To appear in *Interfaces*

Assessing Golfer Performance on the PGA TOUR

Mark Broadie
Graduate School of Business
Columbia University
mnb2@columbia.edu

Original version: April 27, 2010
This version: February 8, 2011

Abstract

The game of golf involves many different types of shots: long tee shots (typically hit with a driver), approach shots to greens, shots from the sand, putts on the green, and others. While it is easy to determine the winner in a golf tournament by counting strokes, it is not easy to assess which factors most contributed to the victory. In this paper we apply an analysis based on strokes gained (previously termed shot value) to assess the performance of golfers in different parts of the game of golf. Strokes gained is a simple and intuitive measure of the contribution of each shot to a golfer's score. Strokes gained analysis is applied to extensive ShotLink™ data in order to rank PGA TOUR golfers in various skill categories and to quantify the factors that differentiate golfers on the PGA TOUR. Long game shots (those starting over 100 yards from the hole) explain about two-thirds of the variability in scores among golfers on the PGA TOUR. Tiger Woods is ranked number one in total strokes gained and he is ranked at or near the top of PGA TOUR golfers in each of the three main categories: long game, short game and putting. His dominance is a result of excelling at all phases of the game, but his long game accounts for about two-thirds of his scoring advantage relative to the field. A similar approach is used to rank PGA TOUR courses in terms of overall difficulty and difficulty in each part of the game. A preliminary analysis shows that the recent change in the groove rule for irons by the United States Golf Association (USGA) has had almost no impact on scores from the rough.

Acknowledgement: Thanks to the PGA TOUR for providing the ShotLink™ data. Thanks also to Kin Lo of the PGA TOUR, Richard Rendleman and Soonmin Ko for helpful discussions and comments, Lou Lipnickey for extensive programming on the project, Alexsandra Guerra for assistance with the data and the USGA for supporting the initial development of the Golfmetrics software.

1. Introduction

A golf score tells how well a golfer played overall, but does not reveal what factors contributed most to that score. The goal of this paper is to analyze the play of PGA TOUR professional golfers in order to understand and quantify the contributions of different categories of golf shots (e.g., long game, short game and putting) in determining a total golf score for an eighteen-hole round. This performance attribution analysis is used to rank golfers in various skill categories. The relative impact of each skill category on overall score is also examined.

Golf fans know that Tiger Woods is the best golfer of his generation, but it is often debated whether his low scores are primarily due to superior putting, wedge play around the greens, driving, or some other factor or combination of factors. Sweeney (2008) writes: "What really differentiates Woods from everyone else is his ability to make more putts from the critical range of 10 to 25 feet." In June 2010, U.S. Open winner Geoff Ogilvy said, "I think by now every player on tour is aware that the biggest reason Tiger is the best is because he putts the best." (Diaz, 2010). In spite of these assertions, it is not clear whether putting is the most important factor contributing to Tiger's scoring advantage.

This paper shows that Tiger Woods' scoring advantage in the years 2003-2010 was 3.20 strokes per round better than an average tournament field. Tiger is ranked at or near the top of PGA TOUR golfers in all three categories (long game, short game and putting), and his dominance is a result of excelling at all phases of the game. But Tiger's long game accounts for 2.08 of the total 3.20 stroked gained per round, so about two-thirds of his scoring advantage comes from shots over 100 yards from the green. Tiger's putting advantage versus the field is 0.70 shots per round, while his short game contributes 0.42 shots per round. Even though he is a phenomenal putter, his gain from putting is less than the 1.01 strokes he gains from shots starting between 150 and 250 yards from the hole, and comparable to the 0.70 strokes he gains from long tee shots.

Performance attribution analysis is difficult using the current standard golf statistics, many of which involve relatively crude counting measures. For example, the fairways hit statistic is the count of the number of fairways hit on a long tee shot (i.e., on par-4 and par-5 holes) divided by the number of tee shots. One problem with this statistic is that it doesn't distinguish between shots which barely miss the fairway, from shots that miss the fairway by a large amount and end up behind trees, in water, or out of bounds. Many standard golf statistics have the drawback that they mix several parts of the game together. For example, the sand save statistic counts the number of times a golfer gets the ball in the hole in one or two shots from a greenside sand bunker divided by the number of attempts. However, this statistic mixes together sand play with putting, making it difficult to isolate sand shot skill from putting skill. It is useful to have shot location information in order to better measure driving skill, sand play and putting skill. The PGA TOUR has collected this type of detailed data using their ShotLink^{TM} system since 2003.

In this paper, detailed shot data is used to assess and rank the performance of PGA TOUR golfers in three main parts of the game: the long game (shots over 100 yards from the hole), the short game (shots under 100 yards from the hole, excluding putts), and putting (shots on the green). The performance analysis is based on the concept of *strokes gained* (see Broadie 2008, where the term shot value was used instead of strokes gained), which measures the quality of each shot based on its starting and ending locations. As pointed out in Broadie and Ko (2009), the strokes gained metric is related to the value function of a dynamic program. If, for example, a golfer hits a poor sand shot followed by great putt, the sand shot will have a negative strokes gained value while the putt's strokes gained value will be positive. This approach allows each shot to be measured on its own merits, which is not possible with the sand save statistic which combines both shots. Just as golf scores are often compared to the benchmark of par, strokes gained represents the quality of a shot relative to a benchmark defined by the average performance of PGA TOUR golfers. Adding strokes gained for shots in a given category gives a performance measure for that category and is useful in understanding a golfer's strengths and weaknesses and in comparing one golfer to another. Strokes gained analysis is used to determine what separates the top golfers on the tour from others.

Since the publication of the landmark book by Cochran and Stobbs (1968), a large literature on the scientific and statistical analysis of golf has developed. Recent surveys include Penner (2003), Farrally et al. (2003) and Hurley (2010). Statistical analysis of amateur golfers was done in Riccio (1990). An early attempt to quantify the value of a shot was given in Landsberger (1994). Several papers have investigated which golf skill factors are most important in determining earnings in professional tournaments. Examples include Davidson and Templin (1986), Shmanske (1992), Moy and Liaw (1998), Berry (1999), Berry (2001), Nero (2001), Callan and Thomas (2007), Shmanske (2008) and Puterman and Wittman (2009). Most of these studies were limited by the lack of detailed shot information and had to rely on standard golf statistics (e.g., putting average, sand save percent, fairways hit, etc.). The strokes gained approach used in this paper directly decomposes a golfer's score by the quality of each shot, and is an alternative to the regression analyses used in many earlier studies.

Strokes gained analysis was introduced in Broadie (2008), primarily to determine which skills most separate the play of professional and amateur golfers. Putting performance on the PGA TOUR was investigated in Fearing et al. (2010), also using strokes gained analysis. In their study, the putting benchmark was adjusted to account not only for the distance to the hole, but also the difficulty of the green on each hole and the quality of putters in each tournament. Larkey (1994) and Berry (2001) represent early efforts to adjust tournament results for course difficulty and golfer skill factors. More recently, Connolly and Rendleman (2008) employed a statistical model in order to investigate golfer skill and streaky play on the PGA TOUR. The important idea in Larkey (1994), Berry (2001), Connolly and Rendleman (2008) and Fearing et al. (2010), is that overall scores and number of putts depend on golfer skill and on the difficulty of the course. Fewer putts are sunk on bumpy greens and scores are higher on more difficult courses, e.g., those with narrow fairways, deep rough and many water hazards. However, discerning the difficulty of a course is problematic when golfer skill is unknown: scores could be high because of less skill or a more difficult course.

The issue of disentangling golfer skill from course difficulty in golf scores also arises in creating golf handicaps for amateur golfers. For issues related to golf handicapping, see Pollock (1974), Scheid (1977) and Stroud and Riccio (1990).

This paper extends the analysis in Broadie (2008) in several ways. First, a benchmark function representing the average strokes to complete a hole is estimated for PGA TOUR golfers. The benchmark is interesting in itself, because it summarizes the skill of PGA TOUR golfers in various shot categories. A component of estimating the benchmark is the automatic identification of recovery shots. An estimation procedure is used to simultaneously estimate the difficulty of each course and round and to adjust the strokes gained results for the difficulty factors. In addition to providing a better measure of golfer performance, this procedure allows courses to be ranked in terms of overall difficulty and difficulty in each part of the game. Finally, the analysis is applied to a database of more than eight million shots by PGA TOUR golfers, leading to many interesting results, including the relative importance of the long game versus the short game.

In the next section the strokes gained concept is defined and illustrated. The construction of a benchmark function representing the average strokes to complete a hole for PGA TOUR golfers is described in Section 3. In Section 4 the strokes gained approach is applied to rank PGA TOUR golfers and analyze the factors that differentiate golfers on the PGA TOUR. PGA TOUR courses are ranked as well. A preliminary analysis of the effect of the USGA's groove rule change is also presented. Brief concluding remarks are given in Section 5.

2. Strokes gained

In this section the strokes gained concept is defined, examples are given and a simple but important additivity property of strokes gained is presented. The connection of strokes gained with dynamic programming is mentioned at the end of the section.

Strokes gained is a simple and intuitive quantitative measure of the quality of a golf shot. Suppose a function $J(d, c)$ has been estimated, where d represents the distance to the hole from the current location (not the distance of the shot), c represents the condition of the current ball location (i.e., green, tee, fairway, rough, sand or recovery) and J is the average number of strokes a PGA TOUR golfer takes to finish the hole from the current location. For brevity, J will often be referred to as the benchmark. Define the *strokes gained* of the ith shot on a hole that starts at (d_i, c_i) and finishes at (d_{i+1}, c_{i+1}) to be

$$g_i = J(d_i, c_i) - J(d_{i+1}, c_{i+1}) - 1. \quad (1)$$

Strokes gained represents the decrease in the average number of strokes to finish the hole from the beginning of the shot to the end of the shot, minus one to account for the stroke taken. For example, suppose the average number of shots to complete the hole is 2.6 from a position in the fairway forty yards from the hole. If the golfer hits the shot to one foot from the hole, where the average number of shots to complete the hole is 1.0, then equation 1 attributes a gain of 0.6 strokes to the shot: it reduced the average number of shots to complete the hole by 1.6 and it took one shot to do so, for a gain of 0.6. In general, a positive g_i indicates that a shot is better than a PGA TOUR golfer's average shot, while a negative g_i indicates that a shot is worse than average.

The units of strokes gained are strokes, e.g., a strokes gained value of -0.1 means the shot is 0.1 strokes worse than the benchmark. Because the units are the same for different types of shots, e.g., long shots and putts, the strokes gained metric offers a consistent method for evaluating different aspects of the game of golf. The strokes gained approach solves the problem of incommensurable measures in standard golf statistics that was pointed out in Larkey and Smith (1999).

To give an example of strokes gained, suppose that PGA TOUR golfer A plays a long par-3 that takes the PGA TOUR field an average of 3.2 strokes to complete the hole. Golfer A's tee shot finished on the green, leaving a 16-foot putt for birdie. From 16 feet, the PGA TOUR field takes an average of 1.8 putts to finish the hole. The PGA TOUR field will one-putt about 20% of the time, two-putt about 80% of the time, and rarely three-putt from 16 feet ($1.8 = 20\%(1)+80\%(2)+0\%(3)$). The ball started in a spot where the benchmark is 3.2 and finished at a position where the benchmark is 1.8, so the strokes gained for the shot is $3.2 - 1.8 - 1 = +0.4$. Golfer A left his birdie putt one inch short. His ball started in a spot where the benchmark is 1.8 and finished in a spot where the benchmark is one (the average number of shots to finish the hole for a tap-in is one), for a strokes gained value of: $1.8 - 1 - 1 = -0.2$. Golfer A's missed putt represents a loss of 0.2 shots relative to the benchmark, because he reduced the average number of strokes to complete the hole by 0.8 but he used one putt to do so. Because a PGA TOUR golfer only sinks 20% of 16-footers, missing this putt doesn't cost a full shot: it really only costs 0.2 strokes. To complete the example, golfer A tapped-in for par. The strokes gained equation (1) gives a value of zero for this putt, because he reduced the benchmark from one to zero using one shot. This makes sense, because sinking a one-inch putt neither gains nor loses shots relative to the benchmark.

The strokes gained metric has a simple but important *additivity property:* the strokes gained of a group of shots is the sum of the strokes gained of the individual shots. Suppose a golfer takes n shots on a hole. Then the total strokes gained for the n shots is:

$$\sum_{i=1}^{n} g_i = \sum_{i=1}^{n} (J(d_i, c_i) - J(d_{i+1}, c_{i+1}) - 1) = J(d_1, c_1) - n \tag{2}$$

because of the telescoping sum and $J(d_{n+1}, c_{n+1}) = 0$ for the last shot which ends in the hole. In the previous example, golfer A's score of $n = 3$ represents a net gain of 0.2 strokes compared to the benchmark of $J(c_1, d_1) = 3.2$ from the tee. Golfer A did this with a great tee shot (+0.4 strokes gained), a disappointing putt (-0.2 strokes gained), and a tap-in (0 strokes gained), for a total strokes gained of 0.2 for the hole, consistent with equation (2).

Let's consider PGA TOUR golfer B playing the same par-3 hole. Golfer B's tee shot missed the green long and left. From this position in the rough, suppose the average number of shots to complete the hole (the benchmark) is 2.6. The strokes gained equation (1) gives: $3.2-2.6-1 = -0.4$, so golfer B lost 0.4 strokes compared to the PGA benchmark. Golfer B hit his second shot from the rough to inside of four feet from the hole, where the benchmark score is 1.1 (a PGA TOUR golfer sinks about 90% of these putts). Applying equation (1) gives: $2.6 - 1.1 - 1 = 0.5$, so golfer B's second shot gained a half-stroke compared to the benchmark. Golfer B sunk the four-footer and the strokes gained equation gives $1.1 - 0 - 1 = 0.1$. Golfer B's score of three also represents a net gain of 0.2 strokes compared to the benchmark value of 3.2 from the tee. Golfer B did this with a poor tee shot (-0.4 strokes gained), a good chip from the rough (0.5 strokes gained), and a one-putt (0.1 strokes gained), for a total of 0.2 strokes gained for the hole.

Golfers A and B had the same score on the hole, but they did it in very different ways. If this was a representative example, we could see that golfer A has a great long game, while golfer B has a great short game. Strokes gained allow us to compare golfer A's game to golfer B's, both in total strokes gained (for the hole, round, or season) and in various categories (e.g., long game, short game, and putting). This observation can be formalized by decomposing the total strokes gained for a round into different categories as follows:

$$\sum_{i=1}^{m} g_i = \sum_{i\in\mathcal{L}} g_i + \sum_{i\in\mathcal{S}} g_i + \sum_{i\in\mathcal{P}} g_i \tag{3}$$

where m is the total number of shots in a round, $\mathcal{L}$ is the set of indices corresponding to long game shots, $\mathcal{S}$ is the set of indices corresponding to short game shots, $\mathcal{P}$ is the set of putts and where $\{1, 2, \ldots, m\} = \mathcal{L} \cup \mathcal{S} \cup \mathcal{P}$. In a similar way, the strokes gained for a given category can be further decomposed into subcategories. For example, the total strokes gained of all long game shots can be split into the sum of strokes gained for tee shots and approach shots from various distance categories. Unlike fraction of greens hit, proximity to the hole, or other statistical measures, the strokes gained approach provides a consistent way to quantify the value of shots in various categories and subcategories.

The game of golf can be modeled as a dynamic program. The score on a hole depends on the strategy and results of each of the shots on the hole. The optimal strategy from the tee depends on all of the possible outcomes of the first shot and the optimal strategy for the second shot, which depends on all of the possible outcomes of the second shot and the optimal strategy for the third shot, etc. The solution of a dynamic program involves starting from the last stage, in this case the shot which ends in the hole, and working backwards to determine the optimal strategy. The Bellman (1957) equation says:

$$J(d_i, c_i) = \min_{\mu} E[J(d_{i+1}, c_{i+1}) + 1 \mid (d_i, c_i, \mu)] \tag{4}$$

where the expectation is taken over (d_{i+1}, c_{i+1}), the random distance and condition of the end of shot i, given its start at (d_i, c_i) and the strategy μ (e.g., target and club) chosen by the golfer.

For more detail, see Broadie and Ko (2009). This paper does not address the strategy choices of golfers, but PGA TOUR golfers are among the best golfers in the world, so it is not unreasonable to assume that they play optimal or nearly optimal strategies and the observed data can be used to estimate $J(d_i, c_i) = E[J(d_{i+1}, c_{i+1}) + 1 \,|\, (d_i, c_i, \mu^*)]$, where μ^* represents an optimal strategy. Now an individual shot can be measured by the difference in the left and righthand sides of the equation for a particular outcome, i.e., by $J(d_i, c_i) - J(d_{i+1}, c_{i+1}) - 1$, which is the strokes gained definition given in equation (1). This dynamic program viewpoint provides the justification for the strokes gained definition.

3. PGA TOUR benchmark

The strokes gained computation is based on a benchmark function that gives the average number of strokes for PGA TOUR golfer to complete a hole. The benchmark typically increases with the distance to the hole and depends on the course condition at the location of the ball, i.e., tee, fairway, rough, green, sand or recovery. Shots from the rough are more difficult than shots from the fairway, and the benchmark is larger as a consequence. There are situations, typically from the rough, where a direct shot to the hole is impossible because the path is blocked by trees or other obstacles. In this case a golfer may elect to play a recovery shot, i.e., a short shot that is hit back to the fairway rather than directly toward the hole. Recovery shots are placed in their own category in order to better estimate the differential effects of fairway and rough. The estimation of the benchmark function, the recovery shot identification procedure and empirical results are given in this section.

The results in this paper are based on the PGA TOUR's extensive ShotLink database, which includes all shots at PGA TOUR tournaments from 2003 to 2010. The data is collected by 250 volunteers at each tournament. The ShotLink database contains more than eight million shots (about one million shots per year), with shot locations measured to within one inch on putts and one foot on other shots. Further information on the ShotLink system is given in Deason (2006). The ShotLink database does not include detailed shot information for the four major tournaments: the Masters, U.S. Open, British Open and the PGA.

The benchmark function (the average number of shots to complete the hole) needs to be defined in terms of observable information recorded in the database. Not all shots from the fairway with 125 yards to the hole are equal in difficulty, since there are many other factors involved: the ball's lie might be perfect or in a divot, the golfer's stance might be level or on a hill, the wind could be calm or gusting, etc. All of these other factors affect the difficulty of a shot and the average number of shots to complete the hole. However, the benchmark can only be computed from observable information, and the ShotLink database includes the most important of these factors: the distance from the hole and the condition of the ball (e.g., tee, fairway, green, sand, or rough). The benchmark function can be interpreted as an average over these other unobservable factors.

The goal is to estimate the benchmark function $J(d, c)$, where d represents the distance to the hole from the current location and c represents the condition of the current location (i.e., green, tee, fairway, rough, sand or recovery). Statistical and model-based approaches are two main ways to accomplish this. Statistical procedures include simple interpolation, linear regression, splines, kernel smoothing and other methods. In model-based approaches a parametric analytical or simulation model is formulated and optimization is used to determine the model parameters that best fit the data. Both approaches attempt to find a benchmark that is close to the data and appropriately smooth to take into the noise in the data.

The large size of the database allows for accurate estimation of the benchmark, since in most distance and condition categories, there are many shots available to estimate the average score to complete the hole. Experimentation with several approaches yielded similar results. Piecewise polynomial functions were used as the form of the benchmark, except for putts on the green. For putts on the green, a model-based approach is used to fit one-putt probabilities based on a simplification of the putting model presented in Broadie and Bansal (2008). This is combined with a statistical model for three-putts to give an average score function for putts on the green. This approach is used to smooth the somewhat limited data for long putt distances.

3.1. Tee shot benchmark

From the tee, a simple linear regression of average score (J) on the distance to the hole (d, measured in yards) for PGA TOUR pros using 2003-2010 data gives: $J = 2.38 + 0.0041d$. The distance to the hole d is measured along the fairway from the tee to the hole (i.e., the dogleg distance, not the direct "as the crow flies" distance). In this regression, the data are grouped into 20-yard distance buckets and the R^2 of the regression is over 98%. The slope of the equation implies that every additional 100 yards of hole distance adds 0.41 strokes to the average score of a PGA TOUR pro. (This regression is similar to the result $2.35 + 0.0044d$ obtained in Cochran and Stobbs (1968), based on a smaller set of data collected from a single British professional tournament in 1964.)

In spite of the high R^2, a linear regression does not provide an adequate fit to the data as shown in Figure 1. In particular, the average score from the tee exhibits a jump between long par-3 holes at 235 yards and short par-4 holes at 300 yards (there is little data between these distances). The computations in the paper are based on a more accurate piecewise polynomial fit to the data. The results are given in Appendix A.

Broadie (2008) finds that the average score from the tee is $2.79 + 0.0066d$ for golfers whose 18-hole average score is 90. The slope implies that every additional 100 yards of hole distance adds 0.66 strokes to the average score of 90-golfers, while for PGA TOUR pros it adds 0.41 strokes. The USGA refers to this slope as the *ability to overcome distance*. For 90-golfers, going from 180 yards (par-3 distance) to 580 yards (par-5 distance) will increase their average score by about 2.6. But the par goes up by 2, so 90-golfers do *worse* relative to par on par-5 holes compared to par-3 holes. Going from a hole of 180 yards (par-3 distance) to 580 yards (par-5 distance), pros will see

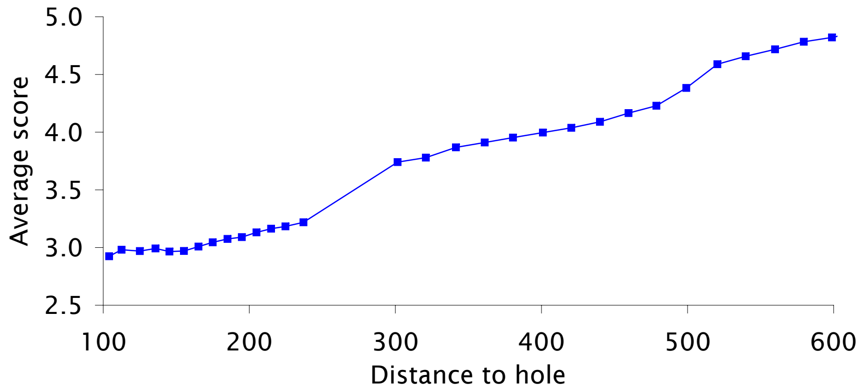

Figure 1: Average score from the tee for PGA TOUR golfers in 2003-2010. Distance to the hole is measured along the fairway from the tee to the hole, not "as the crow flies."

an average score increase of 1.6. The par goes up by two, so the pros do *better* relative to par on par-5 holes compared to par-3 holes. The main reason is the 290-yard average distance that the pros drive the ball, compared to an average drive of about 210 yards for 90-golfers. (Of course, on par-5 holes 90-golfers have more chances to flub shots or hit into trouble.)

Using the Golfmetrics database of primarily amateur data, described in Broadie (2008) an average score from the tee of $2.79 + 0.0066d$ is found for golfers whose 18-hole average score is 90. The slope implies that every 100 yards of hole distance adds 0.66 strokes to the average score, while for PGA TOUR pros it adds 0.41 strokes. The USGA refers to this slope as the *ability to overcome distance*. For 90-golfers, going from 180 yards (par-3 distance) to 580 yards (par-5 distance) will increase their average score by about 2.6. But the par goes up by 2, so 90-golfers do *worse* relative to par on par-5 holes compared to par-3 holes. Going from a hole of 180 yards (par-3 distance) to 580 yards (par-5 distance), pros will see an average score increase of 1.6. The par goes up by two, so the pros do *better* relative to par on par-5 holes compared to par-3 holes. The main reason is the 290-yard average distance that the pros drive the ball, compared to an average drive of about 210 yards for 90-golfers. (Of course, on par-5 holes 90-golfers have more chances to flub shots or hit into trouble.)

3.2. Benchmark within 50 yards of the hole

In this subsection average strokes to complete the hole from the sand, rough and fairway are compared on shots within 50 yards of the hole. It is often claimed that professional golfers are so good from the sand that they would rather be in the sand than in the rough. Figure 2 illustrates the data (bucketed in 5-yard increments) and the fitted curves. The figure shows that when the distance to the hole is less than 15 yards or greater than 34 yards, sand shots have *larger* average strokes to complete the hole than shots from the rough from the same distance. In the range from 15 yards to 34 yards, sand shots are easier than shots from the rough, on average. Conditioned on the shot starting within 50 yards of the hole, the average initial distance to the hole for shots from the sand and rough is 16 yards, just about the distance of equal difficulty for sand and rough shots. The average score can be translated into an up-and-down fraction, i.e., the fraction of the time it takes two or less shots to finish the hole. From 15 yards from the hole, pros get up-and-down 51% of the time from the rough or sand and 69% of the time from the fairway. At 25 yards from the hole, pros get up-and-down 42% of the time from the sand, 35% from the rough and 54% of the time from the fairway. These are averages over all situations; note that the outcome for an individual shot will depend on the ball's lie, the contour of the green near the hole and other factors. However, the distance from the hole and condition of the ball are primary factors in determining the average number of shots to complete the hole.

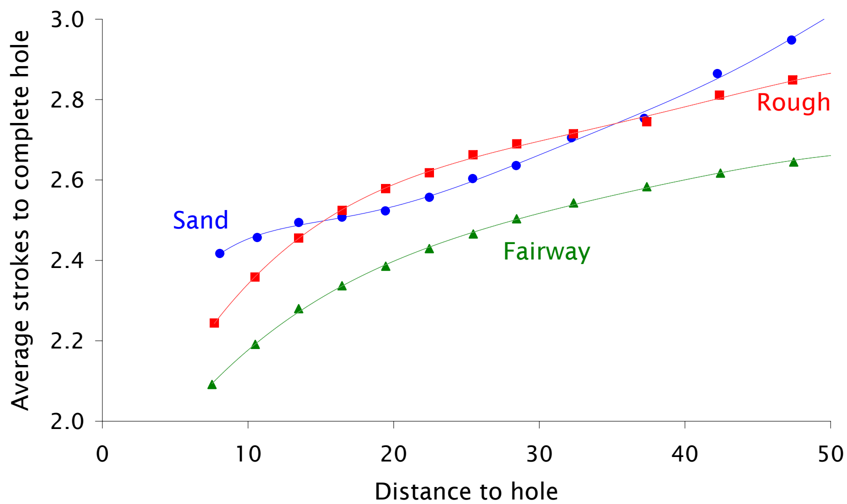

Figure 2: Average strokes to complete the hole from the rough, sand and fairway for PGA TOUR golfers in 2003-2010.

3.3. Putting benchmark

In this subsection the estimation of the benchmark function for putts is discussed. The benchmark is fit in three steps. First, a one-putt probability function is fit, then a three-putt function is fit, and then these two are combined into a benchmark average putts-to-complete-the-hole function. This approach is followed for several reasons. First, the data is sparse and noisy for long putts (e.g., greater than 50 feet from the hole) and so smoothing is necessary. Second, the procedure works well for fitting smaller data sets and it is useful to have a consistent procedure for all sets of data. Finally, golfers think in terms of one-putts and three-putts, so these models and results are of independent interest.

The one-putt probability function is based on a simple physical model for putts. Putting skill is modeled with two components: a random distance and a random direction, both independently distributed normal random variables. The random direction of the putt with respect to the hole is α, with $\alpha \sim N(0, \sigma_\alpha^2)$, so angular putt errors have a standard deviation of σ_α. The putt rolls a random distance l with $l \sim N(d+t, (d+t)^2\sigma_d^2)$, where d is the initial distance to the hole and t is the target distance beyond the hole (all measured in yards). The standard deviation of the distance a putt rolls, $(d+t)\sigma_d$, is proportional to the intended target distance $d+t$. If $t = 1/2$ yards it means the golfer aims to hit the putt 1.5 feet beyond the hole. For the putt to have a chance of finishing in the hole, the angle must satisfy $|\alpha| \leq \alpha_c = \tan^{-1}(r/d)$, where d is the distance to the hole and r is the radius of the hole (2.125 inches). In addition, the distance the putt rolls, l, must be at least d, otherwise the putt will not reach the hole. If the putt is hit too hard (even if hit straight at the hole) and rolls a distance greater than $d+h$, it will also not result in a holeout. A holeout occurs if the putt rolls a distance l satisfying $d \leq l \leq d+h$ and is hit with an angle satisfying $|\alpha| \leq \alpha_c$.

This model is a generalization of the Gelman and Nolan (2002) model which only takes putt direction into account. It is a simplification of Broadie and Bansal (2008), which models distance, direction and green reading errors, but is not analytically tractable and requires simulation to evaluate. The holeout criterion is used for analytical tractability. A detailed physical model for holeouts, i.e., the putt finishing in the hole, was developed in Holmes (1991) and used in Broadie and Bansal (2008). This model has few parameters, has a physical interpretation, is analytically tractable, and it fits the data very well. The probability of a one-putt, $p_1(d; \sigma_\alpha, \sigma_d, t, h)$ is:

$$
\begin{aligned}
P(|\alpha| \leq \alpha_c)P(d \leq l \leq d+h) &= P(|\alpha| \leq \alpha_c) P\left(\frac{-t}{\sigma_d(d+t)} \leq Z \leq \frac{h-t}{\sigma_d(d+t)}\right) \\
&= \left(2\Phi\left(\frac{\alpha_c}{\sigma_\alpha}\right) - 1\right)\left(\Phi\left(\frac{h-t}{\sigma_d(d+t)}\right) - \Phi\left(\frac{-t}{\sigma_d(d+t)}\right)\right),
\end{aligned}
\tag{5}
$$

where Z is a standard normal random variable and Φ is the cumulative distribution of a standard normal. Given a set of one-putt data by distance to the hole, an optimization model is solved to find the best-fit parameters σ_α and σ_d. The model can be fit very quickly because of equation (5) and the readily available routines for computing Φ. (The parameters t and h were fixed at $t = 1/2$ and $h = 2/3$.)

The three-or-more putt probability function is estimated by fitting the equation:

$$
p_3(d; a_0, a_1, a_2, a_3) = \frac{1}{1 + e^{a_0 + a_1 d + a_2 d^2}} + a_3. \tag{6}
$$

for the parameters a_0, a_1, a_2 and a_3. This functional form was chosen because, of the many forms tested, it fit the data very well. Four (or more) putts are observed in the professional data, but are so rare that the fit is not affected. The optimization model to find the best-fit parameters is quick to solve.

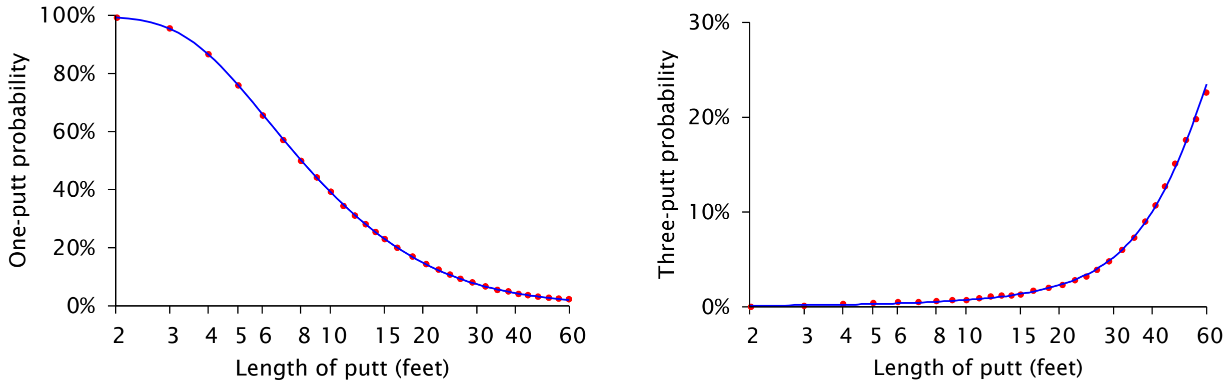

Figure 3: PGA TOUR putting results using 2003-2010 data. Left panel: one-putt probability. Right panel: three-putt probability. Dots represent the data and the curves are the fitted models.

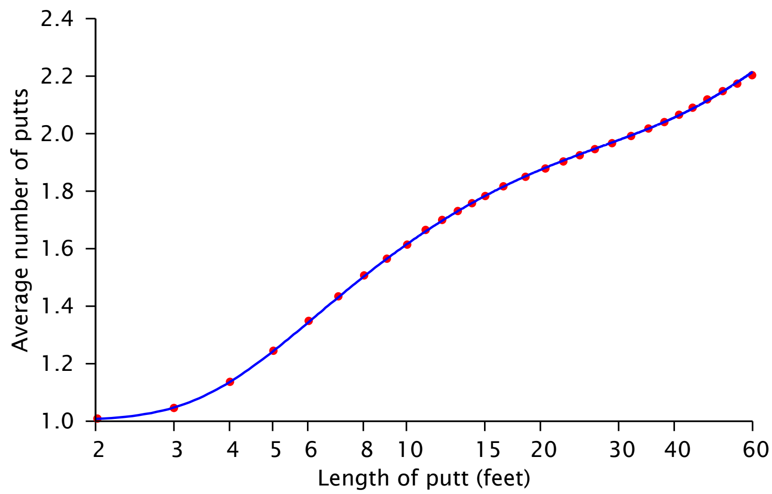

Figure 4: Average number of putts by initial distance to the hole for PGA TOUR golfers in 2003-2010. Dots represent the data and the curve is the fitted benchmark model.

The average number of putts to holeout benchmark function is now easy to compute using

$$(7) \qquad J(d) = p_1(d;\sigma_\alpha,\sigma_d,t,h) + 2\left(1 - p_1(d;\sigma_\alpha,\sigma_d) - p_3(d;a_0,a_1,a_2,a_3)\right) + 3p_3(d;a_0,a_1,a_2,a_3),$$

where the condition of starting on the green is implicit. This approach of separately fitting a physical model for one-putts, a statistical model for three-putts, and combining to give an average number of putts curve is simple, easy to calibrate, and most importantly, fits the data very well. This method has the advantage of being able to see one-putt, two-putt, and three-putt probability functions as well as the average number of putts function. A different approach is used in Fearing et al. (2010), where a statistical model is used for the one-putt probability function and a gamma distribution is fit to the remaining distance of missed putts.

The left panel of Figure 3 illustrates the data and the fitted one-putt probability curve. The parameters of the fitted curve are: $\sigma_\alpha = 1.46$, $\sigma_d = 0.057$ (with $t = 0.5$ and $h = 0.667$). The model probability is almost always within one standard error of the data. (The standard errors are too small to be shown clearly on the graph.) PGA TOUR golfers one-putt 50% of the time from a distance of eight feet. It is interesting to compare this result from 2009 data with earlier results. Cochran and Stobbs (1968, Chap. 29) found that in 1964 pros one-putt 50% of the time from a distance of seven feet, though their sample was fairly small. Soley (1977, Chap. 4) found that pros sunk 50% of their putts from seven feet using data from the early 1960s on regular tournament courses. He found the same result at the 1974 U.S. Open at Winged Foot, but closer to six feet at the 1972 U.S. Open at Pebble Beach. Using hand-collected data from PGA tournaments in the 1980s, Pelz found that pros sunk 50% of their putts from about six or seven feet (Pelz (1989, p.38) and Pelz (2000, p.7)). The increase in the 50% one-putt distance from six or seven feet to the current eight feet could be due to better conditioned greens, better putting skill, or a combination. By contrast, amateur golfers with an average 18-hole score of 90 (90-golfers), one-putt 50% of the time from five feet.

The putting data yields many interesting results. For example, the average number of one-putts for a single pro golfer from 22 feet or over in a *four-round* tournament is only 1.4. This is a much smaller number than most people expect, perhaps because of the bias from watching television highlights of putts that are made, while missed 30-footers rarely make the highlight reel. A pro averages just over five putts per round from 22 feet or more and sinks slightly less than 7% of these long putts.

The right panel of Figure 3 illustrates the data and the fitted three-putt probability curve. The parameters of the fitted curve are: $a_0 = -0.106$, $a_1 = 5.49$, $a_2 = 0.000563$ and $a_3 = -0.00398$. It is not until 40 feet that the three-putt probability for PGA TOUR golfers exceeds 10%. PGA TOUR golfers average 0.55 three-putt greens per round, or about 2.2 per four-round tournament. Amateur 90-golfers three-putt about 2.3 times per round, or four times more often than pros. Figure 4 shows how the average number of putts increases with distance for PGA TOUR golfers. PGA TOUR golfers average two putts from 33 feet, i.e., the fraction of one-putts equals the fraction of three-putts. Amateur 90-golfers average two putts from 19 feet.

3.4. Recovery shots

A shot is called a recovery shot if the golfer's shot to the hole is impeded by trees or other obstacles. Even if the golfer decides to hit toward the hole through a small opening in trees, or attempts a hook or slice around an obstacle, it is still considered a recovery shot because the golfer is recovering from trouble. In this subsection we discuss this category of shots, their importance in the benchmark, and how they are identified.

Suppose a golfer hits a long drive but ends up behind a tree and is forced to chip back out onto the fairway for the second shot, i.e., the second shot is a recovery shot. If the benchmark does not account for recovery shots, the strokes gained of the tee shot may be close to zero, but the second shot will have a very negative strokes gained since it didn't travel very far. This doesn't make sense, because the problem was caused by a poor tee shot, not a poor second shot. The strokes gained equation can account for this situation by identifying the condition of the second shot as a recovery shot, and the benchmark will have a larger average number of strokes to complete the hole than from a comparable distance in the rough. In this example, a using separate benchmark for recovery shots will give a negative strokes gained for the poor tee shot and a strokes gained of close to zero for the second shot. The recovery label is important for correctly allocating strokes gained between the two shots and important for estimating the penalty for being in the rough versus fairway. In this example, the rough was not the direct cause of the increase in score—it was an obstructed route to the hole.

The recovery condition is an important element in assessing the quality of a shot, but the ShotLink database does not have an identifier of this category of shot. The main reason is that labeling a shot a recovery shot is a judgement call, unlike distance to the hole which is an observable and objective quantity. The recovery shot condition needs to be inferred from existing information in the data. Because the database contains millions of shots, a manual identification procedure is not feasible. The automatic recovery shot identification procedure has two steps. The first step finds shots that travel an unusually short distance (e.g., travel less $r_1 = 40\%$ of the distance to the hole) or are hit at a large angle relative to the hole (e.g., an absolute angle greater than $r_2 = 15$ degrees with respect to the ball-hole line). Shots are also screened to started at least $r_3 = 30$ yards from the hole. The parameters are determined by visually inspecting a number of shots which satisfy the criterion. The second step finds shots that start close, e.g., within $r_4 = 3$ yards, to shots by other golfers labeled as recovery shots in step one. For example, suppose two golfers are in nearly identical recovery shot positions obstructed by trees. The first golfer chips back onto the fairway while the second golfer attempts a big slice around the trees. The first golfer's shot would be identified as a recovery shot by step one, but the second golfer's shot might not, even though the shot started in the same position and was significantly affected by trees. The second step of the procedure allows the shot of the second golfer to be labeled as a recovery shot.

Figure 5 illustrates three shots labeled as recovery shots by this procedure.[1] This method of inferring which shots are recovery shots works very well, but two types of errors will occur. Some shots will be labeled as recovery shots that are not, and some shots that are recovery shots will not be labeled as such. Given the current data and judgement involved, it is not possible to design an error-proof procedure. However, the magnitudes of the two types of errors can be controlled by the choice of the parameters. Recovery shot identification is important when comparing the average number of shots to complete a hole from the rough versus the fairway. Once recovery shots are identified, benchmark functions are fit to the data using piecewise polynomials. The results are given in Appendix A. Figure 6 shows the average strokes to complete the hole for recovery shots

[1] Golf course images from Google Earth are used to display the shots. The ShotLink database contains shot starting and ending position using (x, y) coordinates that were translated to latitude and longitude for plotting.

Figure 5: Recovery shot examples. Left panel: Corey Pavin, 6/11/2006, hole 3, Westchester Country Club. The shot indicated is labeled a recovery shot because of the distance criterion. Middle panel: Tim Clark, 6/26/2005, hole 15, Westchester Country Club. The shot indicated is labeled a recovery shot because of the distance criterion. Right panel: Fred Couples, 6/11/2006, hole 15, Westchester Country Club. The shot indicated would not be labeled a recovery shot by the distance or angle criteria, but it is labeled a recovery shot because it is nearby to another golfer's recovery shot (not shown). Arcs show 100-, 150- and 200-yard distances from the hole.

and shots from the rough and fairway.

4. PGA TOUR golfer rankings and results

In this section strokes gained analysis is used to rank PGA TOUR golfers in various skill categories and subcategories. The strokes gained are first adjusted by the difficulty course for that round in order to produce more reliable comparisons between golfers. This section provides details of the adjustment procedure, results and discussion of the rankings, and analyzes which skill factors separate the best golfers on the tour from others.

4.1. Course-round difficulty adjustments

Some four-round PGA tournaments have winning scores of 30 under par while for others it may only be 6 under par. The difference of six shots per round is due to two main factors. First, the course at one tournament may be easier than the course at another tournament. The difference in difficulty can be due to length of the course, width of the fairways, firmness of the greens, height of the rough, severity of bunkers, and other factors. Second, from one round to the next, scores on

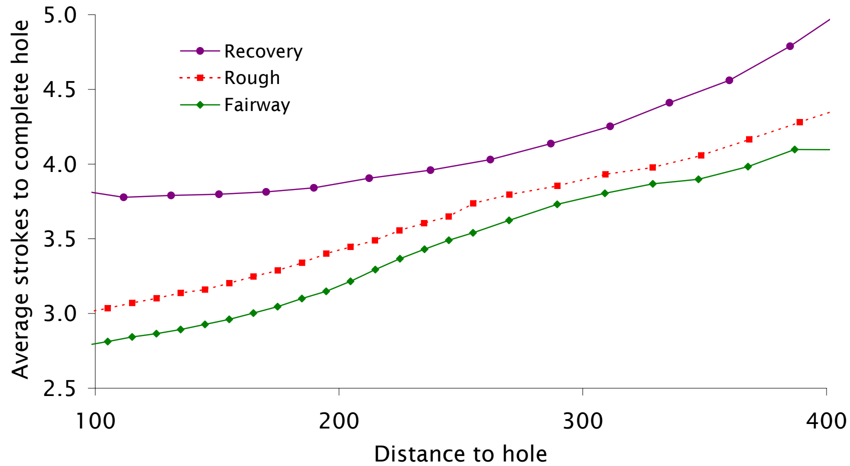

Figure 6: Average strokes to complete the hole for recovery shots and shots from the rough and fairway for PGA TOUR golfers in 2003-2010. Most recovery shots are in the range between 150 and 300 yards from the hole. In this range, the average number of strokes to complete the hole is 0.6 strokes greater from a recovery position than from the fairway and 0.4 greater than from the rough.

the same course can change dramatically because of different weather conditions, especially wind which affects the flight of the ball. Of all of these factors, only the length of the course is directly included in the benchmark average score. So a golfer who shoots 12 under par in the tournament with a winning score of 30 under par is likely to have played relatively worse than a golfer with a score of 4 under par where the winning score is 6 under par. In order to make a direct comparison of two golfers who play in a different set of tournaments, it is necessary to adjust scores and strokes gained for the difficulty of the course for each round.

Let g_{ij} represent the total (18-hole) strokes gained for golfer i playing on a course and round indexed by j. In order to separate golfer skill from the course difficulty for that round, the strokes gained, g_{ij}, is modeled as

$$g_{ij} = \mu_i + \delta_j + \epsilon_{ij} \tag{8}$$

where μ_i represents golfer i's intrinsic skill (i.e., the golfer's average strokes gained on a PGA TOUR course of average difficulty), δ_j represents the intrinsic difficulty of the course-round j, and ϵ_{ij} is a random mean-zero error term.[2] The model is estimated using a standard iterative procedure (see, e.g., Larkey (1994), Berry (2001) and Connolly and Rendleman (2008)).

[2]In the ShotLink data, when a tournament is played on multiple courses, data is collected only at one course. There are examples in the data with two different tournaments played on different courses on the same days, and these are represented by different course-round indices j. The δ_j are estimated as random effects.

4.2. Golfer strokes gained results and rankings

The main golfer results are given in Tables 1−3. Table 1 gives PGA TOUR golfer rankings based on the entire 2003-2010 data which contains over eight million shots. Ranks are relative to the 299 golfers with 120 or more rounds in the data.[3] Tiger Woods' total strokes gained per round is 3.20, which means that he gains, on average, 3.20 strokes per 18-hole round versus an average PGA TOUR field. That is, 3.20 represents the μ for Tiger Woods estimated from equation (8). Tiger is ranked first in this category, with second place occupied by Jim Furyk, who gains 2.12 strokes per 18-hole round versus an average PGA TOUR field. The difference between the two is an enormous 1.08 strokes per round. Differences between lower ranks are much smaller: the average difference is 0.08 strokes between ranks 2 and 10 and 0.01 strokes between ranks 95 and 105. Not only was Tiger the best golfer between 2003 and 2010, but he was the best by a large margin.

The strokes gained approach gives direct insight into where Tiger Woods gained the 3.20 strokes per round. Table 1 shows that 2.08 strokes came from the long game (rank 1), 0.42 strokes from the short game (rank 16), and 0.70 strokes from putting (rank 3). Tiger dominates the competition because he excels in every category, but his long game contributes 65% (2.08/3.20) to his total strokes gained relative to an average field. Many people have commented on Tiger's superior putting, and the strokes gained analysis is consistent with this observation: he is ranked of third with a gain of 0.70 putts per round. However, his gain from putting is less than the 1.01 strokes he gains between 150 and 250 yards from the hole, and comparable to his long tee shots, where he gains 0.70 strokes per round versus the field.

Table 1 shows average strokes gained for the top ten golfers, and the long game contributes 65% (1.20/1.84) to their total strokes gained relative to an average field. The bottom ten golfers, ranks 290-299, lose 71% (−0.97/−1.36) of their strokes in the long game. The top ten golfers in total strokes gained are all ranked in the the top 70 in long game strokes gained, but four of these golfers are not ranked in the top 100 in putting. The bottom ten golfers in total strokes gained are all ranked worse than 200 in long game strokes gained. These results suggest that the long game is the most important factor that differentiates golfers at the elite PGA TOUR level.

Table 2 focuses on the long game, and shows that Tiger is ranked in the top 10 in each of the long game subcategories. Of his 2.08 long game strokes gained, 1.01 strokes are gained between 150 and 250 yards from the hole. Although Tiger is known for his occasional wild drives, in the long tee shot category he is ranked 7 (out of 299), and he gains 0.70 strokes per round versus the field, the same as his gain from putting.

Table 3 gives short game strokes gained results during 2003-2010. Steve Stricker had the best short game overall, while Mike Weir and Luke Donald had the best greenside sand games. Table 4 focuses on putting and shows that David Frost, Brad Faxon and Tiger Woods were the top three

[3] The rankings are based on strokes gained per round. An argument can be made that a better measure of skill is strokes gained per stroke, however both approaches give very similar results. Strokes gained per round is used here because the additivity property makes it easier to see how total strokes gained splits into long game, short game and putting strokes gained.

Table 1: Total strokes gained per round, broken down into three categories: long game (shots over 100 yards from the hole), short game (shots under 100 yards from the hole, excluding putts) and putting (shots on the green, not including the fringe). Ranks are out of the 299 PGA TOUR golfers with at least 120 rounds during 2003-2010.

	Rank				Strokes gained			
Golfer	Total	Long	Short	Putt	Total	Long	Short	Putt
Woods, Tiger	1	1	16	3	3.20	2.08	0.42	0.70
Furyk, Jim	2	10	10	14	2.12	1.13	0.47	0.52
Singh, Vijay	3	2	5	195	2.05	1.63	0.51	−0.09
Els, Ernie	4	4	15	153	1.86	1.40	0.44	0.01
Mickelson, Phil	5	12	12	95	1.72	1.11	0.47	0.15
Donald, Luke	6	65	7	9	1.55	0.46	0.50	0.58
Goosen, Retief	7	19	22	46	1.52	0.90	0.33	0.29
Garcia, Sergio	8	5	60	220	1.47	1.39	0.23	−0.15
Scott, Adam	9	7	53	201	1.46	1.33	0.24	−0.11
Harrington, Padraig	10	54	4	42	1.44	0.57	0.56	0.31
				Average	1.84	1.20	0.42	0.22
Boros, Guy	290	283	292	91	−1.14	−0.87	−0.43	0.16
McGovern, Jim	291	293	158	197	−1.15	−1.05	−0.01	−0.09
Waite, Grant	292	279	120	282	−1.17	−0.79	0.07	−0.45
Begay III, Notah	293	265	194	286	−1.23	−0.67	−0.09	−0.48
Bolli, Justin	294	267	274	264	−1.27	−0.69	−0.25	−0.32
Veazey, Vance	295	294	246	178	−1.33	−1.10	−0.19	−0.04
McCallister, Blaine	296	262	273	294	−1.49	−0.64	−0.25	−0.60
Gossett, David	297	292	103	295	−1.49	−1.01	0.12	−0.61
Duval, David	298	297	219	143	−1.51	−1.41	−0.14	0.03
Perks, Craig	299	298	195	249	−1.79	−1.44	−0.09	−0.26
				Average	−1.36	−0.97	−0.12	−0.27
Notable golfers								
Couples, Fred	29	28	37	209	1.00	0.84	0.28	−0.12
Villegas, Camilo	30	13	126	212	0.99	1.05	0.06	−0.13
Westwood, Lee	43	17	253	129	0.83	0.97	−0.20	0.06
Pavin, Corey	99	252	8	26	0.33	−0.57	0.48	0.42
Durant, Joe	117	9	267	299	0.20	1.14	−0.24	−0.70
O'Meara, Mark	236	284	89	62	−0.49	−0.87	0.15	0.24
Daly, John	238	138	254	272	−0.50	0.08	−0.21	−0.38

putters during 2003-2010. Sergio Garcia is ranked 220 in putting overall: 271 in short putts, 179 in medium putts and 85 in long putts. It is clear that the shorter the putt, the more trouble he has. Sergio's total putting strokes gained is −0.15, so he loses 0.85 strokes per round to Tiger Woods just from putting. Brad Faxon gains 0.71 strokes on the field in putting but loses 0.96 in the long game (see Table 2).

Table 5 shows strokes gained results for Tiger Woods by year. Tiger Woods was ranked first in total strokes gained in each year from 2003 to 2009. Tiger Woods had the worst year of his career in 2010, with his total strokes gained per round decreasing by three compared with 2009. His game faltered across the board, dropping 1.19 strokes in his long game, 0.89 in his short game, and 0.91

Table 2: Long game strokes gained per round, broken down into five categories: long tee shots (tee shots starting over 250 yards from the hole), approach shots 100-150 yards from the hole, approach shots 150-200 yards from the hole, approach shots 200-250 yards from the hole and shots over 250 yards from the hole (excluding tee shots). For space reasons, recovery shots and sand shots greater than 100 yards from the hole are not reported (but are included in the total long game strokes gained). Ranks are out of the 299 golfers with at least 120 rounds during 2003-2010.

	Rank						Strokes gained per round					
Golfer	Long total	Long tee	100−150	150−200	200−250	>250	Long total	Long tee	100−150	150−200	200−250	>250
Woods, Tiger	1	7	8	1	1	1	2.08	0.70	0.20	0.66	0.35	0.14
Singh, Vijay	2	3	20	8	7	14	1.63	0.81	0.16	0.33	0.19	0.07
Allenby, Robert	3	14	4	6	2	47	1.59	0.61	0.25	0.38	0.26	0.05
Els, Ernie	4	16	14	2	15	26	1.40	0.55	0.18	0.41	0.16	0.06
Garcia, Sergio	5	15	13	13	4	17	1.39	0.55	0.18	0.31	0.23	0.07
Perry, Kenny	6	6	25	10	18	95	1.37	0.73	0.15	0.32	0.15	0.03
Scott, Adam	7	18	5	12	48	16	1.33	0.54	0.25	0.31	0.10	0.07
Weekley, Boo	8	2	58	84	25	113	1.19	0.83	0.09	0.09	0.13	0.02
Durant, Joe	9	11	43	22	40	136	1.14	0.67	0.11	0.24	0.11	0.01
Furyk, Jim	10	51	7	4	12	134	1.13	0.32	0.21	0.40	0.18	0.01
						Average	1.40	0.61	0.17	0.35	0.19	0.06
Notable golfers												
Couples, Fred	28	28	22	92	26	117	0.84	0.47	0.16	0.08	0.12	0.02
Daly, John	138	55	242	252	123	128	0.08	0.31	−0.08	−0.15	0.03	0.02
Faxon, Brad	289	297	79	190	223	279	−0.96	−0.82	0.07	−0.05	−0.05	−0.08
Duval, David	297	298	249	211	293	164	−1.41	−1.08	−0.09	−0.08	−0.17	−0.00

Table 3: Short game strokes gained per round, broken down into three distance categories: 0-20 yards from the hole, 20-60 yards from the hole, 60-100 yards from the hole (excluding sand and recovery shots and putts). Greenside sand shots within 50 yards of the hole ('sand') are given in a separate category. For space reasons, 0-100 yard recovery shots and 50-100 yard sand shots are not reported (but are included in the total short game strokes gained). Ranks are out of the 299 golfers with at least 120 rounds during 2003-2010.

	Rank					Strokes gained per round				
Golfer	Short	0-20	20-60	60-100	Sand	Short	0-20	20-60	60-100	Sand
Stricker, Steve	1	7	1	1	59	0.69	0.19	0.22	0.17	0.08
Olazabal, Jose Maria	2	1	27	66	7	0.57	0.30	0.10	0.04	0.15
Riley, Chris	3	9	4	45	3	0.56	0.18	0.15	0.06	0.15
Harrington, Padraig	4	5	11	4	39	0.56	0.21	0.12	0.15	0.10
Singh, Vijay	5	14	9	53	4	0.51	0.15	0.12	0.05	0.15
Weir, Mike	6	53	15	40	1	0.51	0.09	0.11	0.06	0.21
Donald, Luke	7	3	84	49	2	0.50	0.23	0.04	0.06	0.17
Pavin, Corey	8	4	30	20	23	0.48	0.21	0.09	0.08	0.12
Imada, Ryuji	9	18	21	38	19	0.48	0.15	0.10	0.06	0.12
Furyk, Jim	10	6	17	13	66	0.47	0.20	0.11	0.09	0.07
					Average	0.53	0.19	0.12	0.08	0.13
Notable golfers										
Haas, Jay	11	11	143	10	8	0.47	0.17	0.01	0.11	0.14
Mickelson, Phil	12	15	6	42	20	0.47	0.15	0.13	0.06	0.12
Woods, Tiger	16	22	8	47	64	0.42	0.13	0.13	0.06	0.07
Garcia, Sergio	60	64	75	60	101	0.23	0.08	0.05	0.04	0.04
Westwood, Lee	253	260	162	88	286	-0.20	−0.10	−0.00	0.02	−0.14
Daly, John	254	290	182	179	90	-0.21	−0.17	−0.01	−0.01	0.05

Table 4: Putting strokes gained per round, broken down into three distance categories: short putts (0-6 feet), medium length putts (7-21 feet) and long putts (22 feet and over). Ranks are out of the 299 golfers with at least 120 rounds during 2003-2010.

	Rank				Strokes gained per round			
Golfer	Putt	0-6 ft	7-21 ft	22+ ft	Putt	0-6 ft	7-21 ft	22+ ft
Frost, David	1	83	1	1	0.72	0.08	0.42	0.22
Faxon, Brad	2	21	3	2	0.71	0.19	0.31	0.21
Woods, Tiger	3	11	4	3	0.70	0.21	0.31	0.19
Crane, Ben	4	1	10	24	0.67	0.29	0.27	0.11
Roberts, Loren	5	4	13	13	0.65	0.25	0.26	0.14
Baddeley, Aaron	6	9	9	7	0.64	0.22	0.27	0.15
Chalmers, Greg	7	2	14	37	0.62	0.27	0.26	0.09
Parnevik, Jesper	8	3	27	9	0.61	0.25	0.21	0.15
Donald, Luke	9	14	17	16	0.58	0.20	0.24	0.13
Cink, Stewart	10	28	7	22	0.58	0.17	0.29	0.12
				Average	0.65	0.21	0.28	0.15
Notable golfers								
Stricker, Steve	19	15	60	19	0.46	0.20	0.13	0.13
Pavin, Corey	26	97	23	17	0.42	0.06	0.22	0.13
Mickelson, Phil	95	68	139	102	0.15	0.09	0.02	0.04
Singh, Vijay	195	152	252	97	-0.09	0.01	−0.14	0.04
Couples, Fred	209	294	102	41	-0.12	−0.28	0.07	0.08
Garcia, Sergio	220	271	179	85	-0.15	−0.16	−0.03	0.04
Daly, John	272	261	272	247	-0.38	−0.12	−0.19	−0.07

Table 5: Results for Tiger Woods, by year. Ranks for individual years are out of approximately 220 golfers with at least 30 rounds during each year. An exception was made to show Tiger Woods in 2008, even though he only played in three PGA TOUR events. Ranks for 2003-2010 are out of the 299 golfers with at least 120 rounds.

		Rank				Strokes gained			
	Year	Total	Long	Short	Putt	Total	Long	Short	Putt
Tiger Woods	2010	48	28	160	91	0.71	0.83	−0.20	0.08
	2009	1	1	4	2	3.70	2.02	0.70	0.99
	2008	1	1	3	4	4.14	2.56	0.72	0.85
	2007	1	1	24	2	3.68	2.47	0.41	0.80
	2006	1	1	16	21	3.78	2.83	0.45	0.49
	2005	1	1	98	5	2.82	2.03	0.09	0.70
	2004	1	5	11	3	3.07	1.62	0.49	0.96
	2003	1	2	3	16	3.71	2.44	0.72	0.55
	2003-2010	1	1	16	3	3.20	2.08	0.42	0.70

	Rank						Strokes gained per round					
Year	Long total	Long tee	100−150	150−200	200−250	>250	Long total	Long tee	100−150	150−200	200−250	>250
2010	28	123	29	2	44	16	0.83	−0.08	0.16	0.48	0.12	0.10
2009	1	18	25	1	1	2	2.02	0.53	0.16	0.79	0.43	0.15
2008	1	7	9	1	1	51	2.56	0.60	0.25	1.17	0.40	0.05
2007	1	4	1	1	4	1	2.47	0.81	0.38	0.83	0.30	0.17
2006	1	4	52	1	1	1	2.83	0.91	0.13	0.94	0.62	0.16
2005	1	1	6	16	28	3	2.03	1.09	0.29	0.35	0.14	0.15
2004	5	17	54	2	9	7	1.62	0.53	0.13	0.58	0.24	0.12
2003	2	6	38	2	1	3	2.44	0.87	0.14	0.59	0.59	0.15
2003-2010	1	7	8	1	1	1	2.08	0.70	0.20	0.66	0.35	0.14

	Rank					Strokes gained per round				
Year	Short	0-20	20-60	60-100	Sand	Short	0-20	20-60	60-100	Sand
2010	160	169	72	135	173	−0.20	−0.10	0.04	−0.02	−0.11
2009	4	6	1	67	14	0.70	0.25	0.25	0.03	0.17
2008	3	12	1	29	144	0.72	0.23	0.42	0.08	−0.05
2007	24	77	22	22	85	0.41	0.06	0.12	0.10	0.03
2006	16	17	108	18	90	0.45	0.20	0.02	0.12	0.03
2005	98	143	150	67	51	0.09	−0.03	−0.03	0.05	0.08
2004	11	70	2	40	86	0.49	0.08	0.27	0.07	0.04
2003	3	3	34	79	7	0.72	0.41	0.10	0.03	0.17
2003-2010	16	22	8	47	64	0.42	0.13	0.13	0.06	0.07

	Rank				Strokes gained per round			
Year	Putt	0-6 ft	7-21 ft	22+ ft	Putt	0-6 ft	7-21 ft	22+ ft
2010	91	58	98	150	0.08	0.11	0.03	−0.06
2009	2	1	40	1	0.99	0.47	0.20	0.31
2008	4	29	12	5	0.85	0.20	0.40	0.25
2007	2	62	3	4	0.80	0.10	0.44	0.26
2006	21	32	58	17	0.49	0.17	0.12	0.20
2005	5	27	15	10	0.70	0.19	0.31	0.20
2004	3	53	2	9	0.96	0.12	0.62	0.22
2003	16	7	32	61	0.55	0.26	0.23	0.07
2003-2010	3	11	4	3	0.70	0.21	0.31	0.19

Table 6: Total strokes gained per round for selected golfers, by year. Ranks for individual years are out of approximately 220 golfers with at least 30 rounds during each year. Ranks for 2003-2010 are out of the 299 golfers with at least 120 rounds.

		Rank				Strokes gained			
	Year	Total	Long	Short	Putt	Total	Long	Short	Putt
Jim Furyk	2010	3	26	2	22	2.03	0.90	0.64	0.49
	2009	3	31	8	4	2.12	0.80	0.53	0.80
	2008	6	17	61	28	1.62	0.98	0.20	0.44
	2007	10	14	8	105	1.68	1.02	0.62	0.04
	2006	2	3	17	3	2.94	1.69	0.44	0.81
	2005	4	6	8	26	2.27	1.32	0.53	0.41
	2004	22	33	100	16	1.50	0.83	0.06	0.60
	2003	4	13	5	10	2.55	1.31	0.61	0.62
	2003-2010	2	10	10	14	2.12	1.13	0.47	0.52
Vijay Singh	2010	30	3	33	196	1.05	1.42	0.31	−0.68
	2009	70	28	61	186	0.40	0.83	0.18	−0.61
	2008	4	4	4	177	1.80	1.55	0.63	−0.38
	2007	9	6	19	107	1.75	1.25	0.47	0.03
	2006	6	6	9	90	2.07	1.43	0.53	0.10
	2005	2	5	3	63	2.58	1.77	0.61	0.20
	2004	2	1	7	120	2.86	2.28	0.61	−0.03
	2003	2	3	6	63	3.06	2.23	0.60	0.23
	2003-2010	3	2	5	195	2.05	1.63	0.51	−0.09
Ernie Els	2010	7	16	37	28	1.75	1.05	0.28	0.42
	2009	16	6	23	152	1.37	1.35	0.32	−0.30
	2008	25	10	26	190	1.10	1.16	0.40	−0.46
	2007	2	2	27	104	2.16	1.74	0.38	0.04
	2006	8	15	3	96	1.94	1.11	0.75	0.07
	2005	3	3	60	32	2.37	1.80	0.18	0.39
	2004	3	4	5	79	2.48	1.63	0.69	0.16
	2003	10	8	16	159	1.90	1.67	0.51	−0.29
	2003-2010	4	4	15	153	1.86	1.40	0.44	0.01
Phil Mickelson	2010	10	10	15	118	1.49	1.16	0.39	−0.05
	2009	19	23	13	119	1.29	0.92	0.42	−0.05
	2008	1	5	8	50	2.25	1.40	0.57	0.27
	2007	3	11	4	59	2.06	1.14	0.69	0.23
	2006	5	4	32	66	2.13	1.58	0.34	0.20
	2005	8	18	7	49	1.82	0.98	0.54	0.30
	2004	10	8	25	123	1.79	1.44	0.39	−0.04
	2003	46	94	30	54	0.87	0.19	0.39	0.29
	2003-2010	5	12	12	95	1.72	1.11	0.47	0.15
Steve Stricker	2010	1	9	3	15	2.36	1.17	0.64	0.55
	2009	2	9	1	56	2.23	1.18	0.75	0.30
	2008	14	112	1	26	1.31	0.03	0.83	0.46
	2007	5	25	3	25	1.97	0.82	0.73	0.41
	2006	18	113	2	20	1.47	0.15	0.82	0.50
	2005	129	213	2	8	−0.05	−1.38	0.65	0.68
	2004	144	213	9	12	−0.22	−1.41	0.53	0.67
	2003	141	188	11	95	−0.35	−0.91	0.53	0.03
	2003-2010	22	160	1	19	1.13	−0.02	0.69	0.46

in his putting. His combined results for 2003-2010 shows he is the best golfer of his era because of his all-round excellence in every category, with his long game contributing 65% (2.08/3.20) of his total strokes gained versus the field.

Table 6 shows strokes gained results for selected golfers by year. Steve Stricker was the comeback player of the year in 2006 when his total strokes gained increased from -0.05 to 1.47, moving from rank 129 to 18. The improvement was almost entirely due to a better long game, with a long game strokes gained increase from -1.38 to 0.15. He was also the comeback player of the year in 2007 where his total strokes gained increased from 1.47 to 1.97, moving from rank 18 to 5.

4.3. Influence of skill factors on golf scores

Many people claim that the short game and putting are the most important determinants of golf scores. For example, Pelz (1999, p.1) writes, "60% to 65% of all golf shots occur inside 100 yards of the hole. More important, about 80% of the shots golfers lose to par occur inside 100 yards." Several academic studies have reached similar conclusions. In contrast, strokes gained analysis of PGA TOUR data shows that the long game is the most important factor explaining the variability in professional golf scores.

For a single golfer, the relative contribution of each skill category can be assessed directly by comparing strokes gained by skill category. Across golfers the relative contributions can be assessed using variance and correlation analysis. Equation 8 is used to estimate μ_i, the mean total strokes gained of golfer i and also the mean strokes gained of long game shots (μ_i^L), short game shots (μ_i^S) and putts (μ_i^P). Note that $\mu_i = \mu_i^L + \mu_i^S + \mu_i^P$, and all quantities represent 18-hole round averages estimated using equation (8). For notational convenience we drop the golfer subscript i. Then $\mathrm{Var}(\mu) = \mathrm{Var}(\mu^L) + \mathrm{Var}(\mu^S) + \mathrm{Var}(\mu^P) + 2\,\mathrm{Cov}(\mu^L, \mu^S) + 2\,\mathrm{Cov}(\mu^L, \mu^P) + 2\,\mathrm{Cov}(\mu^S, \mu^P)$ (where each of the terms represents the variance or covariance across golfers). A unique decomposition of $\mathrm{Var}(\mu)$ is complicated because of the covariance terms. However, the covariance terms are quite small and $V \equiv \mathrm{Var}(\mu^L) + \mathrm{Var}(\mu^S) + \mathrm{Var}(\mu^P) \approx \mathrm{Var}(\mu)$.[4] So define the contributions of the long game, short game and putting to total strokes gained by: $\mathrm{Var}(\mu^L)/V$, $\mathrm{Var}(\mu^S)/V$ and $\mathrm{Var}(\mu^P)/V$, respectively. More variability in a strokes gained category means that golfers have more opportunity to distinguish themselves as better or worse golfers. Using data from 2003-2010 for golfers with at least 120 rounds, the contributions to total strokes gained are 72%, 11% and 17% for the long game, short game and putting, respectively. By this measure, *the long game explains more than two-thirds of the variation in total strokes gained.*

Correlation results across golfers are summarized in Table 7. When the three broad skill categories are further divided, approach and tee shots in the 150-200 yard range are seen to have the highest correlation with total strokes gained, with a correlation of 74%. Table 7 shows that the correlation of putts gained with long game strokes gained across golfers is -14% (with a standard

[4] Using data from 2003-2010 for golfers with at least 120 rounds gives: $\mathrm{Var}(\mu) = 0.50$, $\mathrm{Var}(\mu^L) = 0.35$, $\mathrm{Var}(\mu^S) = 0.06$, $\mathrm{Var}(\mu^P) = 0.08$, $\mathrm{Cov}(\mu^L, \mu^S) = 0.01$, $\mathrm{Cov}(\mu^L, \mu^P) = -0.02$ and $\mathrm{Cov}(\mu^S, \mu^P) = 0.03$.

error of 7%). At the tournament professional level, these skill factors are nearly uncorrelated, as illustrated in Figure 7. The slight negative correlation can be explained by survivorship bias: golfers with a subpar long game need better than average putting (and/or short game) to survive on the PGA TOUR.

Table 7: Correlation results using 2003-2010 data for all PGA TOUR golfers with 120 or more rounds. Top panel: *Total* refers to the total strokes gained per 18-hole round. *Long* refers to the total strokes gained per 18-hole round for shots over 100 yards from the hole. *Short* refers to the total strokes gained per 18-hole round for shots under 100 yards from the hole excluding putts. *Putt* refers to the total strokes gained for per 18-hole round for shots on the green. The bottom two rows give the average number of shots and fractions of shots in each category. Bottom panel: correlations of subcategories with total strokes gained. Standard errors (computed by with standard bootstrapping procedure) are given in parentheses.

	Total (se)	Long (se)	Short (se)	Putt (se)
Total	100%			
Long	79% (2%)	100%		
Short	54% (4%)	6% (6%)	100%	
Putt	41% (5%)	−14% (7%)	39% (4%)	100%
Number of strokes	71.1	32.2	9.8	29.1
Fraction of strokes		45%	14%	41%

	Long game					Short game				Putt		
	Long tee	100−150	150−200	200−250	> 250	0-20	20-60	60-100	Sand	0-6	7-21	22+
	54%	61%	74%	66%	53%	50%	37%	44%	33%	27%	37%	40%
	(4%)	(4%)	(3%)	(4%)	(4%)	(4%)	(5%)	(5%)	(5%)	(6%)	(5%)	(5%)
Number:	13.9	4.8	7.1	3.2	1.6	4.3	2.1	1.6	1.7	16.0	7.9	5.3
Fraction:	19.6%	6.7%	10.0%	4.5%	2.3%	6.0%	3.0%	2.1%	2.4%	22.4%	11.1%	7.4%

Correlation and variability should not equated with importance. If every professional golfer hit every drive 320 yards in the middle of the fairway, then long tee shots would have zero correlation with score and the variability in long tee strokes gained would be zero. In this example, the golfers are not differentiating themselves with their long tee shots—they all happen to be equally outstanding in this skill category. However, it is still important to be a good driver of the ball: a golfer who doesn't hit his drives 320 yards in the fairway in this example will not survive on the tour for long.

4.4. Course difficulty factors

The estimation of course-round difficulty parameters, δ_j in equation (8), allows courses to be ranked just as golfers were ranked. By using individual shot data, course difficulty be further explained and broken down into difficulty of long game shots, short game shots, and putts. A related question of the difficulty of winning a tournament is studied Connolly and Rendleman (2010). For handicapping purposes, the USGA rates course difficulty for zero-handicap (scratch)

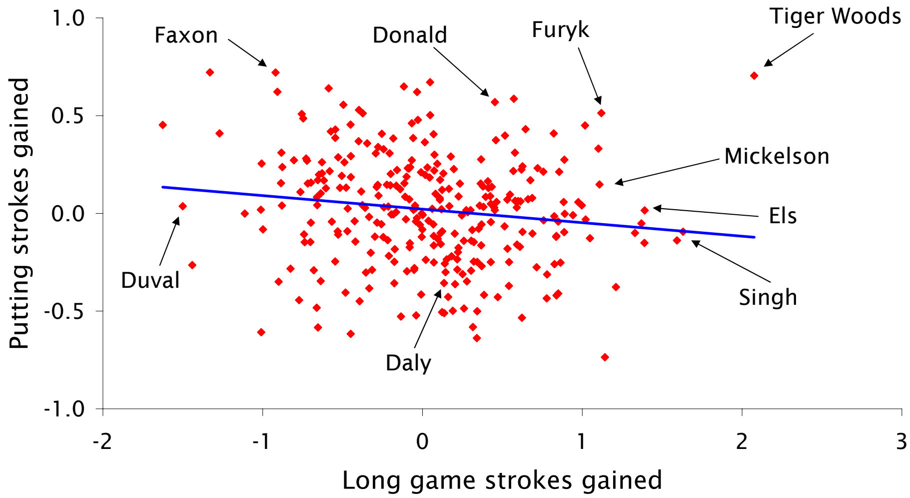

Figure 7: Scatter chart of putts gained versus long game strokes gained using 2003-2010 data. Each data point represents the results for a single golfer; a few golfers are indicated to illustrate. The regression trendline shows a slight negative correlation between the two skill categories (the correlation is -14% with a standard error of 7%).

and bogey golfers by tabulating hole distances, counting the number and severity of bunkers and other hazards, etc. Their method does not use scores nor shot information. To the best of our knowledge, this is the first attempt at ranking courses using shot data and the first to break down course difficulty by shot categories.

Define the difficulty factors for each course to be the average value of $-\delta_j$ for all rounds played at that course. The negative sign is used so that the most difficult courses are ranked at the top. Table 8 shows the ten most difficult and the ten easiest courses that hosted tournaments during 2003-2010 and had at least 12 rounds of data. The TPC Sawgrass course, host of the Players Championship and famous for the island green on its 17th hole, is ranked as the most difficult course on the PGA TOUR. The strokes gained approach explicitly accounts for the length of the course, so courses are rated as more difficult because of trees, hazards, rough height, firmness and contours of the greens, etc. The strokes gained approach enables us to see which part of the course contributes most to its difficulty and allows courses to be ranked for difficulty in the long game, short game and putting. For example, Westchester Country Club is rated as the most difficult course for the short game and putting and Harbour Town Golf Links is the most difficult course in the long game category.

4.5. Effect of the groove rule change

The USGA, golf's rule-making body in the United States, recently changed the rules regarding the grooves in irons. The rules were changed because of the perception that equipment advances in the past decade made shots from the rough easier: clubs with sharper grooves allow highly skilled golfers to impart more spin on the ball from the rough and stop the ball closer to the hole. The

Table 8: Ranking of courses by difficulty factors. Ranks are out of the 45 courses that hosted PGA TOUR tournaments and had at least 12 rounds of data during 2003-2010.

	Rank				Difficulty factors			
Course	Total	Long	Short	Putt	Total	Long	Short	Putt
TPCSawgrass	1	2	3	9	2.41	1.72	0.47	0.23
WestchesterCC	2	19	1	1	1.70	0.16	0.84	0.70
HarbourTownGolfLinks	3	1	27	18	1.69	1.78	−0.12	0.03
MuirfieldVillageGC	4	7	2	10	1.68	0.89	0.64	0.15
BayHillClub	5	3	28	11	1.54	1.58	−0.12	0.09
PebbleBeachGolfLinks	6	6	35	3	1.32	0.91	−0.21	0.62
WestinInnisbrook-Copperhead	7	5	22	12	1.20	1.21	−0.10	0.09
PGANationalChampionCourse	8	4	33	39	0.94	1.38	−0.20	−0.24
QuailHollowClub	9	16	12	4	0.77	0.22	0.05	0.50
TorreyPinesSouthCourse	10	14	25	8	0.48	0.36	−0.11	0.24
				Average	1.37	1.02	0.11	0.24
LaCanteraGC	36	36	32	27	−1.07	−0.80	−0.16	−0.11
TucsonNat'lGolf	37	39	24	14	−1.11	−1.07	−0.10	0.07
WarwickHillsG&CC	38	35	37	29	−1.16	−0.77	−0.26	−0.13
MagnoliaGC	39	33	38	38	−1.19	−0.69	−0.29	−0.21
ForestOaksCC	40	37	30	41	−1.33	−0.94	−0.14	−0.25
AtunyoteGolfClub	41	40	40	17	−1.42	−1.13	−0.32	0.03
TPCSummerlin	42	44	6	43	−1.60	−1.67	0.33	−0.26
TPCDeereRun	43	43	23	44	−1.75	−1.29	−0.10	−0.36
SedgefieldCountryClub	44	45	18	19	−1.97	−1.93	−0.06	0.02
En-JoieGC	45	41	45	30	−1.99	−1.17	−0.69	−0.13
				Average	−1.46	−1.15	−0.18	−0.13

purpose of the new rule is to "roll back" these equipment advances, so that shots from the rough will have less spin and the rough more of a penalty compared to the fairway. The rules were put into place on the PGA TOUR at the start of the 2010 season.

To estimate the effect of the groove rule change, benchmark functions representing the average strokes to complete a hole are estimated for the fairway and rough for each year. Recovery shots (as described in Section 3.4) are excluded, so the rough benchmark functions are not biased by these shots. Define the rough penalty to be the difference in the average strokes to complete the hole between the rough and fairway at comparable distances to the hole. For example, from 120 yards in the fairway the average number of strokes to complete the hole is 2.85 and from the rough it is 3.08. The penalty for being in the rough at 120 yards from the hole is an increase of 0.23 strokes. Since the rough penalty varies slightly by distance, results are given for the average rough penalty between 50 and 150 yards from the hole, where the rule is designed to have the maximum impact.

Figure 8 shows a decline in the rough penalty from 2003-2009. Surprisingly, the figure shows a large drop from 2008 to 2009, *prior* to when the groove rule change went into effect. Figure 8 shows that the rough penalty was unchanged at 0.20 in 2009 and 2010 (with standard errors of 0.004). Tests of ball spin indicate a measurable impact of the rule change, so it is puzzling that there has been little impact on scores. Differences in the height, thickness and moisture of the rough and the firmness of the greens will influence the results and these factors should, if possible, be incorporated in the analysis. Another possible explanation is that the golfers adapted their swings and strategy in order to minimize the impact of the change in ball spin. These issues are left for future research.

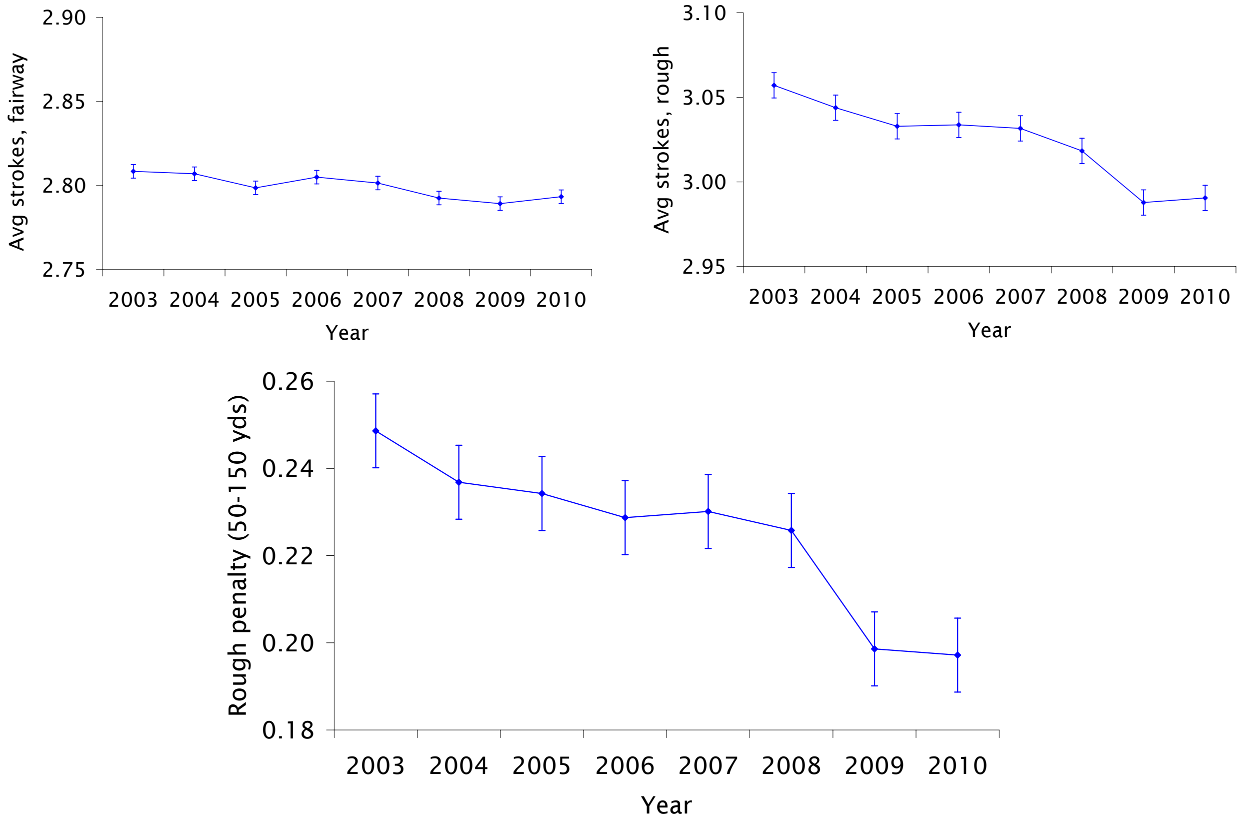

Figure 8: Upper charts: average strokes to complete the hole from the fairway and rough. Lower chart: rough penalty (the difference between the rough and fairway values). All three charts show results by year for shots starting between 50 and 150 yards from the hole. Two standard error bars are shown in each chart (standard errors were computed with a standard bootstrapping procedure).

5. Concluding remarks

The availability of detailed golf shot data makes it possible to create golf measures that allow consistent comparisons between different parts of the game. Using the starting and ending locations of each shot, strokes gained gives the number of strokes a golfer gains or loses relative to an average PGA TOUR tournament field. Analysis of over eight million shots on the PGA TOUR in 2003-2010 shows that the long game (defined as shots starting over 100 yards from the hole) accounts for more than two-thirds of the scoring differences between PGA TOUR golfers. In the 2003-2010 data, Tiger Woods led in total strokes gained, with a gain of 3.20 strokes per 18-hole round. He gained 2.08 strokes (65% of the total) in the long game. A preliminary analysis of the impact of the new groove rule for irons that went into effect on the PGA TOUR in 2010 showed, somewhat surprisingly, that there has been almost no impact of the rule on scores.

References

Bellman, R.E. 1957. *Dynamic Programming*. Princeton University Press, Princeton, NJ.

Berry, S.M. 1999. Drive for show and putt for dough. *Chance* **12**(4) 50–55.

Berry, S.M. 2001. How ferocious is Tiger? *Chance* **14**(3) 51–56.

Broadie, M. 2008. Assessing golfer performance using Golfmetrics. D. Crews, R. Lutz, eds., *Science and Golf V: Proceedings of the World Scientific Congress of Golf*. Energy in Motion, Inc., Mesa, Arizona, 253–262.

Broadie, M., M. Bansal. 2008. A simulation model to analyze the impact of hole size on putting in golf. S.J. Mason, R.R. Hill, L. Moench, O. Rose, eds., *Proceedings of the 2008 Winter Simulation Conference*. The Society for Computer Simulation, 2826–2834.

Broadie, M., S. Ko. 2009. A simulation model to analyze the impact of distance and direction on golf scores. M.D. Rossetti, R.R. Hill, B. Johansson, A. Dunkin, R.G. Ingalls, eds., *Proceedings of the 2009 Winter Simulation Conference*. The Society for Computer Simulation, 3109–3120.

Callan, S.J., J.M. Thomas. 2007. Modeling the determinants of a professional golfer's tournament earnings: A multiequation approach. *Journal of Sports Economics* **8** 394–411.

Cochran, A.J., J. Stobbs. 1968. *Search for the Perfect Swing: The Proven Scientific Approach to Fundamentally Improving Your Game*. Triumph Books, Chicago, Illinois.

Connolly, R.A., R.J. Rendleman. 2008. Skill, luck, and streaky play on the PGA tour. *Journal of the American Statistical Association* **103**(481) 74–88.

Connolly, R.A., R.J. Rendleman. 2010. What it takes to win on the PGA TOUR (if your name is "tiger" or if it isn't). Working paper, University of North Carolina.

Davidson, J.D., T.J. Templin. 1986. Determinants of success among professional golfers. *Research Quarterly for Exercise and Sport* **57** 60–67.

Deason, L. 2006. Shotlink a statistical superstar. URL `http://www.pgatour.com/story/9596346/`.

Diaz, J. 2010. Tiger, 10 years later. URL `http://www.golfdigest.com/golf-tours-news/us-open/2010-06/tiger-diaz-open`.

Farrally, M.R., A.J. Cochran, D.J. Crews, M.J. Hurdzan, R.J. Price, J.T. Snow, P.R. Thomas. 2003. Golf science research at the beginning of the twenty-first century. *Journal of Sports Sciences* **21**(9) 753–765.

Fearing, D., J. Acimovic, S. Graves. 2010. How to catch a Tiger: Understanding putting performance on the PGA tour. *SSRN eLibrary* URL `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1538300`.

Gelman, A., D. Nolan. 2002. A probability model for golf putting. *Teaching Statistics* **24** 93–95.

Holmes, B.W. 1991. Putting: How a golf ball and hole interact. *American Journal of Physics* **59(2)** 129–136.

Hurley, W.J. 2010. Operations research and golf. J. Cochran, ed., *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc., New York, to appear.

Landsberger, L.M. 1994. A unified golf stroke value scale for quantitative stroke-by-stroke assessment. A.J. Cochran, M.R. Farrally, eds., *Science and Golf II: Proceedings of the World Scientific Congress of Golf*. E & FN Spon, London, 216–221.

Larkey, P.D. 1994. Comparing players in professional golf. A.J. Cochran, M.R. Farrally, eds., *Science and Golf II: Proceedings of the World Scientific Congress of Golf*. E & FN Spon, London, 193–198.

Larkey, P.D., A.A Smith. 1999. All around improvements. M.R. Farrally, A.J. Cochran, eds., *Science and Golf III: Proceedings of the 1998 World Scientific Congress of Golf*. Human Kinetics, UK, 377–384.

Moy, R.L., T. Liaw. 1998. Determinants of professional golf tournament earnings. *American Economist* **42** 65–70.

Nero, P. 2001. Relative salary efficiency of PGA tour golfers. *American Economist* **45** 51–56.

Pelz, D. 1989. *Putt like the Pros: Dave Pelz's Scientific Way to Improve Your Stroke, Reading Greens and Lowering Your Score*. Harper & Row, New York.

Pelz, D. 1999. *Dave Pelz's Short Game Bible: Master the Finesse Swing and Lower Your Score*. Broadway Books, New York.

Pelz, D. 2000. *Dave Pelz's Putting Bible: The Complete Guide to Mastering the Green*. Doubleday, New York.

Penner, A.R. 2003. The physics of golf. *Reports on Progress in Physics* **66**(2) 131–171.

Pollock, S.M. 1974. A model for the evaluation of golf handicapping. *Operations Research* **22**(5) 1040–1050.

Puterman, M.L., S.M. Wittman. 2009. Match play: Using statistical methods to categorize PGA tour players' careers. *Journal of Quantitative Analysis in Sports* **5**(1).

Riccio, L.J. 1990. Statistical analysis of the average golfer. A.J. Cochran, ed., *Science and Golf: Proceedings of the First World Scientific Conference of Golf*. E & FN Spon, London, 153–158.

Scheid, F. 1977. An evaluation of the handicap system of the united states golf association. S.P. Ladany, R.E. Machol, eds., *Optimal Strategies in Sports*. North Holland, Amsterdam, 151–155.

Shmanske, S. 1992. Human capital formation in professional sports: Evidence from the PGA tour. *Atlantic Economic Journal* **20**(3) 66–80.

Shmanske, S. 2008. Skills, performance, and earnings in the tournament compensation model: Evidence from PGA tour microdata tour. *Journal of Sports Economics* **9**(6) 644–662.

Soley, C. 1977. *How Well Should You Putt? A Search for a Putting Standard*. Soley Golf Bureau.

Stroud, R.C., L.J. Riccio. 1990. Mathematical underpinnings of the slope handicap system. A.J. Cochran, ed., *Science and Golf: Proceedings of the First World Scientific Conference of Golf*. E & FN Spon, London, 135–140.

Sweeney, M. 2008. Woods's success starts with finishing touch. URL `http://www.nytimes.com/2008/04/07/sports/golf/07woods.html`.

A. Appendix

This appendix summarizes the benchmark average strokes-to-complete-the hole functions from tee, fairway, rough, sand and recovery positions.

Table 9: Average number of strokes to complete the hole for PGA TOUR golfers from various starting positions. Distance to the hole is measured in yards. Values are estimated using over eight million shots during 2003-2010.

Distance	Tee	Fairway	Rough	Sand	Recovery
10		2.18	2.34	2.43	3.45
20		2.40	2.59	2.53	3.51
30		2.52	2.70	2.66	3.57
40		2.60	2.78	2.82	3.71
50		2.66	2.87	2.92	3.79
60		2.70	2.91	3.15	3.83
70		2.72	2.93	3.21	3.84
80		2.75	2.96	3.24	3.84
90		2.77	2.99	3.24	3.82
100	2.92	2.80	3.02	3.23	3.80
120	2.99	2.85	3.08	3.21	3.78
140	2.97	2.91	3.15	3.22	3.80
160	2.99	2.98	3.23	3.28	3.81
180	3.05	3.08	3.31	3.40	3.82
200	3.12	3.19	3.42	3.55	3.87
220	3.17	3.32	3.53	3.70	3.92
240	3.25	3.45	3.64	3.84	3.97
260	3.45	3.58	3.74	3.93	4.03
280	3.65	3.69	3.83	4.00	4.10
300	3.71	3.78	3.90	4.04	4.20
320	3.79	3.84	3.95	4.12	4.31
340	3.86	3.88	4.02	4.26	4.44
360	3.92	3.95	4.11	4.41	4.56
380	3.96	4.03	4.21	4.55	4.66
400	3.99	4.11	4.30	4.69	4.75
420	4.02	4.19	4.40	4.83	4.84
440	4.08	4.27	4.49	4.97	4.94
460	4.17	4.34	4.58	5.11	5.03
480	4.28	4.42	4.68	5.25	5.13
500	4.41	4.50	4.77	5.40	5.22
520	4.54	4.58	4.87	5.54	5.32
540	4.65	4.66	4.96	5.68	5.41
560	4.74	4.74	5.06	5.82	5.51
580	4.79	4.82	5.15	5.96	5.60
600	4.82	4.89	5.25	6.10	5.70