THE HOT HAND IN GOLF

by

Jorge Eduardo Rodriguez-Castillo

M.Sc., University of Connecticut, 2005

M.Sc., Universidad de Guanajuato, 2003

B.Sc., Instituto Tecnológico Autónomo de México, 1999

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Spring 2013

APPROVAL

Name: Jorge Eduardo Rodriguez-Castillo

Degree: Master of Science

Title of project: THE HOT HAND IN GOLF

Examining Committee:

Dr. Yi Lu
Simon Fraser University
Chair

Dr. Tim Swartz
Simon Fraser University
Senior Supervisor

Dr. Derek Bingham
Simon Fraser University
Committee Member

Dr. Dave Campbell
Simon Fraser University
External Examiner

Date Approved: April 26, 2013

Abstract

In this project, an analysis is made to try to determine whether the phenomenon known as the hot hand, exists in golf. Data from a particular golf tournament in 2012 is studied in order to try to find out whether this proposition seems true. For this tournament, the scores for each golfer are split into the number of strokes and the number of putts required to complete the course. The key idea in this project is the substitution of the number of putts with the expected number of putts. The rationale is that putting is a highly stochastic element of golf and that the randomness conceals evidence of the hot hand. This expected value will be based on the distance to the pin once the ball is on the green. This distance to the pin is obtained from the ShotLink website. New scores for all golfers are calculated and consist of the sum of the number of strokes plus the expected number of putts in order to complete a course. The association between said scores in the first round and similar scores in the second round is calculated. The results seem to point to the conclusion that there is no hot hand in golf.

Acknowledgments

I want to thank the members of the Department of Statistics & Actuarial Science in general for their support throughout all the years I spent here.

Specifically I want to thank the members of the committee for reading this project.

Special thanks to Tom Loughin for his help, support and patience in all these years.

Special thanks to Tim Swartz for his help in the completion of this project and for setting the conditions so that this could take place under these special circumstances.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The hot hand belief in sports is the idea held by many fans, athletes and coaches that current performance and future success depends on how an athlete has been performing in the past. It is thought that players who have been performing well will have a tendency to keep performing well, whereas players who have passed through a bad stage will have a tendency to keep performing poorly. The hot hand is in conflict with the assumption of independence which is common in many sports modelling scenarios.

The hot hand phenomenon has been discussed in many sports. The work by Bar-Eli et. al. (2006) is an attempt to provide a review of the hot hand in the 1985-2005 period. In particular they provide a list of previous studies and they classify them as supportive or non-supportive of the hot hand. Actually one of the studies was classified by them as inconclusive. They claim that out of the studies they considered, there are few whose results support the hot hand belief compared to those which are non-supportive.

However, they realized that most of the so called supportive studies provide weak and fairly limited evidence in favor of the hot hand, due to what they refer to as an unrealistic model, questionable data and other issues such as setting questionable definititons of hot and cold players.

The sports covered in the studies considered by Bar-Eli et. al. (2006) are baseball, basketball, bowling, golf, darts, tennis, volleyball, pocket billiards and horseshoe pitching. A quick look at one of the tables they provide, shows that baseball and basketball seem to be the two most analyzed sports regarding the hot hand.

Avugos et. al. (2013) conducted a meta-analysis of the hot hand. They considered around 250 papers studying this phenomenon from 1985 to 2012. Only 22 of those publications met the authors' criteria in order to be considered to be part of the meta-analysis. The conclusion of this meta-study is that no hot hand effect exists in sports.

Tversky and Gilovich (1989a, 1989b) have devoted several papers to the study of the hot hand phenomenon. It is interesting that they also seem to concentrate mostly on the hot hand in baseball and basketball. In Gilovich et. al. (1985), the authors mention that basketball players, as well as fans, believe that a player's chance of hitting a shot are greater following a make than following a miss on the previous shot. Analyzing shooting records of the Philadelphia 76ers gave no evidence of a positive correlation between the outcomes of succesive shots. They also analyzed free-throw records of the Boston Celtics and performed a controlled shooting experiment with Cornell's varisity teams. They didn't find evidence of hot hand in either of these data sets. They attribute the belief in the hot hand on a misconception of chance, where any streak (even small ones) is regarded by the public as clear evidence in favor of this belief.

In 1989, a small debate arose around the hot hand effect in *Chance* magazine. In Tversky and Gilovich (1989a), the authors expose much of the same data sets, results, conclusions and explanations as in the 1985 work by Gilovich et. al.(1985). Also in *Chance*, Larkey et. al. (1989) question the conclusions and explanations of Tversky and Gilovich (1989a) and claim that the hot hand does exist. A further rebuttal by Tversky and Gilovich (1989b) came in the same *Chance* volume, where the authors provide an explanation on why the conclusions by Larkey et. al. (1989) are wrong and that their data actually shows no evidence in favor of the hot hand.

Other examples of papers on the hot hand in basketball are the works by Forthofer (1991), Wardrop (1995), Wardrop (1999), Vergin (2000), Koehler and Conley (2003). The hot hand has also been investigated in other sports. Some of these papers give evidence against the existence of the hot hand, some provide weak evidence in favor of it and some are inconclusive. In general, the evidence seems to suggest that the hot hand doesn't exist and that for most sports this belief is due to a misunderstanding of random sequences on the part of sports fans and professional players. Whenever a member of the general public sees a short streak of wins or losses (scored or missed shots, etc.), it appears unusual. The most simple explanation available to them is given by a hot or a cold hand.

1.1 Hot hand in golf

As with other sports, the hot hand has been studied in golf and its existence has also been dismissed by most studies. Some examples of these studies are those by Gilden and Wilson (1995), Clark (2003a), Clark (2003b) and Clark (2005). In his 2005 paper, Clark investigated the hot hand in golf by analyzing contingency tables. He found that golfers were as likely to score a birdie or better following a par or worse hole than when following a birdie or better hole. He remarked that his results were consistent with those found for individual players in baseball and basketball. His previous papers from 2003 had found some significant streakiness in players' sequence of scores but concluded that the streakiness was due to the difficulty of golf courses.

In the paper by Gilden and Wilson (1995), two putting experiments were conducted with volunteers. The volunteers practiced putting with different difficulty levels. The most relevant conclusion the authors obtained was that outcome sequences are streakiest when the difficulty of the task is commensurate with the ability of the performer.

In golf, there are various distinct skills which contribute to performance. For example, driving, wedge play and putting are all distinct skills. Whereas there are good putters and bad putters, we believe that the luck component in putting is greater than in the other skills. For example, a 20-foot putt which goes close to the hole is skill. Whether it goes in or not is luck. The luck element is due to many uncontrollable features of the green including, spike marks, the setting of the pin, imperfections due to worms, etc. If there is "too much" luck in a game, then the luck (or random component) can overshadow evidence of a hot hand effect.

In this project, we try to determine if the hot hand in golf exists once the luck element is removed from the number of shots required to complete the golf course.

Chapter 2

Data Collection

The data used in this project is obtained from the ShotLink website. ShotLink is a system that collects data related to golf tournaments and supports three tours: PGA, Champions and Nationwide. The 2012 Honda Classic tournament, part of which we analyze is part of the PGA tour. The ShotLink system collects data through a group of volunteers who operate laser devices that can measure with a precision of 2 inches, the position of the ball on the course.

2.1 Data set

The data set used in this project consists of the results of the first two rounds of the 2012 Honda Classic golf tournament. These first two rounds are used because only those players who made the cut are able to play in the last rounds (third and fourth). There were initially 144 players participating but one of them only played the first round. Additionally, 23 other players competed who do not regularly play on the PGA tour. This means that there is no measure of their overall performance for 2012 available to us. We have removed these players from the study since we later require covariates based on PGA tour participation. Therefore, 120 observations are used to conduct the analysis based on the first two rounds of this tournament.

For each of the players, information used in this project is the number of strokes and the number of putts required for each of rounds 1 and 2 and for each of the 18 holes on the course. Also, the distance from the current position of the ball to the pin is utilized, as well as whether the ball is already on the green or not.

Table 2.1 shows for a few players, the type of data collected and used in this project. The numbers under columns G_1 and G_2 represent the score obtained by each golfer in rounds 1 and 2, respectively. Numbers under columns P_1 and P_2 represent the number of putts for each of rounds 1 and 2. For each round we can express the score for each golfer as $G_i = S_i + P_i$, $i = 1, 2$, where S_i represents the number of strokes in round i and can be obtained from G_i and P_i. Again, our intuition is that a golf score G_i consists of two components, strokes S_i and putts P_i where P_i is more variable than S_i.

Player	G_1	G_2	P_1	P_2
Michael Bradley	73	72	29	28
Bob Estes	73	78	31	28
Davis Love III	71	72	32	27
Jeff Maggert	73	73	26	30
Billy Mayfair	70	71	30	29
Rocco Mediate	69	69	32	31
Tom Pernice Jr.	69	69	28	29
Stephen Ames	69	62	27	26
Chris DiMarco	71	70	31	25
Ernie Els	73	69	29	24

Table 2.1: Score and putts in rounds 1 and 2 for some of the golfers in the 2012 Honda Classic.

Chapter 3

Methodology and Analysis

For each of the 120 golfers who participated in the 2012 Honda Classic and for whom an average score on the 2012 PGA tour exists, the score for each of the first two rounds is analyzed. Since we think the luck element is more prominent in putting than in other golf skills, we split the score for each of the first two rounds into the number of strokes and the number of putts needed for each golfer to complete the course.

For golfer i in the first round and using notation previously introduced, this can be expressed as $G_{1i} = S_{1i} + P_{1i}$, where S_{1i} and P_{1i} are the sums of strokes and putts over all 18 holes, $S_{1i} = \sum_{j=1}^{18} S_{1i}^{(j)}$ and $P_{1i} = \sum_{j=1}^{18} P_{1i}^{(j)}$, and where the superscript (j)=1,...,18 indexes the hole. We expect the number of putts P_{1i} to vary from golfer to golfer, mainly because of the luck element rather than because of the inherent skill of the golfer.

Since we think the value P_{1i} is the result of an activity with an important component of randomness, we substitute it with its expected value P_{1i}^{*} in order to analyze a golfer's score which reflects more closely their true skill. The value P_{1i}^{*} is calculated as the sum over all 18 holes, of the expected number of putts required to get the ball into the hole once it reaches the green, $P_{1i}^{*} = \sum_{j=1}^{18} P_{1i}^{*(j)}$.

The data from ShotLink provides the distance (in inches) to the pin at the beginning of each hole and after each stroke or putt. For a particular player and hole, the expected number of putts $P_{1i}^{*(j)}$ is calculated based on this distance once the ball is on the green. The further the ball is from the pin, the bigger this expected number of putts will be. When the distance to the pin approaches 0 inches, the expected number of putts required to reach the hole should approach 1.

Once P_1^* and P_2^* have been calculated for each golfer, a new score for each of them and for each round can be obtained. For round 1, we define this new score as $G_1^* = S_1 + P_1^*$. This new score reflects more accurately the skill of each golfer and reduces the luck component attached to putting.

In order to calculate P_1^* and P_2^* for each player, we need to calculate some probabilities. In golf tournaments on the level of those in the PGA tour, once a golfer has reached the green, he usually requires at most three putts to reach the hole. This means that in order to calculate the expected number of putts, we can (for practical purposes) neglect the probability that a golfer needs more than three putts, since this probability is essentially zero.

In fact, the information provided by the PGA tour website (pgatour.com), only shows the percentage of times that a golfer requires either one or two putts (3-putt avoidance) once the ball is on the green. These percentages are provided for different distance intervals: less than 5 feet, 5 feet or more but less than 10 feet, 10 feet or more but less than 15 feet, 15 feet or more but less than 20 feet, 20 feet or more but less than 25 feet, and 25 feet or more.

Another useful type of information gives the percentage of holes achieved in one putt. This information is also given for the same distances as for the three-putt avoidance. Together with the three-putt avoidance, the probabilities of needing one, two or three putts can be derived. At a particular distance, the probability that a golfer requires exactly two putts can be obtained by subtraction from the information on the PGA tour website.

Let X represent the number of putts needed once the ball is on the green. Then the expected number of putts from a given distance δ can be calculated as

$$1 \times P[X = 1] + 2 \times P[X = 2] + 3 \times P[X = 3].$$

For a specific player i and a specific hole j in round 1, the previous equation provides the expected number of putts $P_{1i}^{*(j)}(\delta)$ from that given distance. Since we only have information available about the probabilities of requiring 1, 2 or 3 putts from specific distances, we need to estimate these probabilities at other distances.

In order to estimate $P_{1i}^{*(j)}(\delta)$ for $\delta > 0$, we fix $P_{1i}^{*(j)}(2.5)$ as $1 \times P[X = 1|\delta < 5] + 2 \times P[X = 2|\delta < 5] + 3 \times P[X = 3|\delta < 5]$.

We also fix $P_{1i}^{*(j)}(\delta = 5m + 2.5) = 1 \times P[X = 1|5m < \delta < 5(m+1)] + 2 \times P[X = 2|5m < \delta < 5(m+1)] + 3 \times P[X = 3|5m < \delta < 5(m+1)]$, for $m = 1, 2, 3, 4$.

Even though we can set up another equation using the information available for distances greater than 25 feet, we don't do so for reasons that will become clear later.

We can also fix the value of $P_{1i}^{*(j)}(0) = 1$ because the closer the ball is to the pin, the expected number of putts required to make a hole approaches 1. Estimating the number of putts via the distance of the putt is in the spirit of Broadie (2008), an idea also used by Fearing et. al. (2011).

We also need to fix $P_{1i}^{*(j)}$ at another distance so that we can then fit an equation to all these points in order to get estimated expected values for the whole range of distances. In order to fix this last point, we look at information specifically from the 2012 Honda Classic tournament and not from the PGA tour website for the whole 2012 year, as we did for the other 6 points. For this tournament, in round one, the longest distance from which a putt was made on a first-on-the-green opportunity is 87.6 feet. In round 2, this distance is 98.0 feet. We use 100 feet as the longest distance since it is a convenient value which is very close to the longest distance from which an initial putt was made in the data set.

We now explain why the 7th point is based on information from only one tournament and not from all 2012 PGA tournaments. Besides providing information for smaller ranges of distances like 5 feet to 10 feet, the PGA tour website provides information for all putts made from distances greater than 25 feet. Since we know putts can be attempted from distances near 100 feet, combined information for all putts from as close as 25 feet or as far as 100 feet may not be precise enough for fitting an equation.

The information from the 2012 Honda Classic tournament can provide a way in which we can fix the 7th point more precisely. It is worth mentioning that more than 70% of the roughly 2,000 putts attempted in each round were attempted from distances of 25 feet or less away from the pin. Less than 5% of the putts were made from distances longer than 50 feet.

So we want to fix $P_{1i}^{*(j)}(100)$ at some sensible value. Table 3.1 shows data for different distances in feet with the frequencies for 1, 2, 3 and 4 putts in round 1. The distance ranges were fixed from the Honda 2012 Classic tournament so that each range contains around 100 values. This is done in order to have enough observations with which probabilities can be estimated in a frequentist way. We can see that for all first putts that were made in round 1 from distances greater than 57 feet, golfers required 1 putt on 1 occasion, golfers required 2 putts on 79 occasions, golfers required 3 putts on 19 occasions, and golfers required 4 putts on 1 occasion. The one 4-putt required, confirms the fact that the probability of more than 3 putts once the ball is on the green is negligible. With the values in Table 3.1 we can calculate expected putts for the various ranges in the table. We observe that these expected values are very close and that they never exceed 2.2.

δ	1 putt	2 putts	3 putts	4 putts	Expected putts
$[57.0, 100.0)$	1	79	19	1	2.20
$[46.9, 57.0)$	5	87	8	0	2.03
$[35.4, 46.9)$	3	88	9	0	2.06
$[31.4, 35.4)$	6	88	6	0	2.00
$[28.3, 31.4)$	9	89	2	0	1.93

Table 3.1: Frequencies of putts from certain distances (feet) in round 1.

In Table 3.2 we find similar values as for Table 3.1 but for round 2 of the tournament. The conclusions that can be drawn are very similar to the ones observed in Table 3.1. The expected number of putts is never above 2.08.

The information in both Tables 3.1 and 3.2 suggests that we fix $P_{1i}^{*(j)}(100) = 2.14$ because 2.14 is the average of 2.20 and 2.08. We can display this value together with the other

δ	1 putt	2 putts	3 putts	Expected number of putts
$[57, 100)$	1	79	19	2.08
$[46.9, 57)$	5	87	8	2.03
$[35.4, 46.9)$	3	88	9	2.04
$[31.4, 35.4)$	6	88	6	2.00
$[28.3, 31.4)$	9	89	2	1.92

Table 3.2: Frequencies of putts from certain distances (feet) in round 2.

6 values we had fixed for $P_{1i}^{*(j)}(\delta)$. Table 3.3 shows the expected number of putts for the 7 distances previously mentioned. We observe that the expected number of putts near the pin is nearly 1.0 and that it is rare (only at large distances) for the expected number of putts for a PGA golfer to exceed 2.0.

δ	Expected number of putts
0.0	1
2.5	1.04
7.5	1.45
12.5	1.71
17.5	1.83
22.5	1.91
100.0	2.14

Table 3.3: Estimated number of putts required to score from various distances (feet).

Having fixed $P_{1i}^{*(j)}(\delta)$ for $\delta = 0.0, 2.5, 7.5, 12.5, 17.5, 22.5, 100.0$, we can estimate $P_{1i}^{*(j)}(\delta)$ for other values by fitting a curve to the 7 values already determined.

Two ways of doing this are described. In the first approach, a quadratic model is fit to the first 6 pairs of data in the previous table and then a line is fitted between the 6th and 7th pairs. One of the reasons why it is preferable to use the data from Table 3.3 rather than that from Tables 3.1 and 3.2 is that the former table uses information from all year 2012 for 6 of the 7 points and not only information from the 2012 Honda Classic.

The resulting equation is $P_{1i}^{*(j)}(\delta) = 1 + 0.0686\delta - 0.00121\delta^2$ for $0 < \delta < 22.5$ and $P_{1i}^{*(j)}(\delta) = 1.929 + \frac{0.211}{77.5}(\delta - 22.5)$ for $22.5 < \delta < 100$. In Figure 3.1 we investigate the fit of the quadratic part relative to the knots in Table 3.3. We observe that the fit is very good, with only the pair (2.5, 1.04) not being accurately described by this curve. However, we have to mention that it makes sense that the expected putts at a distance of 2.5 feet is still close to 1, since it is still most likely that at short distances a professional golfer will only need one putt to make a hole.

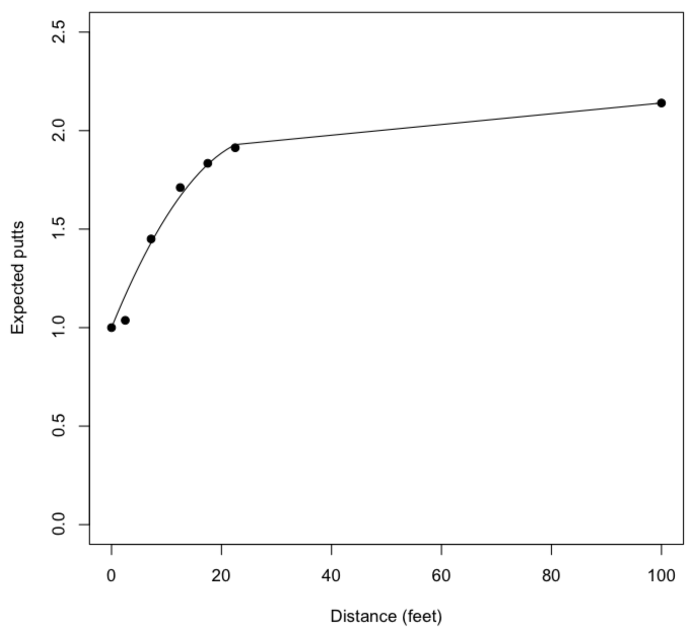

Figure 3.1: Quadratic fit followed by linear fit to the pairs of points in Table 3.1.

With these equations we can calculate the expected number of putts for every hole and for every round and player. We can also calculate P_{ki}^*, the expected number of putts for player i in round k. Table 3.4 shows some of these expected P_{1i}^* values along with the corresponding actual number of putts P_{1i} for 10 golfers who participated at the 2012 Honda Classic tournament. We observe that there is considerable discrepancy in the two columns, and the discrepancies are argued to be largely a function of luck.

Table 3.5 shows some of the actual scores G_{1i} and expected scores $G_{1i}^* = S_{1i} + P_{1i}^*$ for 10 different players. We observe differences in the two values by as many as 6.2 strokes, a

Player	Actual putts P_{1i}	Expected putts P^*_{1i}
Michael Bradley	32	30.4
Bob Estes	29	31.0
Davis Love III	27	26.9
Jeff Maggert	33	31.6
Billy Mayfair	29	29.0
Rocco Mediate	30	27.2
Tom Pernice Jr.	24	25.5
Stephen Ames	29	29.2
Chris DiMarco	32	30.4
Ernie Els	30	30.5

Table 3.4: Actual putts and expected putts for 10 golfers in the 2012 Honda Classic (quadratic interpolation followed by linear interpolation).

very meaningful difference in a round of golf. Again, the intuition is that the expected score G^*_{1i} is a better measure of skill than G_{1i}, which is conjectured to have a large component of luck due to putting.

Player	Actual score in round 1	Expected score in round 1
Michael Bradley	70	68.4
Bob Estes	67	69.0
Davis Love III	64	63.9
Jeff Maggert	71	69.6
Billy Mayfair	72	72.0
Rocco Mediate	69	66.2
Tom Pernice Jr.	67	68.5
Stephen Ames	75	75.2
Chris DiMarco	72	70.4
Ernie Els	70	70.5

Table 3.5: Actual score and expected score for 10 golfers in round 1 of the 2012 Honda Classic (quadratic interpolation followed by linear interpolation).

A second way in which we can estimate $P^{*(j)}_{1i}(\delta)$ is by linear interpolation between the 7 pairs of points in Table 3.3. This type of interpolation will produce the following linear equations, valid on specific ranges of distances.

$$P_{ki}^{*(j)}(\delta) = 1 + \tfrac{0.037}{2.5}\delta, \qquad 0 < \delta \le 2.5$$

$$P_{ki}^{*(j)}(\delta) = 1.037 + \tfrac{0.413}{5}(\delta - 2.5), \qquad 2.5 < \delta \le 7.5$$

$$P_{ki}^{*(j)}(\delta) = 1.450 + \tfrac{0.261}{5}(\delta - 7.5), \qquad 7.5 < \delta \le 12.5$$

$$P_{ki}^{*(j)}(\delta) = 1.711 + \tfrac{0.123}{5}(\delta - 12.5), \qquad 12.5 < \delta \le 17.5$$

$$P_{ki}^{*(j)}(\delta) = 1.834 + \tfrac{0.079}{5}(\delta - 17.5), \qquad 17.5 < \delta \le 22.5$$

$$P_{1i}^{*(j)}(\delta) = 1.913 + \tfrac{0.227}{77.5}(\delta - 22.5), \qquad 22.5 < \delta \le 100$$

The resulting interpolation is shown in Figure 3.2. This obviously provides a better fit to the data in Table 3.3 but the lack of smoothness in the interpolation is not appealing and is its main drawback.

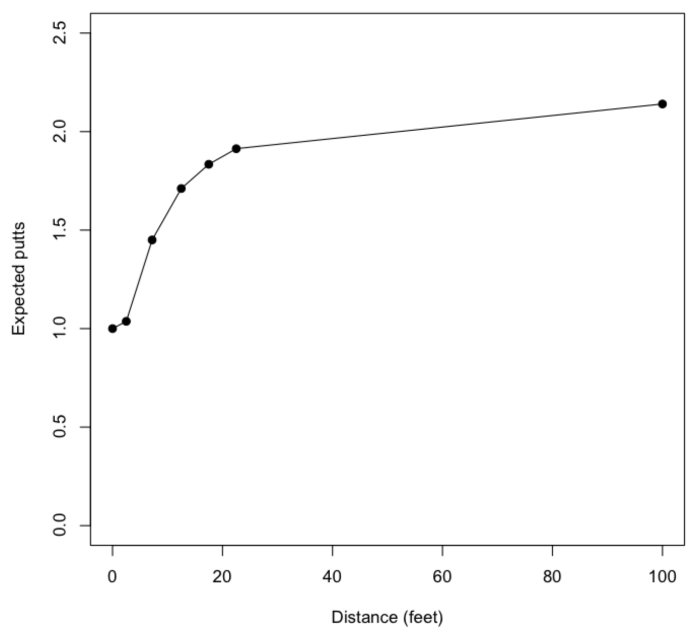

Figure 3.2: Linear interpolation to the pairs of points in Table 3.3.

Table 3.6 is similar to Table 3.4, the putts and expected putts based on the first 10 players in the data set. Table 3.7 is similar to Table 3.5, the score G_{1i} and expected score G^*_{1i} based on the first 10 players in the data set. We observe that for all Tables 3.4, 3.5, 3.6 and 3.7, the expected values are sometimes higher than the actual ones and sometimes smaller than the actual ones.

Player	Actual putts P_{1i}	Expected putts P^*_{1i}
Michael Bradley	32	30.2
Bob Estes	29	30.8
Davis Love III	27	26.6
Jeff Maggert	33	31.3
Billy Mayfair	29	28.5
Rocco Mediate	30	26.9
Tom Pernice Jr.	24	25.0
Stephen Ames	29	29.0
Chris DiMarco	32	30.0
Ernie Els	30	30.2

Table 3.6: Actual putts and expected putts for 10 golfers in the 2012 Honda Classic (linear interpolation).

Player	Actual score G_{1i}	Expected score G^*_{1i}
Michael Bradley	70	68.2
Bob Estes	67	68.8
Davis Love III	64	63.6
Jeff Maggert	71	69.3
Billy Mayfair	72	71.5
Rocco Mediate	69	65.9
Tom Pernice Jr.	67	68.0
Stephen Ames	75	75.0
Chris DiMarco	72	70.0
Ernie Els	70	70.2

Table 3.7: Actual score and expected score for 10 golfers in round 1 of the 2012 Honda Classic (linear interpolation).

For the remainder of the project, we will use the results obtained by fitting the quadratic equation to the first 6 pairs of points in Table 3.3 and a linear equation between the 6th and 7th points, in order to calculate P^*_{ki} and G^*_{ki}. Once $P^{*(j)}_{ki}$ has been calculated for both rounds (k=1,2), all holes (j=1,...,18) and all players (i=1,...,120), the calculation of the expected score G^*_{ki} can be obtained for any player i.

Once we have calculated the expected putts, the actual scores and the expected scores, we want to assess whether there is association between G^*_{1i} and G^*_{2i}, for example. If there is an association between these two quantities, we can interpret it as an indication of a "hot hand" or a "cold hand" effect.

At this point we want to establish the intuition that adding noise (i.e., random putting P) to a skilled activity such as shotmaking (S) decreases correlation.

Proposition: Assume that $Cov(S_1, P_1) = Cov(S_2, P_1) = Cov(S_1, P_2) = Cov(S_2, P_2) = Cov(P_1, P_2) = 0$ and $Cov(S_1, S_2) > 0$, then $Corr(S_1 + P_1, S_2 + P_2) \leq Corr(S_1, S_2)$.

Proof:

$$
\begin{aligned}
& Corr(S_1 + P_1, S_2 + P_2) \\
&= \frac{Cov(S_1 + P_1, S_2 + P_2)}{\sqrt{Var(S_1 + P_1)Var(S_2 + P_2)}} \\
&= \frac{Cov(S_1, S_2) + Cov(S_1, P_2) + Cov(P_1, S_2) + Cov(P_1, P_2)}{\sqrt{(Var(S_1) + Var(P_1) + 2Cov(S_1, P_1))(Var(S_2) + Var(P_2) + 2Cov(S_2, P_2)}} \\
&= \frac{Cov(S_1, S_2)}{\sqrt{(Var(S_1) + Var(P_1))(Var(S_2) + Var(P_2))}} \\
&\leq \frac{Cov(S_1, S_2)}{\sqrt{Var(S_1)Var(S_2)}} \\
&= Corr(S_1, S_2).
\end{aligned}
$$

One way in which we can detect association between G^*_{1i} and G^*_{2i} is by calculating the correlation between G^*_{1i} and G^*_{2i} across the 120 golfers. However, spurious correlation may be observed as a result of some golfers being better than others and not as a result of correlated G^*_{1i} and G^*_{2i} values.

In order to avoid the appearance of this spurious correlation, we subtract a measure of inherent skill possessed by each golfer. The intent is that the adjusted $G^*_{k1}, ..., G^*_{k120}$ forms a sample from a population and that the adjusted $G_{k1}, ..., G_{k120}$ forms a sample from a population. The PGA website publishes information on the average score for golfers in a year. For a particular player, this is calculated as the average of all their scores in all PGA tournaments in which they participated throughout a particular year.

For the year 2012, the average scores for all 120 of the golfers in our data set were available. Let's represent this average score for player i as $\bar{G}_i$. Once these averages are obtained we can calculate the correlation between $G^*_{1i} - \bar{G}_i$ and $G^*_{2i} - \bar{G}_i$. We are also interested in the correlation between $G_{1i} - \bar{G}_i$ and $G_{2i} - \bar{G}_i$. If the former correlation is greater than the latter, this is evidence of the hot hand being more detectable via pure shotmaking.

In Table 3.8, we provide the top 5 $\bar{G}_i$ values and the bottom 5 $\bar{G}_i$ values from the 2012 Honda Classic. We observe that from the PGA golfers who participated in the 2012 Honda Classic, the best in 2012 had an average score which is around 70 and even a bit lower than this. From those golfers who participated in the same tournament, the worst had scores higher than 72. We can see that there seems to be a gap of more than 2 shots in the average scores between the best and the worst golfers.

Player	$\bar{G}_i$
Tom Gillis	69.63
Greg Chalmers	69.78
Brendon De Jonge	70.10
Jim Furyk	70.17
David Hearn	70.22
Jimmy Walker	72.04
John Rollins	72.08
Stuart Appleby	72.20
Jason Bohn	72.34
Blake Adams	72.66

Table 3.8: The top 5 and bottom 5 $\bar{G}_i$ values representing inherent skill amongst players.

In Figure 3.3, we provide a scatterplot of $G_{1i} - \bar{G}_i$ versus $G_{2i} - \bar{G}_i$ over all 120 golfers. In the scatterplot we observe a cloud of points that is essentially random and with no clear pattern. Most of the points gather inside an imaginary circle, except for some in the right-most part of the graph. This shows that there really doesn't seem to be an association between $G_{1i} - \bar{G}_i$ and $G_{2i} - \bar{G}_i$. In other words, there does not seem to be evidence of the hot hand effect given traditional golfing measures.

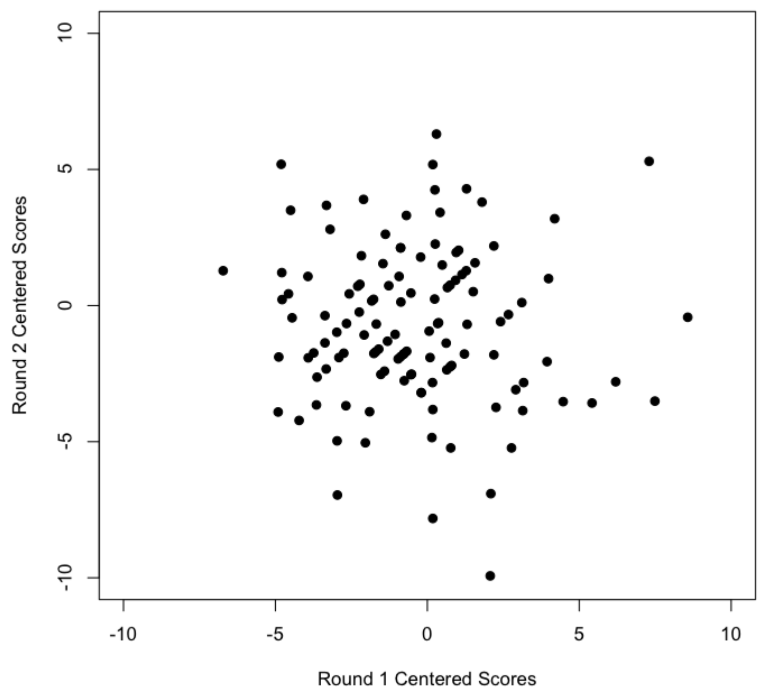

Figure 3.3: Scatterplot of $G_{1i} - \bar{G}_i$ versus $G_{2i} - \bar{G}_i$.

Careful observers may notice that several of the points in Figure 3.3 lie surprisingly on several straight lines which are all parallel to each other and with positive slopes. This is a result of the few different values of the actual scores G_{1i} and G_{2i} obtained by golfers on both rounds. Actually, the differences $G_{2i} - \bar{G}_i - (G_{1i} - \bar{G}_i) = G_{2i} - G_{1i}$, which represent the intercepts of the lines, take only 21 different values. The fact that 120 points have to lie on only those 21 lines makes it clear that several of the lines will necessarily have many points lying on them.

On the other hand, parallel lines with negative slopes cannot be easily detected by the eye in Figure 3.3 because the intercepts of these lines are given by the possible values of $G_{2i} - \bar{G}_i + (G_{1i} - \bar{G}_i) = G_{2i} + G_{1i} - 2 \times \bar{G}_i$, which are 111 different ones. As a result, very few of these lines will be containing 2 or more points.

In Figure 3.4, we provide a scatterplot of $G^*_{1i} - \bar{G}_i$ versus $G^*_{2i} - \bar{G}_i$. This scatterplot still shows almost complete randomness, although a bit less than in Figure 3.3. Again, we do not see an association between the first and second round expected scores once the average score in 2012 is subtracted from each player. In this figure it is also not possible to see the parallel lines pattern displayed by the points in Figure 3.3 because the possible values of $G_{2i} - G_{1i}$ and $G_{2i} + G_{1i} - 2 \times \bar{G}_i$ are all different, so each point lies on only one line, be it with a positive or a negative slope.

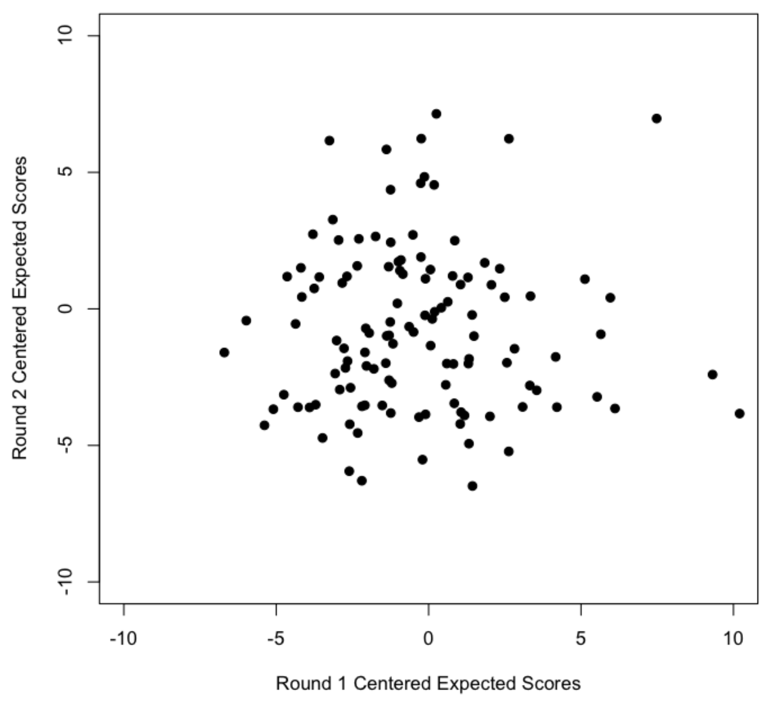

Figure 3.4: Scatterplot of $G^*_{1i} - \bar{G}_i$ versus $G^*_{2i} - \bar{G}_i$.

The graphical displays in Figure 3.3 and Figure 3.4 are supplemented with corresponding numerical statistics. Unfortunately none of the mentioned correlations was significant. For example, the correlation between $G^*_{1i} - \bar{G}_i$ and $G^*_{2i} - \bar{G}_i$ is -0.029, which is not statistically significant at a level of 0.05. The p-value is 0.75. The correlation between $G_{1i} - \bar{G}_i$ and $G_{2i} - \bar{G}_i$ is -0.052, which is also not statistically significant. The p-value is 0.57. Also, the sign of the correlations is negative, the opposite of what we were hoping to discover.

Chapter 4

Discussion

The hot hand effect is in the mind of many sports players and fans around the world, who take its existence as natural. It has been shown in several studies conducted for several sports that the hot hand effect does not exist and that it seems to be created in the imagination of players and fans due to a lack of proper understanding of chance and variability.

In golf, the hot hand phenomenon hasn't been conclusively proven either. However, it was our hope that in studying golf and by removing an element of randomness from the game, we would be able to detect a hot hand effect. We claimed that a random or luck component in the game is mostly present in the task of putting. We removed this luck component by taking into account the expected number of putts based on the distance from the pin, instead of the actual putts that occurred in the 2012 Honda Classic tournament.

By calculating the expected putts at certain distances from the pin, we proposed fitted equations that provide the expected number of putts for any given distance up to 100 feet. We then used these estimated values to obtain new scores for the players who participated in the tournament. We saw that the fit used to calculate the expected putts didn't have much influence as long as sensible equations are fitted. These scores are only based on the number of strokes and the expected putts. We further centered these scores based on the ability of each player.

Unfortunately, we were not able to detect any hot hand effect based on our results. It seems that even after removing one of the golf elements where more luck is involved, there is no hot hand in golf.

Given that no hot hand effect was found in this project, it may be possible to explore other possibilities. Maybe analyzing a different tournament, or data from many tournaments may produce different results. Analyzing more than 2 rounds can be another possibility. Maybe the hot hand effect cannot be seen from round to round but from hole to hole.

In any case, what is true is the faith that people in sports place in the hot hand. For example, while this project is being concluded, the Masters 2013 tournament will be taking place and the pundits have indicated as favorites to win the tournament, those players who have been playing well recently.

Even if the data used for this project doesn't support the existence of a hot hand effect, according to my understanding of golf and other sports, I think that a hot hand effect may be possible and that maybe a different way of analyzing this or other data may provide evidence of such effect.

Bibliography

[1] Avugos, S., Koppen, J., Czienskowski, U., Raab, M. and Bar-Eli, M. (2013). The"hot hand" reconsidered: A meta-analytic approach. *Psychology of Sport and Exercise*, **14**: 21-27.

[2] Bar-Eli, Michael, Avugos, Simcha and Raab, Markus. (2006). Twenty years of "hot hand" research: review and critique. *Psychology of Sport and Exercise*, **7**: 525-553.

[3] Broadie Mark. (2008). Assessing golfer performance using golfmetrics. *Science and golf V: Proceedings of the 2008 world scientific congress of golf*, **chapter 34**: 253-262.

[4] Clark, R.D. (2003a). Streakiness among professional golfers: Fact or fiction? *International Journal of Sport Psychology*, **34**: 63-79.

[5] Clark, R.D. (2003b). An analysis of streaky performance on the LPGA tour. *Perceptual and Motor Skills*, **97**: 365-370.

[6] Clark, R.D. (2005). An examination of the "hot hand" in professional golfers. *Perpetual and Motor Skills*, **101**: 935-942.

[7] Fearing, D., Acimovic, J. and Graves, S. (2011). How to catch a tiger: understanding putting performance on the PGA tour. *Journal of Quantitative Analysis in Sports*, **Vol. 7:** Iss. 1, Article 5.

[8] Forthofer, R. (1991). Streak shooter-The sequel. *Chance*, **4**: 46-48.

[9] Gilden, D.L., and Wilson, S.G. (1995). Streaks in skilled performance. *Psychonomic Bulletin and Review*, **2**: 260-265.

[10] Gilovich, T., Vallone, R., and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, **17**: 295-314.

[11] Koehler, J.J., and Conley, C.A. (2003). The "hot hand" myth in professional basketball. *Journal of Sport and Exercise Psychology*, **25**: 253-259.

[12] Larkey, P.D., Smith, R.A., and Kadane, J.B. (1989). It's okay to believe in the "Hot hand". *Chance*, **2(4)**: 22-30.

[13] Tversky, A. and Gilovich, T. (1989a). The cold facts about the "Hot hand" in basketball. *Chance*, **2(1)**: 16-21.

[14] Tversky, A. and Gilovich, T. (1989b). The "hot hand": Statistical reality or cognitive illusion? *Chance*, **2(4)**: 31-34.

[15] Vergin, R.C. (2000). Winning streaks in sports and the misperception of momentum. *Journal of Sports Behavior*, **23**: 181-197.

[16] Wardrop, R.L. (1995). Simpson's paradox and the hot hand in basketball. *The American Statistician*, **49**: 24-28.

[17] Wardrop, R.L. (1999). Statistical tests for the hot-hand in basketball in a controlled setting. *http://www.stat.wisc.edu/ wardrop/papers/tr1007.pdf.*