STATISTICAL ANALYSIS OF PROFESSIONAL GOLF PLAYER ABILITIES

Based on the selection in the Ryder Cup

Abstract

In this project we will try to establish whether the teams selected for the Ryder Cup 2014 were the correct teams or whether there are flaws in the current selection process. We will analyse whether other statistical approaches would be better to select the teams. We will do this via three main methods; Plackett-Luce, an extension of a Bradley-Terry model, an ANOVA selection process and lastly a dynamic version of this ANOVA selection process which will analyse the form going into the Ryder Cup. We will then finish our study into the Ryder Cup selection process by picking our own team based off our statistical analysis and whether the teams picked are the same ones that we have selected based on our models and if not possible flaws in both our statistical model and the actual selection process used in the Ryder Cup 2014.

Nathan Thomas-Peter (UG)
N.Thomas-Peter@ncl.ac.uk

Contents

1 Introduction

1.0 Golf

First we will talk about the rules and most common formats of golf. Golf is a sport that contains tournaments that are usually played over 4 days, 18 holes a day with a par score over those 18 holes of 72. This means 72 holes are usually played per tournament with a par score of 288. Players compete to get around these holes in the lowest number of strokes possible, for example if it took 270 strokes to get around all 72 holes the score would be 18 under par. The problem arises for analysing the players that play in these tournaments because unlike most others sports where all teams or players compete at the same time, multiple players do not play in certain tournaments, meaning overall the players are competing around different golf courses. To compare these players we will need to way to analyse the abilities of players who do not play the same courses or even the same amount of tournaments. Some of the problems we encounter are the fact some courses are harder than others in general and that different weather conditions can effect a player's score on a given day. We also need to consider that for the tournaments offering more prize money, the calibre of player will be higher than a tournament offering less prize money meaning that a course might look easier just because players in that tournament are better.

Another thing we have to consider when analysing these players is the "cut". This generally means that some players get knocked out of the tournament, generally at the halfway point. This is decided by a predefined condition and is usually set to be a predefined number of players and anyone tied for the last position or players who are lower than a defined position, for example 70th or lower. Some tournaments, for example the Masters tournament is the top 50 golfers or those within 10 strokes of the leader, whichever the higher of the two is. We therefore are going to look at three main methods that will try to overcome some, if not all, of the listed problems and then we will look at other things that could affect our outcome we haven't considered.

The Ryder Cup is a tournament that happens bi-yearly between America and Europe where they pick their best 12 players to face off against each other in a series of golf rounds, some being played as pairs where players take alternating shots and others on their own. This means that picking players isn't just as easy as picking the top 12 players over the last year as some players will cooperate better as a team and boost morale. Another thing to note is how the teams are picked. This is based off prize money in tournaments and in most tournaments first place is awarded double the prize money than second so even though they may complete a tournament in the same number of strokes a play-off arises where the players compete over one more hole to determine a champion (usually the 18th hole of the 4th day). This would reflect much worse for the loser of the playoff on the rankings system. However, there were in fact no difference between the players over the 4 day tournament. This is the reason to investigate further into the player abilities to deem whether the selection process is flawed or whether the process is the most fair process overall. Also note that the two teams use different selection processes, Europe uses their ranking points from tournaments as well as prize money where as the American team uses prize money solely. Another difference is that the European team is mainly picked on the tournaments on the European tour and the official world rankings, whereas the American team is picked mainly from tournaments on the USPGA tour. However one exception is that both teams use Majors in their selection process, these are the four biggest tournaments of the year which earn the highest prize money in general and offer the most ranking points. These are the Masters, USPGA, US Open and The Open. Generally the world's best players all compete in these tournaments hence making it more competitive and scores much lower for these tournaments in general in comparison to how difficult a course really is. Also to note is each team has a number of captain's picks before the Ryder Cup where they are free to choose whomever they would like. These are in general the picks that most go against form. This is mainly because when the captains make selections we can see how much it has been biased by certain players being able to play together in the fourballs than actually selecting people who are at the height of form going into the Ryder Cup. This should be more apparent in the American team where Phil Mickelson will only play with Keegan Bradley, even though he may not be the best use of a captain's pick. They also could be influenced by media pressure into picking a player like Tiger Woods, the best golfer of his generation but has heavily declined in recent years after the scandal of 2009 (David, 2009) but like some other players have a big reputation that could influence captains decision, a factor that will not be incorporated into our statistical analysis.

1.1 Statistical Methods

One main thing to consider before looking at our methods is to know our methods are calculating correctly. We are going to do this by generating our own golf scores based on a player ability we know and generate, we can therefore know that our methods are working and can use the data on real players to make conclusions knowing how well the methods work and how much noise there will be contained in these methods. These statistical methods we will be looking at are the Plackett-Luce method, an extension to the Bradley-Terry method. This will use the ranking system of players, their positions, to generate a player ability value. We will also use an ANOVA method, a method that will use the player's scores to calculate a player ability. The final method we will consider is an extension to the ANOVA model but we will use a dynamic player model which analyses form over time so we can see how well a player performs throughout the year. This could be used to influence captains picks at the end of the year especially when we can see the form of particular players going into the Ryder Cup. One of the main problems we will encounter is the fact some formulas may need certain conditions applied to compensate as not every player plays in each tournament.

2 Plackett-Luce Model

2.0 Introduction to Plackett-Luce

We use this model to generate our toy data, the data we are going to use to check if the model works before using it on real data. To know why we use this model we must first briefly consider the Bradley-Terry model, which is a well more well-known model. The Bradley-Terry model is used to calculate the winner of a sporting event between 2 teams or people. This model could for example be used in boxing to calculate the winner of an upcoming fight based off their ability. There are many generalised forms for this model that have uses in the score difference in American football, football etc. The Bradley-Terry model states:

For a group of m individuals, with $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_m)$, the player/team ability, the probability of player/team i beating player/team j is given by

$$p_{ij} = \frac{\gamma_i}{\gamma_i + \gamma_j} \quad , \quad i = 1, 2, \dots, m \ , \quad j = 1,2, \dots, m \ , \quad i \neq j \qquad (2.0.1)$$

There are many extensions to this model but the one we will be looking at in more detail is the Plackett-Luce model.

2.1 Plackett-Luce

This is a model that is an extension to the Bradley-Terry model which Plackett and Luce both worked on independently. The Plackett-Luce model is defined as follows.

For a group of m individuals, with $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_m)$, the probability of the individuals ranking first to last in a given event is,

$$P[W = w|\gamma] = \prod_{i=1}^{m} \frac{\gamma_{w_i}}{\gamma_{w_i} + \gamma_{w_{i+1}} + \dots + \gamma_{w_k}} \qquad (2.1.1)$$

Where w_i is the i^{th} player in the ranking.

This model works because we determine the probabilities of all the players winning, randomly select a winner of the tournament based on the probabilities then calculate the probability of all the players winning a new tournament with the winner removed to generate second place. This continues until we have drawn all the players so we can rank the players first to last, we can repeat this process to generate results for our toy data for multiple tournaments.

2.2 MM algorithm

In order to estimate the overall player ability from the rankings we use MM algorithm of Hunter (Hunter, 2004). This algorithm runs iteratively by constructing a log-likelihood function and then maximising this. This comes with conditions that no player finishes first or last in every tournament. Providing this hold the algorithm is guaranteed to converge to a unique maximum likelihood of our player ability. The algorithm is as follows:

Consider a total of t players and N tournaments where the j^{th} tournament contains m_j players so each player is denoted by $a(j,1), \dots, a(j,m_j)$ for each tournament from first to last. Assuming i.i.d data then Hunters algorithm runs iteratively using the equation:

$$\gamma_t^{(k+1)} = \frac{w_t}{\sum_{j=1}^{N} \sum_{i=1}^{m_j - 1} \delta_{jit} \sum_{s=i}^{m_j} (\gamma_{a(j,s)}^{(k)})^{-1}} \tag{2.2.1}$$

Where w_t is the number of times the player finishes higher than last and δ_{jit} is an indicator of the event that the player no better than position i in the j^{th} tournament.

2.3 Applying Plackett-Luce to toy data

After we have generated our data using the Plackett-Luce model (2.1.1) we obtained a set of ranking for players throughout multiple tournaments. Figure 1 (below) shows the results from the MM algorithm, via the Plackett-Luce model, in blue and the true gamma values obtained before running our data through this method in red. Note that we have sorted our gamma via increasing values. For this process we ran a simulation for 300 players competing randomly in 80 tournaments with probability 0.4 of competing in each individual tournament. Here the Gamma variable represents the overall chance of a player winning a tournament, so essentially the ability of a player scaled to sum to 1. The algorithm has 25 iterations in total.

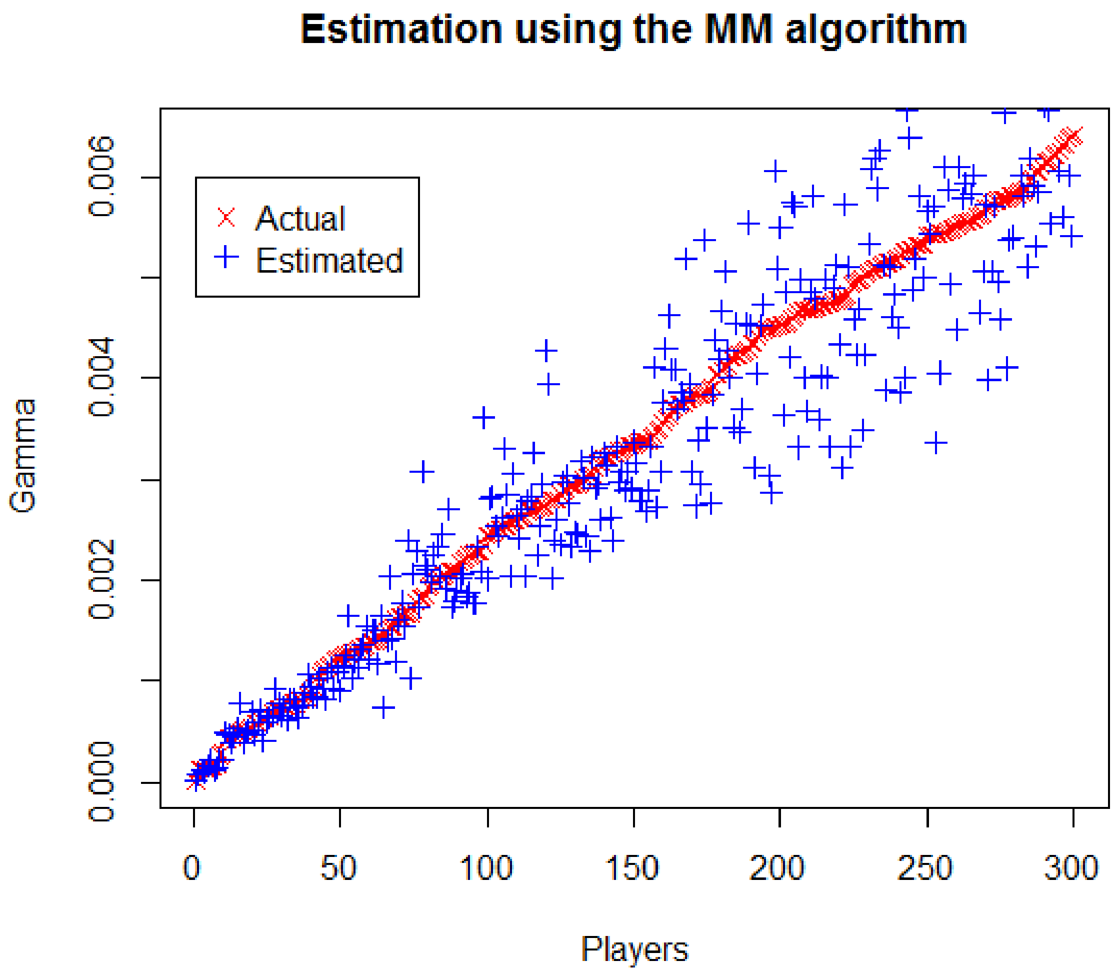

Figure 1. Plot showing results from a Plackett-Luce simulation using real data via the MM algorithm.

We can see from Figure 1 that the players with the worse values of gamma, and thus the worst players in our toy data, are predicted more accurately. This is reasonable as they complete in more of our sub-tournaments than the better players who are generally removed quite quickly. Hence, we have more information about those players. Looking at the plot we can see this more clearly. We can see at the lower end of the data the player's ability is almost predicted perfectly with our actual values of gamma, whereas at the top half of the data we can see that the data fans out and is predicted less accurately as they complete in a far lower number of our sub-tournaments. However, if we consider this fanning effect relatively the overall error still isn't that high even at the top end of the gamma values and the prediction is accurate overall.

3 Linear Model (Two-Way ANOVA)

3.0 Introduction to the linear model

What the Placket-Luce model lacks is a true knowledge of how well a player performs on a particular day. The reason for this is if two players won a tournament with a score of 280, there would be a play-off to determine first place. This would then reflect as the player in second below the person in first even though they completed the round in the same amount of strokes. This also applies if the winner wins a tournament by 10 strokes there would be no difference in the conclusion drawn from this model. Instead we look to consider the players scores instead of their ranking to determine their player ability. The fact we consider the scores retains more information about a player. Instead we consider a two-way analysis of our data by first considering the standard statistical model:

$$y_{ij} = \boldsymbol{\mu} + \beta_i + \alpha_j + \varepsilon_{ij} \qquad (3.0.1)$$

Where y_{ij} is the outcome of the event dependent on a mean effect $\boldsymbol{\mu}$, two independent effects, β_i and α_j and a noise term ε_{ij}. However, we notice that the independent terms both sum to 0 over i and j respectively. Now if we apply this to golf we can remove the mean effect by incorporating it into the $\boldsymbol{\beta}$ parameter and form a similar equation that reads:

$$y_{ij} = \beta_i + \alpha_j + \varepsilon_{ij} \qquad (3.0.2)$$

Where the y_{ij} now represents the score of player i in the j^{th} tournament, $= (\beta_1, \beta_2, \dots, \beta_t)$ can be interpreted as the player ability, similar to the gamma variable in the Plackett-Luce model. $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N)$ term can be thought of the course effect and if the difference between player's score if all the tournaments where competed in on a level playing field, that includes how difficult a course is, the ability of the other players in a tournament and the weather for instance. Notice that now the $\boldsymbol{\beta}$ vector doesn't sum to give 0 as we have incorporated the mean effect into this the $\boldsymbol{\alpha}$ vector sums to give 0 still, as overall the course differences are based around 0. This means effectively that we are interested in the beta parameter and we can consider alpha as a noise parameter we are not interested in the purpose of trying to form our own Ryder Cup teams. This means that our overall aim is to calculate this noise parameter and take it out to leave our overall player ability. We can generate scores for our toy data by simulating them from a normal distribution with mean $\beta_i + \alpha_j$ and variance ${\sigma_i}^2$ that is:

$$y_{ij} \sim N(\beta_i + \alpha_j, {\sigma_i}^2) \qquad (3.0.3)$$

Where $\beta_i + \alpha_j$, the score a player should obtain is based on his ability and how the course conditions and ${\sigma_i}^2$ represents the noise term of our model, the variance of a given player.

3.1 Simulation of our linear model

The overall aim as referred to before in the introduction to this chapter is to calculate alpha and remove it to leave the overall player ability. We generate a score matrix from the normal distribution (3.0.3) and use this to be our toy data. We run this for 325 players playing in 440 tournaments with a probability of playing in each tournament to be 0.4. We also calculate a score standard deviation for each player and this is drawn from a $U(5,15)$ distribution, this can be thought of a standard deviation of 5 to be a consistent player and 15 to be an erratic player, where erratic players are capable of having very good days and very bad days so their variance is high, the consistent player usually performs around their true value.

Difficulties arise in stripping alpha out of our model because even though we have deemed it to be a course effect we cannot measure this effect on players scores, this is due to the fact it will effect some players more than others, some players may not find playing in windy conditions as hard as others meaning the course effect for that individual is much higher than for the average person. We should therefore view alpha to be a variable that states how difficult it was for the average player playing in the tournament around the course. To determine the difficulty of the course we can only use the scores.

Consider all players playing in every tournament, this means that a tournament won with a score of 282 will be considered more difficult than a tournament where the winner scored 260. Comparing these directly would look like the player winning with 260 was a lot better than the player scoring 282. However, after removing alpha we may see that 260 was good enough to score 275 on an average tournament and 282 was actually good enough to score 270 on most courses. This means it's vital we remove alpha first before analysing the data.

Now considering the usual situation where not every player plays in every tournament, we would also need to be able to take into account that a Major tournament would attract the best players around the world, so a score of 280 in that tournament and an average score of 289 would not mean as much a tournament offering less prize money though a score of 280 and an average score of 289 would look the same on paper.

This means overall we need alpha to take account for a lot of noise terms such as field strength, weather, course difficulty etc. We can obtain an estimate of how good the players are in a tournament by from the beta values, but beta cannot be truly known until alpha is. Hence the algorithm is reciprocal. Thus we make an estimate of beta from our initial alpha values (0), then form a new alpha from the newly formed beta vector. This is repeated until they converge. We then estimate sigma from a matrix that has been adjusted by the alpha estimates.

3.2 Applying the linear model

After we have run this we can plot graphs of our estimated alpha to our true known value of alpha and the same for the beta values to see how close our iterative scheme is performing to the true value. Below (Figure 2) is a plot that shows the true beta values in red and our estimates in blue. Notice that the values for beta are sorted into ascending order. We have ran this simulation for 400 players over 80 tournaments with the chance of an individual playing in a tournament to be 0.4. The reason being that over the Ryder Cup year we expect somewhere near 300 players to compete in a sensible number of tournaments over that time period, with approximately 80 tournaments to be played during the qualification process that count towards being picked for a team. Most tournaments consist of approximately 120 players so assuming an individual has a probability of 0.4 to appear means we have 120 players in each tournament on average, but not always.

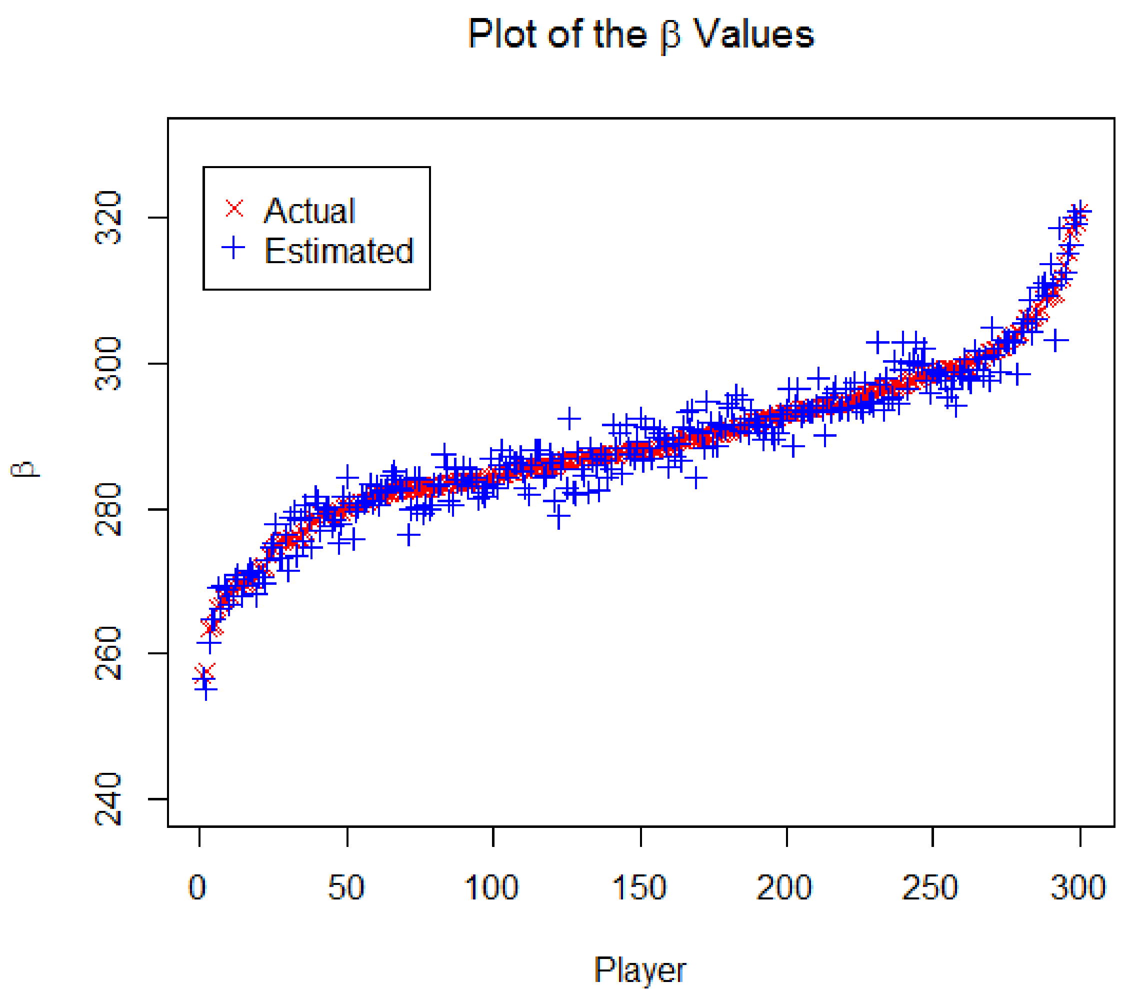

Figure 2. Plot showing results from our two-way linear model for the beta variable.

As we can see from Figure 2, we get a very consistent prediction of the true player ability where our estimation “curve” almost exactly follows the exact value “curve”, meaning our algorithm is a good estimator of true player ability. We can also see why the normal distribution is better than drawing from a uniform. We get a result with a lot of people close to the average score with a few players being very good and a few that are very bad. The same can be applied to the alpha variable.

Below (Figure 3) shows the same graph for the alpha values (or course effect), true values in red and our estimates in blue, to see how close we are predicting the value. Again the alpha values have been ordered in ascending order.

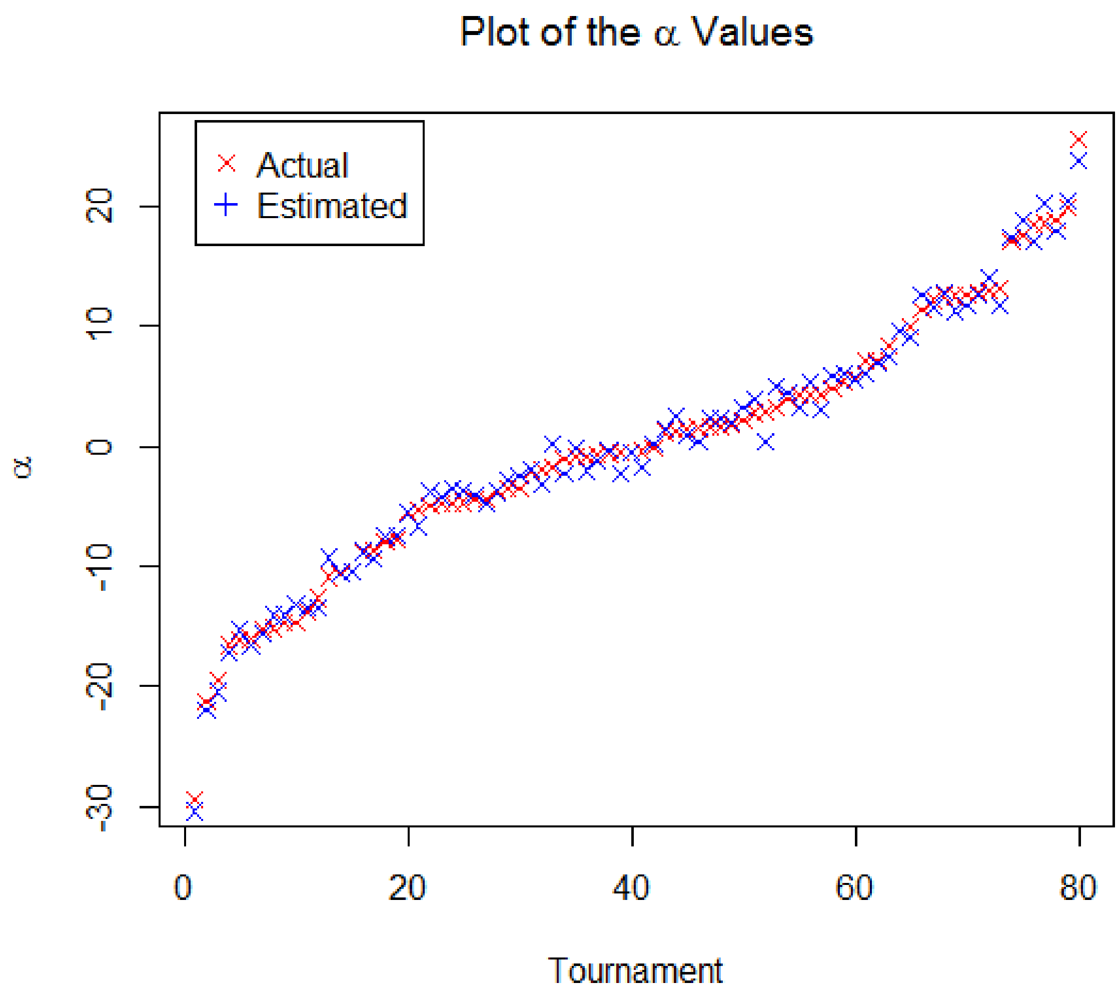

Figure 3. Plot showing results from our two-way linear model for the alpha variable.

From Figure 3 we can see that the tournament difficulty is predicted very well, again the estimated points almost follow the same “curve” as the true values of the course effect. And the estimation even though not as exceptional as the beta values is still very good. The reason for this is the fact we are running less tournaments in total, we only have 80 tournaments compared to the 400 players we are running for the beta values, the values, however, do not look too far away from their true values.

Below (Figure 4) is our estimates of the standard deviation, now we have to note that we do not expect this estimate to be as accurate as our previous two. The reason for this is that erratic players will be much harder to predict as their scores will fluctuate more over the different tournaments, whereas more constant players will be easier to predict due to the fact their scores will be fairly near each other more often and we can estimate a more accurate standard deviation. Again the values of sigma have been ordered in ascending order to give a clearer idea of how well our prediction is working.

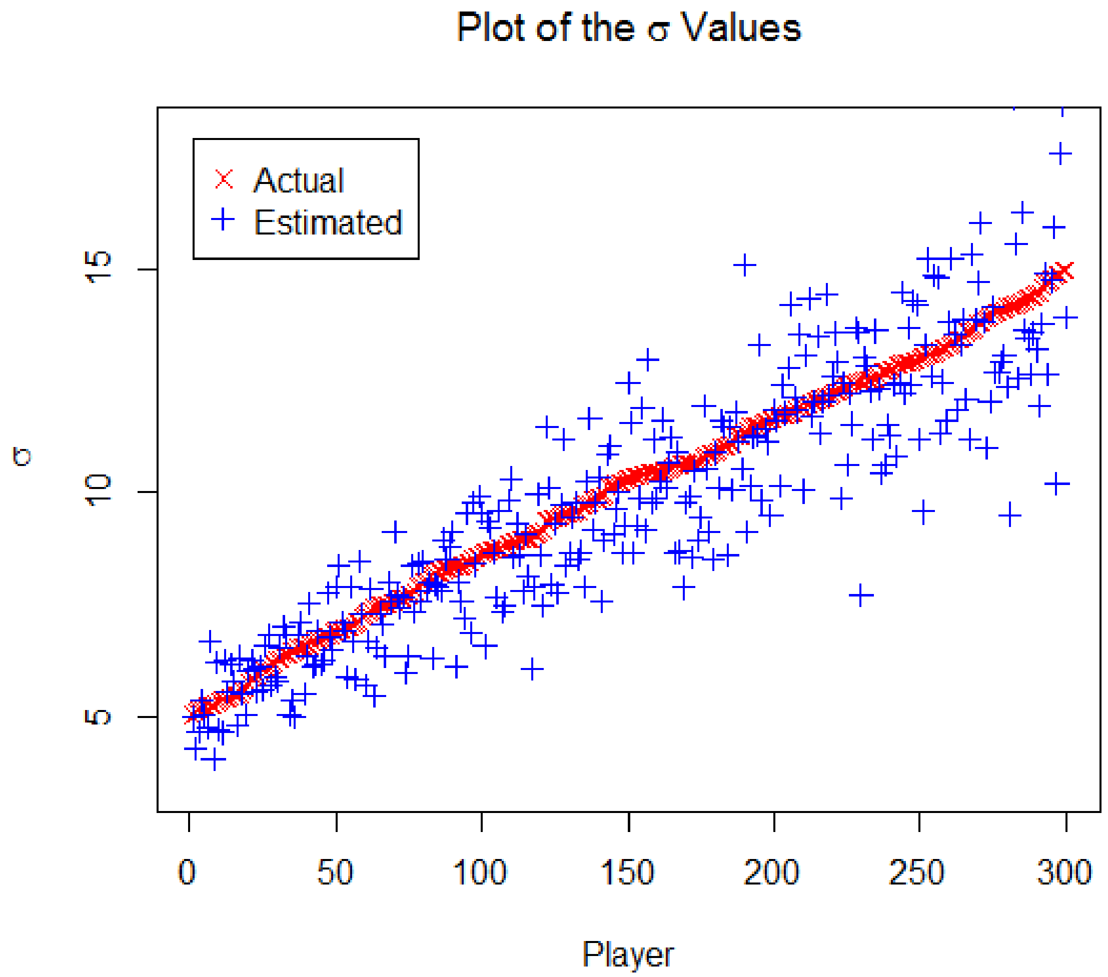

Figure 3. Plot showing results from our two-way linear model for the sigma variable, the standard deviation.

As expected the graph isn't as easy to predict as the other variables in the two-way model. There is a clear fan shape to the graph where we are predicting the standard deviation very accurately at lower values of the standard deviation and getting less accurate as the standard deviation increases, a very similar graph to figure 1, but with more variance at the start and a less clear fan shape overall.

4 Dynamic Model

4.0 Introduction to the dynamic model

So far we have averaged our player ability over the whole period in which players have competed equally and have the same player ability throughout the time period. However in a real life situation it is more likely players increase and decrease with form and confidence in their ability among other things. Take Tiger Woods for instance, a few years ago he was at the top of golf winning many tournaments and majors he was involved with then a scandal hit the press and he has not been as good since. If you averaged his performance around that time he's likely to appear quite good, whereas in fact he was exceptional before the scandal hit the newspapers and distinctly average after. This is where the dynamic model comes in. It analyses players form over time to see when they are at the peak of their game. This is important as if a player is very good entering the Ryder Cup but been distinctly average all season he is unlikely to show up on the static cases where as this will pick him up at the peak of his form and we can consider whether he deserves a place in a Ryder Cup team based on his latest form. To do this we will extend our linear model from before to take into account player ability over time rather than over the whole year. I would rather have a player in my Ryder Cup team that has been bad at the beginning of the year but hit peak of form at the right time, rather than a player who has been distinctly average all season.

4.1 Generating data

We defined how to generate data for the static case back in equation (3.0.3). To generate toy data for the dynamic case we use the same simulation as the static case but generate a new beta vector after every tournament played. Therefore the beta vector takes values that follow a random Gaussian walk around the previous player ability. For the toy data we can assume that tournaments do not take place at the same time so one tournament proceeds the other. We generate the data from the following:

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

4.2 Estimating alpha in the dynamic case

As before we need to estimate a course effect before we can directly compare the player abilities. But since there are enough players in each tournament we can assume that the variances in the beta values cancel out and we can use the same process as before for estimating our alpha values. This has been tested before and seems to be a very accurate assumption. It is true that calculating alpha from the static case generates more accurate values of alpha than using the dynamic data but it is close enough that we can proceed with the assumption we have made, as the algorithm before assumes static data was used. Even though we do lose some accuracy in calculating our beta values it is not significant enough to calculate the player abilities dynamically as this will overcomplicate our model without enough benefit to warrant it.

4.3 Estimating beta dynamically

To estimate our player abilities we use a Kalman filter. This method essentially contained the method of least squares recursively. Say we are given tournament scores that have already been adjusted for the course effect. Having applied the two-step ANOVA process to our model the estimation of beta essentially becomes an estimation of a 1-D Kalman filter for each player. The only thing we need to consider specially that isn't standard in this method is missing observations, ie. When a player doesn't compete in a tournament. For any tournaments that are missed we just use the beta value from the last time point for that player. We do have to consider that there will be a larger noise terms for these players however, whether their form decreases due to lack of match practice or whether they becomes better due to being rested. In our chosen coding package, R, we require using a package called 'dlm' (Petris, 2009) (dynamic linear model) to analyse it as we cannot complete the analysis of the standard R functions alone.

4.4 Interesting Dynamic graphs

In this section we will consider some interesting plots for the toy data in the dynamic case. This essentially will highlight how form can change for a player in a particular tournament. For instance in the Ryder Cup year we will see shortly that Rory McIlroy won back to back Majors. This will mean his form will dramatically increase (and hence have a lower beta value meaning a better overall player ability). We will look at a couple of graphs with similar static beta values but one which increases and decreases over the tournaments played and one that stays relatively similar over the time and see how form makes a difference to his overall performance. We will look at the beta value at the end of the time period which will weigh more recent tournaments more heavily on the player's ability to see which of our players in our toy data are in the best recent form, something that applies more to our real data but interesting nevertheless.

To simulate the data we have presumed the standard deviation has been drawn on a $U(7,13)$ distribution then presumed ${\sigma_{eval}}^2 = 1$ (Farmer, 2003) for the iterative updating process of beta. We have to also set the start point for our time series and we have set this to be 288 for all players so the first point on every graph reads 288 for this reason. The actual time points come at every time interval after this. Since this is toy data we can assume we the tournaments take place one after the other so there is no time overlap. This is a problem we will have to consider however when running the data for real players.

Below is our first graph, this graph is the most standard we will see for the dynamic result, however it shows what a dynamic graph looks like if someone were to have a lower standard deviation. Here we would expect to have low spread of our beta estimates over the time points and a fairly consistent estimation of our beta value.

The result of player 295 in our toy data of our dynamic model is shown below (Figure 4), it was picked out of the players as the player seemed to average the same player ability throughout the year and therefore highlighted the point about what we would see from a consistent player in our Ryder Cup qualification data.

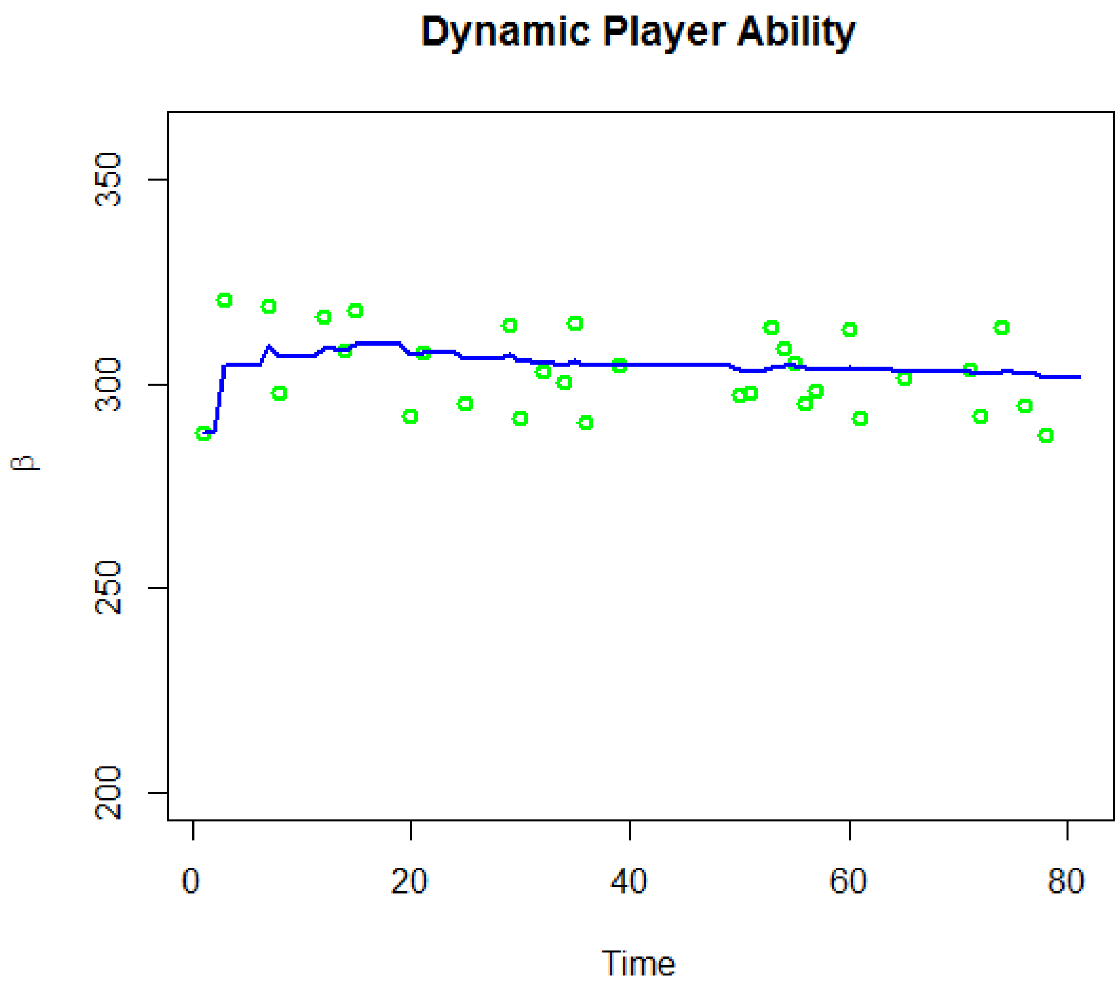

Figure 4. Plot showing results from one player's dynamic model for a fairly consistent season.

As we can see in Figure 4 above apart from the initial sharp rise in our beta estimate, due to us estimating the first point of our dynamic model to be 288 as alluded to before. We can see the results for our beta estimate over the time period are fairly close to our overall filter with equal spread. This is because the player is fairly consistent at a beta value of around 310 so we get a fairly straight line through all of the points. This is something we will look for later in the real data to indicate consistency of a player over the year, something we may prefer in a player going into the Ryder Cup as it may indicate an ability not to feel pressured. However, we will be looking for a better overall average player ability as this player looks to be fairly consistent at being 10 above our average mean level of player ability, where a higher value indicates a worse player.

Another interesting plot would be someone with a mixed year where they were good and bad throughout the season and how their beta estimate would change according rather than just estimate it to be an average of the two.

In Figure 5 (below) we have picked player 245 from our toy data and he exhibited a huge form increase something we would be delighted to consider in any Ryder Cup player.

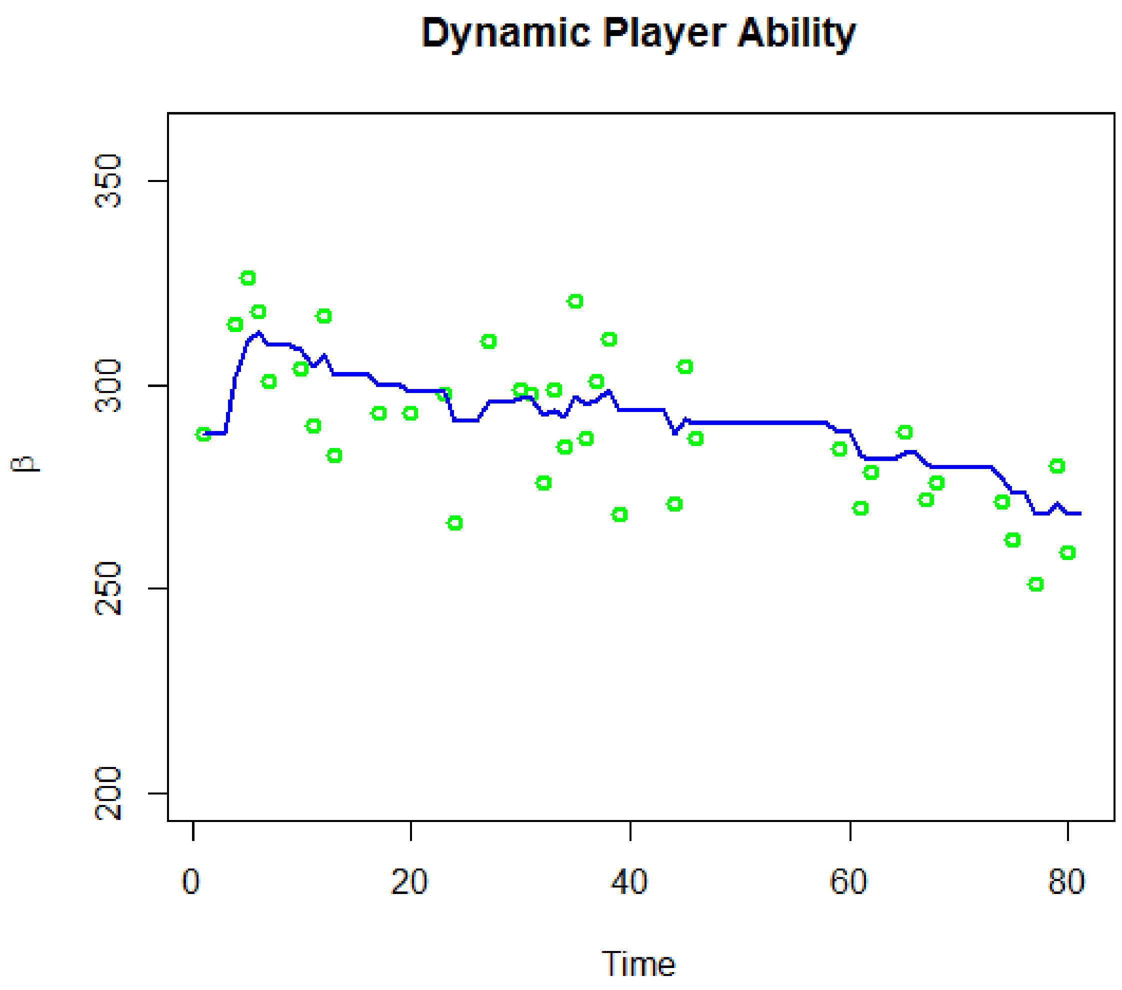

Figure 5. Plot showing results from one player's dynamic model for a form affected season.

Looking at Figure 5 above we see that (ignoring the initial value) we have values that are instantly fairly bad to start the qualification his beta values where first initially estimated to be over the 300 mark showing a very bad start to the season with the average being 288. He then had a dramatic form increase and improved consistently, possibly by taking some tournaments away from the tour to improve much more coming into the actual process. If you take his average beta value from the static case you would imagine it to be estimated at somewhere around 290 looking like he had an average season. Whilst this maybe the case on further inspection to his dynamic result we see this player has hit really good form and even had a beta value that was estimated to be around 255. A definite consideration to be in our team if this was run for the real data.

Below we look at a graph showing the players final dynamic form to end the season to see which players have done the best from the toy data. This allows us analyse players form going into the selection process all at once by looking at their weighted form all at the same time. Below is Figure 6 that shows all the players weighted form that has been put into ascending order.

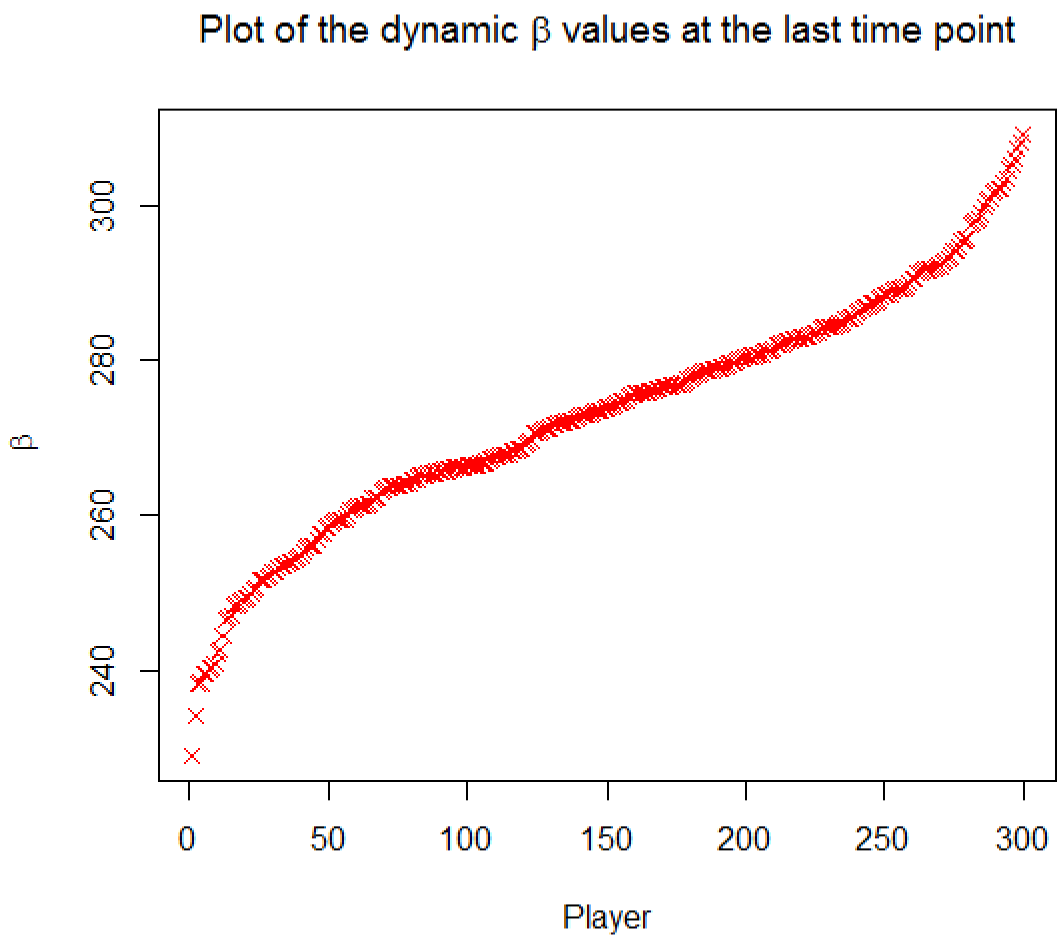

Figure 6. Plot showing results of the weighted beta estimation at the last time point of our dynamic model.

We note here that due to the volatile nature of our generating model we got some beta values that can spiral one way a bit more than we traditionally would like. However this allows us to analyse everything we would expect from the real players of the Ryder Cup in one simulation rather than running the data many times and generating more scores to highlight the points we may see having run the real data.

Here we would look into more detail at the players in the height of form of which there are a fair amount in this model. We would also look at players who have a good static beta value but do not appear as much on the dynamic case, this could indicate a drop in form over the year and we would have to analyse that before we pick them for our overall team.

We have already looked at one player in the height of their form coming into the selection process and analysis for the other players would be similar apart from the slight drop in forma at the beginning of the year perhaps so for the toy data we will not run all of these instances to make the same conclusions as they have no real impact on what this project truly wants to look at which is the Ryder Cup team selections.

4.5 Conclusion to toy data

Overall in this section we have looked at three models all of which will be required to get as much information as possible before making our final decision of our Ryder Cup teams for Europe and America. We have also seen some advantages and disadvantages to each model, their limitations and drawbacks. The main drawback we have seen is with the Plackett-Luce model. The problem with this model is that we lose information in converting scores into a ranking system which would be more helping in making our decision about the teams overall. This is clear from comparing Figure 1 to Figure 2, Figure 1 formed a fan shape around as the gamma variable grew, where gamma measures the overall player's ability. However when processing the beta values for our two-way linear model we clearly saw a much more precise estimate overall, with no fan shape. Briefly comparing the dynamic model and static model of the two way ANOVA models, the difference isn't that high as will be alluded to later.

5 Real Data

5.0 Introduction to using real data

Now we have analysed the toy data and found that our models work for this data and explained why we are using each one we will apply it to real data instead. There are a few differences here mainly based on conditions of the formula, difficulties we will have to overcome to process our model. We will run the same models as before and try to draw conclusions for our Ryder Cup teams from them and mention any problems we may encounter.

5.1 Data collecting

Firstly we have to consider a few issues that we have in real data that we would not have in the toy data. I collected data (Anon., 2015) from all of the tournaments that qualified players for the Ryder Cup teams for 2014 (Anon., 2014) (Anon., 2014). This involved collecting data on 1404 players over 78 tournaments. One special mention is for the world match play tournament where they use a different scoring system and even though it does qualify players for the Ryder Cup teams we will not consider this tournament as it is based on data we cannot use as well, even though we would use it for the Plackett-Luce model we cannot use it for our linear model and the traditional score is not recorded due to the ranking system they use. Another problem is that many of the 1404 players only compete in one or two tournaments as a one off throughout the year. People who played very few tournaments would not be in our Ryder Cup selection process anyway. It is also worth pointing out that the data we gather from them is very little as there is not much to analyse. We therefore retain any player who has played in more than 15 or more tournaments, due to the fact this is amount of tournaments you have to compete in to retain your card for the following year. This reduces the players to 323. This will cause problems mainly in the

Plackett-Luce model where for instance we may have lost the person who finished 6th in one tournament and this is something we will address later.

Also as discussed in our introduction another problem is the 'cut' where a number of the field are removed halfway through a tournament, this is usually about half of the starting players in a tournament. The problem this causes is that their score is recorded over 2 rounds and not 4, this is clearly not acceptable for our ANOVA model where they will appear to have got ridiculously low scores, where in fact they are so far behind the field they have been dropped at the halfway point. We cannot discount their score when they are cut, otherwise we may mask a lot of the players very poor performances just because he didn't make the cut and wouldn't give a true value of the players ability. To adjust for this one way we can deal with this is to simply double the other players total that have missed the cut, this is not without drawbacks such as if the weather greatly improves or gets worse in the final two rounds. It does not account for the fact the latter half of the course could be a lot harder than the rest of the course. We accept these drawbacks for the processing of this data as some of it will be removed through the course effect value but it is a point worth noting in our final results.

Another problem to note is tournaments that have been cut short due to adverse weather usually consisting of only 3 rounds. These tournaments have not been included in this model. Most of these tournaments where the less known ones and there were not many over the year of the Ryder Cup 2014 selection process.

5.2 Plackett-Luce applied to real data

The first model we consider is the Plackett-Luce model. This model as referred to in chapter 2 ranks players by their final position in the tournament. Now the problem, since we have removed players who play in 5 tournaments or less, is that our rankings do not go numerically, some values are missing due to removed data. We can sort this issue by producing new rankings for the data which players are ranked numerically out of those that played. So for instance if the 6th place player played in 5 or less tournaments and was removed, the 7th ranked player would become 6, the 8th ranked player would become 7th and so on and so forth. This process is repeated for all 78 rankings until we obtain a ranking system of those players left in our dataset. We also have a problem that is more likely to occur in a real dataset than in our toy data. The problem is that the Plackett-Luce model will not work if a player finishes in first or last in all tournaments he competed in, after the other players have been removed. We will adapt our model to search for this unlikely outcomes before running the data. Luckily this didn't occur. If it did we would have had to remove the player from this dataset and run the data with the remaining players as Hunters algorithm will not proceed correctly if this occurs.

In Figure 7 (below) we can now go on to produce a plot of the real player abilities determined via this method and have a closer look at which players came out as the best players over the year for both the American and European teams just looking at this model. This will be used for comparison later.

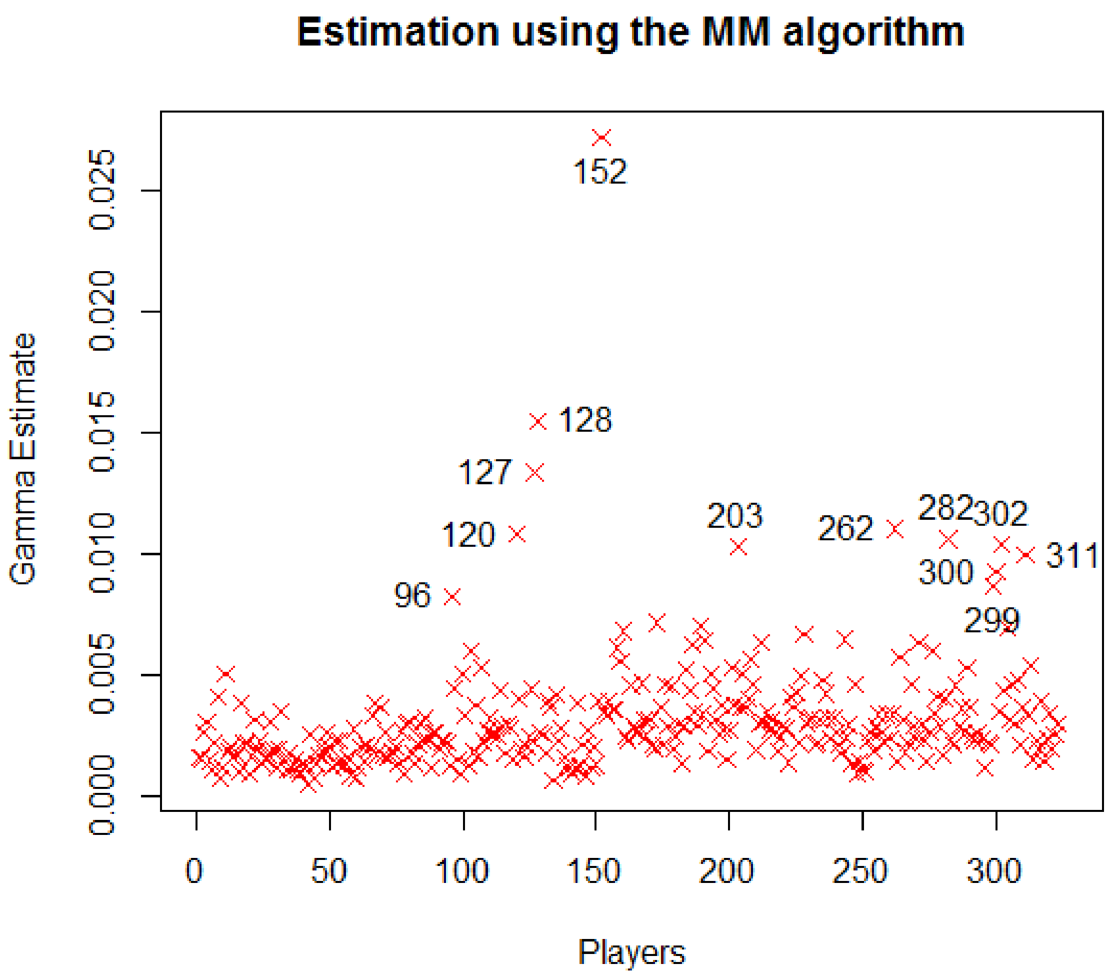

Figure 7. Plot showing results from our Plackett-Luce method through the MM algorithm.

Here we can see a highlighted group of players who seem to have a higher gamma estimate than others. Player 152 is Adam Scott and this very high value indicates he has been the best player in terms of final finishing positions in the Ryder Cup this year. However, since he is Australian he is not considered eligible for the American or European teams. Excluding these non-eligible players the list of players from best to worst for this model are:

128(E), 127(E), 262(A), 120(A), 282(A), 203(E), 311(A), 300(A), 299(A), 96(E), 173(A), 189(E), 304(A), 160(A), 228(A), 243(A), 191(A), 271(A), 186(A), 158(A), 103(E), 276(A), 159(A), 313(A), 289(A), 107(E), 184(E), 193(E), 100(E), 205(A), 227(A), 235(A), 308(A), 167(A), 247(E), 306(A), 176(A), 268(A), 209(E).

Where E means the player is European and A means the player is American, the highest 8 for each team will be on our list for our team from this model and hence the teams are

European	**American**
Sergio Garcia	Dustin Johnson
Rory McIlroy	Phil Mickelson
Justin Rose	Matt Kuchar
Henrik Stenson	Charley Hoffman
Graeme McDowell	Bubba Watson
Luke Donald	Bill Haas
Martin Kaymer	Chris Kirk
Francesco Molinari	Jim Furyk
Ian Poulter	Brendon Todd
Joost Luiten	Ryan Moore
Victor Dubuisson	Zack Johnson
Lee Westwood	Harris English

5.3 Static Linear model applied to real data

Again we have stripped the data to people who have competed in 15 or more tournaments. We have doubled the score of every player who missed the cut and then run it through the algorithm described in chapter 3. We have then obtained graphs showing the beta, alpha and score standard deviation values. We have then plotted these to see which players come out with the highest score and used this model to predict both the European and American teams for this model which we will use for comparison later.

Below is the player ability estimates for the scores using our two-way model. Figure 8 (below) shows the player abilities out of our dataset that has been reduced down to players that have played in 15 or more tournaments.

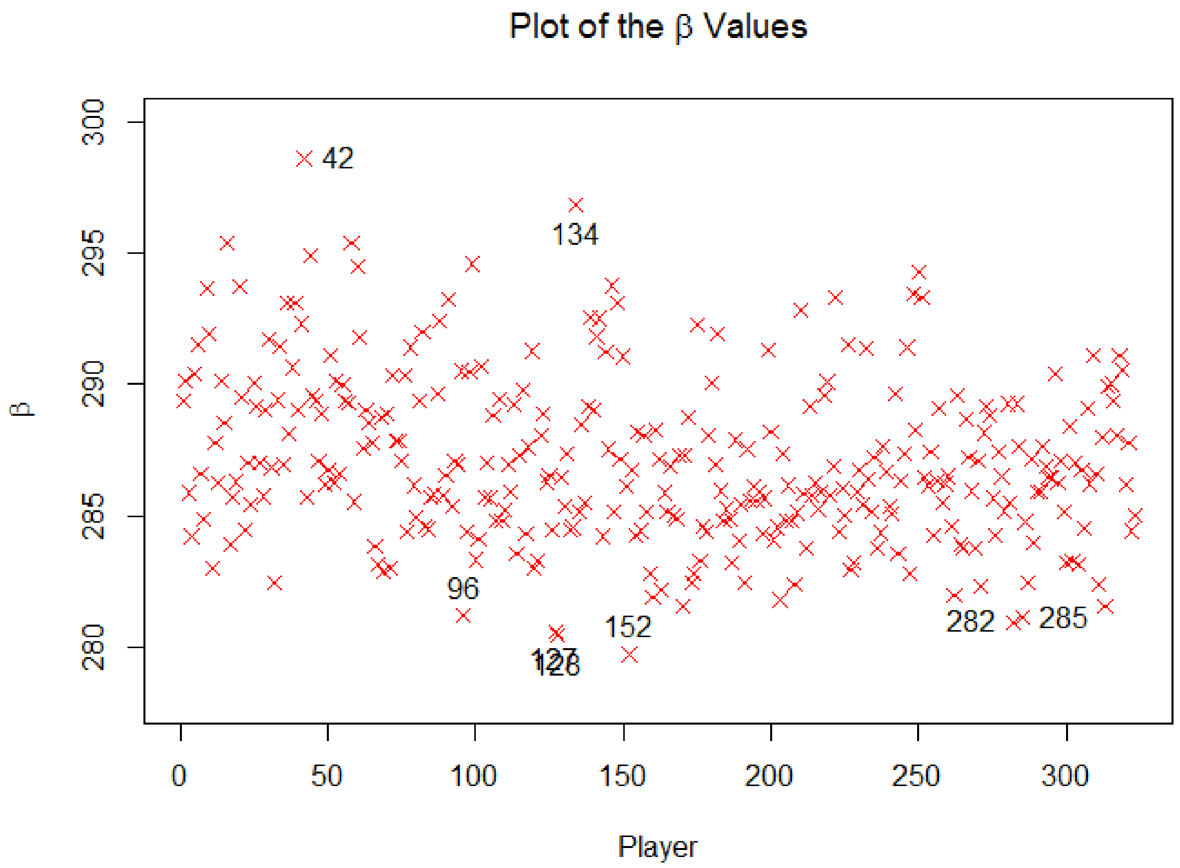

Figure 8. Plot showing results from our static model for beta for real data.

From Figure 8 we can see that player ID 152 is the best player estimate player ability in our static case. This is closely followed by players 127 and 128. Other players that are high up are 282, 285 and 96. On the other hand, the worst players are players 42 and 134. Player 152 is Adam Scott but since he is an Australian he cannot play for either Ryder Cup team so we look at players 128 and 127 who are Sergio Garcia and Rory McIlroy both of whom can represent Europe and in our analysis of the static case, providing they are have reasonable standard deviations they will probably make our European team. The 34 player's ID who are best after them are as follows(not including ineligible players); 282, 285, 96, 313, 170, 203, 160, 262, 163, 271, 311, 173, 191, 287, 159, 247, 174, 69, 227, 71, 120, 67, 304, 228, 187, 300, 176, 121, 100, 114, 243, 269, 265, 66. These players can be found below in the appendix. These 34 players, after Adam Scott could all potentially be selected for the Ryder Cup team if they do not have a standard deviation which is too high. Therefore our overall top list in terms of player IDs are as follows:

127(E), 128(E), 282(A), 285(A), 96(E), 313(A), 170(A), 203(E), 160(A), 262(A), 163(A), 271(A), 311(A), 173(A), 191(A), 287(A), 159(A), 247(E), 174(A), 69(E), 227(E), 71(E), 120(A), 67(E), 304(A), 228(A), 187(A), 300(A), 176(A), 121(E), 100(E), 114(E), 243(A), 269(A), 265(A), 66(E).

Where E means the player is European and A means the player is American. Unless the standard deviation is too high for the highest 8 golfers on this list from each team they will make our team for this model.

We have then done the same for the alpha values to see which tournaments where the hardest throughout the qualification process. Even though this has no bearing on the Ryder Cup selections (for our model) we do note that it has an effect on some decisions such as which are most like Gleneagles (location of Ryder Cup 2014) and this is shown in Figure 9.

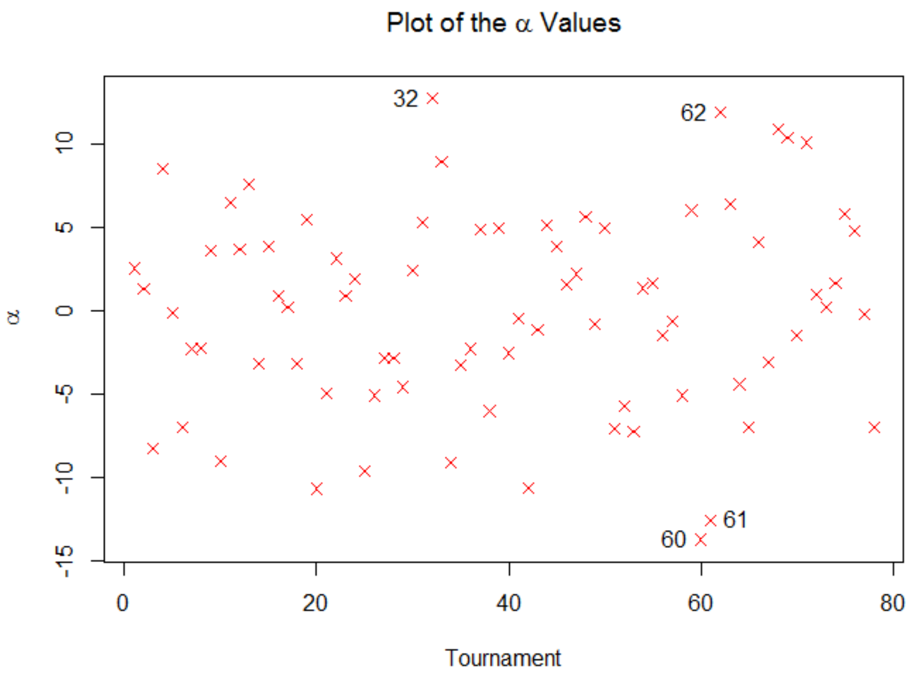

Figure 9. Plot showing results from our static model for alpha for real data.

The above, Figure 9 shows that the hardest tournaments where tournament ID's 32 and 62 and could be caused by a multitude of factors such a field strength, course difficulty etc. that was previously alluded to in chapter 3. Conversely tournament ID's 60 and 61 where the easiest again probably for similar reasons. These tournaments 32,62,60,61 correspond to the 2013 Masters, Open de Espana, Shriners Hospitals for Children Open and Sony Open in Hawaii respectively. The reason that 32 and 62 are the hardest looks to be specifically because in the Masters the field strength is really strong, added to some adverse weather conditions that year makes the tournament the hardest according to our estimates. The Spanish Open seems to have just been a tricky course for the players that year, nothing wrong with the weather and Sergio Garcia scored 4 over par so it cannot have been that easy, the overall field strength didn't seem to high either. The main reason tournaments 61 and 60 seem to be on the easier side is down to one major factor, field strength, the players playing those days do not have especially good beta estimates and yet the winning score was still quite low, therefore the courses doesn't seem to be too difficult.

We now plot the same data for the score standard deviation. This is important as we can see which players can be more consistent allowing them to keep their cool under pressure situations for instance and could have some bearing on our selection if their overall player abilities are similar. Figure 10 (below) shows us the standard deviations and we will look to see if any of our original list of best players for the static two-way ANOVA table have come too high on this list. The players we have the most concerns about consistency are marked on the graph.

Looking at Figure 10, below, we see that the players with a potentially concerning level of standard deviation are players with IDs of; 2, 20, 20, 42, 78, 102, 146, 172, 197, 220, 224, 296, 304. Comparing this to our list from our player ability estimates we notice that only one of these, 304, appear players on our original ordering of our beta ability estimates, since he is not in the top 12 American players for our estimates of player abilities, we can conclude this models overall team selection to be the best 12 players from each team from our original player ability list made from calculating our beta estimates.

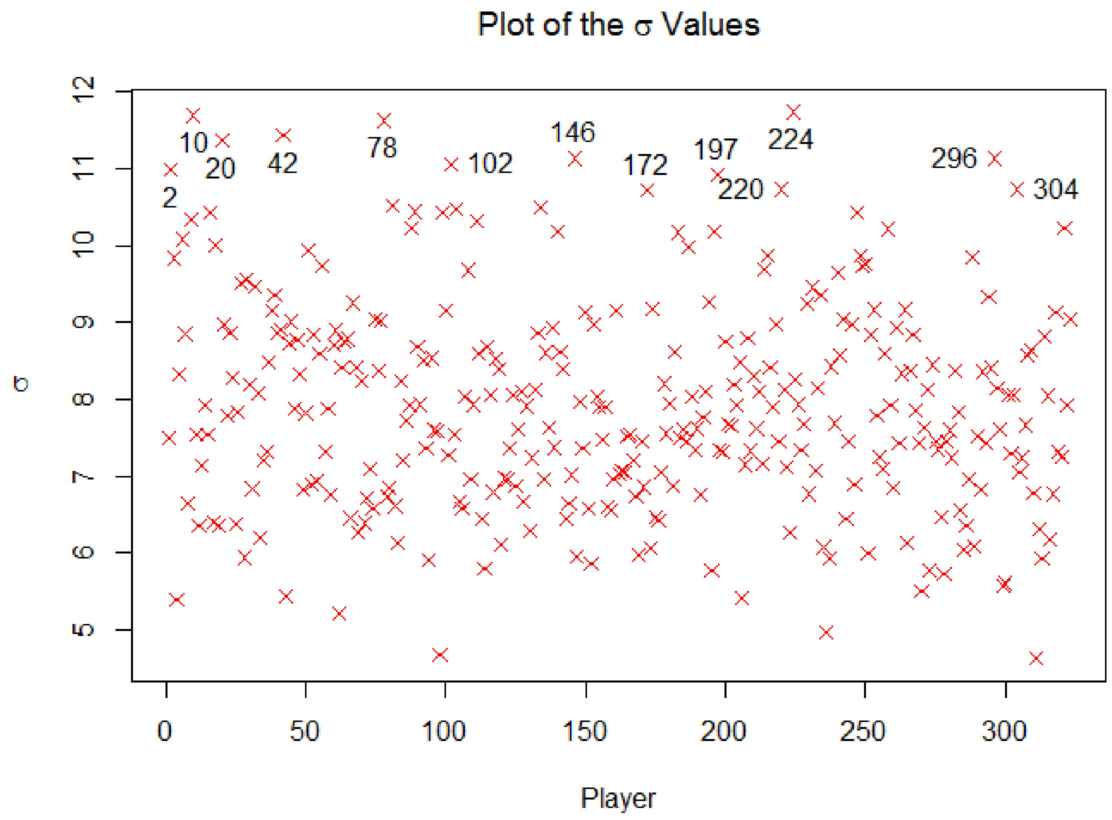

Figure 10. Plot showing results from our static model for sigma for real data.

Therefore our overall conclusion for this model is to select the following players:

Europe	**America**
Sergio Garcia	Matt Kuchar
Rory McIlroy	Ricky Barnes
Henrik Stenson	Jerry Kelly
Justin Rose	Charles Howell-III
Seve Benson	Brendon Todd
Russell Knox	Dustin Johnson
Simon Dyson	Brian Harman
Ross Fisher	Jimmy Walker
Rafael Cabrera-Bello	Charley Hoffman
Joost Luiten	Chris Kirk
Paul Casey	Harris English
Romain Wattel	Robert Streb

It is interesting to note here that Phil Mickelson, 120, is missing from the American team since he has been a star performer for them over the last Ryder Cup meetings.

5.4 Dynamic two-way model applied to real data

We now look at players form going into the Ryder Cup selection process and take note of players with exceptional form going into the Ryder Cup. For this model the data had to be ordered in time order as it is computed as a time series. Apart from that point the data is calculated in a very similar way as it is to the static case as alluded to in chapter 4. We will show the plots of a few more interesting players that have better form going into the Ryder Cup selection. We will then consider whether these players deserve a spot for Europe or America solely based on form going into the Ryder Cup and see if this differs significantly from our other two models which will analyse overall later. Therefore, we will look at all the players last dynamic score (score based more heavily on recent results than the others) to see the players with the lowest score and therefore the best overall ability level.

Figure 11 (below) shows this last weighted estimate of player ability of all the players and after we will see the best 12 players from each team going into the Ryder Cup.

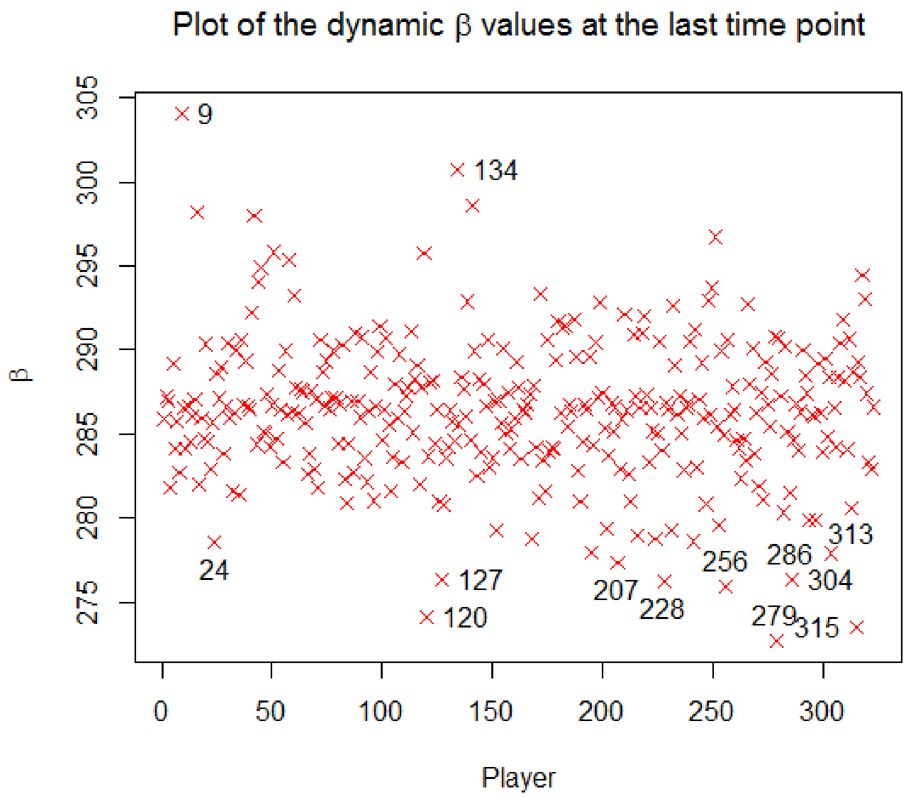

Figure 11. Plot showing results from our dynamic model, weighted estimates of beta at the last time point

Here we can see the players who have the lowest weighted value of player ability the last time point as use this to calculate Ryder Cup potential based on their form, so we note that players 279, 315 and 120 look particularly impressive. Therefore we will take a closer look at 2 of the player's dynamic models. The reason we will not look closer at player 120 is that Phil Mickelson came up before in our Plackett-Luce model and wasn't too far off making the cut in the static two-way ANOVA model, therefore we will probably make the team off this basis without the need to look closer at his dynamic model. We also note that we do not need to reorganise tournaments that run at the same time as we can treat them the same as if they directly run one after the other. This is because no player can play in the two ongoing tournaments at the same time, especially as they usually take place in a different country or state. Player 279's dynamic model is shown below in Figure 12

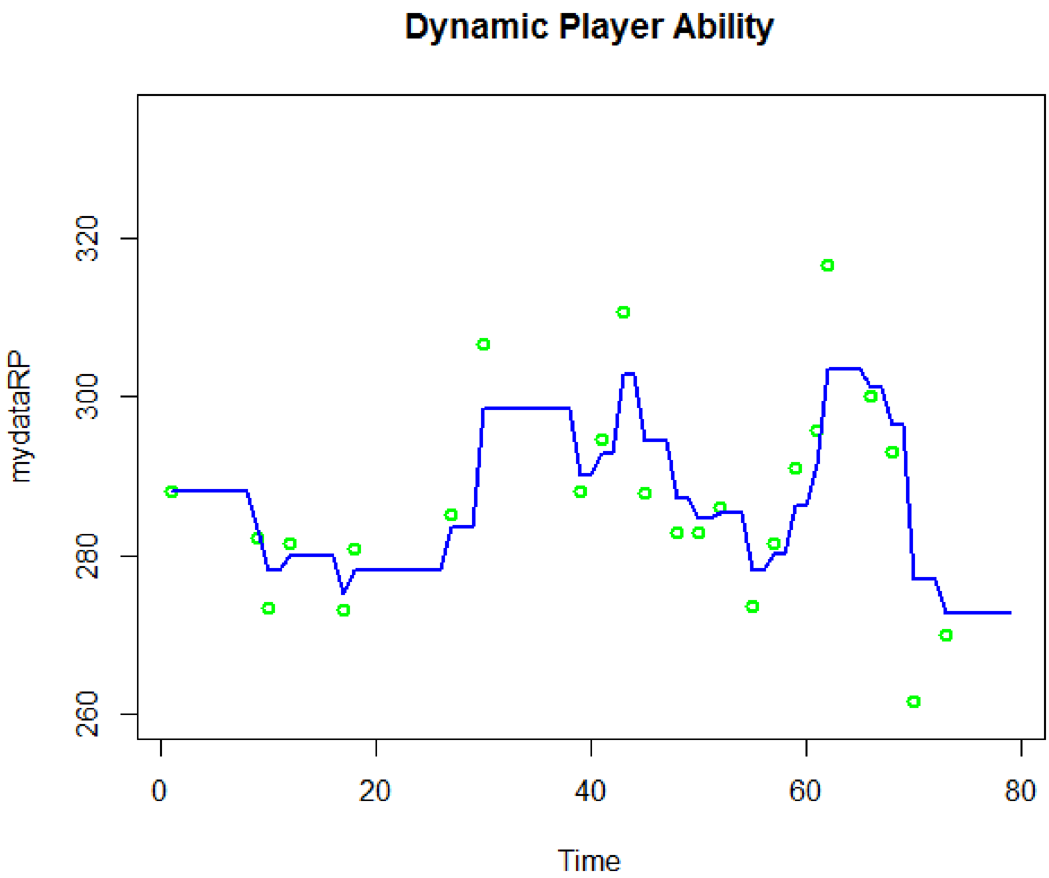

Figure 12. Plot showing results from our dynamic model, weighted estimates of player 279.

In Figure 12 above we can see that player 279, Kevin Kisner who is American, has fairly good form to begin with obtaining an estimated beta value of around 280 through the first 20 tournaments. Remember that the first time point of 288 is an initial value picked because it is par and can be ignored in the overall scheme of things. The players form then drops quite dramatically towards the midpoint of the year and even though the player obtained some average (288) beta estimates the weighted model shows that he went through fairly bad form through this period and this form, even though it got better briefly, continued until the 70th tournament until there was a sharp increase in form hence why he was analysed further. Overall I wouldn't consider picking this player 2 very good scores are skewing out dynamic data down which is showing the player is going through very good form towards the end.

The other player we chose to consider in further detail is the player with the ID 315 whose dynamic model is shown in Figure 13 below.

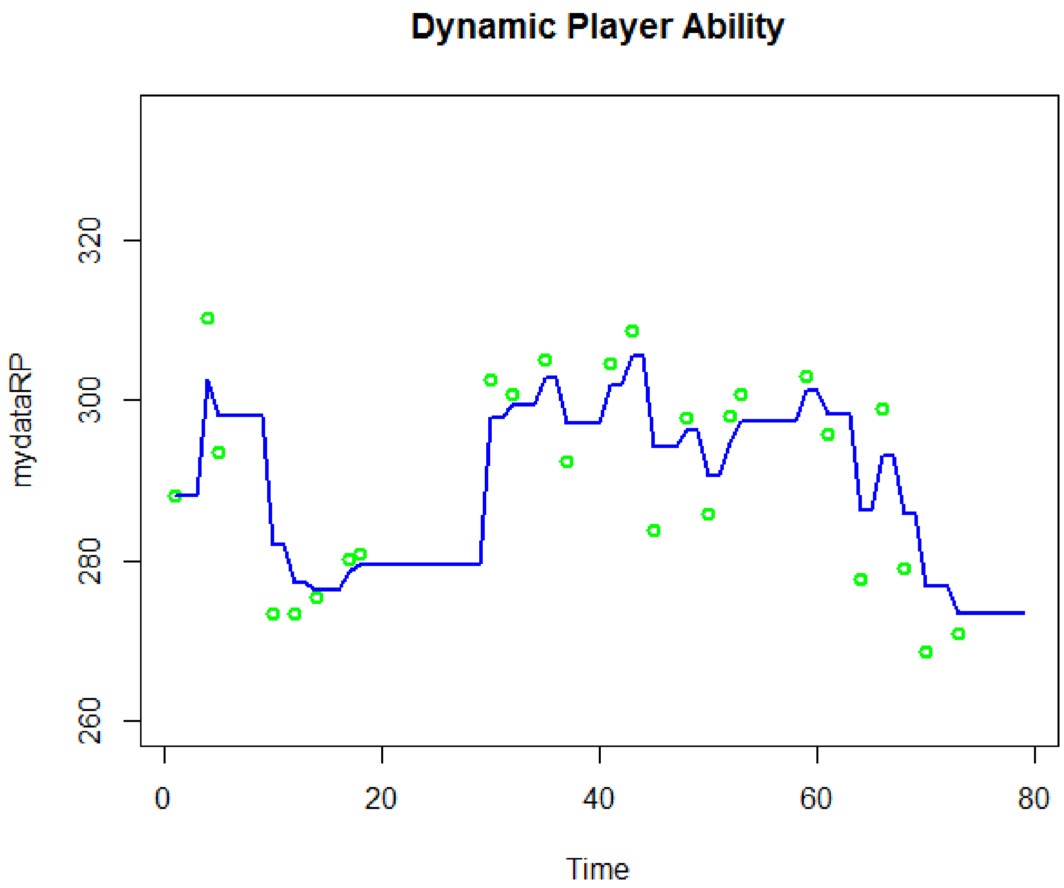

Figure 12. Plot showing results from our dynamic model, weighted estimates of player 315.

Again ignoring our initial value of 288 we can see the player started off the season in fairly good form overall even if there looks to be a bad score in one particular tournament overall. He then doesn't play for 10 tournaments or so before coming back showing many dips and rises in the form of which overall seem to be averaging around 288 which is the sign of an average player. However from tournament 60 onwards the player just gets better with the tournaments they play.

Overall, I would consider this player for the Ryder Cup team. The increase in form is undeniable and for most of the year seemed to be in fairly good form. There is a bad patch of form half-way through the year highlighting the reasons why this player didn't come up in other models and why it is important to do this model.

Therefore the Ryder Cup teams would only change due to this model based on player 315, Kyle Stanley who is American.

5.5 Comparisons of team outcomes

In this section we will discuss the outcome of the teams in all the tournaments for the Ryder Cup and try to settle on an overall team for the Ryder Cup for both Europe and America. For this we have to consider the pros and cons to each model. The flaws of Plackett-Luce are the most apparent, the fact that, as alluded to earlier in chapter 3, a player can be 1 stroke or 20 strokes behind a player and as long as they are second on the ranking system it have no difference is a clear flaw. It is one of the main reason there seemed to be flaws in the overall selection process with prize money although at least this data seems to take into account how bad an off day is for a player if you missed out of prize money by a place or finished last it would make difference to most of the selection process for most teams, whereas it does make a difference to the Plackett-Luce models the clear flaws are apparent.

There are some pro points to the actual selection system and the Plackett-Luce in the static ANOVA model. But it still lacks looking at the form of a player. It would not consider if a player has hit a very bad patch of form just before the Ryder Cup as an average is taken across all tournaments.

This is proved on again in the dynamic ANOVA model, but again there are flaws such as not using a dynamic update to the beta in the iterative scheme but using the same one as the static case is the stand out flaw, even if it has little impact on our overall outcome. Overall it is probably the best model to use is the dynamic model. However not many people stood out that were not standing out before. I would include player 315 in the American team solely based on form but then backtrack to the other models for our picks. As mentioned before even though player 120 was not analysed further he seemed to have good form and was picked up in both other models so he definitely makes the cut for the American team.

Overall looking at all the team outcomes from all three models, taking some output from them all would be warranted. We would weight more emphasis as the models get progressively better to our overall decision. If it was up to me to captain both teams the teams and I got 12 picks. The ones I would choose are:

Europe	**America**
Sergio Garcia	Kyle Stanley
Rory McIlroy	Phil Mickelson
Justin Rose	Matt Kuchar
Henrik Stenson	Jerry Kelly
Martin Kaymer	Brendon Todd
Joost Luiten	Charley Hoffman
Seve Benson	Chris Kirk
Russell Knox	Harris English
Graeme McDowell	Dustin Johnson
Luke Donald	Charles Howell-II
Simon Dyson	Brian Harman
Ross Fisher	Bubba Watson

These teams are my opinions based on the models that have been run. I have weighted more towards the static two-way model than the Placket-Luce as it retains more information of the players by using the scores over the overall rankings. However I have considered all three models overall and the reasons for including the players or not based on the dynamic model where all mentioned in the previous section.

6 Conclusion and Improvements

6.1 Improvements

There are clearly some improvements we could make to a few of the models and things we could go further with if we had the time. The first major improvement we could make is one to do with what we did with the cut in the two-way ANOVA model in the real data case. We used a prediction that doubled a players score based on the score he had when he was cut. We instead could use a model such as adding the worst score over the next two rounds obtained by the remaining field as a prediction, as clearly they were not performing as good as the lowest scoring player before the cut so that seems a sensible model to use.

Another investigation we could make is to look at similar courses to that which is played at Gleneagles, the location of the 2014 Ryder Cup, and weight those score more heavily in our model so we could pick players that are better suited for that course. For instance if the fairways are long we would want players who could drive further, were as if the greens are short and narrow we would want a player who is more accurate in their strokes. Maybe it's both in which case we want a compromise between the two and we could look at courses similar to this over the past year to get an indication of how well they would do at the Ryder Cup and use that to influence our overall decision on our teams.

Another compromise we made is to assume alpha is constant for players throughout a tournament. Maybe some players struggle a lot more in adverse weather and maybe some players best results come in those conditions. The best way to take this into account in our data is to look at the weather forecast for the Ryder Cup for instance and see if selections of players that have played on courses of similar difficulty and weather conditions to see if their scores do not fluctuate if the forecast is bad and similarly if they perform better when the weather is very nice. We could use this to influence our decision on the teams if we could investigate further.

Another thing we could do as alluded to in chapter 4 is to use a dynamic estimate of beta, but as explained then the difference wasn't too much, maybe it was enough to push a player over the edge into our team on their form if we had gone ahead and used the more complicated model however.

Overall, these things are very hard to quantify though and wouldn't have a huge effect on our overall outcome. We could in fact look at more statistics that are available such as accuracy of shots from the tee and successful putt distances and how far away from the hole they were on their approach shot. We could then create a very specific profile of each golfer we could then use to relate to the course at Gleneagles and make a very informed decision on our team, by seeing how good a player should perform a course like the one used for the Ryder Cup, also known as a player compatibility coefficient.

6.2 Conclusion

As mentioned at the end of chapter 5 the teams I would pick are (next to the actual team picked (Anon., 2014) (Anon., 2014)):

Europe	Europe (actual)	America	America (actual)
Sergio Garcia	Sergio Garcia	Kyle Stanley	Rickie Fowler
Rory McIlroy	Rory McIlroy	Phil Mickelson	Phil Mickelson
Justin Rose	Justin Rose	Matt Kuchar	Matt Kuchar
Henrik Stenson	Henrik Stenson	Jerry Kelly	Jordan Speith
Martin Kaymer	Martin Kaymer	Brendon Todd	Patrick Reed
Joost Luiten	Jamie Donaldson	Charley Hoffman	Zach Johnson
Seve Benson	Thomas Bjorn	Chris Kirk	Hunter Mahan
Russell Knox	Ian Poulter	Harris English	Keegan Bradley
Graeme McDowell	Graeme McDowell	Dustin Johnson	Webb Simpson
Luke Donald	Lee Westwood	Charles Howell-III	Jim Furyk
Simon Dyson	Stephen Gallacher	Jimmy Walker	Jimmy Walker
Victor Dubuisson	Victor Dubuisson	Bubba Watson	Bubba Watson

The differences between my selection of players and the selection actually made is more apparent in the American team. As alluded to in the introduction there was a few strange picks mainly as wild card selections by the American captain. For instance Hunter Mahan was actually in bad form entering the Ryder Cup in 2014 but was picked as he played well throughout the year, something out dynamic model would have highlighted. Another questionable decision was the inclusion of Keegan Bradley whom was only picked to partner Phil Mickelson In the fourballs, where as a selection such as Kyle Stanley who was in form may have worked. The European team is more similar to my selections this could be the fact they don't just work off prize money alone, unlike the American's and an actually ranking system would be a fairer and better method than using just prize money won over the year. The main differences come in the captains picks. The reason they have picked different to me is I have used statistics alone without considering factors such as experience, which seemed to play a heavy role in the selections such as 41 year old Lee Westwood who usually has good rounds at the Ryder Cup.

Overall I think the Ryder Cup selection process is flawed for many reasons. This is mainly because money lists rather than player skill are used to select the teams. This means a European golfer playing in America for a long portion of the season is unlikely to get picked for the European team as he has not competed in enough tournaments in Europe to earn the prize money of someone who is competing and finishing in average positions all season long. The fact someone can earn double the money from winning a playoff and this having a huge bearing on whether he is selected for the Ryder Cup team is also very flawed. There should not be so much emphasis on a one hole shootout to determine your Ryder Cup spot. The data was noisy enough over many tournaments consisting of 72 holes each without double money being on offer for one hole at one particular course that can favour one player more than another depending on his playing style.

Overall I think the whole selection process needs a huge rework, the fact players can get in because a player they are good friends with have been elected, such as Phil Mickelson did in the actual 2014 selection process, therefore denying a player who has rightfully earned there spot in all but prize money is another thing that needs to be stamped out of the game, even if that means a slightly worse performance on one round of fourballs, it could potentially improve the scores in 3 other rounds of golf.

In fact the selection process for the American team has changed as they must have thought it needed a rework too. For the 2016 Ryder Cup the selection process has changed (Anon., 2015) to a process which is still based on prize money earned but is not based more on tournaments which are likely to attract the major players such as the Majors, Players Championship events and World Golf Championship events and runs over a longer time period. The process now allows the captain to pick 4 wildcards, 3 of which will take place earlier than the final pick, presumably to pick a player who is in the best form going into the 2016 Ryder Cup.

7 Bibliography

Anon., 2014. *BBC.* [Online]
Available at: http://www.bbc.co.uk/sport/0/golf/29033517
[Accessed 28 April 2015].

Anon., 2014. *BBC.* [Online]
Available at: http://www.bbc.co.uk/sport/0/golf/29035788
[Accessed 26 April 2015].

Anon., 2014. *Ryder Cup.* [Online]
Available at: http://www.rydercup.com/usa/teams/2014-ryder-cup-team-usa-qualification-process
[Accessed 6 January 2015].

Anon., 2014. *Ryder Cup.* [Online]
Available at: http://www.rydercup.com/europe/teams/2014-ryder-cup-team-europe-qualification-process
[Accessed 11 January 2015].

Anon., 2015. *OWGR.* [Online]
Available at: http://www.owgr.com/events
[Accessed 12 Janurary 2015].

Anon., 2015. *Ryder Cup.* [Online]
Available at: http://www.rydercup.com/usa/news/2016-ryder-cup-team-usa-qualification-process
[Accessed 28 April 2015].

David, G., 2009. *Daily Mail.* [Online]
Available at: http://www.dailymail.co.uk/news/article-1231857/Tiger-Woods-crash-mystery-deepens-revelations-BOTH-passenger-windows-car-smashed-in.html
[Accessed 27 April 2015].

Farmer, C. J., 2003. *Edinburgh University.* [Online]
Available at: http://www.inf.ed.ac.uk/publications/thesis/online/IM030033.pdf
[Accessed 15 October 2014].

Hunter, D. R., 2004. *Project Euclid.* [Online]
Available at: http://projecteuclid.org/download/pdf_1/euclid.aos/1079120141
[Accessed 13 October 2014].

Petris, G., 2009. *cran.r-project.* [Online]
Available at: http://cran.r-project.org/web/packages/dlm/index.html
[Accessed 11 Febuary 2015].

8 Appendix

Below is a list of all players with their corresponding player ID for the reduced player list of 323, those who have competed in 15 or more tournaments.

1 Adrian Otaegui
2 Adrien Saddier
3 Alejandro Canizares
4 Alexander Levy
5 Andrea Pavan
6 Andreas Hart
7 Andy Sullivan
8 Brendon de Jonge
9 Brinson Paolini
10 Carlos del Moral
11 Charl Schwartzel
12 Chris Doak
13 Craig Lee
14 Daan Huizing
15 Damien McGrane
16 Daniel Brooks
17 Danny Willett
18 Darren Fichardt
19 David Drysdale
20 Dawie van der Walt
21 Eddie Pepperell
22 Edoardo Molinari
23 Emiliano Grillo
24 Fabrizio Zanotti
25 Francois Calmels
26 Gareth Maybin
27 Garth Mulroy
28 Gary Stal
29 George Coetzee
30 Graeme Storm
31 Gregory Havret
32 Hennie Otto
33 James Heath
34 James Kingston
35 James Morrison
36 Jamie McLeary
37 Jbe' Kruger
38 Jens Dantorp
39 Joachim B Hansen
40 Johan Carlsson
41 John Daly
42 John Parry
43 Jorge Campillo
44 Jose-Filipe Lima
45 Justin Walters
46 Kevin Phelan
47 Kristoffer Broberg
48 Lee Slattery
49 Lucas Bjerregaard
50 Magnus A Carlsson
51 Marco Crespi
52 Matthew Baldwin
53 Matthew Nixon
54 Maximilian Kieffer
55 Mikael Lundberg
56 Morten Orum Madsen
57 Nacho Elvira
58 Niclas Fasth
59 Oliver Fisher
60 Peter Lawrie
61 Peter Whiteford
62 Ricardo Santos
63 Richard Finch
64 Richard Sterne
65 Robert Rock
66 Romain Wattel
67 Ross Fisher
68 Scott Jamieson
69 Seve Benson
70 Sihwan Kim
71 Simon Dyson
72 Simon Thornton
73 Simon Wakefield
74 Soren Hansen
75 Steve Webster
76 Stuart Manley
77 Thomas Aiken
78 Thomas Pieters
79 Tom Lewis
80 Tyrone van Aswegen
81 Tyrrell Hatton
82 Victor Riu
83 Wade Ormsby
84 Alvaro Quiros
85 Anthony Wall
86 Bernd Wiesberger
87 Branden Grace
88 Brett Rumford
89 Brooks Koepka
90 Chris Wood
91 Darren Clarke
92 David Horsey
93 David Howell
94 Eduardo De la Riva
95 Gaganjeet Bhullar
96 Henrik Stenson
97 Jamie Donaldson
98 Jeev Milkha Singh
99 Jin Jeong
100 Joost Luiten
101 Julien Quesne
102 Kiradech Aphibarnrat
103 Luke Donald
104 Marc Warren
105 Marcel Siem
106 Mark Foster
107 Martin Kaymer
108 Matteo Manassero
109 Michael Hoey
110 Miguel A Jimenez
111 Mikko Ilonen
112 Pablo Larrazabal
113 Padraig Harrington
114 Paul Casey
115 Paul Lawrie
116 Paul McGinley
117 Paul Waring
118 Peter Hanson
119 Peter Uihlein
120 Phil Mickelson
121 Rafael Cabrera Bello
122 Raphael Jacquelin
123 Ricardo Gonzalez
124 Richard Bland
125 Richard Green
126 Robert Karlsson
127 Rory McIlroy
128 Sergio Garcia
129 Shane Lowry
130 Shiv Kapur
131 Soren Kjeldsen
132 Stephen Gallacher
133 Thomas Bjorn
134 Thomas Levet
135 Thongchai Jaidee
136 Thorbjorn Olesen
137 Tommy Fleetwood
138 Adam Gee
139 Alastair Forsyth
140 Daniel Im
141 David Higgins
142 Estanislao Goya
143 Gregory Bourdy
144 Jack Doherty
145 Jason Knutzon
146 John Hahn
147 Mikko Korhonen
148 Patrik Sjoland
149 Roope Kakko
150 Sam Walker
151 Aaron Baddeley
152 Adam Scott
153 Angel Cabrera
154 Ben Martin
155 Billy Horschel
156 Billy Hurley-III
157 Boo Weekley
158 Brandt Snedeker
159 Brendan Steele
160 Brendon Todd
161 Brian Davis
162 Brian Gay
163 Brian Harman
164 Brian Stuard
165 Brice Garnett
166 Bryce Molder
167 Cameron Tringale

168 Camilo Villegas
169 Chad Collins
170 Charles Howell-III
171 Charlie Beljan
172 Chesson Hadley
173 Chris Kirk
174 Chris Stroud
175 D.A. Points
176 Daniel Summerhays
177 Danny Lee
178 David Hearn
179 David Lingmerth
180 David Lynn
181 Davis Love III
182 Derek Ernst
183 Erik Compton
184 Francesco Molinari
185 Freddie Jacobson
186 Gary Woodland
187 George McNeill
188 Gonzalo Fdez-Castano
189 Graeme McDowell
190 Greg Chalmers
191 Harris English
192 Hudson Swafford
193 Ian Poulter
194 J.B. Holmes
195 J.J. Henry
196 Jason Bohn
197 Jason Kokrak
198 Jhonattan Vegas
199 Jim Renner
200 John Merrick
201 John Senden
202 Justin Hicks
203 Justin Rose
204 K.J. Choi
205 Keegan Bradley
206 Ken Duke
207 Kevin Chappell
208 Kevin Na
209 Lee Westwood
210 Lucas Glover
211 Luke Guthrie
212 Marc Leishman
213 Martin Laird
214 Matt Every
215 Matt Jones
216 Michael Putnam
217 Morgan Hoffmann
218 Nicholas Thompson
219 Nicolas Colsaerts
220 Pat Perez
221 Patrick Reed
222 Paul Goydos
223 Retief Goosen
224 Rickie Fowler
225 Robert Garrigus
226 Russell Henley
227 Russell Knox
228 Ryan Moore
229 Ryo Ishikawa
230 Sangmoon Bae
231 Scott Brown
232 Scott Stallings
233 Sean O'Hair
234 Seungyul Noh
235 Stewart Cink
236 Stuart Appleby
237 Tim Wilkinson
238 Trevor Immelman
239 Vijay Singh
240 Will Mackenzie
241 William McGirt
242 Woody Austin
243 Zach Johnson
244 Felipe Aguilar
245 Louis Oosthuizen
246 Simon Khan
247 Victor Dubuisson
248 Alexandre Kaleka
249 Anders Hansen
250 Jose M Olazabal
251 Jose Manuel Lara
252 Richie Ramsay
253 Andres Romero
254 Andrew Loupe
255 Andrew Svoboda
256 Ben Crane
257 Ben Curtis
258 Bud Cauley
259 Carl Pettersson
260 Charlie Wi
261 David Toms
262 Dustin Johnson
263 Edward Loar
264 Graham Delaet
265 Heath Slocum
266 James Driscoll
267 James Hahn
268 Jason Dufner
269 Jeff Overton
270 Jim Herman
271 Jimmy Walker
272 John Huh
273 John Peterson
274 John Rollins
275 Jonathan Byrd
276 Jordan Spieth
277 Josh Teater
278 Justin Leonard
279 Kevin Kisner
280 Kevin Tway
281 Martin Flores
282 Matt Kuchar
283 Mike Weir
284 Richard H. Lee
285 Ricky Barnes
286 Robert Allenby
287 Robert Streb
288 Rory Sabbatini
289 Ryan Palmer
290 Shawn Stefani
291 Spencer Levin
292 Steven Bowditch
293 Ted Potter-jr
294 Tim Clark
295 Tommy Gainey
296 Troy Matteson
297 Troy Merritt
298 Wes Roach
299 Bill Haas
300 Bubba Watson
301 Ernie Els
302 Hideki Matsuyama
303 Hunter Mahan
304 Jim Furyk
305 Jonas Blixt
306 Kevin Stadler
307 Kevin Streelman
308 Webb Simpson
309 Roberto Castro
310 Bo Van Pelt
311 Charley Hoffman
312 Geoff Ogilvy
313 Jerry Kelly
314 Johnson Wagner
315 Kyle Stanley
316 Mark Wilson
317 Nick Watney
318 Stephen Ames
319 Y.E. Yang
320 Michael Thompson
321 Rikard Karlberg
322 Robert-Jan Derksen
323 Scott Langley