reduction in the estimation accuracy. Provided this reduction is not too significant, this may be a sacrifice that is acceptable, as being to able to estimate $\boldsymbol{\alpha}$ in this way will considerably reduce the task of estimating $\boldsymbol{\beta}$ dynamically. In order to quantify the effect on the accuracy, consider that for the 50 trials the mean summed absolute error in the static $\boldsymbol{\beta}$ case is 215.56±40.36, while for the dynamic $\boldsymbol{\beta}$ case it is 257.70±43.18, where the errors quoted are one standard deviation. Thus the mean difference in the summed absolute error in $\boldsymbol{\alpha}$ for the two cases over the 50 trials is 42.14. As these values are summed over all the tournaments for each trial, of which there are 325, the mean difference in error per tournament is 0.1297. The mean absolute value of $\boldsymbol{\alpha}$ over the 50 trials was 3.77, thus the average percentage reduction in accuracy from having dynamic data compared to static data is just 3.43%.

This is a reasonable trade to make for the benefits which estimating $\boldsymbol{\alpha}$ in this way will provide, thus in the analysis of dynamic data in this project, $\boldsymbol{\alpha}$ is always estimated using the two step algorithm assuming static data.

4.3 Dynamic estimation of β

The task of estimating $\boldsymbol{\beta}$ dynamically is performed using a Kalman filter approach. The Kalman filter is a set of mathematical equations that provide an efficient recursive solution of the least-squares method. It was described first by Kalman[17] in 1960 and since has been developed and applied in a broad range of contexts. Thorough reviews of the theory and application are available from many sources[18,19] .

Given tournament scores which have already been adjusted to account for $\boldsymbol{\alpha}$, using the two-step algorithm, the task of dynamic estimation of β reduces to a 1-dimensional Kalman filter for each player. Thus the implementation is essentially standard[20], with a slight modification to account for the possibility of missing observations – i.e. where players don't play in a tournament. If there is a missing observation at time t, the estimate of β is just the estimate at time t - 1, while the uncertainty in the value is increased from that at time t - 1 to account for the potential of β changing during the missing observation. The parameters used in the Kalman filter are obtained by EM estimation[20,21,22]. This procedure uses the dataset to obtain maximum likelihood estimates of the system and observation variance and the initial state vector and variance used in the filter. In this 1-d Kalman filter the system matrix and the observation matrix are both identity.

4.4 Comparison of β estimation in dynamic and static models

To evaluate the relative performance of the static and dynamic models their predictions of $\boldsymbol{\beta}$ are considered. Given a dataset of tournament scores, generated with dynamic $\boldsymbol{\beta}$, where each tournament is considered to be one time step, estimates of $\boldsymbol{\beta}$ from both the static and dynamic models were obtained for each time step. In the case of the dynamic model these values were obtained using the Kalman filter approach described in section 4.3. For the static model, the estimate of $\boldsymbol{\beta}$ was obtained from the two step iterative algorithm detailed in section 3.2. 50 trials were carried out with data generated using a 'realistic' number of players and tournaments. The results for one player from one trial are shown in fig 12.

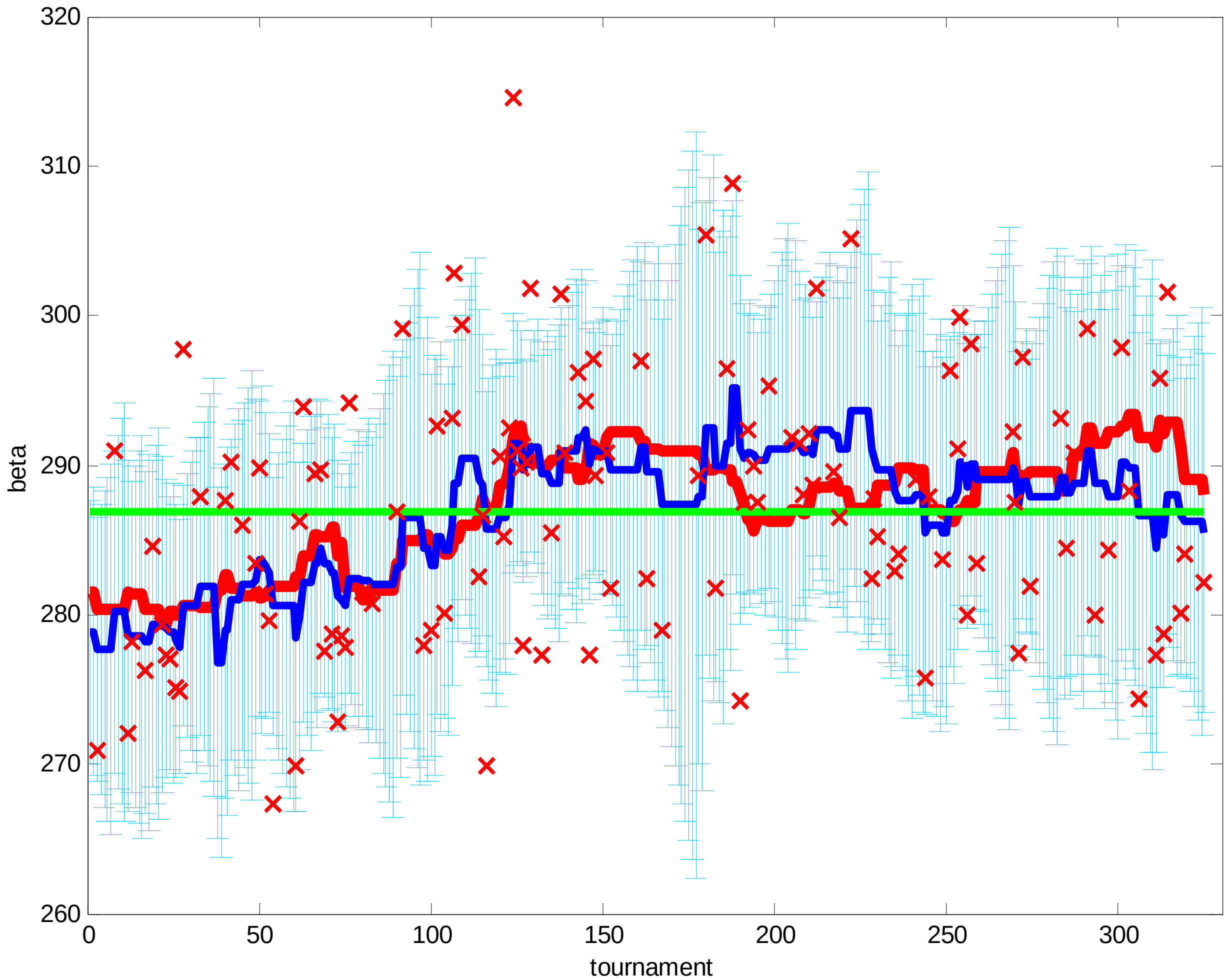

Fig 12. Plot showing $\boldsymbol{\beta}$ estimation for one player. The red crosses are the observed scores (adjusted to account for $\boldsymbol{\alpha}$). The red line is the true value of β for the player, which is seen evolving over time. The green line is the static estimate of β, which is essentially just the mean of the observed scores. The blue line is the dynamic estimate of β, while the light blue one standard deviation error bars represent the uncertainty in β from the Kalman filter.

To compare the two models the cumulative squared error in $\boldsymbol{\beta}$ is calculated for each. With reference to fig 12 this is the squared distance between the red line and the green line for the static case, and the red line and the blue line for the dynamic case, calculated at each time point – i.e. each tournament on the x-axis, and these values summed. This was done for all the players to give a total accumulative error over all players over all time. The results of repeated trials of this are presented in Fig 13.

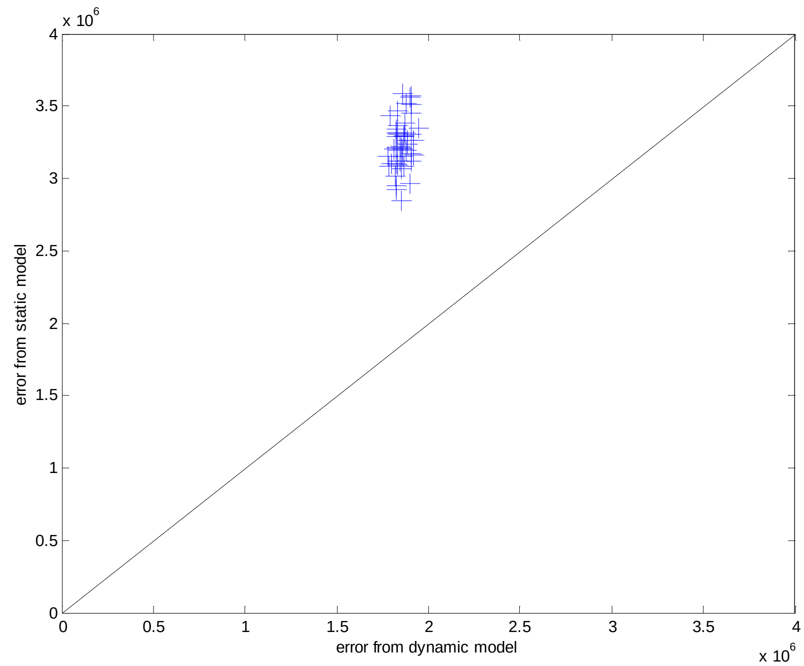

Fig 13. Comparison of β estimation in static and dynamic case. Each cross, of which there are 50, represents the accumulative error from one trial.

From fig 13 it is clear that the dynamic Kalman filter approach produces better estimates of β than the static model.

4.5 Estimation of score standard deviation in dynamic case

For the purposes of score standard deviation estimation in the dynamic case, a Kalman smoother[18, 23, 24] approach is used. As the β value of a player is changing, simple estimation of score standard deviation assuming static data will logically produce values that are too high as the evolution variance of β, σ_{evol}^2 will essentially act as an additional source of variance. A better estimate of score standard deviation can be obtained by estimating $\boldsymbol{\beta}$ dynamically and adjusting the scores to account for this. Thus a dynamic estimate of score standard deviation for a player is found by first obtaining the vector of β evolving over time for that player, $\boldsymbol{\beta}^{evol}$, using a Kalman smoother. This is then subtracted from the non-zero elements of the corresponding row of the $\boldsymbol{\alpha}$-adjusted scores matrix, $\boldsymbol{A}$, to form the vector $\boldsymbol{s}$. The estimate of score standard deviation is given by calculating the standard deviation of the elements of $\boldsymbol{s}$.

To confirm that, in the dynamic β case, the Kalman smoother approach to score standard deviation estimation gives better results than a static approach both methods were tested on dynamically generated data using a ‘realistic’ number of players. The players were assigned score standard deviations from the uniform distribution with upper and lower limits of 13 and 7 respectively. The results from one such trial are shown in fig 14.

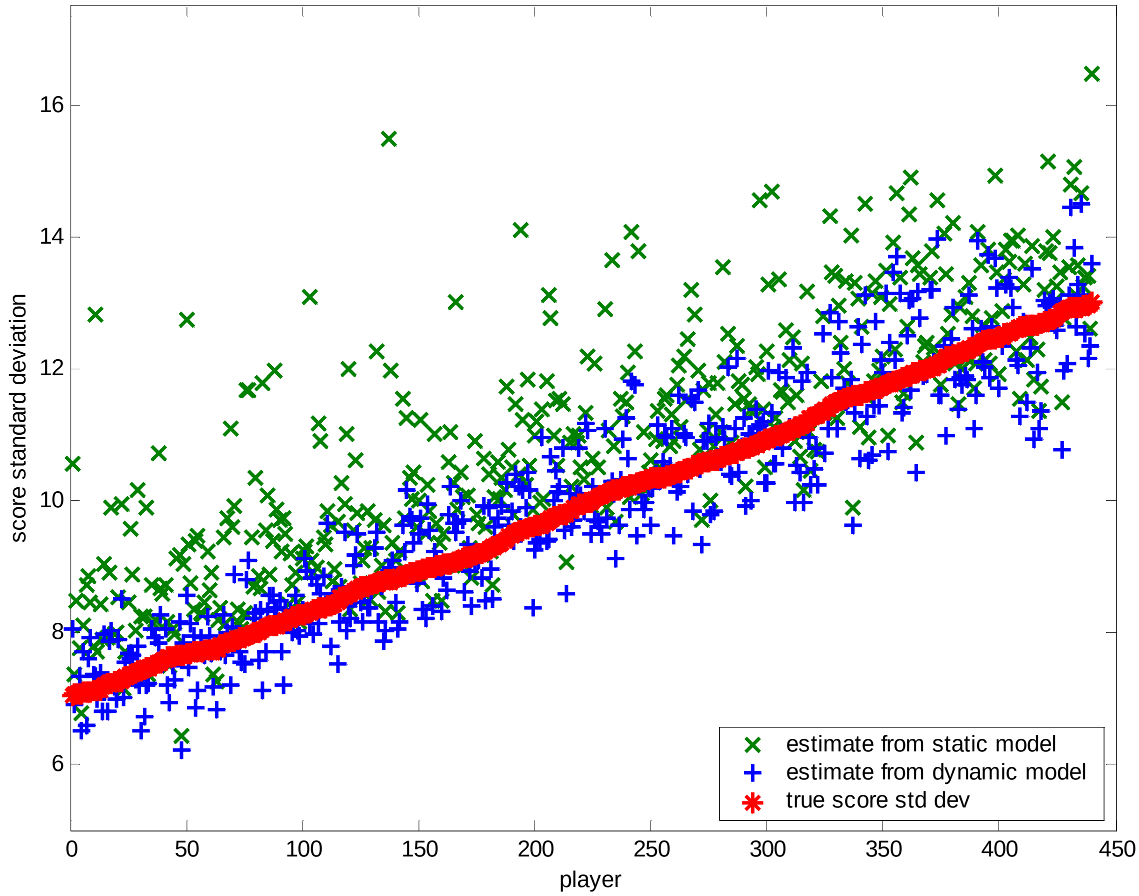

Fig 14. Estimate of score standard deviation for one trial. The true score standard deviation of each player is shown in red, while the estimate from the dynamic Kalman smoother approach is shown in blue and the static estimate shown in green. Note that the players have been sorted in to ascending true score standard deviation

Fig 14 shows that, as expected, the static method generally gives estimates which are too high, while the Kalman smoother approach produces better estimates which are centered around the true value as the additional source of variance caused by the dynamic β values has been corrected for. In total 50 trials like the one displayed in fig 14 were carried out. In each case the squared error between the true score standard deviation and the two estimates was calculated and this summed over all the players to get an accumulative error. The mean accumulative error when using the Kalman smoother method was $1.8561 \times 10^6 \pm 3.8443 \times 10^4$. For the static method it was $3.2382 \times 10^6 \pm 1.6900 \times 10^5$. The dynamic Kalman smoother approach clearly produces the better estimates, and thus is the method used for score standard deviation estimation in the analysis of dynamic data in this project.

5

5.0 Application to real world data

In previous sections the models investigated in this project have been applied to artificially generated ‘toy’ data. The ultimate intention when developing any model is to apply it in a real world scenario. In this section the findings from application of the models to 7 years worth of US PGA tour records are presented. Section 5.1 discusses how the data was collected and preprocessed. Section 5.2 covers the findings from obtaining α values for the tournaments. Section 5.3 contains an analysis using static models, which is compared with Official World Rankings in section 5.4. In section 5.5 the results from the dynamic analysis of real world data is presented in the form of a case study of two players. The models were applied to a gambling scenario for evaluation purposes and the findings from this are presented in section 5.6.

5.1 Data Issues

This study uses data from the period 7th Jan 1996 – 20th July 2003 inclusive. In this time there were 332 US PGA Tour tournaments played, in which a total of 1911 players competed. Many of these players only played in a limited number of tournaments and never made any impact in terms of successful performances. As these players are not of any particular significance to the study, only players who had played in at least 20 tournaments in this period were retained, leaving 438 players. As discussed in section 2, many golf tournaments include a ‘cut’ after the second round, whereby a number of the lowest placed players are removed - often around half of the starting field - and don’t continue on to play the final two rounds. In this case of players who miss the cut the final score recorded is their two-round total. This clearly can not be used as the player’s score for the purposes of this study as this would result in players who missed the cut –

by playing poorly – having much better scores than all the players who made the cut. Also, the player's participation in the tournament can not simply be discounted, as doing this would fail to recognise poor performances of players. One way to obtain a final score for these players would be to simply double their two-round total, which would give an estimate of what they would have scored over four rounds. This however does not take in to account any differences in the playing conditions between the first two rounds of a tournament and the last two. If the weather significantly improves or deteriorates it can have a significant effect on the average score. Taking this in to account, to obtain a final score for players who missed the cut a projection is made using:

$$\begin{array}{c}\text{Final score for}\\ \text{player missing}\\ \text{cut}\end{array} = \begin{array}{c}\text{Highest (worst) score}\\ \text{over four rounds}\\ \text{obtained by the other}\\ \text{players who made cut}\end{array} - \begin{array}{c}\text{Score required to}\\ \text{make cut after}\\ \text{two rounds}\end{array} + \begin{array}{c}\text{The player's}\\ \text{score after two}\\ \text{rounds}\end{array} \qquad (11)$$

The golf season generally runs from early January to early November, thus there is an 'off-season' period of roughly two months each year. As time series analysis is to be performed, these periods require to be accounted for in the data. The data is stored in a matrix, whereby each row corresponds to one player, and each column to one tournament. The *(i,j)* th element of the matrix is the score of player *i* in tournament *j*, and a score of zero indicates non-participation in the tournament. There is generally one tournament played per week during the season, so each column of the matrix can be considered to represent a time period of one week. Thus to represent the off-season, 8 or 9 (dependent on the length of the off-season that year) tournaments in which nobody participates were inserted in the relevant places in the data.

The key which matches the indices of the rows and columns to the corresponding real world players and tournaments is included in appendix 1

5.2 Calculated α values for tournaments

Fig 15. shows the α values obtained for the 325 tournaments in the study using the two-step algorithm.

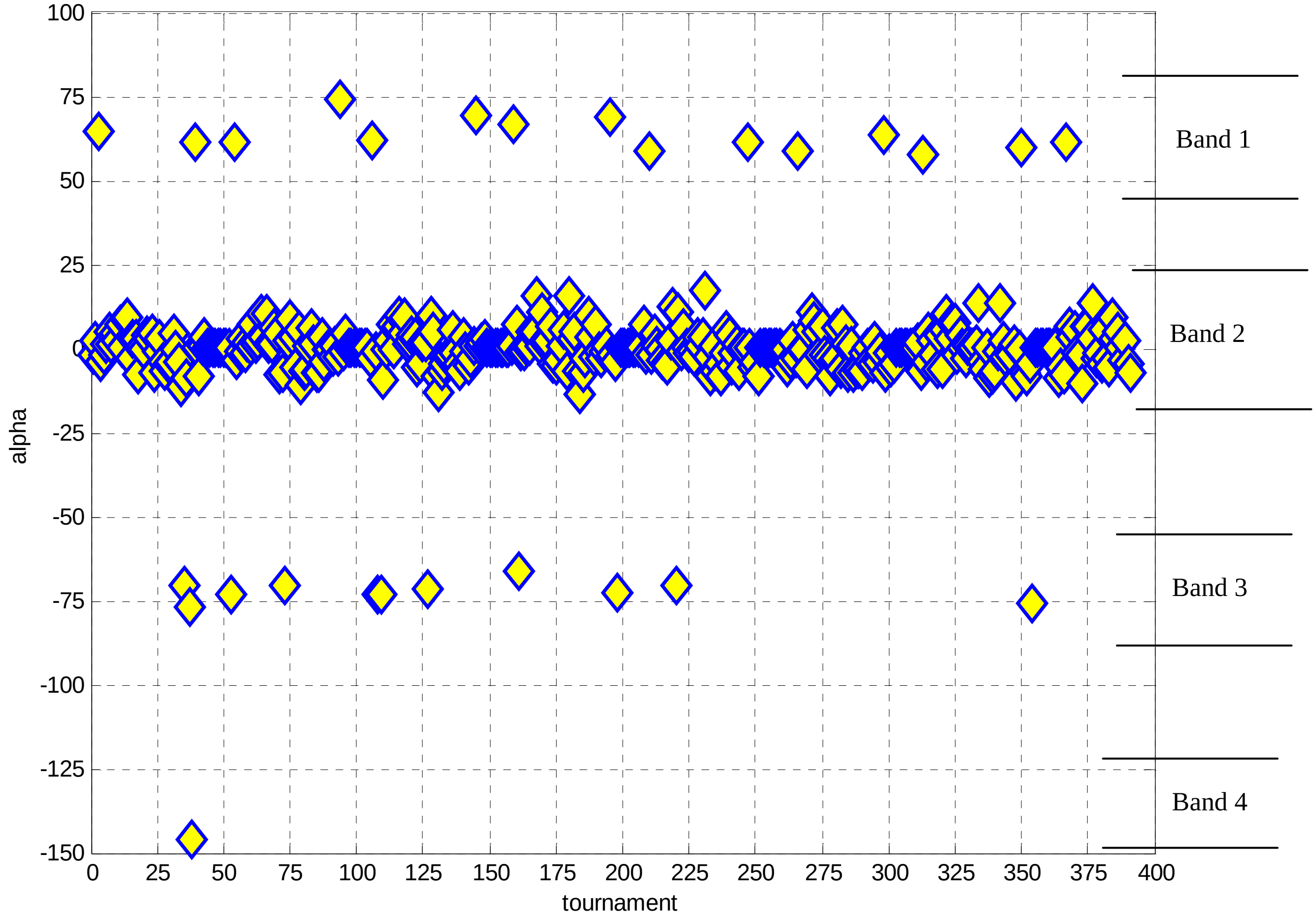

Fig 15. α values obtained for each real world tournament.

The tournaments in fig 15 can be seen as being dispersed in to four horizontal 'bands', with means of around +65, 0, -70 and -145 on the α-axis. This effect is due to large structural differences between the tournaments. The majority are played over four rounds, i.e. 72 holes, however for each year of the study two are played over five rounds, ie 90 holes. These tournaments, The Bob Hope Classic and The Las Vegas Invitational are tournament numbers 3 and 39 respectively for the 1996 season, and can be seen in Band 1 on the plot repeated at regular intervals of roughly 52 tournaments as these tournaments are played around the same date every year. The α values obtained for these tournaments are higher as the total final scores will be greater due to the extra round played.

The tournaments in Band 3, beginning with number 35, the 1996 Canadian Open and 37, the 1996 BC Open, are all tournaments during which the number of rounds was reduced to three due to adverse weather conditions. Tournament number 38, in Band 4 with an α value of 145.84, is the 1996 Buick Challenge, which as with most other tournaments is normally played over four rounds but in this year particularly bad weather resulted in it being shortened to two rounds. Unlike in Band 1, there is no regular pattern for the occurrence of tournaments in Bands 3 and 4 as, obviously, playing tournaments over three rounds because of inclement weather is not something which is deliberate or planned. Although these tournaments do seem to occur at around the beginning and end of seasons – corresponding to winter months

The par score for one round on PGA tour golf courses is generally in the region 70-72, thus with one additional round on top of the normal four, the scores will be greater by roughly this value, and in tournaments with less than four rounds the scores will be reduced by a corresponding multiple. This explains the α-axis separation distance of the bands. The tournaments with five rounds have positive α values because, as discussed in section 3, in order to form the adjusted scores, α is subtracted from the true observed scores. In effect this adjustment obtains, for tournaments not played over four rounds, what the scores would have been had the tournament been played over four rounds. Note that this is not the same as just throwing away the scores from the final round of a five round tournament and using the 4 round aggregate, which would lose the information contained in the final round performances of the players. Similarly, the α values of tournaments played over fewer than 4 rounds are negative, as these scores must be increased to obtain an equivalent '4 round estimate'.

The smaller scale variations in α between tournaments within the same band represent the differences in course difficulty, weather effects and field strength as previously discussed in section 2 or 3?. As discussed in section 5.1, in order to create a continuous, roughly regular timeline, additional tournaments where all players don't take part are inserted to represent the off-season. These can be seen on the plot as a group of 8 or 9 successive tournaments with α values of zero. The first such region, representing the off-season period between the Sarazen World Open on 3rd Nov 1996 and the Mercedes Championship on 12th Jan 1997, runs from tournament number 44 to number 52. The later off-season periods follow at regular intervals on the plot.

5.3 Static model analysis of real world data

Fig 16 shows the static β values for the players obtained from the two-way anova model using the two step algorithm.

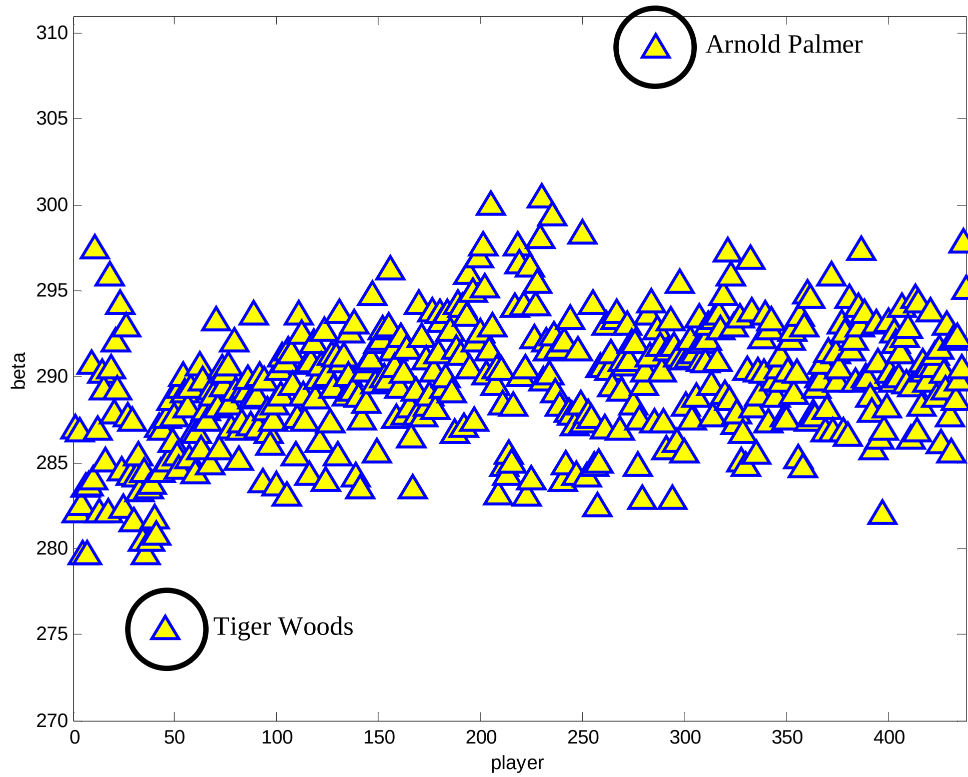

Fig 16. Static β values obtained from 2 step algorithm for all players over the entire period of study from 7th Jan 1996 – 20th Jul 2003. The players with the best and worst values of β are labelled.

Fig 16 shows that the best player, Tiger Woods (β = 275.16), has a considerably better β value than any of the other players. This is not a surprising finding: Tiger Woods has been described as "undeniably the best player of his generation"[25] and at the date of the final tournament in this study he had been ranked number 1 in the Official World Rankings for 206 consecutive weeks – nearly four years. That Arnold Palmer (β = 309.06) should be ranked as the worst player here is also unsurprising. Palmer is undoubtedly one of the greatest and most famous golfers of all time; however he was at his prime in the early 1960s and was 67 years old at the beginning of the period of study. During this period he played most of his competitive golf on the Seniors Tour, where the courses are perhaps better suited to players of his age. Although he could not realistically be considered as a competitive player, for reasons of nostalgia, and maybe because tournament organisers and sponsors knew what effect a household name such as Arnold Palmer, whatever his age, playing would have on audience figures, he played in twenty PGA Tour tournaments in the period, thus qualifying him for inclusion in this study.

Fig 17 shows the γ values for each player obtained from the Plackett-Luce model using the MM algorithm over the entire period of study.

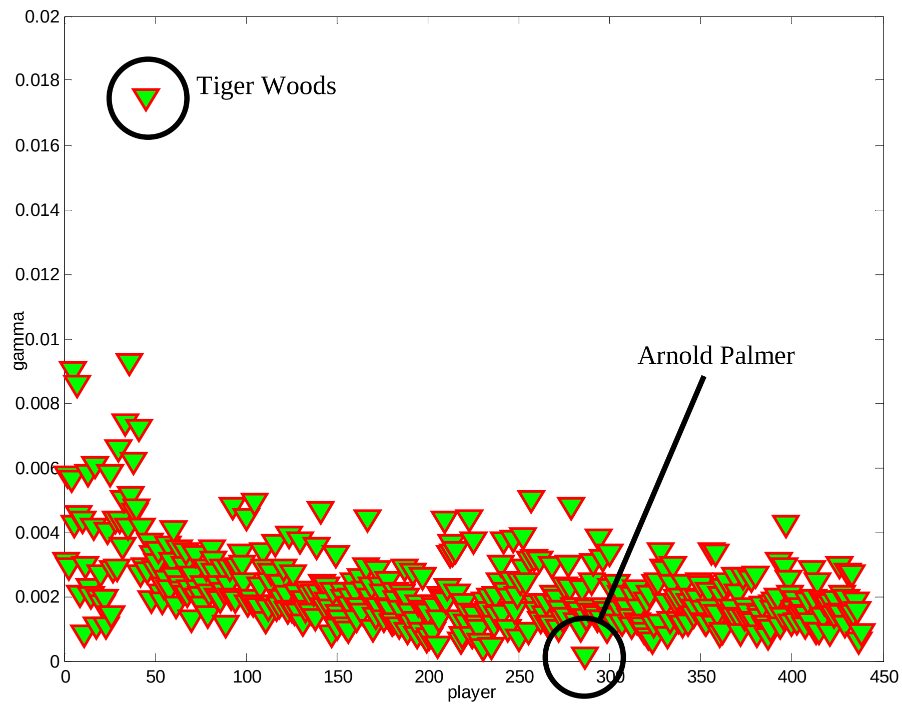

Fig 17. γ values obtained from the Plackett-Luce model. The best and worst players are the same as for the β ranking of the two-way anova model in Fig 16: Tiger Woods (γ = 0.017409) and Arnold Palmer (γ = 0.00020637).

5.4 Comparison of static models with Official World Rankings

The Official World Golf Ranking [26] is updated and issued weekly following the completion of the previous week's tournaments in five professional tours; the PGA Tour, PGA European Tour, Southern African PGA Tour, PGA Tour of Australasia, and PGA Tour of Japan. Points are awarded according to the players' finishing positions in tournaments. These points are accumulated over a two-year "rolling" period. Each player is then ranked according to his average points per tournament, which is determined by dividing his total number of points by the number of tournaments he has played over that two-year period, with a minimum requirement of 20 tournaments for each year. Tournaments that take place on different tours are given higher or lower ratings dependent on the strength of the tour. This means there are more points awarded for winning on the PGA Tour than in any of the other tours. In addition to this, some tournaments which are deemed to be of higher importance such as the Masters, U.S. Open, British Open, and PGA Championship carry more points.

Table 1 shows a comparison between the Official World Rankings and rankings obtained from the models studied in this project.

Player	Official World Ranking Points Avg. (ranking)	β value from two-way anova model (ranking)	γ value from Plackett-Luce model (ranking)	γ value from two-way anova model (ranking)
Tiger Woods	16.81 (1)	273.91 (1)	0.043038 (1)	0.10581 (1)
Ernie Els	9.48 (2)	277.99 (2)	0.015775 (2)	0.038194 (3)
Vijay Singh	8.06 (3)	278.63 (3)	0.014664 (4)	0.018224 (14)
Mike Weir	7.96 (4)	280.27 (12)	0.0076143 (17)	0.033096 (5)
Davis Love III	7.59 (5)	279.47 (8)	0.0099172 (6)	0.021937 (11)
Jim Furyk	7.34 (6)	279.97 (11)	0.0093904 (10)	0.015698 (15)
David Toms	6.70 (7)	279.37 (10)	0.008587 (13)	0.033036 (6)
Kenny Perry	6.29 (8)	280.38 (13)	0.0092066 (11)	0.015495 (16)
Padraig Harrington	6.05 (9)	279.10 (5)	0.0089318 (12)	0.030911 (7)
Nick Price	5.73 (10)	278.86 (4)	0.015522 (3)	0.02141 (12)
Phil Mickelson	5.60 (11)	279.46 (7)	0.0078575 (15)	0.034685 (4)
Retief Goosen	4.83 (12)	279.12 (6)	0.014601 (5)	0.050186 (2)
Justin Leonard	4.28 (13)	280.71 (15)	0.0098476 (8)	0.015091 (17)
Sergio Garcia	3.79 (14)	283.75 (39)	0.0043067 (65)	0.010503 (24)
Robert Allenby	3.64 (15)	281.8 (18)	0.0068998 (21)	0.018602 (13)

Table 1. The top 15 players in the Official World Rankings as of 20th Jul 2003 together with the player's two-way anova β and Plackett-Luce γ value and a γ value obtained from Monte Carlo sampling of the two-way anova model.

All of the values in Table 1 were obtained by using only data from the preceding two year period, as is the case in the Official World Rankings. Results from tours other than the PGA Tour are incorporated in to the Official World Rankings while, due to data constraints only results from the PGA Tour were used here to obtain the β and γ values. This may be reflected in the discrepancies in the ranking of Sergio Garcia, 14th in the Official World Rankings, but ranked much lower where only PGA Tour data is considered. Garcia is Spanish and has divided his time between the European and US

PGA Tours, with most of his success enjoyed in his home continent – boosting his Official World Ranking, but not the rankings based only on PGA Tour data.

Despite the fact that both are obtained from the same model, the rankings of players according to β (column 3) and the Monte Carlo estimate of γ (column 5) are not the same. This is because the Monte Carlo estimation uses both the player's score standard deviation as well as their β value.

Fig 18 shows a comparison between the γ estimates obtained from the Plackett-Luce model and by sampling of the two-way anova model – columns 4 and 5 of Table 1.

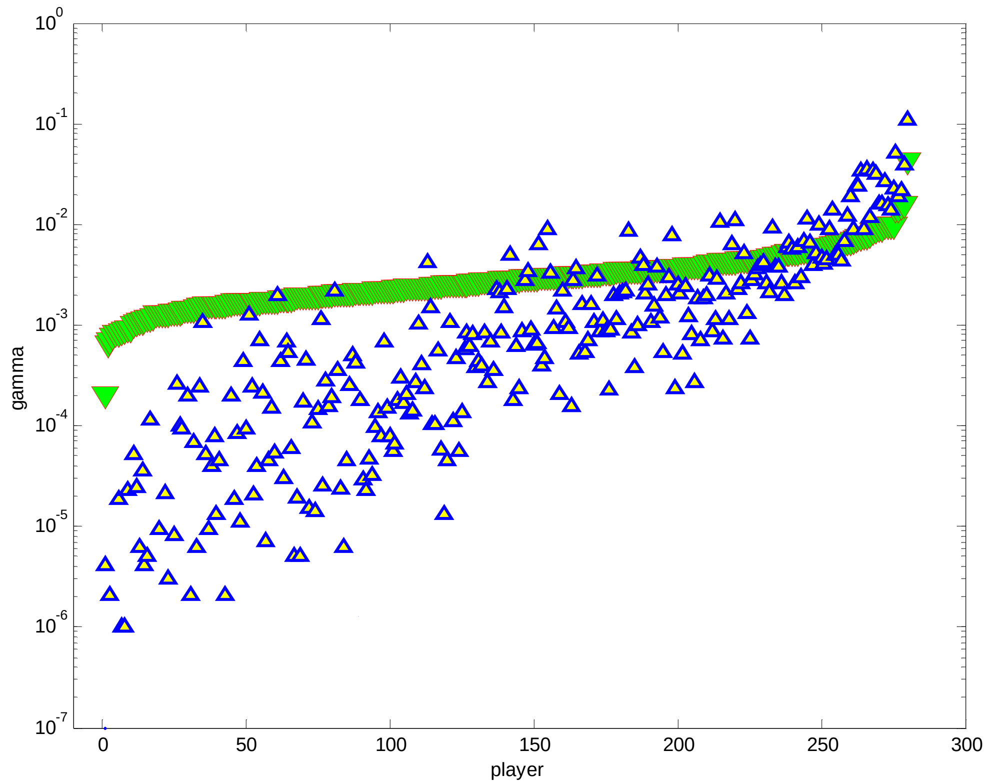

Fig. 18. Plot for all players of the results given in columns 3 and 4 of Table 1. The number of players is reduced from 438 to 282, as the period of study is reduced to two years, with the requirement of playing in 20 tournaments retained. γ is shown on a log scale for visibility. The values obtained from the Plackett-Luce model are shown in red/green, while the values from Monte Carlo sampling of the two-way anova model are shown in blue/yellow. The players have been sorted in to increasing Plackett-Luce γ order.

The same effect as observed for the toy data (fig 9 in section 3.3) is seen for the real world data in fig 18: The Plackett-Luce γ's are more 'bunched together', whilst there is a greater spread of the two-way anova γ's. It is possible that this effect could be an artefact of the way the two-way anova model estimate was obtained by Monte Carlo sampling. However the results in the above figure are based on 10^8 generated sample tournaments, so the values obtained should be accurate estimates. A player who receives a γ value of 10^{-6}, i.e. a one in a million chance of winning, has won 100 of these sample tournaments.

5.5 Dynamic two-way anova model analysis of real world data

Here a case study of two players is presented. The dynamic β values obtained from the Kalman filter for the players Tiger Woods and David Duval are plotted in Fig 19. These are two players about whom plenty of analysis and opinion has been written by sports journalists, allowing the results obtained to be compared and contrasted with ‘expert’ opinion.

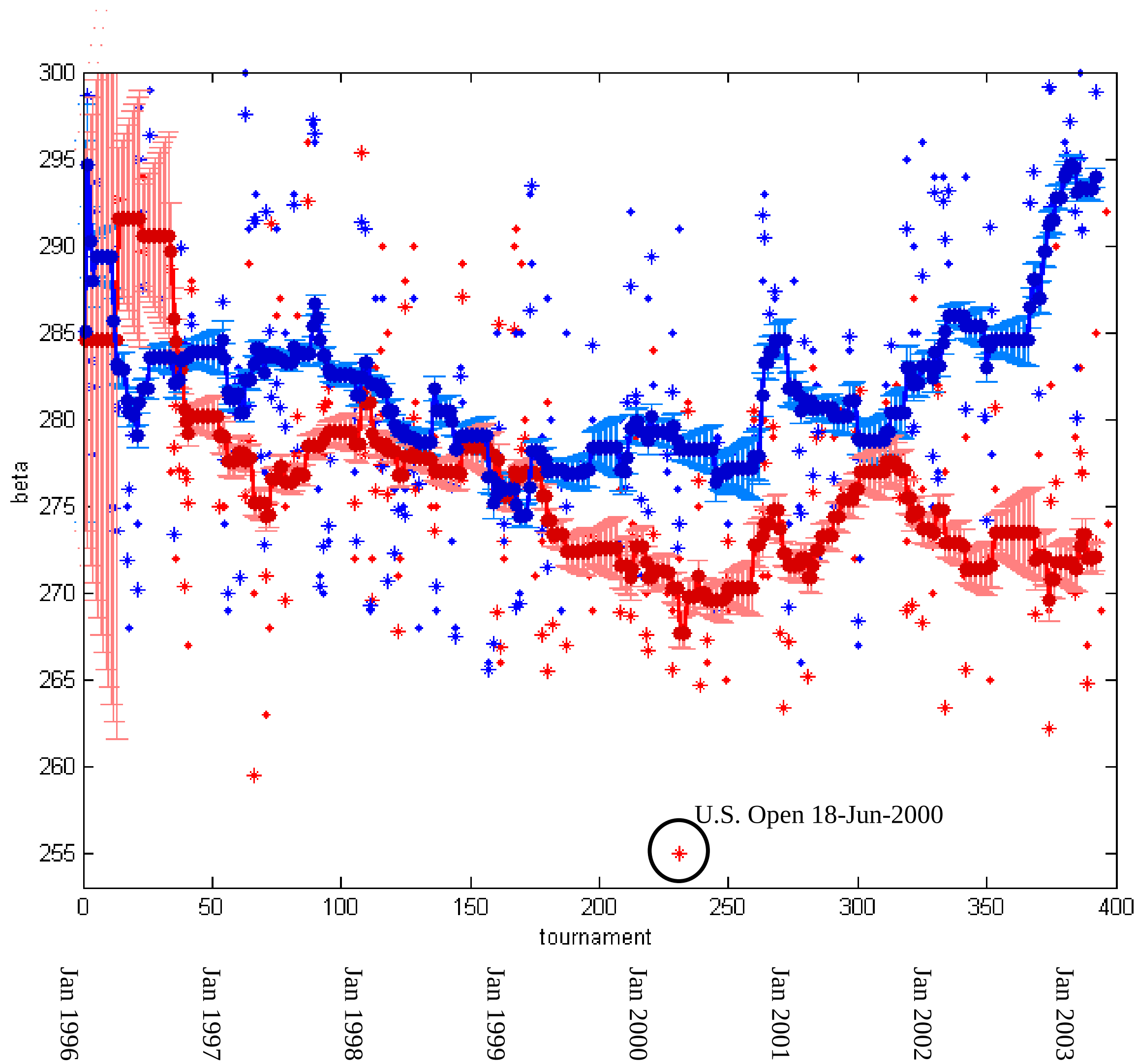

Fig 19. Plot of dynamic β obtained from Kalman filter for two players. Shown here in red is the β value evolving over time for Tiger Woods. The red dots are the actual scores for each of the tournaments, while the asterisks are the α-adjusted scores. In blue, is the equivalent for David Duval.

An article[27] by Ed Smerman in *The Australian* details the perceived fall from grace of David Duval stating that "*His decline is stunning*". It goes on to detail his collapse in form: "*His fall started in 2002, when he dropped to 80th on the money list with earnings* *of $838,045. This year he has made the cut in only four of 17 tournaments, with earnings of $84,708. His world ranking is 119*."

Looking at Fig 19, his β value (in blue) seems to mirror this analysis, a sharp increase beginning at around the start of 2002 and continuing to the present day. It wasn't always doom and gloom for Duval however, as Smerman notes "*Duval won 11 tournaments from 1997 to 1999.*" Indeed Duval's success was so pronounced that he was considered by many to be the main contender to Tiger Woods' crown: "*It wasn't that long ago that Duval was hovering in Tiger Woods' stratosphere. In fact, he was the last player to be ranked No.1 in the world before Woods claimed the top spot in 1999.*"

Again this assertion is echoed in the analysis of Fig 19. Duval's best β rating came early in 1999, when for a brief period it was indeed superior to that of Tiger Woods (shown in red).

Looking now at the development of Tiger Woods' career[28], early in the period of study – during 1996 – Woods was only 20 years old and making his first steps in the world of the PGA Tour. Indeed he only left college and turned professional in August of that year, his first professional tournament being tournament number 34, The Greater Milwaukee Open, in which he finished tied 60th with a score of 277 (which was α-adjusted here to 288.63). This bedding in period is reflected in his high β rating at the start of the study. As the plot shows, Woods quickly found his feet and went on to establish himself as one of the top golfers in the world, achieving his first number one place in the Official World Rankings in June 1997.

Undoubtedly the pinnacle thus far of his career was in the period June 2000 – April 2001, when he became the first player in history to complete golf's grand slam by holding all four 'major' championships at one time. His lowest β value came at around this period. His win at the US Open on the 18th of June 2000 was by an unprecedented 15 strokes, which remains the largest margin of victory ever recorded at a major tournament. Wood's final score was 272, which is α adjusted here to a score of 254.9, which is the lowest α adjusted score by any player in the period of study. The next best score was also by Woods: his α adjusted 259.4 (actual score = 270) at The US Masters tournament in Apr 1997. He won it by 12 strokes, a record for the tournament. The best α adjusted score by a player other than Woods was Mark Calcavecchia's 259.8 to win the Phoenix Open on the 28th Jan 2001 by 8 strokes. His actual score of 256 broke the PGA Tour record for the lowest 72-hole score that had stood for 46 years.

There has been some debate in the golfing press over a perceived fall in Woods' form of late. Mark Reason, writing in the Daily Telegraph[29] says "*Woods's gradual decline in form started in 2001*". He attributes the blame to physiological ailments, something Woods himself admits suffering from, saying "*my swing is unique. I put a lot of forces through the knee area - and the hip area and the back area… That's what made it sore in the first place*". Recent articles written about Woods invariably contain the word 'slump', and despite the protestations of Woods ("*It's been more of an annoyance than anything because I have to keep answering it*") as Reason puts it: "*by normal people's standards, he is not in a slump. But Woods is not normal and neither are his standards*."

The dynamic β analysis of Fig 19 would seem to show that Woods did indeed suffer a decline in form starting around 2001. However, his performances in mid to late 2002 and in 2003 would seem to indicate a recovery and that he is performing at a good level, although not quite reaching the heights of around 3 years ago.

5.6 Evaluation of models by application to gambling scenario

While sections 5.4 and 5.5 have included qualitative evaluation of model performance, in the absence of knowledge of the 'true' values of parameters, as were available for the toy data, there is no obvious method for quantitative evaluation of the models on real world data. As an attempt to do this, and also to gain a measure of the predictive power of the models, a study of their performance in the context of gambling was carried out, the results of which are presented here.

There are many ways offered by bookmakers to gamble on golf such as outright betting - betting on the eventual winner of a tournament before it starts - match betting - given two or more players, pick who will finish better placed and in-running betting - where the odds are updated in real time as the tournament progresses – to name a few.

Due to time constraints and for the sake of simplicity, for the purposes of this study only outright betting is considered. Odds were collected[30] for 12 tournaments: from the HP Classic on the 4th May 2003 till the BC Open on the 20th July 2003. The odds are traditional fractional odds, whereby for example if a bet of 1 unit on a player who's odds are 20/1 is placed and that player wins, it will result in a return of 20 units, plus the 1 unit stake, giving a net profit of 20 units on the transaction. The odds also provide for 'each-way' bets whereby half of the stake goes on the player winning outright as usual, but the other half is bet on the player finishing in a given number of the top places (in golf this is generally 5 places – which is the case for all the odds collected in this study) at reduced odds (1/4 of the winning odds here). Thus for an each way bet of 2 units on a player at 20/1, if the player finishes in the top five, but does not win, the returns will be the winning part of the stake, 1, + 1 * 20 / 4 = 6, i.e. a total profit of 4 units. If the player wins outright the return is (1 + 1 * 20) + (1 + 1 * 20 / 4) = 27, i.e. a profit of 25 units. In cases where there is a tie for 5th place dead heat rules apply, whereby the stake is proportionally reduced for every player in the tie.

As this study is performed on a relatively small sample size of only 12 tournaments, by simply using outright betting on winners, there would only be 12 possible bets that could win in the whole study. By using each-way betting, the number of potentially winning bets is increased to at least 60 (for the top five places), and more in the case of ties for 5th place. Thus all bets placed in this study are made on an each way basis.

For this study a system is required by which to convert the probabilistic predictions of the models in to actual bets to place on players, for this the Kelly Criterion[31] was considered. A system originally designed for information rates, it operates by taking in to consideration the probability of the player/team winning and also the overlay (advantage) that you have over the bookmaker based on the available odds. The overlay for a player, *i*, is calculated as:

$$\text{overlay for player } i = (\ P(\text{player } i \text{ wins}) * \text{bookmaker's odds for player } i) - 1. \quad (12)$$

The Kelly Criterion states that bets should only be placed where the overlay is positive and the stakes should be weighted proportionally as:

$$\text{bet amount} = \text{overlay} \,/\, (\text{bookmaker's odds} - 1). \quad (13)$$

In this study, 3 models are tested on their performance: the static Plackett-Luce and two-way anova models and the dynamic two-way anova model. Also, as a further basis for comparison, the predictions of a human expert[32] over the period are tested. Use of the Official World Rankings would also have been of interest here, however there is no principled way to convert these in to anything meaningful about how to bet: they were not designed with the ability of producing probabilistic predictions in mind.

Under the Kelly Criterion, all three of the models returned negative overlays for virtually every player in every tournament of the study. Thus the odds offered are almost always too small and Kelly Criterion is not a feasible system under which to bet. The proliferance of negative overlays would indicate that the odds offered by bookmakers for golf tournaments are of particularly poor value.

A measure of the 'value' offered in bookmakers odds can be obtained by calculating the sum of the corresponding probabilities. Given fractional odds for a player, the probability of that player winning, according to the odds, is given by 1 / (odds + 1). For example if a player is offered at odds of 40/1 then the corresponding probability is 1 / (40 + 1) = 0.0224. Given odds of winning for all the players competing in a tournament, the sum of the corresponding probabilities, all things being equal, should be 1. If the sum is greater than 1, then this shows that the odds offered are too small to be considered completely ‘fair’. The value of this sum of probabilities is often referred to as the ‘bookie’s cut’ as this is in fact how bookmakers make their money, by setting the odds in their favour. The extent to which the sum of the probabilities is greater than 1 indicates how much of a cut the bookmaker is taking from the money bet on a tournament. An ideal situation for a bookmaker is for all the money bet on a tournament to be evenly distributed proportional to the odds offered, such that the total to be paid out in winnings is the same no matter which player wins. In this case the proportion of the total money bet on the tournament retained by the bookmaker is given by (<bookie’s cut> - 1) / <bookie’s cut>. So for example a bookie’s cut value of 2 means that the bookmaker retains half of the total stake money.

For the odds obtained for the twelve tournaments in this study, which are the combination of the best odds from three different bookmakers, the bookie’s cut is on average 1.4609, with a high of 1.5450 and a low of 1.2871. This means that the bookmaker retains on average 31.6% of the total money bet on a tournament. Thus any potential for making a profit by betting on golf is somewhat limited.

For the purposes of this evaluation, a system is required by which a reasonable number of bets are actually placed, which was not the case for the Kelly Criterion. Provided all models use the same system, the results obtained will still give an indication of which model performs better – even if this is defined in terms of the least money lost, as opposed to the most money gained.

In the system used, each model is assigned 100 units of virtual money to bet on each tournament. The amount to bet on each player (the stake) is calculated as proportional to the probability of that player winning according to the model under investigation. The number of players that can be backed in each tournament is limited to five so as to avoid spreading the stake money too thinly throughout the field. The study was carried out with a shifting time line of the data available to the models: in calculating the probabilities the models can only use data up to and including the previous week’s tournament. The human expert predictions are in the form of three players to back each tournament, so the 100 unit stake money was divided equally among the three players. The results are shown in fig 20.

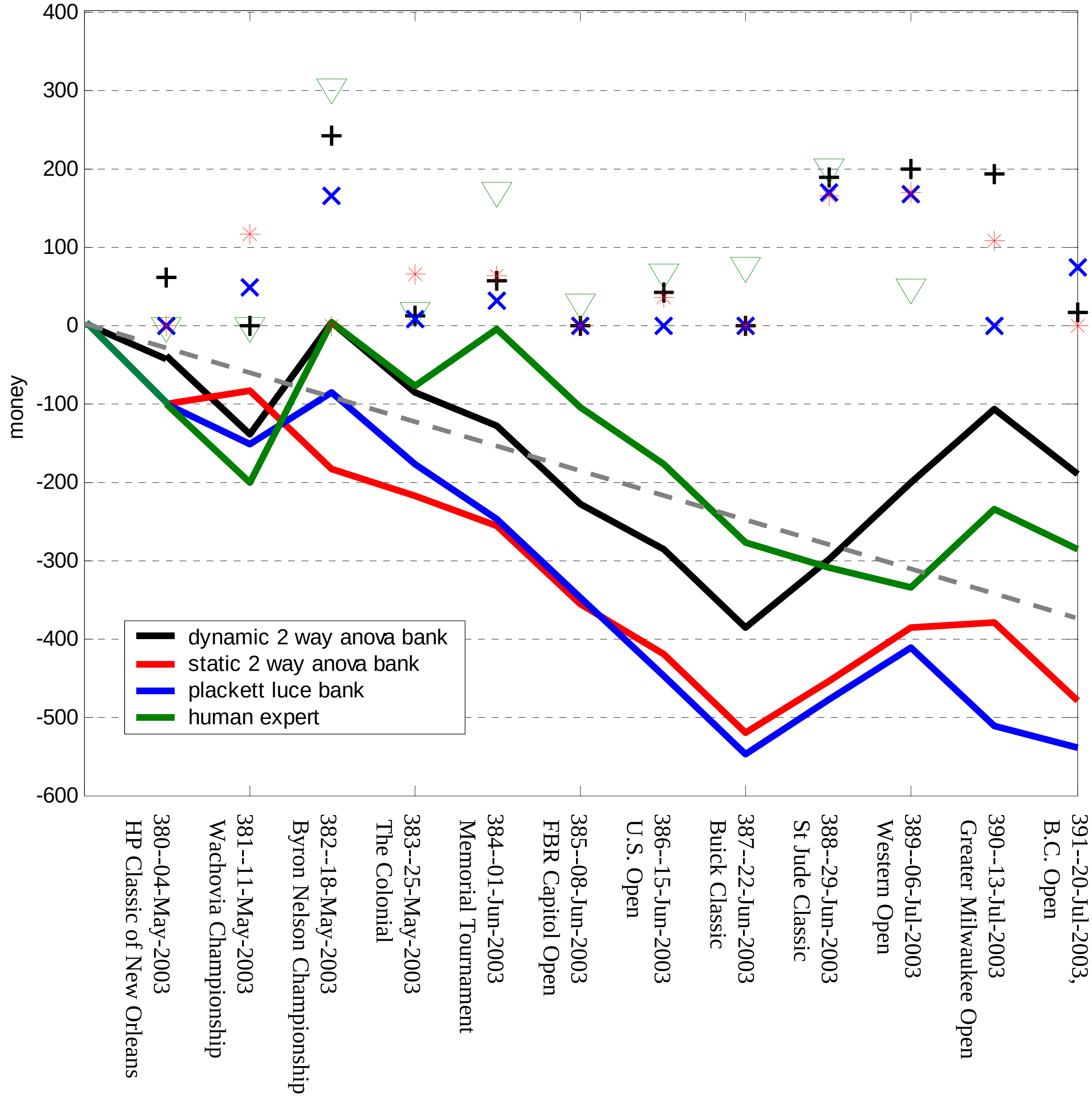

Fig 20. The results from application of the models to a gambling scenario.

The colour coded symbols towards the top of fig 20 are the returns from the 100 unit stake invested in each tournament, while the lines show a running total of profit/loss for each of the models and the human expert. The dashed line shows the expected effect of the 'bookie's cut': a 31.6% loss in money.

The obvious observation to make from fig 20 is that all of the methods, including the human expert, lose money over the period – from 15.9% of the total investment for the dynamic two-way anova model to 44.9% in the case of the Plackett-Luce model. As mentioned earlier, the way in which odds are set means that any betting strategy will almost certainly lose money in the long run: assuming that the bookmakers are doing their jobs properly, any monetary gains are usually due to short term fortuitous runs. As discussed previously the 'bookies cut' gives a measure of this disparity. The dashed line on the plot in fig 20 shows the performance of a system, which in the absence of any bookie's cut, would perform at a break-even level.

In terms of relative performance of the models, the dynamic model (black) clearly outperforms the two static models and ends up with the most money at the end. There is very little to separate the two static models, although the static two-way anova model (red) does finish on a slightly higher amount than the Plackett-Luce model (blue). The study was performed on quite a small sample size of only twelve tournaments, so any assertions made from it are of little statistical significance, however it can be concluded, on the basis of this evidence, that whilst the dynamic two-way anova model loses money, this can possibly be attributed to the bookmaker's profit margins rather than poor predictions by the model. It is also shown to be competitive with a human expert in its predictions.

6

6.0 Discussion and Conclusions

In this section the results and findings from this project are discussed. Section 6.1 considers the main difference between the Plackett-Luce and two-way anova models: the use of rankings or scores. In section 6.2 the potential for use of the models in other sports is briefly discussed, while 6.3 looks at the issues involved in obtaining ‘fair’ ranking systems and how the models could be applied to this in golf. In section 6.4 improvements to the work in this project are suggested and possible future work outlined. The conclusion in section 6.5 gives a brief retrospective overview of the project as a whole.

6.1 Scores vs Rankings

The key difference between the Plackett-Luce model and the two-way anova model proposed in this project is that the latter uses scores, whilst the former uses rankings. It is clear that a set of scores for a tournament contains more information than a ranking: the ranking can be obtained from the scores, but the reverse – obtaining the scores from the ranking – is not possible. Thus, it would seem that a model based on scores should produce a more accurate reflection of the true underlying model. However, it may be that the additional information contained in the scores is not relevant, or even misleading. This is probably an application-specific consideration.

Multi-competitor sports could perhaps be further sub-divided in to categories where either ranking or score is of greater importance. Consider motor racing such as Formula One or Nascar - multi-competitor sports in which all the drivers are on the track at the same time, and directly influencing each other’s performance. Motor racing is dominated by tactics associated with over-taking. A driver will attempt to physically block the driver behind from overtaking. Thus in formula one, the level of performance is not so directly related to the actual time a driver finishes in, but more to the ranking they finish in.

In golf, the actions of one player can not physically affect the performance of any other player in a comparable way. Although there is the possibility of an interactive effect between players for psychological reasons – a player may focus themselves to play better if they see they are only a few shots behind the leader, or a leader may panic and play worse if they look at a scoreboard and see a player rapidly catching them up. However, psychological effects aside, there is a clear distinction between multi-competitor sports where physical interaction, although not necessarily physical contact, between competitors is possible such as motor and horse racing, and where it is not possible such as golf and time trial skiing.

The relative importance of scores and rankings may well be strongly related to the degree of interaction between players. This project was carried out specifically with application to golf data in mind, which resulted in the focus being shifted more towards the score-based two-way anova model as opposed to the rank-based Plackett-Luce model. While the two-way anova model was shown to outperform the Plackett-Luce model on the toy data, it must be remembered that the toy data was generated using the two-way anova model – so the comparison could be expected to be somewhat biased. In terms of performance on the real world data, the difficulties in evaluating the models make it hard to firmly conclude in favour of a model one way or the other. The static two-way anova model does slightly outperform the Plackett-Luce model in the gambling study, and the dynamic two-way anova model more so, however the reality is that neither are the true underlying model of PGA Tour golf, which in terms of the relative importance between scores and rankings probably lies somewhere between the two.

6.2 Application to other sports

While the two-way anova model developed here is quite specific to golf, there is certainly potential for it to be applied to other sports. The key idea of establishing an ‘event effect’ which can be estimated and then used to adjust scores – allowing results from different days and places to be directly compared – is applicable to many other multi-competitor sports. Also the time dependent model based on random Gaussian walks and the Kalman filter approach to fitting such a model is more or less directly applicable to obtaining a dynamic model in other sports.

The two-way anova model is not necessarily the best model to use in any situation. As mentioned previously, there are other sports which intuitively would seem better suited to a rankings based analysis as afforded by the Plackett-Luce model.

6.3 Ranking systems

There is often debate in sports about what constitutes a fair long term ranking system – often this is important because it is the ranking which decides who wins trophies or how the money is distributed. In golf this isn’t so much the case – money and trophies are awarded for each tournament individually – but there are other requirements for a fair ranking system. Aside from prestige reasons, there is the question of Ryder Cup qualification. The Ryder Cup is held every two years and pitches twelve golfers from the United States against twelve from Europe in direct team competition. The golfers involved are ostensibly the best players over the previous season, who have ‘earned’ the right to play. Being selected in a Ryder Cup team is a high accolade and very important for golfers for patriotic and, perhaps more cynically, exposure and sponsorship reasons. Currently the sides are selected using the US PGA Tour and European Tour money lists. This method of selection can lead to ‘unfair’ inclusions and exclusions from the teams. Players who play in lots of tournaments getting average results will often end up with more money despite not being as good players as those who are more selective about which tournaments they compete in. Also many of the top European golfers play on the higher profile US Tour; however the European team is selected only on the basis of the player’s ranking on the European Tour money list. Often the top Europeans who play in America go back to Europe to play in a few high profile big money tournaments from which alone they hope to earn enough money to qualify.

Clearly, the ethos of the Ryder Cup tournament is to have the best twelve players available playing on each side, which under the current system is far from guaranteed. A ranking system based on a Plackett-Luce or static two-way anova would circumvent the problems associated with a prize money accumulation system and thus provide a fairer system for player ranking and Ryder Cup team selection. Using the existing Official World Rankings would also be an improvement on the current system, although these are produced in a somewhat ad hoc fashion, compared to the more principled approach of the Plackett-Luce and two-way anova models.

Having said this, there is also a desire for sports ranking systems to be transparent and largely understandable by the lay person. It is also perhaps useful to be able to state before a tournament that if a player finishes say above 10^{th} place that they will move up in to the top 20 players in the world. Or more importantly, if a player needs to know exactly what place they must finish in order to qualify for the Ryder Cup team. In systems such as the Plackett-Luce and static two-way anova the performance of all players, even those seemingly unrelated to the situation, is taken in to consideration. This can mean that making such statements prior to a tournament becomes very difficult or impossible.

6.4 Future work and scope for improvements

Much of the conventional wisdom about golf holds that some players are better suited to certain types of courses: a very long course would be more likely to produce a winner who is a 'big hitting' powerful player, while a short course with tight fairways and awkwardly placed greens would be better matched to a player, who's game is based more on accuracy than length. Human expert predictions are often based on such factors: whether or not a player has won or played well on the course, or similar courses, previously can form a large part of the analysis.

The models in this project do not make any allowance for factors of this type. The fact that the dynamic two-way anova model is shown to be at least competitive with a human expert in the gambling study shows that there is merit in a model based purely on quantitative results such as player scores. This in itself is an interesting point – it could be that the conventional wisdom is unfounded and that the type of course is not of such importance. However, if there is truth in it, the models could feasibly be improved by building in such considerations. In the case of the two-way anova model, the course effect term, α, is assumed to be constant for all players for a given tournament. If this assumption was relaxed by allowing it to be different for different players, this could allow for effects due to different types of course.

The problem with effects such as these is that they are very difficult to quantify. A possible means for this purpose could be a statistical analysis of tournament records in far more detail than just the final scores of players. There is a huge amount of statistical information recorded in each tournament, such as the number of times the player found the fairway from the tee, the number of greens they reached within two or more shots of par for the hole (greens in regulation), the number of putts for the round, the average length of drive and so on. This information could be used to establish a profile of each golfer: possibly a cluster analysis could classify them in to groups such as 'long hitting and not very accurate', 'short hitting and accurate' and 'good at putting'. These profiles could then be used to produce a profile of each course, whereby it could be possible to show that players who are good drivers or good putters for example perform better than would otherwise be expected on that course. If these effects could be quantified, then something akin to a 'player-course' compatibility co-efficient could be calculated and incorporated in to predictions.

Another way to improve the predictions of the models would be to expand the data to include other tours as well as the PGA Tour. In this project this was not possible due to time constraints. Some players, who are amongst the best in the world, choose to play a lot of golf on other tours such as the European tour, as they wish to remain based in Europe for personal reasons. These players often only play in a limited number of the larger tournaments on the US PGA Tour and can be considered as realistic contenders to win these. In terms of the models in this project, only using US PGA Tour data means the information on these players is limited, which leads to greater error and uncertainty in their ratings. In the 12 tournament period of the gambling study, there were 7 occasions when a player not included in the data, through having not played in 20 tournaments, finished in the top 5. On every occasion it was not that this player had just appeared from nowhere, they had been playing on other tours – in most cases the European PGA Tour.

Inclusion of other tours, such as is done in the Official World Rankings would mean more information is made available on players. In order for the models to work properly using data from different tours there would need to be a reasonable amount of mixing of players between the tours, as all ratings are made relative to other players. The greater the degree of mixing, the more accurate the relative weightings of players who only play on one tour would be.

In the two-way anova model the Score standard deviation has an important effect on predictions. This can be seen in the rankings comparison (table 1). Retief Goosen is ranked 6th in terms of β, but when Monte Carlo sampling is performed, using the estimated score standard deviation, to obtain γ estimates, his ranking increases to 2nd. In the model, the distribution of players scores is assumed to be Gaussian. This allows for easier analysis, but it could be possible that these distributions are not truly Gaussian. The use of a kernel estimator could possibly produce better representations of the distributions.

The two step algorithm used for the estimation of α in the two-way anova model is shown empirically to produce good estimates on the toy data. A possible improvement would be a more principled approach to this problem which is essentially the optimisation of the objective function $\sum_{i,j}\left(y_{ij}-\beta_i-\alpha_i\right)^2 z_{ij}$ subject to the constraint $\sum_j \alpha_j = 0$ where the elements of z can take the value 1 or 0 to indicate participation or non-participation in a tournament.

It would be interesting to test the models on real world data from other multi-competitor sports. Particularly to see if the suggested relationship between the degree of competitor interaction in the sport and relative importance of scores and rankings is observed. For these purposes the development of a dynamic extension of the Plackett-Luce model and accompanying procedure for fitting data to this model would be credit-worthy work.

6.5 Conclusion

The central objective of the project was to obtain a model for multi-competitor sports, which given data in the form of previous performance of players can make probabilistic predictions on the outcome of future tournaments. This has certainly been achieved. Both the existing Plackett-Luce model and the developed static two-way anova model can be used for this purpose. A further proviso would be that the predictions of the models are of merit. The performance in the gambling study of the dynamic extension of the two-way anova model, which is more tuned towards predictive power, would seem to indicate that the predictions are of reasonable quality, while there is still potential for improvement of the model.

More generally the aim to apply the models to the real world data was achieved and was successful in terms of producing findings which intuitively seem correct. This indicates that there is potential for using statistical modelling and machine learning in sport for the confirmation, and quantisation of existing ‘anecdotal’ theories, and possibly for the discovery of previously unseen relationships and trends. The ability afforded by the two-way anova model to directly compare the performance of a competitor or competitors in different events, held on different dates and in different locations in a principled manner is certainly of great practical benefit.

Acknowledgements

Many thanks to my supervisor Chris Williams for much helpful advice and feedback both on the work done and in the preparing of this document.

The computational work was carried out using MathWorks' MATLAB version 6. Thanks also to Dave Hunter of Penn State University for providing his (far more efficient) MATLAB implementation of the MM algorithm.

The data used in the project was collected from a wide variety of sources, the principle of which was the golf database at www.tour-tips.com, who also provided the betting tips used as the 'human expert' in the gambling study.

My studies in Informatics at the University of Edinburgh were funded by EPSRC.

Bibliography

[1] Warren D. Smith *Rating Systems for Gameplayers*, and Learning NEC Institute Technical Report (1993)

[2] Soren P. Sorensen *An overview of some methods for ranking sports teams*, http://www.phys.utk.edu/sorensen/ranking/Documentation/Sorensen_documentation_v1.pdf (accessed 10/9/2003)

[3] David Wilson *Bibliography on College Football Ranking Systems*, http://www.cae.wisc.edu/~dwilson/rsfc/rate/biblio.html (accessed 10/9/2003)

[4] R. A. Bradley and M. E. Terry *Rank analysis of incomplete block designs*, Biometrika, 39, 324–345 (1952)

[5] Zermelo, E. *Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung*, Mathematische Zeitschrift, 29, 436–460 (1929)

[6] Leonard Knorr-Held *Dynamic Rating of Sports Teams*, (1999) http://citeseer.nj.nec.com/knorr-held99dynamic.html (accessed 10/9/2003)

[7] D.A. Harville *Predictions for National Football League games via liner-model methodology*, Journal of the American Statistical Association, 75, 516-524 (1980)

[8] H. A. David *The Method of Paired Comparisons*, 2nd edition, Oxford University Press, New York (1988)

[9] G. Simons and Y.C. Yao. *Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons,* Annals of Statistics, 27, 1041–1060 (1999)

[10] R. N. Pendergrass and R. A. Bradley *Ranking in triple comparisons*, Contributions to Probability and Statistics, O. Olkin et al., eds. Stanford University Press (1960)

[11] John I. Marden *Analyzing and Modeling Rank Data*, 1st Edition, Monographs on Statistics and Applied Probability 64, Chapman & Hall (1995)

[12] R. L. Plackett *The analysis of permutations*, Applied Statistics, 24, 193–202. (1975)

[13] R. D. Luce *Individual Choice Behavior*, Wiley, New York. (1959)

14 D.R. Hunter *MM algorithms for generalized Bradley-Terry models*, Annals of Statistics, to appear (2004)

Available from http://www.stat.psu.edu/~dhunter/papers/bt.pdf (accessed 10/9/2003)

[15] Christopher Chatfield *Statistics for technology*, Third edition (revised), A course in applied statistics, Chapman & Hall (1992)

[16] D.R. Hunter *MATLAB code for Bradley-Terry models*,
Available from http://www.stat.psu.edu/~dhunter/btmatlab/ (accessed 10/9/2003)

[17] R. E. Kalman. *A New Approach to Linear Filtering and Prediction*, Transaction of the ASME-Journal of Basic Engineering, March, pp. 35-45 (1960).

[18] Grewal, S. Mohinder and Angus P. Andrews *Kalman Filtering Theory and Practice*, Upper Saddle River, NJ USA, Prentice Hall. (1993)

[19] Peter S. Maybeck *Stochastic Models, Estimation, and Control*, Volume 1, Academic Press, Inc. (1979)

20 Code used included code from and modified code from Kevin Murphy *Kalman filter toolbox for Matlab* (1998)
Available from http://www.ai.mit.edu/~murphyk/Software/Kalman/kalman.html (accessed 10/9/2003)

[21] Ghahramani and Hinton *Parameter Estimation for LDS*, U. Toronto tech. report (1996)

[22] Digalakis, Rohlicek and Ostendorf *ML Estimation of a stochastic linear system with the EM algorithm and its application to speech recognition*, IEEE Trans. Speech and Audio Proc., 1(4),431-442 (1993)

[23] Durbin and Koopman *Time Series Analysis by State Space Models*, Oxford Statistical Science Series (2002)

24 Fahrmeir and Tutz *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer Series in Statistics (1994)

25 Sandy Zinn *Lesson learned at Masters: Tiger is human*, SportsTicker
Available from http://www.allsports.com/cgi-bin/showstory.cgi?story_id=40601 (accessed 10/9/2003)

[26] *OWGR* Compiled by IMG, Pier House, Strand-on-the-Green, Chiswick, LondonW4 3NN Information adapted from http://www.officialworldgolfranking.com (accessed 10/9/2003)

[27] Ed Smerman *Duval wandering in wilderness*, The Australian (1/8/2003)
Available from http://foxsports.news.com.au/story/0,8659,6845838-23213,00.html (accessed 10/9/2003)

28 Mike Morrison *Tiger Woods Timeline*,
Available from http://www.infoplease.com/spot/tigertime1.html (accessed 10/9/2003)

[29] Mark Reason *Woods is plagued by pain*, Daily Telegraph (22/06/2003)
Available from
http://www.telegraph.co.uk/sport/main.jhtml?xml=/sport/2003/06/22/sgwood22.xml
(accessed 10/9/2003)

[30] Odds collected from three bookmakers: TotalBet (http://www.totalbet.com), SportingOdds (http://www.sportingodds.com) and Victor Chandler (http://www.victorchandler.com), compared, and the best available odds for each player retained for the calculation of winnings.

[31] J.L.Kelly,Jr. *A new interpretation of information rate*, Bell Syst. Tech. J.,35,917-926 (1956)

[32] Weekly tips provided free by golf information website http://www.tour-tips.com and archived in the golf forum at http://www.madjacksports.com

A1.1 Player IDs

1 Nobilo Frank
2 Hoch Scott
3 Stadler Craig
4 Stewart Payne
5 Price Nick
6 Calcavecchia Mark
7 Love Davis
8 Cabrera Angel
9 Barranger Todd
10 Flesch Steve
11 Donald Mike
12 Jimenez Miguel Angel
13 Harrington Padraig
14 Romero Eduardo
15 Fryatt Edward
16 Cook John
17 Couples Fred
18 Parsons Lucas
19 Zoeller Fuzzy
20 Cejka Alex
21 Nicklaus Jack
22 Daly John
23 Crenshaw Ben
24 Goosen Retief
25 Leonard Justin
26 Curry Paul
27 Johansson Per-Ulrik
28 Perry Chris
29 Ogrin David
30 Lehman Tom
31 Faxon Brad
32 Stricker Steve
33 Perry Kenny
34 Els Ernie
35 Watson Tom
36 Singh Vijay
37 Sluman Jeff
38 Mickelson Phil
39 Funk Fred
40 Duval David
41 Furyk Jim
42 Jones Steve
43 O'Meara Mark
44 Pavin Corey
45 Woods Tiger
46 Bradley Michael
47 Faldo Nick
48 Tolles Tommy
49 Waldorf Duffy
50 Brooks Mark
51 Maggert Jeff
52 Norman Greg
53 Gamez Robert
54 Henke Nolan
55 Clements Lennie
56 Fehr Rick
57 Haas Jay
58 Black Ronnie
59 Carter Jim
60 Hart Dudley
61 Kelly Jerry
62 Springer Mike
63 Day Glen
64 McGovern Jim
65 Nelson Larry
66 Sindelar Joey
67 Azinger Paul
68 Bryant Brad
69 Fabel Brad
70 Jurgensen Steve
71 Lancaster Neal
72 McCarron Scott
73 Purtzer Tom
74 Uresti Omar
75 Wadkins Bobby
76 Allem Fulton
77 Kraft Greg
78 Mattiace Len
79 Sullivan Mike
80 Tryba Ted
81 Appleby Stuart
82 Blake Jay Don
83 Edwards Joel
84 Paulson Carl
85 Pooley Don
86 Scherrer Tom
87 Browne Olin
88 Burke Patrick
89 Kamm Brian
90 McCallister Blaine
91 Mize Larry
92 Sauers Gene
93 Tway Bob
94 Bryant Bart
95 Fleisher Bruce
96 Lickliter Frank
97 Lowery Steve
98 Martin Doug
99 Reid Mike
100 Roberts Loren
101 Sasaki Hisayuki
102 Blackmar Phil
103 Gullion Joey
104 Hulbert Mike
105 Parnevik Jesper
106 Standly Mike
107 Stockton, Jr. Dave
108 Claar Brian
109 Herron Tim
110 Jacobsen Peter
111 Swartz Mike
112 Acosta Joe
113 Dawson Marco
114 Cochran Russ
115 Tennyson Brian
116 Huston John
117 Smith Taylor
118 Green Ken
119 Gump Scott
120 Maginnes John
121 Mayfair Billy
122 Pohl Dan
123 Simpson Tim
124 Toms David
125 Wiebe Mark
126 Byrum Tom
127 Gallagher Jeff
128 Gibson Kelly
129 Rymer Charlie
130 Sutherland Kevin
131 Beck Chip
132 Fergus Keith
133 Gallagher Jim
134 Henninger Brian
135 Rose Clarence
136 Royer Hugh
137 Austin Woody
138 Daley Joe
139 Estes Bob
140 Langham Franklin
141 Mediate Rocco
142 Stankowski Paul
143 Brisky Mike
144 Dunlap Scott
145 Freeman Robin
146 Heinen Mike
147 Lohr Bob
148 Morse John
149 Parry Craig
150 Wrenn Robert
151 Boros Guy
152 Clearwater Keith
153 Doyle Allen
154 Edwards David
155 McCumber Mark
156 Mahaffey John
157 Pruitt Dillard
158 Delsing Jay
159 Goydos Paul
160 Rinker Lee
161 Wadkins Lanny
162 Simpson Scott
163 Waite Grant
164 Pride Dicky
165 Chamblee Brandel
166 Forsman Dan
167 Janzen Lee
168 Armour Tommy
169 Lewis J.L.
170 Brown Billy Ray
171 Carnevale Mark
172 Hammond Donnie
173 Murphy Sean
174 Frost David
175 Williamson Jay
176 Whittaker Ron
177 Adams John
178 Berganio David
179 Gilder Bob
180 Bean Andy
181 Randolph Sam
182 Weibring D.A.
183 Wilson John
184 Bertsch Shane
185 Tyner Tray
186 Smith Chris
187 Andrade Billy
188 Elliott John
189 Hallberg Gary
190 Rinker Larry
191 Bratton Alan
192 Magee Andrew
193 Rintoul Steve
194 Rodriguez Anthony
195 Aubrey Emlyn
196 Julian Jeff
197 Peoples David
198 Inman John
199 Sills Tony
200 Barr Dave
201 Edwards Danny
202 Medlin Scott
203 Jordan Pete
204 Rusnak Gary
205 O'Grady Mac
206 Gorman Bryan
207 Wood Willie
208 Hart Jeff
209 Cink Stewart
210 Levi Wayne
211 Lietzke Bruce
212 Sutton Hal
213 Triplett Kirk
214 Riley Chris
215 Elkington Steve
216 Kite Tom
217 Thorpe Jim
218 Wadsworth Fred
219 Schulz Ted
220 Strange Curtis

221 Fiori Ed
222 Hughes Bradley
223 Howell Charles
224 Sandelin Jarmo
225 DiMarco Chris
226 Mast Dick
227 Utley Stan
228 Dougherty Ed
229 Halldorson Dan
230 Hayes Mark
231 Ogilvie Joe
232 Hensby Mark
233 Zokol Richard
234 Briggs Danny
235 Burns George
236 Hill Guy
237 Dodds Trevor
238 Morland David
239 Leggatt Ian
240 Weir Mike
241 Wilson Mark
242 Kendall Skip
243 Hayes J.P.
244 Wurtz Mark
245 Tanaka Hidemichi
246 Ozaki Joe
247 Maruyama Shigeki
248 Lyle Sandy
249 Begay Notah
250 Hinkle Lon
251 Jobe Brandt
252 Lonard Peter
253 Woosnam Ian
254 Watts Brian
255 Grady Wayne
256 Langer Bernhard
257 Montgomerie Colin
258 Allenby Robert
259 Booker Eric
260 Campbell Michael
261 Bates Pat
262 Sisk Geoffrey
263 Johnson Kevin
264 Dunakey Doug
265 Kresge Cliff
266 Demsey Todd
267 Estes Jim
268 Petrovic Tim
269 Barron Doug
270 Gotsche Steve
271 Byrum Curt
272 Christie Michael
273 Jones Kent
274 O'Keefe Jack
275 Pernice Tom
276 Armstrong Ty
277 Clarke Darren
278 Gay Brian
279 Verplank Scott
280 Tataurangi Phil
281 Dennis Clark
282 Hill Jason
283 Porter Lee
284 Twitty Howard
285 Glasson Bill
286 Palmer Arnold
287 Nicklaus Gary
288 Rummells Dave
289 Yokoo Kaname
290 Pate Steve
291 Kaye Jonathan
292 Allen Michael
293 Hart Steve
294 Westwood Lee
295 Wentworth Kevin
296 Choi K.J.
297 Long Michael
298 Mednick Adam
299 Willis Garrett
300 Geiberger Brent
301 Quigley Brett
302 Hjertstedt Gabriel
303 Johnson Eric
304 Durant Joe
305 Green Jimmy
306 Sutherland David
307 Wolcott Bob
308 Alarcon Rafael
309 Skinner Sonny
310 McRoy Spike
311 Johnston Jimmy
312 Kase Hideki
313 Horgan P.H.
314 Damron Robert
315 Burton Kevin
316 Claxton Paul
317 Bowden Craig
318 Sullivan Chip
319 Silveira Larry
320 Micheel Shaun
321 Mollica Tony
322 Tewell Doug
323 Dowdall John
324 Kanada Craig
325 Beckman Cameron
326 Hnatiuk Glen
327 Herrera Eduardo
328 Franco Carlos
329 Bjorn Thomas
330 Olazabal Jose Maria
331 Bateman Brian
332 Gooch Lan
333 Kribel Joel
334 Sposa Mike
335 Ames Stephen
336 Clark Michael
337 Gogel Matt
338 Hosokawa Kazuhiko
339 Muehr Michael
340 Grunewald Kelly
341 Frazar Harrison
342 Conley Tim
343 Lardon Brad
344 Friend Bob
345 Isenhour Tripp
346 Allan Steve
347 Coughlan Richie
348 Bates Ben
349 Riegger John
350 Toledo Esteban
351 Cheesman Barry
352 Small Mike
353 Loustalot Tim
354 Barlow Craig
355 Eaks R.W.
356 May Bob
357 Veazey Vance
358 Campbell Chad
359 Gage Bobby
360 Steel Iain
361 Rollins John
362 Nolan Keith
363 Kuchar Matt
364 Henry J.J.
365 Kuehne Hank
366 Fasth Niclas
367 Elder Brad
368 Goggin Mathew
369 Chalmers Greg
370 Gossett David
371 Martin Casey
372 Walcher Rocky
373 Baird Briny
374 Gore Jason
375 Moss Perry
376 Tidland Chris
377 Wetterich Brett
378 Clark Tim
379 Howison Ryan
380 Paulson Dennis
381 Purdy Ted
382 Pappas Deane
383 Miyamoto Katsumasa
384 Raulerson Charles
385 Couch Chris
386 Warren Charles
387 Seawell David
388 Van Pelt Bo
389 Brehaut Jeff
390 Burns Bob
391 Beem Rich
392 Sabbatini Rory
393 Donald Luke
394 Spence Craig A.
395 Staton Kenneth
396 Pampling Rod
397 Garcia Sergio
398 Van De Velde Jean
399 Lawrie Paul
400 Heintz Bob
401 Molder Bryce
402 Schwarzrock Brent
403 Cochran Bobby
404 Buha Jason
405 Perks Craig
406 Gove Jeff
407 Bengoechea Aaron
408 Caron Jason
409 Smith Jerry
410 Anderson Jeremy
411 Byrd Jonathan
412 Baddeley Aaron
413 Restino John
414 Scott Adam
415 Haas Hunter
416 Coceres Jose
417 Gow Paul
418 Gangluff Stephen
419 Slocum Heath
420 Ellis Danny
421 Ferguson Ben
422 McLardy Andrew
423 Wilson Brian
424 Ogilvy Geoff
425 Daley Jess
426 Senden John
427 Pappas Brenden
428 Peterson Matt
429 Weekley Boo
430 Perez Pat
431 Crane Ben
432 Ozaki Jumbo
433 McGinley Paul
434 Torrance Sam
435 Rocca Costantino
436 Floyd Ray
437 Ballesteros Seve
438 Player Gary

A1.2 Tournament IDs

ID--Date,Name,score to make cut, highest score after four rounds
Blank entries represent off-season periods.

1--07-Jan-1996, Mercedes Championship,0,299
2--14-Jan-1996, Nortel Open,143,295
3--21-Jan-1996, Bob Hope Classic,282,356
4--28-Jan-1996, Phoenix Open,142,289
5--11-Feb-1996, Buick Invitational,140,291
6--18-Feb-1996, Hawaiian Open,147,295
7--25-Feb-1996, Nissan Open,145,298
8--03-Mar-1996, Doral-Ryder Open,144,295
9--10-Mar-1996, Honda Classic,145,299
10--17-Mar-1996, Bay Hill Invitational,146,296
11--24-Mar-1996, Freeport-McDermott Classic,145,301
12--31-Mar-1996, The Players Championship,143,291
13--07-Apr-1996, BellSouth Classic,148,301
14--14-Apr-1996, The Masters,146,298
15--21-Apr-1996, MCI Classic,142,290
16--28-Apr-1996, Greater Greensboro Classic,146,295
17--05-May-1996, Houston Open,144,296
18--12-May-1996, Byron Nelson Classic,140,283
19--19-May-1996, MasterCard Colonial,144,294
20--26-May-1996, Kemper Open,143,295
21--02-Jun-1996, Memorial Tournament,147,297
22--09-Jun-1996, Buick Classic,145,295
23--16-Jun-1996, U.S. Open,148,297
24--23-Jun-1996, St Jude Classic,140,284
25--30-Jun-1996, Greater Hartford Open,143,291
26--07-Jul-1996, Western Open,143,294
27--14-Jul-1996, Michelob Championship,142,292
28--21-Jul-1996, Deposit Guaranty Classic,142,289
29--28-Jul-1996, CVS Charity Classic,141,288
30--04-Aug-1996, Buick Open,142,289
31--11-Aug-1996, PGA Championship,145,297
32--25-Aug-1996, NEC World Series of Golf,0,301
33--25-Aug-1996, Greater Vancouver Open,143,290
34--01-Sep-1996, Greater Milwaukee Open,139,279
35--08-Sep-1996, Canadian Open,144,221
36--15-Sep-1996, Quad City Classic,142,284
37--22-Sep-1996, B.C. Open,143,216
38--29-Sep-1996, Buick Challenge,143,143
39--06-Oct-1996, Las Vegas Invitational,207,354
40--13-Oct-1996, Texas Open,146,293
41--20-Oct-1996, Walt Disney World Classic,208,282
42--27-Oct-1996, Tour Championship,0,295
43--03-Nov-1996, Sarazen World Open,147,301
44
45
46
47
48
49
50
51
52
53--12-Jan-1997, Mercedes Championship,0,224
54--19-Jan-1997, Bob Hope Classic,281,354
55--26-Jan-1997, Phoenix Open,141,288
56--02-Feb-1997, AT&T Pebble Beach National Pro-Am,212,288
57--09-Feb-1997, Buick Invitational,142,296
58--16-Feb-1997, Hawaiian Open,140,290
59--23-Feb-1997, Tucson Open,145,293
60--02-Mar-1997, Nissan Open,144,292
61--09-Mar-1997, Doral-Ryder Open,144,299
62--16-Mar-1997, Honda Classic,142,293
63--23-Mar-1997, Bay Hill Invitational,144,296
64--30-Mar-1997, The Players Championship,146,301
65--06-Apr-1997, Freeport-McDermott Classic,146,297
66--13-Apr-1997, The Masters,149,301
67--20-Apr-1997, MCI Classic,146,291
68--27-Apr-1997, Greater Greensboro Classic,143,292
69--04-May-1997, Houston Open,146,298
70--11-May-1997, BellSouth Classic,144,298
71--18-May-1997, Byron Nelson Classic,138,283
72--25-May-1997, MasterCard Colonial,140,286
73--01-Jun-1997, Memorial Tournament,147,221
74--08-Jun-1997, Kemper Open,143,292
75--15-Jun-1997, U.S. Open,147,303
76--22-Jun-1997, Buick Classic,145,298
77--29-Jun-1997, St Jude Classic,140,286
78--06-Jul-1997, Western Open,146,296
79--13-Jul-1997, Quad City Classic,140,285
80--20-Jul-1997, Deposit Guaranty Classic,141,287
81--27-Jul-1997, Greater Hartford Open,143,287
82--10-Aug-1997, Buick Open,143,292
83--17-Aug-1997, PGA Championship,146,296
84--24-Aug-1997, NEC World Series of Golf,0,302
85--24-Aug-1997, Greater Vancouver Open,139,286
86--31-Aug-1997, Greater Milwaukee Open,140,285
87--07-Sep-1997, Canadian Open,145,295
88--14-Sep-1997, CVS Charity Classic,141,291
89--21-Sep-1997, Texas Open,143,294
90--28-Sep-1997, B.C. Open,143,294
91--05-Oct-1997, Buick Challenge,142,288
92--12-Oct-1997, Michelob Championship,141,292
93--19-Oct-1997, Walt Disney World Classic,208,288
94--26-Oct-1997, Las Vegas Invitational,212,365
95--02-Nov-1997, Tour Championship,0,292
96--09-Nov-1997, Sarazen World Open,148,301
97
98
99

100
101
102
103
104
105--11-Jan-1998, Mercedes Championship,0,294
106--18-Jan-1998, Bob Hope Classic,278,352
107--25-Jan-1998, Phoenix Open,142,291
108--01-Feb-1998, AT&T Pebble Beach National Pro-Am,139,213
109--08-Feb-1998, Buick Invitational,142,217
110--15-Feb-1998, Hawaiian Open,139,283
111--22-Feb-1998, Tucson Open,143,291
112--01-Mar-1998, Nissan Open,146,295
113--08-Mar-1998, Doral-Ryder Open,145,299
114--15-Mar-1998, Honda Classic,144,290
115--22-Mar-1998, Bay Hill Invitational,146,298
116--29-Mar-1998, The Players Championship,146,299
117--05-Apr-1998, Freeport-McDermott Classic,144,297
118--12-Apr-1998, The Masters,150,300
119--19-Apr-1998, MCI Classic,144,294
120--26-Apr-1998, Greater Greensboro Classic,143,293
121--03-May-1998, Houston Open,146,294
122--10-May-1998, BellSouth Classic,142,294
123--17-May-1998, Byron Nelson Classic,139,285
124--24-May-1998, MasterCard Colonial,141,287
125--31-May-1998, Memorial Tournament,145,293
126--07-Jun-1998, Kemper Open,142,293
127--14-Jun-1998, Buick Classic,147,222
128--21-Jun-1998, U.S. Open,148,299
129--28-Jun-1998, Western Open,146,296
130--05-Jul-1998, Greater Hartford Open,139,283
131--12-Jul-1998, Quad City Classic,138,281
132--19-Jul-1998, Deposit Guaranty Classic,143,288
133--26-Jul-1998, CVS Charity Classic,141,291
134--02-Aug-1998, St Jude Classic,142,289
135--09-Aug-1998, Buick Open,142,289
136--16-Aug-1998, PGA Championship,145,297
137--30-Aug-1998, NEC World Series of Golf,0,304
138--30-Aug-1998, Greater Vancouver Open,141,289
139--06-Sep-1998, Greater Milwaukee Open,139,285
140--13-Sep-1998, Canadian Open,146,296
141--20-Sep-1998, B.C. Open,144,293
142--27-Sep-1998, Texas Open,142,287
143--04-Oct-1998, Buick Challenge,141,287
144--11-Oct-1998, Michelob Championship,142,292
145--18-Oct-1998, Las Vegas Invitational,214,361
146--25-Oct-1998, National Car Rental Golf Classic,144,291
147--01-Nov-1998, Tour Championship,0,297
148--08-Nov-1998, Sarazen World Open,148,303
149
150
151
152
153
154
155
156
157--10-Jan-1999, Mercedes Championship,0,301
158--17-Jan-1999, Sony Open in Hawaii,142,290
159--24-Jan-1999, Bob Hope Classic,284,357
160--31-Jan-1999, Phoenix Open,147,299
161--07-Feb-1999, AT&T Pebble Beach National Pro-Am,220,220
162--14-Feb-1999, Buick Invitational,141,291
163--21-Feb-1999, Nissan Open,142,289
164--28-Feb-1999, Tucson Open,144,293
165--07-Mar-1999, Doral-Ryder Open,144,293
166--14-Mar-1999, Honda Classic,143,298
167--21-Mar-1999, Bay Hill Invitational,146,298
168--28-Mar-1999, The Players Championship,148,306
169--04-Apr-1999, BellSouth Classic,142,291
170--11-Apr-1999, The Masters,148,300
171--18-Apr-1999, MCI Classic,143,292
172--25-Apr-1999, Greater Greensboro Classic,143,293
173--02-May-1999, Houston Open,143,297
174--09-May-1999, Compaq Classic,142,287
175--16-May-1999, Byron Nelson Classic,138,284
176--23-May-1999, MasterCard Colonial,145,290
177--30-May-1999, Kemper Open,145,291
178--06-Jun-1999, Memorial Tournament,148,300
179--13-Jun-1999, St Jude Classic,138,285
180--20-Jun-1999, U.S. Open,147,307
181--27-Jun-1999, Buick Classic,146,295
182--04-Jul-1999, Western Open,143,295
183--11-Jul-1999, Greater Milwaukee Open,141,287
184--25-Jul-1999, John Deere Classic,138,279
185--01-Aug-1999, Greater Hartford Open,139,284
186--08-Aug-1999, Buick Open,142,287
187--15-Aug-1999, PGA Championship,146,300
188--29-Aug-1999, Reno-Tahoe Open,143,293
189--05-Sep-1999, Air Canada Championship,142,290
190--12-Sep-1999, Canadian Open,146,300
191--19-Sep-1999, B.C. Open,144,291
192--26-Sep-1999, Texas Open,142,288
193--03-Oct-1999, Buick Challenge,142,289
194--10-Oct-1999, Michelob Championship,145,292
195--17-Oct-1999, Las Vegas Invitational,206,360
196--24-Oct-1999, National Car Rental Golf Classic,141,288
197--31-Oct-1999, Tour Championship,0,287
198--31-Oct-1999, Southern Farm Bureau Classic,145,218
199
200
201
202
203
204
205
206
207
208--09-Jan-2000, Mercedes Championship,0,307

209--16-Jan-2000, Sony Open in Hawaii,140,290
210--23-Jan-2000, Bob Hope Classic,276,348
211--30-Jan-2000, Phoenix Open,142,290
212--06-Feb-2000, AT&T Pebble Beach National Pro-Am,216,291
213--13-Feb-2000, Buick Invitational,142,291
214--20-Feb-2000, Nissan Open,141,289
215--27-Feb-2000, Tucson Open,142,289
216--05-Mar-2000, Doral-Ryder Open,141,286
217--12-Mar-2000, Honda Classic,140,287
218--19-Mar-2000, Bay Hill Invitational,151,298
219--26-Mar-2000, The Players Championship,150,301
220--02-Apr-2000, BellSouth Classic,144,220
221--09-Apr-2000, The Masters,148,300
222--16-Apr-2000, MCI Classic,144,290
223--23-Apr-2000, Greater Greensboro Classic,144,298
224--30-Apr-2000, Houston Open,142,291
225--07-May-2000, Compaq Classic,143,290
226--14-May-2000, Byron Nelson Classic,143,291
227--21-May-2000, MasterCard Colonial,143,290
228--28-May-2000, Memorial Tournament,148,296
229--04-Jun-2000, Kemper Open,143,293
230--11-Jun-2000, Buick Classic,144,293
231--18-Jun-2000, U.S. Open,149,306
232--25-Jun-2000, St Jude Classic,142,291
233--02-Jul-2000, Greater Hartford Open,139,283
234--09-Jul-2000, Western Open,143,291
235--16-Jul-2000, Greater Milwaukee Open,140,286
236--23-Jul-2000, B.C. Open,143,289
237--30-Jul-2000, John Deere Classic,139,283
238--13-Aug-2000, Buick Open,142,289
239--20-Aug-2000, PGA Championship,147,299
240--27-Aug-2000, Reno-Tahoe Open,145,295
241--03-Sep-2000, Air Canada Championship,141,288
242--10-Sep-2000, Canadian Open,144,290
243--17-Sep-2000, Pennsylvania Classic,143,293
244--24-Sep-2000, Texas Open,140,284
245--01-Oct-2000, Buick Challenge,143,292
246--08-Oct-2000, Michelob Championship,141,292
247--15-Oct-2000, Invensys Classic,207,351
248--22-Oct-2000, Tampa Bay Classic,143,291
249--29-Oct-2000, National Car Rental Golf Classic,140,285
250--05-Nov-2000, Tour Championship,0,296
251--05-Nov-2000, Southern Farm Bureau Classic,140,284
252
253
254
255
256
257
258
259
260--14-Jan-2001, Mercedes Championship,0,299
261--14-Jan-2001, Tucson Open,144,293
262--21-Jan-2001, Sony Open in Hawaii,139,285
263--28-Jan-2001, Phoenix Open,140,285
264--04-Feb-2001, AT&T Pebble Beach59ational Pro-Am,214,291
265--11-Feb-2001, Buick Invitational,141,289
266--18-Feb-2001, Bob Hope Classic,277,349
267--25-Feb-2001, Nissan Open,142,290
268--04-Mar-2001, Genuity Championship,141,289
269--11-Mar-2001, Honda Classic,142,285
270--18-Mar-2001, Bay Hill Invitational,145,296
271--25-Mar-2001, The Players Championship,147,300
272--01-Apr-2001, BellSouth Classic,143,297
273--08-Apr-2001, The Masters,145,294
274--15-Apr-2001, WorldCom Classic,142,287
275--22-Apr-2001, Houston Open,145,296
276--29-Apr-2001, Greater Greensboro Classic,142,289
277--06-May-2001, Compaq Classic,142,287
278--13-May-2001, Byron Nelson Classic,139,282
279--20-May-2001, MasterCard Colonial,142,287
280--27-May-2001, Kemper Insurance Open,141,286
281--03-Jun-2001, Memorial Tournament,147,298
282--10-Jun-2001, St Jude Classic,139,283
283--17-Jun-2001, U.S. Open,146,296
284--24-Jun-2001, Buick Classic,144,292
285--01-Jul-2001, Greater Hartford Open,140,285
286--08-Jul-2001, Western Open,143,290
287--15-Jul-2001, Greater Milwaukee Open,141,286
288--22-Jul-2001, B.C. Open,141,290
289--29-Jul-2001, John Deere Classic,141,287
290--12-Aug-2001, Buick Open,140,283
291--19-Aug-2001, PGA Championship,141,286
292--26-Aug-2001, Reno-Tahoe Open,143,291
293--02-Sep-2001, Air Canada Championship,141,286
294--09-Sep-2001, Canadian Open,141,287
295--23-Sep-2001, Pennsylvania Classic,144,292
296--30-Sep-2001, Texas Open,140,287
297--07-Oct-2001, Michelob Championship,141,289
298--14-Oct-2001, Invensys Classic,209,354
299--21-Oct-2001, National Car Rental Golf Classic,139,282
300--28-Oct-2001, Buick Challenge,142,289
301--04-Nov-2001, Tour Championship,0,287
302--04-Nov-2001, Southern Farm Bureau Classic,143,289
303
304
305
306
307
308
309
310
311--06-Jan-2002, Mercedes Championship,0,304
312--13-Jan-2002, Sony Open in Hawaii,140,283
313--20-Jan-2002, Bob Hope Classic,275,346
314--27-Jan-2002, Phoenix Open,141,285
315--03-Feb-2002, AT&T Pebble Beach National Pro-Am,216,292

316--10-Feb-2002, Buick Invitational,143,294
317--17-Feb-2002, Nissan Open,142,288
318--24-Feb-2002, Tucson Open,140,284
319--03-Mar-2002, Genuity Championship,144,294
320--10-Mar-2002, Honda Classic,140,283
321--17-Mar-2002, Bay Hill Invitational,146,296
322--24-Mar-2002, The Players Championship,146,298
323--31-Mar-2002, Houston Open,143,292
324--07-Apr-2002, BellSouth Classic,144,299
325--14-Apr-2002, The Masters,147,295
326--21-Apr-2002, WorldCom Classic,143,288
327--28-Apr-2002, Greater Greensboro Classic,144,294
328--05-May-2002, Compaq Classic,144,289
329--12-May-2002, Byron Nelson Classic,139,285
330--19-May-2002, MasterCard Colonial,143,292
331--26-May-2002, Memorial Tournament,146,290
332--02-Jun-2002, Kemper Insurance Open,142,291
333--09-Jun-2002, Buick Classic,143,292
334--16-Jun-2002, U.S. Open,150,303
335--23-Jun-2002, Greater Hartford Open,141,286
336--30-Jun-2002, St Jude Classic,140,285
337--07-Jul-2002, Western Open,143,292
338--14-Jul-2002, Greater Milwaukee Open,139,283
339--21-Jul-2002, B.C. Open,141,286
340--28-Jul-2002, John Deere Classic,139,286
341--11-Aug-2002, Buick Open,142,290
342--18-Aug-2002, PGA Championship,148,302
343--25-Aug-2002, Reno-Tahoe Open,144,295
344--01-Sep-2002, Air Canada Championship,141,290
345--08-Sep-2002, Canadian Open,144,291
346--15-Sep-2002, Pennsylvania Classic,144,290
347--22-Sep-2002, Tampa Bay Classic,143,293
348--29-Sep-2002, Texas Open,138,281
349--06-Oct-2002, Michelob Championship,141,291
350--13-Oct-2002, Invensys Classic,206,348
351--20-Oct-2002, Walt Disney World Resort Golf Classic,138,283
352--27-Oct-2002, Buick Challenge,139,282
353--03-Nov-2002, Tour Championship,0,289
354--03-Nov-2002, Southern Farm Bureau Classic,142,215
355
356
357
358
359
360
361
362
363
364--12-Jan-2003, Mercedes Championship,0,290
365--19-Jan-2003, Sony Open in Hawaii,140,287
366--26-Jan-2003, Phoenix Open,138,283
367--02-Feb-2003, Bob Hope Classic,275,351
368--09-Feb-2003, AT&T Pebble Beach National Pro-Am,219,295
369--16-Feb-2003, Buick Invitational,143,294
370--23-Feb-2003, Nissan Open,145,295
371--02-Mar-2003, Chrysler Classic of Tucson,144,293
372--09-Mar-2003, Ford Championship,143,290
373--16-Mar-2003, Honda Classic,138,280
374--23-Mar-2003, Bay Hill Invitational,147,299
375--30-Mar-2003, The Players Championship,144,295
376--06-Apr-2003, BellSouth Classic,147,297
377--13-Apr-2003, The Masters,149,302
378--20-Apr-2003, The Heritage,142,288
379--27-Apr-2003, Houston Open,141,287
380--04-May-2003, HP Classic of New Orleans,139,286
381--11-May-2003, Wachovia Championship,145,294
382--18-May-2003, Byron Nelson Championship,140,288
383--25-May-2003, The Colonial,141,286
384--01-Jun-2003, Memorial Tournament,146,301
385--08-Jun-2003, FBR Capitol Open,144,295
386--15-Jun-2003, U.S. Open,143,293
387--22-Jun-2003, Buick Classic,142,289
388--29-Jun-2003, St Jude Classic,141,290
389--06-Jul-2003, Western Open,143,292
390--13-Jul-2003, Greater Milwaukee Open,142,288
391--20-Jul-2003, B.C. Open,141,286
392--20-Jul-2003, British Open,150,300
393--27-Jul-2003, Greater Hartford Open,140,286
394--03-Aug-2003, BuickOpen,142,288
395--10-Aug-2003, The International,Stableford
396--17-Aug-2003, PGA Championship,149,300
397--24-Aug-2003, NEC Invitational,308,308
398--24-Aug-2003, Reno-Tahoe Open,146,297
399--01-Sep-2003, Deutsche Bank Championship,143,290