Drive for Show, Putt for Dough: Rates of Return to Golf Skills, Events Played, and Age on the PGA Tour

John R. Watkins, Jr.[1]
Davidson College

Abstract

The "winner take all" structure of professional golf, where there is substantial incentive to perform at one's highest ability, invites investigation into the rates of return to golf skills on the PGA Tour. The biggest earner on tour in 2006, Tiger Woods, won $9,941,563 on the PGA Tour, while the 227th earner made just $65,494. Because professional golfers are thought to produce earnings by means of skill in each area of the game, I shall employ a production function to model the marginal effect of golf skills on earnings. I intend to determine if the advancement of golf technology in the last 12 years has affected the impact of each golf statistic on earnings. It turns out that the marginal effect of driving distance and driving accuracy is negligible, but players with exceptional ability in other areas of the game of golf are handsomely rewarded. Age also has a negligible marginal effect on earnings, while events played does have a noticeable effect. Greens in regulation, putting, and the short game are important in determining the magnitude of player's earnings.

---

1   John R. Watkins, Jr. is a third year student concentrating in economics and history at the Davidson College. He would like to thank Dr. Kelly Chaston for her guidance. John may be reached at jowatkins@davidson.edu.

I. Introduction

Professional golf attracts fans that pay to watch golf at the highest level. Unlike team sports such as baseball or football, however, tournament officials determine player compensation according to a structure of remuneration akin to a winner-take-all format. It is worth noting for this analysis that tournament purses vary in size according to the magnitude of the event. Events that are historically more prestigious tend to have a large television viewership and a strong field of players. These events reward each player who "makes the cut" with more money than smaller tournaments are able to do. To make the cut is to shoot a low enough score so that you are invited to play the final two rounds of the tournament, usually held on Saturday and Sunday. The champion of a major PGA Tour event, such as the Players Championship, receives more money for his performance in that event than the champion of a lesser-known event such as the John Deere Classic. Besides this convenient compensation scheme, the PGA Tour warrants analysis for another reason. In 2006, Tiger Woods was the leading money winner, making $9,941,563. Conversely, the 227th earner on tour in 2006 made just $64,494. I intend to determine the factors of earnings production on the PGA Tour which prompt this striking range between the most and least successful players on tour. In fact, $9,941,563 represents only a fraction of Tiger Woods' income including his endorsements. A player who regularly finishes among the top 10 in PGA Tour events can expect lucrative endorsement deals, which only increases the incentive to perform at one's peak. The rewards of outstanding performance on the PGA Tour exceed the sum of players' tournament earnings.

In this study, I shall relate certain factors of golf production to money earnings on the PGA Tour. Golfers produce earnings, fame and recognition, and enjoyment by golf success. Of these, earnings is the dependent variable in this analysis because it is the only variable of the group which can be found in existing literature. The object of the study is to determine the skills with the greatest associated marginal effects. In other words, I hope to discover the skills in which an improvement is associated with the highest returns.

I shall model the factors affecting earnings production as a function of data relating to events, age, driving distance, driving accuracy, greens in regulation, putting, and sand saves. I shall use OLS regression to isolate the effect that each type of playing style produces. If the data demonstrate that performance in statistical categories helps to produce earnings on the PGA Tour, we can surmise the skills that distinguish Tiger Woods and other elite players on the PGA Tour from the players who earn five-digit salaries.

This analysis should provide us with knowledge about which skills are most valuable on the PGA Tour.

II. Data Analysis

I gathered the data from ESPN.com and PGATOUR.com. ESPN provided a table with basic golf statistics for players ranked from 1st to 227th in tour earnings. There were a few gaps in the data for players who play primarily on the European Tour and for a handful of PGA Tour members toward the bottom of the rankings, but I filled that part of the data out using PGATOUR.com. I failed to obtain the statistics of six players ranked above 227 on the PGA Tour. Observations of statistics were consistent between those sources for players whose statistics were provided on both sources. The consistency between the two sources affirms the accuracy of the data. The data sample I shall use for this analysis includes most of the players for whom ESPN included performance statistics. Beyond the 227th player on the money list, ESPN has only recorded a handful of players' statistics. The data set represents the magnitude in variability in earnings on the PGA Tour. The players included in the data set have all played at least eight events in 2006. I have excluded players who participated in just a few events because it is unlikely their tour performance is an accurate reflection of PGA Tour performance as a whole. Players who participated in one or two events received a sponsor's exemption in most cases, signifying they should not be considered regular tour players. It happens that the PGA Tour publishes a list of exempted players for the PGA Tour season, all of whom are included in the analysis. An exempted player is entitled to play in the events of his choosing. I considered including the Nationwide and LPGA Tours, but their data is more limited than the data on the PGA Tour.

The explanatory variables are natural choices for the things which they are meant to represent. At least one sports economist has employed each variable in his analysis; I found most of the variables I chose in multiple empirical models. A more thorough discussion of the existing economic literature about golf can be found in Section II, entitled "Model / Ex Ante Predictions."

Earnings are a sufficient measure of success in professional golf competition. Although scoring average may be a more pure measure of golf skills, the metric of earnings captures the prestige of a tournament along with the performance of the observed golfer. Generally speaking, the size of the purse reflects the prestige of the tournament, because the biggest tournament purses attract the best golfers. Thus, players who perform well in prestigious events against a relatively strong field are rewarded accordingly.

Scoring average indicates a player's proficiency at golf, but can vary with the difficulty of the courses he played or the strength of the field. A player who competes in tournaments held at the easiest courses on tour may have a lower scoring average than a better player who enters tournaments at the hardest venues. Moreover, a player may do very well against weaker fields, but perform poorly in major events against particularly strong fields. The earnings statistic captures the strength of the tournament field and removes the influence of course difficulty. Because everyone competes on the same course in a tournament, the earnings metric captures the quality of a player's performance regardless of the difficulty of the course. Tournament purses are public information, thus individual player's earnings for 2006 are readily available.

Age is a proxy for experience with, or exposure to, the game of golf, although it is hardly a perfect measure because players pick up the game at different stages of their life. Moreover, some players develop their ability more rapidly than others, and so it is likely that age benefits some players more than others. Events is quite simply the number of events in which a player competed during the 2006 season.

The rest of the explanatory variables vary in nature and in the degree to which they can be known. Putting performance on individual holes is a discrete statistic (a player can take 0, 1, 2, 3, etc. putts). Thus, putting average is a precise statistic of a player's putts per hole. Driving distance, on the other hand, is vulnerable to measurement error and is only recorded on some holes. Driving accuracy, greens in regulation, and sand saves are all percent measurements of success rates. I have rounded the percentage frequency variables to one decimal place, except for means and standard deviations, which I have rounded to two decimal places.

The explanatory variables are, for the most part, good proxies for the part of golf which they are meant to represent. Sand save percentage[2] is the most dubious explanatory variable in its ability to represent some larger aspect of a player's ability. Sand save percentage is supposed to represent a player's overall short game ability, in that it is the frequency with which a player is able to hit the ball out of the bunker and make the subsequent putt. If a player has failed to convert a sand save, he has required more than two shots to escape from the bunker and finish the hole. The short game is considered those shots within 90 or so yards of the green but not yet on the green. Unfortunately, a skilled short game player is not necessarily a strong player from the bunker, and vice-versa. Generally speaking, however, a good

2   Moy, Ronald L., and Thomas Liaw. "Determinants of Professional Golf Tourna-
        ment Earnings." *American Economist,* 1998, *volume* 42(1), pp. 65-70.

bunker player tends to have similar ability in his overall short game. Another problem with the sand save percentage is that it incorporates a player's putting ability having put the ball on the green from the bunker. A player may be skilled from the bunker, but be an abysmal putter. This player would struggle in sand save percentage, even though his bunker game is strong. Conversely, a great putter can have a high sand save percentage because he is able to simply put the ball anywhere on the green and make the putt. I have chosen the statistic over another short game statistic, scrambling. Scrambling is the percentage frequency with which a player is able to get "up and down," that is, to hit the ball onto the green from within 90 yards and make the subsequent putt. I decided not use this statistic as a proxy for short game ability because a player who two-putts from just off the edge of the green has completed an "up and down." In reality, he has done nothing more than two-putt, but he has just increased his scrambling statistic. The sand save metric precludes this two-putt from contributing to a player's short-game metric.

Earnings range from $65,494 to $9,941,563, with a mean of $1,117,264 and a standard deviation of $117,391 (Table 1). The magnitude of the standard deviation suggests that a seemingly insignificant difference in skill creates an enormous discrepancy in compensation, although that is likely the product of a few very good players who dominate the purses at the most prestigious events. The most highly correlated variable with earnings is greens in regulation, suggesting the PGA Tour puts a premium on having the ability to hit good approach shots into the green.

The average event participation for tour players in 2006 was 24.66, with a standard deviation of 5.96. One player only played in eight events, while the most active player participated in 36. Remarkably, the correlation coefficient between events and earnings is only .04, suggesting greater participation does not at all produce more earnings.

The average age on tour is 36.21 years old, with a standard deviation of 6.59 years. This comes as no surprise, as golfers tend to lose their ability to compete at the high level of the PGA Tour by their mid-40s. It is also of note that players become eligible to play on the Champions Tour at 50, considerably reducing the incentive for older players to continue to try and compete on the PGA Tour. Members of the Champions Tour are excluded from this analysis. Not surprisingly, driving distance has a fairly negative correlation with age, having a Pearson Correlation Coefficient of -0.50. The youngest PGA Tour player is 23. A likely explanation for the high minimum in golf is the common practice of professional golfers playing in college before they try to earn membership to the Tour.

Average driving distance is 288.85 yards, with the minimum and

maximum being 265 and 319 yards, respectively. The standard deviation of driving distance is 8.56 yards.

PGA Tour players hit 63.90% of fairways, with a standard deviation of 5.46%. The most accurate driver of the golf ball hit 78.4% of fairways, while the most erratic driver only hit 49.8%.

Players hit slightly more greens in regulation than they did fairways, averaging 64.94% with a standard deviation of 2.79%. It is of note that the standard deviation of greens hit in regulation is just under half of the standard deviation of fairways hit in regulation, suggesting that players who tend to drive the ball more erratically off the tee still manage to hit greens as consistently as the straight drivers of the golf ball. In other words, players may use the fairway, the rough, or even the woods as a path to the hole, but every tour player tends to put the ball on the green in close to the same number of shots. The most accurate iron player (the player who hits the most greens in regulation), Zach Johnson, hit 74.1% of greens, while the least accurate iron player, Aaron Baddeley, only hit 56.9%. There is a Pearson correlation coefficient of 0.38 between driving distance and driving accuracy. The correlation coefficient is significant at $\alpha = .01$ and indicates that the most accurate drivers of the golf ball are also the more consistent iron players. The correlation between driving distance and driving accuracy supports the notion that players tend to have strong "long games"(i.e., they are either good at driving and approach shots or bad at driving and approach shots), rather than be simply strong drivers or strong iron players.

Players averaged 1.7802 putts per hole, which implies that the average player took well under the par number of putts on each hole. That this average is well under the expected or par putts per hole does not only indicate that tour players are good putters, but also that they hit approach shots close enough to the hole that they can take only one putt to finish the hole. The standard deviation of putts per hole is .0253, while the minimum and maximum observations of putts per hole are 1.712 and 1.851, respectively. Of note, Aaron Baddeley made up for his weak iron play with the lowest putting average on tour. Conversely, Zach Johnson had the very worst putting average. One explanation for the reversal of these two players is that, because Zach Johnson hits so many greens in regulation, he frequently faces long putts. Conversely, Aaron Baddeley hits very few greens in regulation but hits the ball to a short distance from the hole. Thus, he leaves himself many relatively short putts. Generally speaking, having a low average number of putts per hole indicates a player is either talented at positioning the ball near the hole on his approach shot so that his first putt is relatively short or he makes relatively longer putts more frequently than other

tour players.

The average rate of sand save conversion was 48.66% with a standard deviation of 6.47%. The minimum and maximum sand save conversion rate is 16.7% and 63.8%. The sand save conversion rate is our best measure of short game proficiency. It has a large range, reflecting a large variation in short game ability across the tour. It has a correlation with earnings of 0.2433.

It may seem curious that I have excluded the variable "scoring average." The existing literature calls for it, except for Patrick James Rishe in his 2001 article, entitled "Differing Rates of Return to Performance: A Comparison of the PGA and Senior Golf Tours." I have excluded the variable because it does not reflect any one golf skill, but rather a player's overall ability. While scoring average has immense explanatory power, it does not represent a specific golf skill. There are certain variables other than scoring average, such as cuts made, top 10 finishes, and wins, which are obviously correlated with earnings, but they would likely capture the mental side of golf. Of course, the mental aspect of the game could explain a good deal of variation in earnings, but the literature which I have used as a foundation for my choice in variables largely ignores the mental part of the game. Age is included to capture the change in one's mental maturity that comes with experience.[3]

III. Literature Review

In the context of previous empirical work dealing with this subject, this analysis is in many ways a replication of the work Ronald L. Moy and Thomas Liaw did in their 1998 article, "Determinants of Professional Golf Earnings." They included all of the variables I shall include in my analysis with the exception of age, a variable for which Gerald W. Scully's 2002 article, "The Distribution of Performance and Earnings in a Prize Economy," provided the basis. I selected the independent variable and dependent variables, with the exception of scoring average, based on the empirical work they did in 1998 on PGA Tour results from 1994. In 2000, Stephen Shmanske created a similar model to Moy and Liaw's for data from 1998, so similar that my analysis could be considered a replication of his work as well. I shall try to determine if the advancement of golf technology has significantly affected player performance on the PGA Tour. Driving distance is the most likely statistic to have changed dramatically between 1994 and 2006. I intend to determine if the greater driving distance created by more advanced

---

3   Shmanske, Stephen. "Gender, Skill, and Earnings in Professional Golf." *Journal of Sports Economics,* 2000, *volume* 1(4), pp. 385-400.

technology has caused driving distance to have a more significant effect on tour winnings in 2006 than it did in 1994.

Each of the existing models employs some of the same variables, but a couple of the models use variables that, to the casual observer, would seem unconventional. In Patrick Rishe's 2001 article "Differing Rates of Return to Performance: A Comparison of the PGA and Senior Golf Tours," he introduces a few innovative variables, birdie conversions, bounce back, and scrambling among more conventional proxies for golf talent. Stephen Shmanske, Ronald L. Moy and Liaw, Gerald W. Scully, and Patrick James Rishe's variable choices all provide scholarly authority with respect to the explanatory variables I have chosen to represent player skill (Table 1). Since their studies, however, technology has improved to the extent that there may be a noticeable change in the returns to driving statistics. I shall see if driving, both driving distance and driving accuracy, has had any noticeable effect on player performance overall. Driving distance is the most likely statistic to have changed dramatically since Moy and Liaw's study in 1994. The introduction of graphite technology in golf clubs preceded their study, but golfers have enjoyed more and more advanced technology in the last 12 years. I hope to add to the empirical studies since Moy and Liaw's investigation in 1998 by drawing something from each of the authors I mentioned.

The primary study on which my empirical model is based is Stephen Shmanske's model of the relationship between skills and earnings, published in 2000. Shmanske examined both the PGA and LPGA Tours in order to determine why such a large gap exists between purses on those tours and the skills that had the greatest effect on earnings. Shmanske's empirical model varies only in that he has selected a special variable to represent putting proficiency. He considered total putts a misleading measure because a player who hits few greens in regulation is able to play an extra approach shot from close range and set himself up with a shorter, more easily holed, first putt. His point is well taken, but Moy and Liaw's results suggest that Shmanske underestimates the power of a statistic which averages a player's putts per hole.

Both Shmanske and Moy and Liaw compared the PGA and LPGA tours by examining the relationship between skill and earnings. My empirical model is as much a replication of their empirical model as it is a replication of Shmanske's model, except I shall exclude one variable Moy and Liaw include: scoring average. Since scoring average is the result of the level of skill a player exhibits in all the other statistical categories, I have left it out. Rishe is the authority behind my decision to leave out scoring average. Moy

and Liaw published their empirical work in 1998 based on PGA Tour results from 1994. Aside from the authority they give my variable choice, their results should come in handy because technology has changed significantly since 1994. They found ln(DDIS) and DACCUR to be statistically insignificant. ln(DDIS) has a p-value of .220, while DACCUR has a p-value of .350. Conversely, their greens in regulation, putting, and sand statistics were all significant at $\alpha = .01$, with significant large parameter estimates.[4] If technology has affected the game, we might expect those variables to be significant in 2006. Contrasting the results of their 1994 data and my 2006 data could demonstrate a transformation in the relationship between different skills and earnings.

In "Differing Rates of Return to Performance: A Comparison of the PGA and Senior Golf Tours," Patrick James Rishe sought to determine "whether the earnings gap between Professional Golf Association and Senior Tour golfers is due to differences in average skill levels or the rates of return to these skills."[5] He concluded that the earnings differential across tours is a consequence of different rates of return to performance. The relevant part of Rishe's paper is his empirical model, which has driving accuracy, greens in regulation, sand saves, driving distance, birdie conversions, bounce back, and scrambling as explaining the variation in the natural log of Y, the earnings per event for a given golfer. Birdie conversions, bounce back, and scrambling, as far as previous literature is concerned, are all innovative explanatory variables. The birdie conversions statistic is the percentage of times a player, having reached the green in regulation, is able to convert a birdie or better. Rishe has intended bounce back as a measure of resiliency. He defined it as, "the percentage of times a player, after shooting over par on the previous hole, comes back to shoot under par on the next hole." He expected that erratic players, who frequently make bogeys followed by birdies and consequently have high bounce back statistics, will finish on the lower end of the money list. He defined scrambling as the percentage of times a player can make a par or better having failed to reach the green in regulation.

Rishe's bounce back ex ante hypothesis turned out to be significant at the .11 level. The rest of Rishe's coefficients had their expected sign and were significant at the .05 level, with the exception of driving distance, which was only significant at the .11 level. Conspicuously absent from his

4   Moy, Ronald L., and Thomas Liaw. "Determinants of Professional Golf Tournament Earnings." *American Economist,* 1998, *volume* 42(1), pp. 65-70.

5   Rishe, Patrick James. "Differing Rates of Return to Performance: A Comparison of the PGA and Senior Golf Tours." *Journal of Sports Economics,* 2001, *volume* 2(3), pp. 285-96.

empirical model, however, is some measure of putting ability. This omission could cause bias in the coefficients if there is correlation between it and the included variables. There is likely a correlation between sand saves and putts and scrambling and putts. Because of this correlation, excluding a putting variable could be detrimental to the accuracy of the model.

Gerald W. Scully links prize money on the PGA Tour to scoring average in his paper entitled "The Distribution of Performance and Earnings in a Prize Economy." Scully does not write purely about the determinants of golfer winnings, but he does dedicate a section of his paper to that topic. Scully takes issue with Shmanske's model specification from 1992. Scully contends that one extra average yard of driving distance gains a player nothing in earnings. He argues that performance measures such as driving distance affect scoring average and scoring average affects tournament earnings. Scully's article is of particular relevance to mine because he considered age a determinant of earnings. He believes that the difficulty of replicating a golf swing and the pressure of competing at a high level give an advantage to older player, i.e., players with more experience. He also notes that age becomes a detriment because strength and eye-hand coordination recede with time.

Shmanske, Moy and Liaw, Rishe, and Scully provide the theoretical framework for the relationship between golf skills and earnings. I shall incorporate to some degree each of their contributions to this study. Shmanske's use of performance measures, along with Moy and Liaw's older data of the same statistics, will provide the backbone of the model specification. Rishe has a number of proxies worth consideration, and Scully provides the authority to include age into the model.

IV. Theoretical Model Exposition

At least five economists have considered the effect of golf skills on golf earnings. Earnings on the PGA Tour can be modeled as a production function. Shmanske, Moy and Liaw created empirical models which had one variable acting as a proxy for each part of the game: driving, approach shots, short game, and putting. I shall include all of the variables which they employed, with the additional variable of age. Scully was the first to use that variable.

Production is typically conceived as a function of labor and capital. The production function is relevant to modeling earnings in professional golf because professional golfers produce earnings. Production of golf earnings is, for the most part, a function of labor. Just as some workers are more capable of producing a good in an ordinary production function, so too is

there variance in the abilities of golfers to produce earnings on the PGA Tour. We can include capital in the production function in the sense that a player's physical attributes might represent the capital stock he possesses. There is no evidence in previous literature and little anecdotal evidence to suggest that any particular physical attribute produces earnings. Some observers of the game believe that being "too tall" is a disadvantage, but actual evidence of that claim is scanty. I have included one explanatory variable that could be considered a measure of human capital investment. As a player ages, we expect his level of human capital to increase in proportion to the amount of time he has spent learning how shoot low scores. At a certain point, a player risks loss in hand-eye coordination and physical strength. As it pertains to a production function, this could be conceived of as capital depreciation.

Each skill for which I have included a proxy variable has a detectable effect on earnings.[6] Driving distance, driving accuracy, greens in regulation, and sand save percentage should all be positive, because driving distance, driving accuracy, greens in regulation, and sand saves should all have a positive effect on earnings. The ex ante hypothesis for PUTT, the putting statistic, is negative.[7] Common sense suggests that a player's earnings will decrease as his putting statistic increases. Since the best putters have the lowest average putts, we expect the player with the lowest putting average to have the highest earnings. The ex ante hypothesis has the potential to fall short in explaining the variation in earnings in the case of a player who hits relatively few greens in regulation and consequently has a low putting statistic. Such a player is usually just off the green and is able to give himself a short first putt. He is not necessarily a particularly good putter; rather, his putting statistic is lower as a consequence of having more short putts than a player who hits relatively more greens in regulation.

Once again, I shall use Ordinary Least Squares Regression in order to estimate the true value of each of my explanatory variables. This technique will give me the best (smallest variance), linear, unbiased, estimate. The following equation represents the earnings return to golf skills I expect.

(1) $EARN = a + b_1 EVENT + b_2 AGE + b_3 AGE^2 + b_4 DD + b_5 DA + b_6 GIR + b_7 PUTT + b_8 SS + e$

The statistics represented are event, age, driving distance, driving accuracy, greens in regulation, putting, and sand saves, respectively. EVENT is the number of tournaments in which a player competed during the 2006 season, AGE is the age of the player, DD is the average driving distance a

---

6   Shmanske, Stephen. "Gender, Skill, and Earnings in Professional Golf." *Journal of Sports Economics,* 2000, *volume* 1(4), pp. 385-400.

7   Moy, Ronald L., and Thomas Liaw. "Determinants of Professional Golf Tournament Earnings." *American Economist,* 1998, *volume* 42(1), pp. 65-70.

player achieved during the season, DA is the percentage of times a player hit the fairway on par 4s and 5s over the course of the season, GIR is the percentage of the greens a player hit in no more than two less than par of a given hole, PUTT is the putting average a player compiled over the course of the year, and SS is the percentage frequency with which a player landed a ball on the green from the bunker and holed out the subsequent putt.

Now I shall describe the marginal effects which I expect each variable to have on earnings, beginning with EVENT (Table 2).

$$(2) \qquad \frac{\partial EARN}{\partial EVENT} = b_1; b_1 > 0.$$

β_1 is the marginal effect of the number of events in which a player competes on his earnings. It is expected to be greater than zero, corresponding to the notion that the more events in which a player competes, the more earnings he can produce. As a player competes in more events he increases the number of opportunities he has to earn. Moreover, competing in more tournaments increases a player's experience over the course of that season. Consequently, we can expect a positive β_1 based on the notion that players tend to "get the rust off" when they play in more events.

$$(3) \qquad \frac{\partial EARN}{\partial AGE} = b_2 + b_3 age; b_2 > 0; b_3 < 0.$$

It is difficult to develop a consistent ex ante hypothesis of earnings with respect to age across an entire sample of tour players because there is a good deal of variance in skill at any given age. Players improve at different rates at each point in their careers. Some players improve by leaps and bounds within a year of joining the tour, while others require 10 years on tour before their first PGA Tour win. I have squared the AGE term because I expect greater age to contribute to greater earnings to a point, but to lower earnings after that point. The loss of strength and hand-eye coordination that affects players after their peak causes decreasing earnings with increasing age.[8]

$$(4) \qquad \frac{\partial EARN}{\partial DD} = b_4; b_4 > 0.$$

β_4 is the marginal effect of driving distance on earnings. Driving distance is expected to have a positive effect on earnings.[9] As Shmanske

---

8   Scully, Gerald W. "The Distribution of Performance and Earnings in a Prize Economy." *Journal of Sports Economics,* 2002, *volume* 3(3), pp. 235-45.

9   Shmanske, Stephen. "Gender, Skill, and Earnings in Professional Golf." *Journal of Sports Economics,* 2000, *volume* 1(4), pp. 385-400.

does, I shall use a linear functional form for it. Anecdotal evidence suggests, however, that as a player's driving distance increases, he benefits less from one average extra yard of driving distance. An interview from Tiger Woods shows that longer hitters benefit less from a marginal increase in driving distance. When asked why he was not playing more aggressively by trying to hit the ball farther, Woods responded that he did not know what advantage that would allow him. He claimed that, even on the longest whole of the course, he had only a short distance to the whole and that he endangered his position for no apparent reward.[10] Tiger is an exceptionally long driver of the golf ball, but his answer reflects a recurring sentiment among professional golfers. When Woods hits a 3-wood on the longest hole on a golf course (a long hole, to be sure), he must not benefit a great deal from extra distance. This anecdotal evidence suggests that the improvement in technology that has occurred since Moy and Liaw's data were recorded in 1994 will not change the results Moy and Liaw found in their study. If driving distance is statistically significant in the correct direction, I expect it to have hardly any real effect on earnings.

$$(5) \qquad \frac{\partial EARN}{\partial DA} = b_5 ; b_5 > 0.$$

β_5 is the marginal effect of driving accuracy on earnings. I expect driving accuracy to have a positive effect on earnings.[11] Greater driving accuracy can only improve a player's earnings, because it means he has more shots from a more desirable position, the fairway.

$$(6) \qquad \frac{\partial EARN}{\partial GIR} = b_6 ; b_6 > 0.$$

β_6 is the marginal effect of greens in regulation on earnings. I expect greens in regulation to have a positive on earnings.[12] Once again, I am uncertain about the proper functional form. Previous literature has chosen a linear form, but I believe that a player who hits greens with a relatively greater frequency benefits more from a one percentage point increase in greens in regulation than a player who hits greens with a relatively smaller frequency. Since the player who hits relatively more greens in regulation is more familiar with the situation, he has a higher likelihood of succeeding in

10   USOpen.com. http://www.usopen.com/2003/press/interviews/woods-6-14-flash. html, accessed 3/25/07.

11   Shmanske, Stephen. "Gender, Skill, and Earnings in Professional Golf." *Journal of Sports Economics,* 2000, *volume* 1(4), pp. 385-400.

12   Ibid.

that situation. A more sound explanation requires an example. Let us take two players, one with a higher GIR and one with a lower GIR. The one with the higher GIR hits a green in regulation, as he is accustomed to doing. The one with the lower GIR fails to hit the GIR, and requires an extra shot to put the ball closer to the hole than the higher GIR player did. This situation is typical; the higher GIR player will be more accustomed to having to hit longer putts than the player with the lower GIR. When the lower GIR player does hit a green in regulation and faces the longer putt, he will not have refined the skill and will not succeed as often as the higher GIR player. To make matters worse for the lower GIR player, he requires a broader skill set to make the same score on a hole as the higher GIR player. Because he misses the green more frequently than the higher GIR player, he is forced to draw from a skill set the higher GIR player need not worry about as often.

$$(7) \qquad \frac{\partial EARN}{\partial PUTT} = b_7; b_7 < 0.$$

B_7 is the marginal effect of putting average on earnings. A lower putting average should increase earnings; putting average has a negative relationship with earnings.[13] As demonstrated with Aaron Baddeley and Zach Johnson, the putting statistic reveals more than a player's putting proficiency. Putting is inexorably linked with GIR. Generally speaking, we might expect a player with a higher GIR to have a higher average putting statistic than a player with a lower GIR. As previously discussed, this is because the player with the higher GIR makes fewer of his first putts, simply because they are longer, not necessarily because he is a worse putter. Interacting the terms would reflect their interrelatedness, but the literature does not interact them and we would learn more by being able to observe each variable's marginal effect.

$$(8) \qquad \frac{\partial EARN}{\partial SS} = b_8; b_8 > 0.$$

β_8 is the marginal effect of sand save percentage on earnings. Certainly, increased frequency of sand saves will positively affect earnings.[14] A player's ability from the sand, and more broadly, from anywhere around the green, should produce returns to earnings at the same rate. That is, a one unit increase in sand save percentage is no more beneficial to a player with a low sand save percentage than a player with a high one, or vice versa.

13   Shmanske, Stephen. "Gender, Skill, and Earnings in Professional Golf." *Journal of Sports Economics,* 2000, *volume* 1(4), pp. 385-400.

14   Moy, Ronald L., and Thomas Liaw. "Determinants of Professional Golf Tournament Earnings." *American Economist,* 1998, *volume* 42(1), pp. 65-70.

I have provided an explanatory variable to represent each major skill in the game of golf in order to determine the competence of a player in that area. Economic literature[15] suggests that each of these skills has an effect on production of earnings. I have included age because I believe that golf skills increase up to a certain age, where they actually begin to decrease.[16] I expect events played, age, driving distance, driving accuracy, greens in regulation, putting average, and sand saves to have a statistically and economically significant marginal impact upon earnings.

V. Empirical Model and Estimation

The theoretical framework on which my original model was founded yielded an unimpressive adjusted R^2 and prompted me to reconsider the functional form of a number of variables. After revisiting the literature, I decided that the linear functional form fails to sufficiently represent returns to skill.

$$(9) \qquad \ln(EARN) = a + b_1 \ln(EVNT) + b_2 AGE + b_3 AGE^2 + b_4 DDDA + b_5 \ln(GIR) + b_6 \ln(PUTT) + b_7 \ln(SS)$$

It is different from my original specification in that it has predominantly log-log forms, with an interaction term for driving distance and driving accuracy.

The Ordinary Least Squares Regression produced t-values for each parameter estimate (Table 3). One column of the table, labeled Significant?, indicates whether the hypothesis test for that variable was statistically significant. The results are based on one of the following hypothesis tests, depending on the expected sign. In other words, all hypothesis tests of individual significance were one-tailed and with respect to zero.

$$(10) \qquad \begin{array}{ll} H_0 : b \leq 0; & \qquad H_0 : b \geq 0; \\ H_a : b > 0. & \text{OR} \qquad H_a : b < 0. \end{array}$$

The model represented above has an R^2 of 0.5048 and an adjusted R^2 of 0.4885. Naturally, the adjusted R^2 is somewhat lower because it is adjusted for the intercept term and seven independent variables I have included in the model. Even so, the coefficients of determination above are substantially higher than the R^2s that the first specification of the model yielded. In that specification, all functional forms but two were linear (Equation 1).

---

15   Shmanske, Stephen. "Gender, Skill, and Earnings in Professional Golf." *Journal of Sports Economics,* 2000, *volume* 1(4), pp. 385-400.

16   Scully, Gerald W. "The Distribution of Performance and Earnings in a Prize Economy." *Journal of Sports Economics,* 2002, *volume* 3(3), pp. 235-45.

Moreover, the first specification did not call for an interaction term of DD and DA. It had an R^2 of 0.3926 and an adjusted R^2 of 0.3697.

The AIC is another tool that aids in comparison between models, especially those employing different specifications. It adjusts the RSS for the sample size and number of independent variables, allowing comparison between models with different specifications. It is 498.947 in the log-log model, which is substantially lower than 6709.459, the AIC that the first specification produced. The improved specification of the model explains the vast improvement between the original and final models. The F-statistic for the final model is 31.02, which exceeds the F-critical of 2.10. The F-critical has 7 d.f. in the numerator and ∞ d.f. in the denominator, based on 221 observations. The magnitude of the F-statistic indicates that we have evidence to reject the null that $b_1 + b_2 + \ldots + b_7 = 0$. Therefore, all of our variables jointly have an affect on ln(EARN).

The parameter estimates in the original model for driving distance and driving accuracy caused me to reconsider the specification of the model. Neither was significant, and both parameters had the opposite marginal affect from what I expected. When prompted to consider the variables further, however, I realized that they have little meaning when considered ceteris paribus. When a player steps up to the tee, he considers both driving distance and driving accuracy together, not as distinct entities. For example, when a player wishes to hit the ball relatively far, he has to compromise accuracy. When he is intent on hitting the fairway off the tee, he must compromise distance. The new parameter estimate, which consists of the interaction of driving distance and driving accuracy, is both statistically and economically insignificant for an α of .05. The parameter estimate for DDDA is .000076, meaning a one percent increase in DDDA produces a .000076 percent increase in earnings, i.e., a negligible sum. This could indicate that the tee shot is relatively unimportant, as long as the player puts himself in position to hit the green on his approach shot.

Initially failing the Ramsey RESET Test with a linear model prompted me to reconsider the functional form. The Ramsey RESET is a blunt instrument for detecting omitted variable bias or some other specification error. I believed the variables sufficiently represented every golf skill, so I concluded only a problem with functional forms could cause my specification error. Literature suggests that the linear functional form most accurately represents the true functional form of golf production, but economic theory of production suggests that logs are more appropriate. Having failed the Ramsey RESET Test, I resorted to a log-log functional form. The model employing the log-log form passed a modified Ramsey

RESET Test which excluded yhat3. Inclusion of yhat2 and yhat4 indicates that the model, should it pass, has some level of robustness. It turns out that the Ramsey RESET Test without yhat3 yielded an F-statistic of 0.38 which is not significant at any level of α. The critical F with 2 d.f. for the numerator and ∞ d.f. for the denominator is 4.61. Because $0.38 < 4.61$, we do not have evidence to reject the null hypothesis that F is significantly different from 0. The model failed the traditional Ramsey RESET Test with yhat2, yhat3, and yhat4 included. The inclusion of those variables created substantial multicollinearity and altered parameter estimates (Appendix 1).

VIFs for the parameter estimates were all fairly low with the exception of AGE and AGE^2. ln(EVENT), DDDA, ln(GIR), ln(PUTT), and ln(SS) all have VIFs less than 1.5, with ln(EVENT), ln(PUTT), and ln(SS), all within .1 of 1. AGE and AGE^2 each have VIFs of about 105, suggesting they are highly multicollinear with one another. Given AGE^2 is a function of AGE, we would naturally expect substantial multicollinearity between them. However, theory behind modeling AGE as a polynomial is strong enough such that we should be willing to accept the multicollinearity. Even so, it is distressing that the functional form we have selected for AGE has turned out to be empirically wrong. AGE and AGE^2 are significant in the wrong direction. It is possible that the polynomial form is not appropriate after all given the small variance in ages of PGA Tour players. For now, I shall leave AGE as a polynomial because it does not noticeably cause other variables to be biased or reduce the adjusted R^2 of the model.

As we might expect, the correlation between AGE and AGE2 is strong, 0.995. The next highest simple correlation coefficient is 0.523, between ln(GIR) and DDDA. The correlation between these variables indirectly supports the hypothesis that DD and DA should be an interactive term. If DDDA is a measure of overall driving proficiency, we might expect it to be correlated with GIR, because both are measures of ability in what as known as the "long game." The long game is generally defined as those shots taken from outside of 90 yards from the green. Because they are so highly correlated, we have some reason to believe that DDDA is a good measure of overall driving ability. The correlation coefficient between ln(PUTT) and ln(SS) is another one worthy of our attention. That correlation coefficient is -0.236. The explanation for this is that when a player successfully converts a save from the sand, he has taken only one putt. Consequently, his putting average will decrease. We might expect players who have a high sand save conversion percentage to have a relatively low putting average, because each time they convert a sand save, they only actually take one putt. None of these correlations was as pronounced with linear functional forms. Finally, the

correlation between DDDA and ln(EVENT) is -.204. This correlation is not high, but it is higher than almost every other correlation. There is no obvious reason for the correlation, but it is low enough such that we can dismiss it.

Driving accuracy was significant in the wrong direction when I tried to isolate its effect in the first specification. The significance in the wrong direction might be attributed to the unimportance of putting the ball in the fairway on the tee shot. It may be that players who are too cautious and play very timidly off the tee fail to put themselves in a good position to hit the green in regulation. The simple correlation statistic certainly supports this hypothesis. We might expect serious multicollinearity between driving accuracy and greens in regulation, because they are both measures of accuracy. They, however, have a simple correlation of only .38. Therefore, players who do hit the green in regulation are accurate enough to hit the fairway with some regularity, but hitting the fairway does not necessarily lead to hitting the green in regulation. The evidence suggests that hitting the fairway is not important to earnings success.

Before discussing the extent to which there is autocorrelation in the model, it is worth noting that I sorted the observations in alphabetical order by players' last name. Our null and alternative hypothesis is:

$$(11) \quad \begin{aligned} H_0 &: r \leq 0; \\ H_a &: r > 0. \end{aligned}$$

The autocorrelation procedure yields a DW-d statistic of 1.98. Since 1.98 > 1.83, the critical DW-d statistic (k = 7, n > 100), we do not have evidence to reject the null hypothesis. We have no evidence of serial correlation in the model. Therefore, Ordinary Least Squares regression is still the most appropriate way to regress skills from earnings.

The results of the White Test indicate we do not have evidence for heteroskedasticity.

$$(12) \quad \begin{aligned} H_0 &: Homoskedasticity; \\ H_a &: Heteroskedasticity. \end{aligned}$$

The White Test yields R^2 of .0896, .0917, .1593, respectively when X_1, X_1^2, and $X_1 X_2$ are used to model the error term. When we multiply these by the degrees of freedom, it is apparent that the chi-square critical value far exceeds each of our statistics. We fail to reject the null hypothesis at $\alpha = .10$, even if we compare our chi-square statistics to a chi-square critical value with only 100 degrees of freedom.

Having passed the White Test, for which I did not have to specify a factor of proportionality, I proceeded to the Park Test. Only one variable,

ln(PUTT), yields a significant t-value in the Park Test. Since I passed the White Test, I do not believe the cost of weighting the results by ln(PUTT) is worth the correction I might be able to make to the error term.[17]

In order to determine the marginal effect of each variable on earnings, we must consider the meaning of the variables and their functional forms. First, it is worth noting that I have excluded slope values in X-Y space, $\frac{\partial Y}{\partial X}$, of each parameter, because the marginal effects keeping the original functional form (e.g., $\frac{\partial \ln Y}{\partial \ln X}$), reveal more about the impact of an explanatory variable on the independent variable. All of the values I have provided as marginal effects are in the space of their functional forms. For example, the marginal effect of a one percent increase in PUTT is a -25.45138 percent increase in EARN. Put another way, the marginal effect of ln(PUTT) on ln(EARN) is -25.45138. The marginal affect of a percent change in X on a percent change in Y is simply β for all log-log forms. Let us take another example, ln(EARN). The marginal impact of a one percent increase in the number of events played produces a .88608 percent increase in earnings (Table 3). Considering that a one percent increase could be thousands of dollars, this figure is economically significant statistic. The number of events only needs to be increased by one percent in order to produce a noticeable increase in earnings. ln(GIR) and ln(SS) are expected to produce a 12.96432 and 1.23497 percent increase in earnings. Age is a more complicated case to interpret because the partial derivative of age is represented in two terms. The marginal effect of age on earnings is,

$$\frac{\partial Y}{\partial X} = .15676 + .00195 * 2X$$

At the mean of age, we expect a .298 percent increase in earnings for a one year increase in age.

Finally, the marginal effect of the interaction term DDDA is -0.00007649. We would expect a tiny number for the DDDA interaction term because average driving distance has a range of only 54 yards and DA is only a percentage. In any case, DDDA is not statistically significant, so we need not worry it is in the wrong direction.

For log-log forms, the elasticity for each individual variable is simply β. Thus, elasticity is constant. That is, it is the same at the mean as everywhere else (Table 4). The elasticity for age at the mean is a bit

---

17   I did run a Weighted Least Squares Regression using ln(PUTT) as the factor of proportionality. The parameter estimates were largely unaffected, leading me to believe there is negligible heteroskedasticity in the model.

challenging to calculate (Appendix 2).

The main lesson of actually running the OLS regression is that returns to golf skills on the PGA Tour is not demonstrated best by a linear relationship between golf skills and earnings. Rather, a much sounder specification employs a log-log relationship. Moreover, it seems that an interaction term between DD and DA resembles most closely the considerations of a professional golfer when he steps onto the tee and makes a decision about how he wants to approach the golf shot. The interaction of those terms is not statistically significant, but it does improve the overall explanatory power of the model. However reasonable our theoretical grounds, it seems that driving distance and driving accuracy are all but irrelevant on the PGA Tour. Within reason, a player need not worry too much about what he does off the tee, because it hardly has any impact at all upon his earnings.

VI. Conclusion

Given what I have found, I would like to do a study that includes events played, greens in regulation, a putting statistic, and a short game metric but excludes DDDA, AGE, and AGE2. Such an adjustment could increase the adjusted R^2, and more importantly, given this empirical analysis, it might be theoretically sound as well. It seems that players' driving ability is so similar at the highest level of the game that a marginal increase in driving proficiency has a negligible effect on earnings. Moreover, whatever power a player loses as he gets older is either not important or is made up by that player's experience. The statistical insignificance of DDDA leads one to believe that the change in golf technology, since Moy and Liaw's data were collected in 1994, has not affected the outcome of golf tournaments. Such an outcome is hardly surprising. The average tour player drives the ball far enough that he can put himself in an excellent scoring position without the advantage of the few extra yards that new technology brings. With added distance comes the added risk of driving the ball into the rough, or worse. The cost associated with risk of driving the ball into the rough is greater than the reward of gaining a few extra yards. Moreover, players who are able to drive the ball farther than their opponents also tend to miss the fairway more frequently. The simple correlation coefficient between DD and DA is -.59; there is a strong, negative correlation between driving distance and driving accuracy. Just as Moy and Liaw found ln(DDIS) and DACCUR to be statistically insignificant in 1998, I found the interaction of those variables to be insignificant in 2006. We can confidently say that the most significant determinants of earnings on the PGA Tour are still greens in regulation and

putting.

Appendix

Functional Form

$$EARN = a + b_1 EVENT + b_2 AGE + b_3 AGE^2 + b_4 DD + b_5 DA + b_6 GIR + b_7 PUTT + b_8 SS + e$$

Description of Data Set

The data have been taken from ESPN.com. The data are a cross-section of the sum of earnings in 2006 and observations of each variable listed below. I have drawn some of the lower earners' data from PGATOUR.com. The observational unit is one PGA Tour player.

Variable Definitions

ln(EARN) = The natural log of the observed (ith) player's money earnings on the PGA Tour during 2006 season.

ln(EVENT) = The number of tournaments in which the ith player has participated, including events in which he missed the cut to play all four days.

AGE = The age of the ith player.

AGE^2 = The squared age of the ith player.

DDDA = The interaction of the ith player's average driving distance on par 4 and par 5 holes and the percentage frequency with which the ith player hits the fairway.

ln(GIR) = The natural log of the percentage frequency with which the ith player green in regulation. A player hits a green in regulation when he hits the ball onto the green par for the hole minus two. If a player should hit a green in fewer shots than the number of shots that is considered a green in regulation, it is also considered a green in regulation.

ln(PUTT) = Having hit the ball onto the green, the natural log of the average number of putts the ith player requires to finish a hole.

ln(SS) = The natural log of he percentage frequency with which the ith player finishes a hole from the bunker in two shots or fewer.

Tables

Table 1: Existing Models for Returns to Skill on PGA Tour	
Author(s)	Model
Moy and Liaw	$\ln(EARN) = a + b_1 \ln(DDIS) + b_2 DACCUR + b_3 GIR + b_4 \ln(PUTT) + b_5 SAND + b_6 SCORE + e$
Shmankse	$Y98 = a + b_1 TOTPUTT + b_2 EVENTS + b_3 EVENTS98 + b_4 SCOREAVE + b_5 DRIVDIST + b_6 DRIVACC + b_7 TOTDRIV + b_8 GIR + b_9 PUTTPER + b_{10} PUTTPRED + b_{11} SANDSAVE + b_{12} WINPER + e$
Scully	$\ln PRIZE = a + b_1 \ln SA + b_2 \ln EVENTS + b_3 AGE + e$
Rishe	$\ln Y = b_0 + b_1 DA + b_2 GIR + b_3 SS + b_4 DD + b_5 BC + b_6 BB + b_7 SCRAM + e$

Table 2: Ex Ante Hypotheses	
Explanatory Variable	Marginal Effect of X on EARN
Number of Events	$\dfrac{\partial EARN}{\partial EVNT} = b_1 ; b_1 > 0.$
Age	$\dfrac{\partial EARN}{\partial AGE} = b_2 + 2b_3.$
Driving Distance	$\dfrac{\partial EARN}{\partial DD} = b_4 ; b_4 > 0.$
Driving Accuracy	$\dfrac{\partial EARN}{\partial DA} = b_5 ; b_5 > 0.$
Greens in Regulation	$\dfrac{\partial EARN}{\partial GIR} = b_6 ; b_6 > 0.$
Putting Average	$\dfrac{\partial EARN}{\partial PUTT} = b_7 ; b_7 < 0.$
Sand Saves	$\dfrac{\partial EARN}{\partial SS} = b_8 ; b_8 > 0.$

Table 3: Parameter Estimates and T-Scores				
Explanatory Variable Name	Parameter Estimate (Marginal Effect, except for age)	T-Score	Correct Direction?	Significant?
Intercept	-29.10987	4.93	--	--
ln(EVENT)	0.88608	4.88	Yes	Yes
AGE	-0.15676	2.03	No	No
AGE^2	0.00195	1.86	No	No
DDDA	-0.00007649	1.67	No	No
ln(GIR)	12.96432	9.47	Yes	Yes
ln(PUTT)	-25.45138	7.02	Yes	Yes
ln(SS)	1.23497	3.52	Yes	Yes

Table 4: Elasticity Calculations at the Mean	
Variable	Elasticity
ln(EVENT)	0.88606
age	-0.56248
ln(DDDA)	-0.10469
ln(GIR)	1.43414
ln(PUTT)	1.08171
ln(SS)	1.08088

Table 5: Simple Correlation Coefficients								
	AGE	DD	DA	GIR	PUTT	SS	EVENT	EARN
AGE	1	-0.49583	0.27781	-0.02197	0.04035	-0.04876	-0.13106	-0.17861
DD	-0.49583	1	-0.5998	0.14876	0.06366	-0.15149	0.08518	0.18336
DA	0.27781	-0.5998	1	0.3839	-0.0006	0.00248	-0.17702	-0.02082
GIR	-0.02197	0.14876	0.3839	1	0.06611	-0.03842	-0.00519	0.42915
PUTT	0.04035	0.06366	-0.0006	0.06611	1	-0.26356	-0.0111	-0.31764
SS	-0.04876	-0.15149	0.00248	-0.03842	-0.26356	1	0.07846	0.24331
EVENT	-0.13106	0.08518	-0.17702	-0.00519	-0.0111	0.07846	1	0.03757

Table 6: Means and S.D.s		
Variable	Mean	Std Dev
lEARN	13.46416	1.02777
lEVENT	3.16972	0.28402
AGE	36.21267	6.59373
AGE2	1355	483.7311
DDDA	18429	1323
lGIR	4.17256	0.04333
lPUTT	0.57662	0.01422
lSS	3.87496	0.14682
wlPUTT	1.73528	0.04279

Mathematical Appendix

Appendix 1

$EARN = 896.160 - 23.573 * lEVENT + 4.166 * AGE - 0.052 * AGE2 +$

$0.002 * DDDA - 344.713 * lGIR + 677.115 * lPUTT - 32.804 * lSS$

$+ 2.067 * yhat2 - 0.015 * yhat3 + 0 * yhat4.$

Appendix 2

$$\frac{\partial \ln Y}{\partial X} = \frac{\dfrac{dY}{Y}}{dX} = (b_1 + 2b_2 X) * Y$$

$$Elasticity = \frac{\partial Y}{\partial X} * \frac{X}{Y} = ((b_1 + 2b_2 X) * Y) * \frac{X}{Y} = (b_1 + 2b_2 X) * X$$

References

PGA Tour Statistics-2006, http://sports.espn.go.com/golf/statistics, accessed, 2/13/07.

PGATour.com. http://www.pgatour.com, accessed 2/22/07.

USOpen.com. http://www.usopen.com/2003/press/interviews/woods-6-14-flash.html, accessed 3/25/07.

The Sports Network. *Nationwide Tour Player Profile- B.J. Staten, www.sportsnetwork.com*, accessed 2/22/07.

University of Vermont, http://www.ento.vt.edu/~sharov/PopEcol/tables/chisq.html, accessed 4/5/07.

Moy, Ronald L., and Thomas Liaw. "Determinants of Professional Golf Tournament Earnings." *American Economist,* 1998, *volume* 42(1), pp. 65-70.

Rishe, Patrick James. "Differing Rates of Return to Performance: A Comparison of the PGA and Senior Golf Tours." *Journal of Sports Economics,* 2001, *volume* 2(3), pp. 285-96.

Scully, Gerald W. "The Distribution of Performance and Earnings in a Prize Economy." *Journal of Sports Economics,* 2002, *volume* 3(3), pp. 235-45.

Shmanske, Stephen. "Gender, Skill, and Earnings in Professional Golf." *Journal of Sports Economics,* 2000, *volume* 1(4), pp. 385-400.