Effect of Skills on Player Rank: An In-depth Look at the PGA Tour

Desirae Haselwood
12/13/2016
Senior Capstone
Department of Economics
Pacific Lutheran University

Introduction

The FedEx Cup is a season-long race for male professional golfers and is run by the Professional Golf Association (PGA) Tour. Points are earned throughout the season based on finishes in tournaments and every player is given a FedEx Cup rank based on these points. It is assumed that the rank is related to earnings as well, as a better finish results in more points and a higher payout for that tournament. On average, a Tour player plays 25 events in the newly adopted wrap around October-September season (PGA Tour, 2016). Numerous components contribute to the game of golf and this paper explores a variety of factors in forecasting the determinants of a player's rank at the end of a season. This capstone will first examine previous literature and use a theoretical production model to consider the relationship between skills and rank. It will then investigate empirically using multiple regression analysis factors to consider this same relationship.

Previous research has examined earnings of professional golfers and can be used to study ranking directly. Connolly and Rendleman Jr. (2008) found that "mean skill alone is insufficient to win a golf tournament" and that luck is also needed. Luck is a random component unexplained by factors in their model. Shmanske (2014) conducted a study that tested skill and performance results on earnings for PGA Tour players. He found that "driving distance, approach shot accuracy, and putting prowess are shown to have significant measurable effects on earnings" (Shmanske 2014). While his work was directly about earnings, this finding can still be useful for this project because winning, or finishing closer to the leader in a tournament, is related to higher earnings. Moy and Liaw (1998) found that the number one ranked player on the official money list in their data set was "excellent in all categories but did not lead the tour in any performance category," suggesting that you do not need to be number one at any given part of the game in order to have the highest earnings. All of these studies provide insight on what determines how much prize money a player earns. These insights will be used to further look at how much the different variables effect how a player ranks at the end of the season.

Data from the statistics section on PGAtour.com for the top 125 players on the PGA Tour for the 2014-2015 season is used. Several performance statistics and player characteristics were gathered and a player's long game as well as short game is represented. Long game pertains to shots that are hit when a player is farthest from the hole, such as shots using a driver or other metal woods, hybrids, or long irons. Short game statistics represent skills that a player uses when on or around the green such as chip shots, bunker shots, and putting. Some player characteristics included in this paper are age and marital status.

First, a Cobb-Douglas Production function will be introduced to model the marginal effect of different inputs on the output (rank). We then conduct a multiple regression study using statistics from the entire season to test which variables are statistically significant in contributing to a better rank at the end of the season. Ultimately, earnings was used as a proxy for rank in the theoretical and empirical analysis, as rank is not a continuous variable. A model using ln(earnings) as a dependent variable as well as a model using ln(earnings) per event was used to study the effect of skills on earnings. From each regression, a forecasted rank was created and tested against the actual rank. It was found that the regression using ln(earnings) was highly correlated with actual rank. A Spearman test showed the ln(earnings) model was able to forecast, using skills, a rank that correlates with the actual rank at better than 60%. Specifically, scoring average and putting percentage inside ten feet were statistically significant in receiving higher log earnings at the end of the season, which is comparable to a higher rank.

Literature Review

There are several factors that could define performance of professional golfers – their skill set, money incentives, and their returns to each skill, to name a few. Based on these factors, this literature review will consider several items. It will first consider the relationship between a player's inputs and output, as modeled with a Cobb-Douglas production function. Second, this literature review will consider the skills-earnings relationship that is recognized in professional golf as well as in other sports. Thirdly, the incentive effects of professional golf tournaments and the relationship between earnings and a player's rank will be discussed. Finally, this literature review will examine the methods of previous empirical studies.

Cobb-Douglas Production Function

A Cobb-Douglas production function tells us about the relationship between inputs and an output, marginal product, and returns to scale. Inputs that are commonly used are physical capital and labor for manufacturing applications. Alcantara and Prato (1973) support the idea that the combination of inputs do yield an output, and sometimes a certain input may have a greater effect than another. Shah (1982) conducted a study using a production function to represent the dairy industry of Gujarat using total capital, total labor, and milk supply as inputs and gross value as output. Milk supply was defined in terms of "the quantity of raw milk procurement of the dairy plants," and the output was determined by "adding sales of milk and dairy products and closing stock and depreciation and subtracting opening stock of the milk and dairy products" (Shah, 1982). It was concluded that milk supply was the most significant input influencing the gross value of output, as milk supply increased, value of the dairy plant increased as well. (Shah, 1982). In Shah's, the Cobb-Douglas production function assumes that an increase in milk supply would result in increasing plant value up to a specific point, at which plant value would continue to increase but at a decreasing rate (Shah, 2982). Returns to scale can also be determined using a Cobb-Douglas production function. Shah's (1982) study showed increasing returns to scale, denoting that with a proportional increase in inputs, the gross output of milk production increased more than proportionally. A Cobb-Douglas production function can also model constant or decreasing returns to scale and allow for production at the optimal output level.

While most Cobb-Douglas production functions model manufacturing or agriculture production, Cobb-Douglas production functions have sometimes been used to model the relationship between inputs and outputs in sports settings. Considering golf specifically, researchers have used a Cobb-Douglas production function as a model to examine productivity and the relationship between the inputs and the performance of the golfer (Ehrenberg and Bognanno, 1990; Watkins, 2008). Ehrenberg and Bognanno (1990) present a production function showing a basic positive relationship between rank and payout using this production function. Their study uses nonexperimental data to test whether tournaments elicit effort responses. Their finding shows that a higher or better finish (first place) gives a player the highest percent of earnings. Ehrenberg and Bognanno's (1990) function also models an effect of marginal return for players. Improving a player's rank by one spot has a much greater marginal return if that player was closer to the leader already than a player who was farther down in the rankings (Ehrenbern and Bognanno, 1990). Using a Cobb-Douglas production function to model professional golf scenarios can be helpful in studying marginal returns and returns to scale in the sport.

Payout Incentives

Much of the literature regarding professional golf pertains to earnings received by players (Moy and Liaw 1998; Ehrenberg and Bognanno 1990; Watkins, 2008; Nero 2001). There is a question as to whether payout is seen as an incentive to play well since professional golfers "do not have guaranteed contracts and their tour earnings are based solely on current performance" (Moy and Liaw, 1998). Ehrenberg and Bognanno (1990) found that professional golf tournaments do have incentive effects. They found that "higher prize levels do lead, other things equal, to lower scores" primarily in later rounds of the golf tournament. By design, professional golf tournaments pay heavily to the top finishers as the winner takes home 18% of the total prize fund (PGA Tour, 2016).[1]

Research conducted about workplace settings has shown that higher wages leads to higher productivity and better performance (Aslam, 1983). This relationship has been seen in labor markets as well as in sports scenarios (Ehrenberg and Bognanno, 1990). Some have argued that if earnings are determined by performance, then professionals "should invest their time in areas of the game that will help maximize their wealth" (Moy and Liaw, 1998). Watkins Jr. (2008) studied the rates of return to golf skills on the PGA Tour and found that driving distance has virtually no impact on earnings. If some areas of the game lead to higher earnings then it is suggested that players should focus on those specific parts of the game. Nero (2001) proposes that the best golfers are the ones who are able to utilize their strengths and avoid their weaknesses. Although better performance does bring higher earnings, they are only positively, but not perfect correlated.

Skills-Earning Relationship

The payout incentives of PGA Tour events leads us to explore a relationship between a player's skill set and their earnings. When it comes to driving the golf ball off of the tee, measured by either accuracy or distance, there are mixed findings on how it affects player's earnings. Moy and Liaw (1998) argue that to be successful on the PGA Tour a player must be

[1] The average purse amount for a PGA Tour event is $7,000,000. That puts the first place prize at about $1,260,000. The purse ranges anywhere from 2,500,000 to $10,500,000 depending on the event (PGA 2016).

able to drive the golf ball a long way, in their study driving distance had a p-value of 0.000 and was therefore was statistically significant. Similarly, another study found that driving distance, as well as approach shot accuracy and putting proficiency, have “significant measurable effect on earnings” (Shmanske, 2014). Based on Moy and Liaw (1998) and Shmanske’s (2014) findings, players should focus on being a good driver of the golf ball to yield higher earnings. A study by Watkins (2008) found that male professionals drive the ball far enough to put them in good scoring position with the distance they hit the ball, and that increasing distance brings higher risk of lower accuracy. In this case, the cost of possibly missing the fairway is greater than the reward of extra yardage. Usually very long hitters of the golf ball are not the most accurate hitters (PGA Tour, 2016). While researchers hold opposing views on the importance of driving distance and accuracy, it is still deemed one part of the game that should be considered in the investigation of higher earnings.

Hitting a higher percentage of greens in regulation is deemed to have a positive effect on player’s earnings (Shmanske, 2014; Moy and Liaw 1998; Watkins, 2008).[2] This idea is intuitive for golfers, since hitting a green in regulation gives a golfer a better chance at one-putting for birdie or two-putting for par. Missing the green puts pressure on a player’s skills around the green and increases the chances of a player making a bogey or worse on a given hole. Skills around the green, or short game skills, are skills such as chip shots and bunker shots. Any shots from inside 100 yards are considered part of a player’s “short game.” Shmanske (2014) in his study about PGA Tour player’s compensation, found that hitting more greens has a positive effect on earnings. Shmanske’s study goes further to include “approach shot accuracy” as a skill that yields higher earnings as well. Both greens hit in regulation and approach shot accuracy are

[2] Moy and Liaw’s (1998) study had a p-value of 0.000 for greens in regulation and is therefore significant at all levels of significance.

significant at the one percent level. However, it should be noted that a one percentage point increase in approach shot accuracy is associated with a 9.3 percent increase in earnings, while the same increase in greens in regulation percentage is associated with a 10.3 percent increase in earnings (Shmanske, 2014). This suggests that merely hitting the green has a greater positive effect on a players earnings compared to the accuracy of the shot. Watkins Jr. (2008) also finds that greens in regulation are “one of the most significant determinants of earnings on Tour” and shows that being good at short game skills can increase earnings. Pelz (1999) states that “60% to 65% of golf shots occur inside 100 yards of the hole.” Short game skills such as bunker shots and chip shots would come into play when a player misses the green in regulation and are most commonly measured by scrambling and sand save percentages.

Beyond being a good driver of the ball and hitting greens in regulation, being a good putter is also considered an important skill to have. Putting skills can be measured in numerous ways but is often shown using average putts per round or average putts per hole. Nero (2001) and Shmanske (2014) found that putting has a “much greater impact on earnings than additional [driving] distance.” Their findings lead us to conclude that the best putters on Tour are most likely the highest money earners as well. Shmanske (2014) found that by taking “one fewer putt per round (four strokes per tournament) a golfer will increase their earnings by about 119 percent.” Thus, having refined putting technique has a positive effect on the skills-earning relationship (Shmanske, 2014; Nero, 2001).

Earnings-Rank Relationship

Professional golf tournaments are designed to pay the most money to the player with the best finish. While payouts are distributed from players ranked one to seventy at the end of a tournament, the payouts heavily favor the top finishers. For example, the top five finishers of the

2015 Players Championship earned 18%, 8.8%, 8.8%, 4.4%, and 4.4% of the total prize respectively, with a total purse of $10,000,000.[3] Ehrenberg and Bognanno (1990) showed that a better rank is related to higher earnings. Beyond this, researchers have debated as to whether streaky play can affect a player's rank and therefore, their earnings (Connolly and Rendleman, 2008). Streaky play is defined as when a player faces a period of repeated success or failure.

There is abundant literature regarding streaky play and whether it happens in sports or not (Connolly and Rendleman; Albright, 1993; Bower, 2011). Dorsey-Palmateer and Smith (2004) found that in bowling, rolling a strike in the current period is not independent of whether a strike was bowled in the previous period, suggesting that bowling a strike in the previous period increases the likelihood of bowling a strike in the next period. This finding supports the theory that streaks can happen in sports and that streaky play can lead to better or worse performance. Connolly and Rendleman Jr. (2008) found some evidence of streaky play in golf, stating a "small number of PGA Tour participants experience statistically significant streaky play." While there was no major evidence, as streaky play was only evident in 23 players within the sample of over 1,400 players, Connolly and Rendleman Jr. (2008) did find a small number of players experiencing streaky play. Evidence of streaky play in sports would result in a weaker relationship between rank and earnings. A player may get on a hot streak and finish better in a tournament than they would have without the streak. In this case their hot streak in addition to their skill earns them a better rank and therefore higher earnings in the tournament. Thus the hotter the streak, the more someone's earnings may increase (Connolly and Rendleman, 2008).

Other researchers found that streaky play does not exist in sports. In a study regarding hitting streaks in baseball, Albright (1993) found that hitting is more consistent with a model of

---

[3] There was a tie for second and fourth place which explains the duplicate percentages of payout amount. First place won $1,800,000, the tied for second place won $880,000, and the tie for fourth place won $440,000.

randomness. Albright's (1993) study also found that there may be noticeable streakiness within a season, but that it is in a random sequence, "just as one would expect a certain proportion of people flipping fair coins to produce streaky sequences of heads and tails" (Albright, 1993). Camerer (1989) investigated a "hot hand" in basketball and found no evidence of streaky play. It was found that even the best player's shooting records follow a random sequence and that making one shot does not influence whether the next shot the player takes is made (Bower, 2011). Although there is no evidence of a player having a "hot hand," sports fanatics and gamblers still believe it is relevant (Camerer, 1989). Bower (2011) argues that the human brain is attuned to finding patterns and streaks in sporting events, even if they are not actually occurring. The argument that sports follow a random sequence would potentially make the relationship between a player's rank and earnings stronger because there would be no factor of "hot hands" or streaky play to explain why a player may have performed better during a particular event.

Previous Empirical Studies

Numerous studies have modeled various factors pertaining to professional golf and their potential effects on earnings and performance. Many of them looked at how different variables influence the dependent variable "earnings" (Nero, 2001; Moy and Liaw, 1998; Watkins Jr., 2008). Each of these studies used a logged version of earnings as a dependent variable and used a least squared regression. Nero's (2001) study differed in that he used a semi-logged function, as the independent variables used were not logged while the dependent variable was. Independent variables were comprised of skill determinants, such as driving distance, putts per green, and number of fairways hit (Nero, 2001). Moy and Liaw (1998) and Watkins Jr., (2008) both included independent variables that were logged, such as number of putts. Watkin Jr.'s (2008) study was unique in that age was included because he theorized that as age rises so do skill levels, up to a certain point. Shmanske (2014) conducted a simple study with the log of yearly earnings per tournament depending on scoring average as a measure of performance. While each of these studies used a variation of similar variables, the methods and findings were often alike, in that there are a few aspects of the golf game that are critical to success such as number of putts and scoring average (Nero, 2001; Moy and Liaw, 1998; and Watkins, 2008; Shmanske, 2014). Moy and Liaw (1998) found number of putts to be significant with a p-value of 0.000

Professional golf is an interesting case study in sports economics as there are numerous tournaments that impact a player's season-long rank and many skills that contribute to the game of golf. Performance can be based on earnings, skills, and time invested in developing skills. The relationship between these inputs and earnings received can be modeled by a Cobb-Douglas production function. In addition, there is a relationship between a player's skills and earnings as well as rank that should be analyzed. Those who empirically study professional golf should also take incentive effects and the possibility of streaky play into consideration. Finally, looking at previous empirical studies is useful when studying this topic because it gives a strong base for the research question of this capstone.

Theoretical Model

The Cobb-Douglas production function is used to show the relationship between different inputs and an output defined to be professional golfers' earnings for the season. This model was produced by Cobb and Douglas (1928) and is generally used to show the effect of labor and capital on the production of some output, usually a good or service. It is traditionally used in the manufacturing sector in examining what happens when different levels of labor and capital are used to produce a good. A Cobb-Douglas production function can also be used to look at the relationship between one input and an output. For the purpose of this paper, the Cobb-Douglas production function will be used to look at the effect that a particular golf skill has on the earnings of a player on the PGA Tour. The goal of using this model is to consider the relationship between a player's rank as their skill level changes. Earnings is used as a proxy for rank and it is important to note that the theoretical model and empirical model in this paper will use earnings as the dependent variable. The nature of the relationship between rank and earnings will be discussed later.

Cobb and Douglas (1928) use the following equation to approximate output:

$$Y = AL^{\alpha}K^{\beta} \tag{1.1}$$

where Y is the amount of output, L is the amount of labor, K is the amount of capital, α is the output elasticity of labor, β is the output elasticity of capital, and A is the total factor productivity. Output elasticity measures the responsiveness of output to a change in levels of an input, all else equal. The variable A pertains to total-factor productivity, and is determined by how efficiently the inputs are utilized in production. It is often an intangible variable such as technology or worker human capital. Cobb and Douglas (1928) used this specific function to estimate the theory of production, relative to the indices of production, labor, and capital during the period of 1899-1922 in the United States. They found that labor contributed to about 75% and capital about 25% of the increase in the manufacturing production.

This model assumes that the inputs of labor and capital positively affect output of the product. It also assumes that labor and capital can be used in variable proportions, unlike fixed-proportion production functions where inputs are perfect complements. Additionally, the rate at which labor can be substituted for capital is not constant along the isoquant, unlike in a linear production function. From this model we can learn about output elasticity of inputs, returns to scale, elasticity of substitution, and marginal product. Returns to scale are constant if output elasticities of labor and capital equal 1, as found by Cobb & Douglas (1928). When comparing two inputs and their impact on a single output, the marginal product of a Cobb-Douglas function is positive and decreasing. Additional inputs can be added and output will continue to increase up to a certain point where additional inputs will no longer increase output and may even decrease it. There is an upward sloping total product curve when the effect of a single input on output is studied.

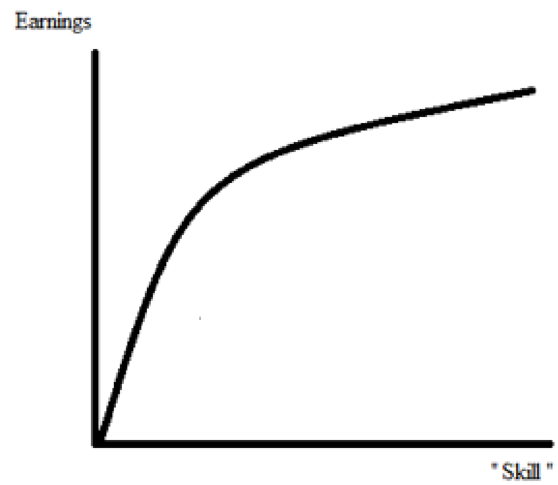
For the purpose of this paper, labor and capital will not explicitly be used as inputs. Instead, different golf skills are used as inputs and a proxy version of rank will be used as the output in the form of earnings. Earnings is the season-long sum of prize money from tournament play for a particular player and excludes sponsorships and contract deals. The inputs included in equation 1.2 are driving accuracy percentage, greens hit in regulation percentage, average putts per round, and scrambling percentage. The equation takes form of:

$$\text{Earnings} = AD^{\beta 1}G^{\beta 2}P^{\beta 3}S^{\beta 4} \tag{1.2}$$

where A is total factor productivity, D is driving accuracy percentage, G is greens hit in regulation percentage, P is average putts per round, and S is scrambling percentage. The econometric model of this Cobb-Douglas equation will include additional inputs.

The use of the specific variables in equation 1.2 come from previous literature. Moy and Liaw (1998) in a similar study found that driving accuracy and greens hit in regulation were significant skills that effect professional golfer’s earnings. In a different study, Shmanske (2014) found that putting has significant effects on earnings. Both Moy and Liaw (1998) and Shmanske (2014) found sand save percentages to be significant, so for the Cobb-Douglas production equation derived above, a variation of this statistic was included: scrambling percentage. The difference is that sand save percentages measure how many attempts were successful when attempting to get up and down from the sand, and scrambling looks at successful attempts when trying to get up and down from anywhere around the green. The graphical expression of the Cobb-Douglas equation that was derived for this project is configured by showing any one of the specific skill inputs and their effect on earnings:

Figure 1. Cobb-Douglas Function with Effect of One Input on Output



The upward sloping curve on Figure 1 demonstrates the same positive and diminishing marginal product that manufacturing may see with inputs of labor and capital. A similar relationship is expected here if we look at the relationship between two specific golf skills and their effect on the output of earnings. Because Figure 1 specifically shows the effect of an increase in skill on earnings, the curve is upward sloping at a decreasing rate because as a player’s skill improves, their earnings are expected to rise.

Empirical Analysis

I. Data

A significant portion of the variables used in this project were gathered from the PGA Tour official website – PGAtour.com. Data was acquired for the 2014-15 FedEx Cup season for the top 125 players. The FedEx Cup happens every year and is what the PGA Tour calls their season. The season starts in October and goes through September of the following year. The data for this project is from the season that began in October 2014 and ended in September of 2015. During the FedEx Cup season players are awarded points based on how they finish in tournaments, with emphasis placed on wins and high finishes. The player whom has the most points at the end of the season wins the FedEx Cup and is ranked number one. The variables that represent number of wins and number of top ten finishes were gathered for this capstone to show how much success a player had during the season. Driving accuracy percentage is the percentage of fairways hit by a player and is a season long statistic. The higher the percentage, the more fairways that player has been recorded to hit which is expected to improve their chances of a win. Greens in regulation percentage is a variable that measures how accurate a player is at hitting the green. A green in regulation is successful when a player gets their ball onto the putting green in two less strokes than par, so for a par-3 a player would hit the green in one stroke, two strokes on a par-4, and three strokes on a par-5.

Scoring average is an independent variable that is an average of the score a player makes each time they play a tournament round. Scores from every round they play throughout the season are used to find their average. For example, Jordan Spieth had a scoring average of 69.9 during the 2014-2015 season, the lowest scoring average on Tour that season. A lower average is better because in golf you want to shoot a lower score rather than a higher one. While this variable is not exactly a “skill” it is an important golf statistic. Average putts per round is a number that represents how many times a player putts the ball on average. A lower amount of putts is better than a higher amount. Putting inside ten feet is different in that it is a percentage of how many putts are made from inside ten feet by a player. The putting inside ten feet statistic was included to offset any bias there may be between some of the golf statistics. For example, a player may struggle with hitting greens in regulation but may be very good at making putts inside ten feet to save the hole and keep their scoring average low.

Sand save percentage is a statistic that represents how often a player is successful at getting a sand save. A sand save means a player who misses the green is able to get the ball out of the sand onto the green and putted into the hole in two or less strokes. The next variable, scrambling percentage, is similar but represents chip shots or putts from less than 50 yards off the green with an additional one putt on the green resulting in par or better. A chip in is also counted as a successful "scramble." The variable for official money represents the earnings a player made from tournament play during the season. This variable will ultimately be logged as the dependent variable. An indicator variable for married is included to represent whether the player is married or not, a 1 meaning they are and a 0 meaning they are not. A Google search was done on each player to verify if they were married or not, most players have their own official website that includes personal information about themselves. Age represents how old a player is in years. The short version of these variable names, as well as their description and source can be found in Table 1.

Table 1. Variable Definitions.

Variable	Abbreviated Variable	Description	Source
2015 FedEx Cup Ranking	Rank	Rank given to players on the PGA tour. Players earn points throughout the season based on how well they perform in tournaments. Strong emphasis is put on winning and high finishes in tournaments. At the end of the season the player with the most points wins the FedEx Cup.	PGA Tour Website
# of Wins	Win	Number of tournaments a player won during the PGA tour season (October-September).	PGA Tour Website
# of Top 10's	TopTen	Number of times a player finished in the Top-10 at a tournament during the PGA tour season.	PGA Tour Website
Events Played	EvntsPld	Number of events a specific player participated in during the PGA tour season.	PGA Tour Website
Driving Accuracy %	DrivAcc	Percentage of drives landing in the fairway on par 4 and par 5 holes.	PGA Tour Website

Greens in Regulation %	GrnReg	Percentage of greens hit in regulation. A golfer earns a green in regulation by getting their ball onto the putting green in one stroke on a par-3, two strokes on a par-4, or three strokes on a par-5.	PGA Tour Website
Scoring Average	ScrAvg	This variable is an average of the score a player makes each time they play a round of golf. Scores from every round are used to find their average. A lower average is better because in golf you want to shoot a lower score rather than a higher one.	PGA Tour Website
Average Putts per Round	AvgPuttRd	How many putts on average a player has during a round. Putts are only counted when they are on the putting surface.	PGA Tour Website
Putting inside 10 Feet	PuttInsdTen	This is shown as a percentage. It is determined by measuring how many putts that are inside 10 feet go in the hole out of how many are attempted. A higher percentage means that a player has made more putts inside 10 feet.	PGA Tour Website
Sand Save %	SndSv	The percentage of times a player gets the ball into the hole in two shots (or less) from a greenside sand trap.	PGA Tour Website
Scrambling %	Scmrblng	This is a percentage of how often a player successfully “scrambles.” The scrambling statistic is defined as “a chip shot or putt from less than 50 yards off the green with an additional one putt on the green resulting in par or better on a hole.”	PGA Tour Website
Official Money Made	OffMney	The prize money a player earned during the season. This is a numeric dollar amount. This is for PGA tour players only. (so it excludes European tour, Champions tour, web.com tour)	PGA Tour Website
Married	Married	Variable showing if a player is married. 1=yes, 0=no.	Player’s Official Website
Age in 2015 Season	Age	Numerical value that shows the age of a specific player during the 2015 season.	PGA Tour Website

The variables were chosen in order to connect this project with the economic literature. Most papers looked at earnings as a dependent variable. Because the dependent variable for this capstone had to be altered from rank to earnings, as rank is not a continuous variable, the economic literature found was useful when deciding what variables to include. The variables driving accuracy percentage and greens hit in regulation percentage were included because similar studies (Moy and Liaw 1998, Shmanske 2014) found these skills to have a significant impact on earnings. There is similar reasoning behind including sand save percentage, scrambling percentage, and average putts per round (Moy and Liaw 1998; Shmanske 2014; Watkins, 2008; Nero, 2001). I decided to include a variable representing putting inside ten feet to see if there was bias in the statistics of average putts per round and scrambling. A variable for marriage was included to see if that has any effect on a player's rank, as having a partner may increase the stress of having to play well to support a family while being single may leave a player with a clearer head. The variables used in our estimation are defined explicitly in table 1.

Table 2 shows the descriptive statistics for the variables acquired for this project. Their mean and standard deviations are computed and reported in the table. The majority of the variables had means and standard deviations that made sense. There are two skill statistics that are expected to have a lower mean, and those are scoring average and average putts per round. With a lower mean, a player is better at those skills, which leads to a better golf score and therefore higher earnings. None of the descriptive statistics stand out as extremely unusual or volatile.

Table 2. Descriptive Statistics.

Variable	**Mean**	**Standard Deviation**
rank*	63	36.228
win*	.368	.7883
topten*	3.808	2.644
evntspld*	25.496	4.52
Drivacc	62.268	4.99
Grnreg	66.848	2.58
Scravg	70.669	.5468
avgputtrd	29.009	.4778
puttinsdten	87.46	3.74
Sndsv	51.1	6.062
scrmblng	59.6	2.729
married	.744	.4381
Age	33.512	6.311

*variable omitted and not used in final regression

II. Econometric Model

This study uses cross-sectional data to analyze the effect of different skills and player characteristics on a player's given rank. The econometric model stems from the Cobb-Douglas function discussed earlier in the paper. The econometric model also takes a form similar to other economic papers (Shmanske, 2014; Moy and Liaw; 1998, Watkins; 2008). It was decided to do two different regression analysis. One would use ln(earnings) as a dependent variable while the other would use ln(earnings) per golf event played as a dependent variable. The reasoning for using ln(earnings) per event in addition to the ln(earnings) model is similar to using GDP versus GDP per capita in a macroeconomic scenario. Overall GDP may be large because the country is large, whereas to study production on a per person basis we need to use GDP per capita as a dependent variable. This can be translated into terms for this project as looking at ln(earnings) per tournament may give a better explanation as to specific player skills and payoffs. It is an empirical question to which model performs better. After each regression is complete, a forecasted rank from the skills regression will be predicted and then correlated with the actual rank to determine the extent to which skills impact rank indirectly. The econometric model is the natural log form of the Cobb-Douglas production function from earlier in the paper:

$$\text{Ln(Earnings)} = \beta_1 + \beta_2 \ln(\text{DrivAcc}) + \beta_3 \ln(\text{GrnReg}) + \beta_4 \ln(\text{ScrAvg}) + \beta_5 \ln(\text{AvgPuttRd}) + \beta_6 \ln(\text{PuttInsdTen}) + \beta_7 \ln(\text{SndSv}) + \beta_8 \ln(\text{Scrmblng}) + \beta_9 \text{ Married} + \beta_{10} \ln(\text{Age}) + u_i \quad (1.2)$$

Where ln(earnings) is log earnings[4], ln(DrivAcc) is log driving accuracy percentage, ln(GrnReg) is log greens in regulation percentage, ln(ScrAvg) is log scoring average, ln(AvgPuttRd) is log average putts per round, ln(PuttInsdTen) is log putts inside ten feet percentage, ln(SndSv) is log sand save percentage, ln(Scrmblng) is log scrambling percentage, Married is an indicator variable for whether a player is married or not, defined as 1 is yes and 0 is no, and ln(Age) is age

[4] When we use this same equation to look at ln(earnings) per event as a dependent variable instead of just ln(earnings), the only thing to change is in fact the dependent variable. It changes from ln(earnings) to ln(earnings)/event.

of the player during the 2014-15 season. It is important to note that the indicator variable for marital status is not logged, as an indicator variable cannot be logged or it will be omitted from the regression. Equation 1.2 was also used for the model using log(earnings) per event as a dependent variable.

If a player is better at a certain skill it is expected to have a positive effect on their earnings. Coefficients for driving accuracy percentage, greens hit in regulation percentage, putting inside ten feet percentage, sand save percentage, and scrambling percentage are all predicted to have positive coefficients. As a player's skill increases for any of these skills, their earnings should increase. The coefficients for scoring average and average putts per round are predicted to be negative. This is because having a lower scoring average, and thus playing better golf, leads to higher earnings. Likewise, making more putts during a round leads to a lower score and thus higher earnings.

It is not as easy to predict the sign for the indicator variable of marital status. For the sake of this model I predict that the coefficient for the married variable will be negative. I predict that if a player is married it will lead to lower earnings as they have more than just themselves to worry about which may increase stress and lead to poor play. Age is another difficult variable to predict the coefficient for. A paper that conducted a similar study concluded that as a player ages, it is expected that his level of human capital will increase in proportion to the amount of time he has spent practicing and playing golf and at a certain point, the player will lose some physical strength and hand-eye coordination (Watkins, 2008). With this, I predict that age will have a positive effect on earnings because the mean age for a PGA tour player from this sample is 33, so we can assume that they have not hit the point of capital depreciation.

Several preliminary regressions were run before figuring out the final regression equation.[5] The regression results from equation 1.2 will be used to generate a player's predicted rank. This predicted value will then be sorted by earnings and used to generate a new variable that is called "rank 2" which is a forecasted rank. This will be done for the ln(earnings) model as well as for the log(earnings) per event model. A correlation test will be performed between the forecasted rank and the actual rank for each model. A higher correlation would mean that the rank that was predicted using a player's earnings is highly associated with their actual rank. A lower correlation would conclude that the model's predicted rank based on the skills and demographics for the players as determinants of earnings does not give a rank that is close to their actual rank.

One thing that could have been an issue during this study was multicollinearity.[6] There were several additional variables that were considered for this study, but they were omitted so as to not cause multicollinearity. The ranks that a player held for each skill were not included for this reason. Jordan Spieth, for example had the lowest rank (so best skill statistic) for average putts at 27.82 average putts per round.[7] This gave him the number one rank for this skill. All of the skills had this accompanying variable that represented the players rank for that specific skill, but these were omitted because they were defined using the same skill and were highly correlated with the skill variable it was accompanying. Other variables such as the indicator

[5] The issue of using a non-continuous dependent variable could either have been resolved by using the ordered-probit model or by modifying the dependent variable. The latter solution was used, and the dependent was modified from being rank to earnings.

[6] Variables such as number of wins, number of top ten's, number of years on Tour, official world golf ranking, number of events played, and an indicator variable relaying whether a player has kids or not were all originally going to be part of this model. Each variable was omitted because it was either too highly correlated to another independent variable or was too closely related to the dependent variable ln(earnings). A VIF and Corr function were used in Stata to determine whether variables needed to be omitted or not. It is already assumed that number of wins and top ten's highly attribute to higher earnings and therefore rank.

[7] These statistics were pulled from pgatour.com and is part of the underlying data set used for the econometric model.

variable for kids and the variables for number of wins and top ten finishes were omitted because they were too correlated or closely related to other variables. The variable representing whether a player had children or not was highly correlated with the variable regarding marital status. Number of wins and number of top tens were highly correlated with rank. Multicollinearity is an important econometric concern that was predicted to be an issue in this paper and by excluding variables that were highly correlated, this specific issue has been fixed.

III. Results

The first regression that was run used ln(earnings) as a dependent variable. Table 3 shows the coefficient estimates from this regression.

Table 3. Coefficient Estimates using Ln(Earnings) as Dependent Variable[8]

Variable	Coefficient	Robust Standard Error	P > $\|t\|$
lndrivacc	-1.0992**	.5501501	0.048
lngrnreg	.6495438	2.575359	0.801
lnscravg	-49.90802***	7.728056	0.000
lnavgputtrd	-9.435281*	5.256725	0.075
lnputtinsdten	1.198971***	.1747908	0.000
lnsndsv	-.1565449	.3562211	0.661
lnscrmblng	-.4702036	1.262839	0.710
married	.190372*	.1018526	0.064
lnage	-.4835975**	.2444158	0.050

The R-squared of this regression was 0.5529. The independent variables ln(scoring average) and ln(putts inside ten feet) were significant at the one percent level, as their p-values were less than 0.01.The variables ln(driving accuracy percentage), and ln(age) were significant at the five percent level and ln(average putts per round) and the married variable were statistically significant at the ten percent level. This indicates that each of these variables has a significant effect on ln(earnings). The results for ln(scoring average) and ln(average putts per round) have

[8] One star means the variable was statistically significant at the 10% level, two stars means it was significant at the 5% level, and three stars means it was significant at the 1% level.

the expected coefficient sign of negative. As a player's scoring average goes down by one percent, their earnings are expected to rise by 49.9 percent, and as their average putts per round goes down by one percent, their earnings are expected to rise by 9.43 percent. These parameters align with what was expected because intuitively a lower score should lead to higher earnings since professional golfers are paid based on performance and in the sport of golf a lower score is better than a higher score. Likewise, a lower amount of putts lead to lower scores which leads to higher earnings.

The coefficients of the variables ln(driving accuracy percentage)and ln(age) are also negative, meaning they have a negative impact on ln(earnings).[9] It was unexpected for these variables to have negative coefficients. It would suggest that one percent increase in driving accuracy would produce a 1.09 percent decrease in ln(earnings). While this result seems contrary to popular belief, the effect on earnings is very small, implying that while ln(driving accuracy percentage) came out significant, it still has a small impact on total earnings or that there are other skill statistics that are more important.

The coefficients for ln(green in regulation percentage), ln(putts inside ten feet percentage), and married had positive coefficients. Out of these, ln(putts inside ten feet percentage) was the only one that came out significant. The model showed that a one percent increase in putts inside ten feed percentage would lead to a 1.19 percent increase in earnings.

After the regression was run and analyzed, forecasted log earnings were predicted. These earnings were sorted such that the highest earnings would yield the number one rank to generate a forecasted rank, rank2. Next, a correlation test was done between the player's new predicted rank that was determined by ln(earnings) and their actual rank that was pulled from pgatour.com.

---

[9] Ln(sand save percentage), ln(scrambling percentage) also had negative coefficients, but were not statistically significant.

The sample correlation was 0.7382. Knowing the forecasted rank explains 73.82% of the variation in the actual rank. If this was all we knew to make a guess for actual rank, it is highly correlated.

After completing the regression and correlation test for the model using ln(earnings) as a dependent variable, the procedure was completed using ln(earnings per event) as a dependent variable. Table 4 shows the coefficient estimates.

Table 4. Coefficient Estimates using Ln(Earnings) Per Event as Dependent Variable

Variable	Coefficient	Robust Standard Error	$P > \|t\|$
lndrivacc	-.0037052	.2396024	0.988
lngrnreg	-1.615191	1.074714	0.136
lnscravg	-20.02787***	4.703847	0.000
lnavgputtrd	4.314634**	2.004373	0.033
lnputtinsdten	-.0093542	.0939172	0.921
lnsndsv	.4824711***	.1784463	0.008
lnscrmblng	-1.294103***	.4717393	0.007
married	-.0170105	.0395339	0.668
lnage	.3121201**	.102873	0.003

The R-squared of this regression was 0.4202. The independent variables ln(average putts per round) and ln(age) are significant at the five percent level of significance. The variables ln(scoring average), ln(sand save percentage), and ln(scrambling percent) are significant at the one percent level of significance. Of these statistically significant variables, ln(scoring average), ln(putts inside ten feed percentage), and ln(scrambling percentage) had negative coefficients. It was only expected to see a negative coefficient for the variable ln(scoring average), a one percent drop in scoring average is predicted to raise earnings by 20 percent. The negative coefficient of 1.29 on ln(scrambling percentage) would suggest that a one percent increase in scrambling percentage would lead to a 1.29 percent drop in earnings. While this does not align with what was predicted, its effect on earnings is only slight.

The variables ln(average putts per round), ln(sand save percentage), and ln(age) all had a positive coefficient and were statistically significant. The model found that a one percent drop in average putts per round would result in a 4.31 percent increase in earnings. It was not expected to see a positive coefficient for average putts per round, because intuitively one would expect a player's earnings to rise if they have less putts on average during a round because having less putts leads to a better score which leads to higher earnings. The coefficient for ln(sand save percentage) showed that a one percent increase in a player's sand save percentage would lead to a 0.48% increase in earnings.

As was done for the first regression, forecasted log earnings were predicted. They were sorted such that highest earnings yielded the number one rank and were used to create a rank2. A correlation test was again conducted between actual rank and predicted rank using ln(earnings per event). The correlation was 0.259 and shows a very weak relationship between actual rank and predicted rank, even though it could be anticipated that using this dependent variable would give a more accurate rank prediction the ln(earnings) model. It was decided to use ln(earnings) per event in this study because the number of events played varies greatly among PGA Tour players. Some players play in all 36 events during the season and some play in as little as 12. This variable was used to account for those players who maybe play in few events but finish very well and also for those who play in many events but maybe have mediocre performances all year. Earnings per event ended up being less related to a player's real rank than the ln(earnings) dependent variable was. The actual ranking is correlated both with earnings per event and with the total number of events played. Only the test using ln(earnings) as a dependent variable captures both of these, hence the higher correlation.

Conclusion

Besides the dependent variable being tested, both of these regressions and correlation tests were run the same way and used the same independent variables. While neither model was chosen to be used as a final model to draw conclusions from, the model using ln(earnings) was clearly stronger. For this model, a Spearman correlation test was conducted to see how good the model was as predicting a rank. The high correlation of .73 showed us that it predicted a rank that was highly correlated with actual rank, but the Spearman test would give insight on how good the ln(earnings) model was at forecasting the rank (Kendall, 1970). Using the correlation t-test and the standard deviation, it was found that the predicted model, using skills, could forecast the actual rank with a better than 60 percent correlation, significant at the one percent level. If the model was able to perfectly predict rank the correlation would have been 100 percent.

The goal of this project was to determine which skills or characteristics held true to lead to a better rank at the end of the PGA Tour season. In order to keep the dependent variable as a continuous one we could not look at rank directly. We know that skills predict earnings because players are paid based on performance. We can then infer that earnings correlate to rank. This is why we predicted a new rank based on ln(earnings) and then ln(earnings) per event. These tests can then determine the extent to which skills impact rank indirectly. Although the test using ln(earnings) had a much higher correlation to the actual rank there are takeaways from both predicted ranks. Both regressions resulted in scoring average and average putts per round being significant independent variables. Among both tests, age was found to be significant, although the age variable in the ln(earnings) model had a negative coefficient.

The Cobb-Douglas production function and multiple regressions presented in the model section stem from the ideas presented in the literature review and explore the relationship between a professional golfer's skills and their rank. The results found were generally consistent with what was expected: scoring average and average putts per round are the most significant skills that lead to better log earnings. The ln(earnings) multiple regression model generated a predicted rank that was found to significantly predict actual FedEx Cup rank at better than 60 percent.[10]

Further Applications

If more time was available to work on this project I would include more variables in the regressions such as additional skill statistics. I would also research whether the state or country a player is from or which they currently reside, impacts their earnings or rank. A test using an ordered-probit regression while using rank as a dependent variable would also be performed. This currently goes beyond the scope of this project as I lack the knowledge of how to complete this test. A Ramsey Reset test was performed during this project and it was discovered that each regression had an omitted variable. I felt that the variables included in this project sufficiently covered the major parts of the game of golf but it would be interesting to add more variables and see if the regressions could finally pass the Ramsey Reset test.

[10] This 60% corresponds to the Spearman correlation test conducted for the ln(earnings) model.

References

Albright, S. Christian. "A Statistical Analysis of Hitting Streaks in Baseball." *Journal of the American Statistical Association* 88, no. 424 (1993): 1175. Accessed October 02, 2016. doi:10.2307/2291254.

Alcantara, Reinaldo, and Anthony A. Prato. "Returns to Scale and Input Elasticities for Sugarcane: The Case of Sao Paulo, Brazil." *American Journal of Agricultural Economics* 55, no. 4 (1973): 577. Accessed October 02, 2016. doi:10.2307/1238342.

Bower, Bruce. "In the Zone: Evolution May Have Trained the Mind to See Scoring Streaks-even Where They Don't Exist." *Science News* 179, no. 4 (2011): 26-28. Accessed October 02, 2016. doi:10.1002/scin.5591790423.

Broadie, Mark. "Assessing Golfer Performance on the PGA TOUR." *Interface*s 42, no. 2 (2012): 146-65. Accessed October 02, 2016. doi:10.1287/inte.1120.0626.

Camerer, Colin F., "Does the Basketball Market Believe in the 'Hot Hand,'? *The American Economic Review* 79, no. 5 (1989): 1257-1261. Accessed October 02, 2016.

Cobb, Charles W., and Paul H. Douglas. "A Theory of Production." *The American Economic Review* 18, no. 1 (1928): 139-65. http://www.jstor.org/stable/1811556.

Connolly, Robert A., and Richard J. Rendleman. "Skill, Luck, and Streaky Play on the PGA Tour." *Journal of the American Statistical Association* 103, no. 481 (2008): 74-88. Accessed October 02, 2016. doi:10.1198/016214507000000310.

Dorsey-Palmateer, Reid, and Gary Smith. "Bowlers' Hot Hands." *The American Statistician* 58, no. 1 (2004): 38-45. Accessed October 02, 2016. doi:10.1198/0003130042809.

Ehrenberg, Ronald G., and Michael L. Bognanno. "Do Tournaments Have Incentive Effects?" *Journal of Political Economy* 98, no. 6 (1990): 1307-324. Accessed October 02, 2016. doi:10.1086/261736.

"Frequently Asked Questions." PGA Tour. http://www.pgatour.com/company/pga-tour-faqs.html.

MG, Kendall. *Rank correlation methods* (4th ed.) (1970): London: Griffin. ISBN 978-0-852-6419-96. OCLC 136868

Moy, R. L., and T. Liaw. "Determinants of Professional Golf Tournament Earnings." *The American Economist* 42, no. 1 (1998): 65-70. Accessed October 02, 2016. doi:10.1177/056943459804200106.

Nero, Peter, "Relative Salary Efficiency of PGA Tour Golfers." *The American Economist* 45, no. 2 (2001): 51-56. Accessed October 02, 2016. doi:10.1177/056943450104500206.

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

Shah, Dilip. "Production Functions in Co-operative Dairy Industry of Gujarat." *Indian Journal of Industrial Relations* 17, no. 3 (1982): 395-408. Accessed October 02, 2016.

Shmanske, Stephen. "Skills, Performance, and Earnings in the Tournament Compensation Model: Evidence from PGA TOUR Microdata." *Super Golfonomics*, 2014, 99-120. Accessed October 02, 2016. doi:10.1142/9789814612555_0006.

Watkins, John R., Jr. "Drive for Show, Putt for Dough: Rates of Return to Golf Skills, Events Played, and Age on the PGA Tour." *Michigan Journal of Business* 1 (2008): 35-60. Accessed October 02, 2016.