C.J. Dickman

Professor Griffith

Applied Econometrics

April 27, 2016

# **"Put the Driver Away!": An Analysis of Determinants in Annual Earnings of PGA Tour Players.**

**Abstract**

This paper examines the determinants of the earnings of PGA Tour golfers from the period 2012-2015. My goal in this paper is to determine the relative importance of each statistical category relevant to a golfer's average round. Golf analysts and economists have argued for well over two decades that putting is what wins championships. Recently however, that theory has come under fire with a leaning more on distance over accuracy and putting. With the addition of two key variables to my model (Strokes Gained Tee to Green and Strokes Gained Putting), courtesy of the PGA Tour, I am able to support the original claim that accuracy and putting are the largest determinants of PGA tour earnings.

**Introduction**

I have been an avid golfer since I was around 6 years old. I have always been fascinated by the game of golf, but moreover with the atmosphere that surrounds golf; being with nature, on a beautiful summer day, with the people that I love. Some of my fondest memories have been shared with my father on a golf course, or with my friends at a golf camp, like winning the long drive competition when I was 10, it's the little things. Clearly golf is an important part of my life when I'm physically on the golf course, but is also important when my family is huddled around the TV, while Jordan Spieth lines up his final putt to solidify his place in Augusta's record books. Watching the sport, practicing myself, and watching PGA Tour players practice, I only see them and myself spending time on the driving range, working on a variety of club selections and shots. Rarely do I see them in the practice bunkers, or on the putting green. I began to ask myself, what is truly important to the PGA Player's earnings, a measure of their performance, when they are practicing? In asking this question, I am analyzing what exactly PGA Tour players should focus on statistically during their practice, in order to improve their overall play, leading to higher earnings.

The implications for the findings in this paper are not only for PGA players, but amateur and beginning golfers as well. In analyzing what areas earn PGA players the most money, I will be able to see where players should allocate their practice time in order to improve those statistics. Furthermore, the results will show how two statistics have impacted previous models in the literature (Nero 2001) and made models more accurate in determining PGA tour earnings. Strokes Gained Tee to Green and Strokes Gained Putting were introduced in 2011 and as this paper will explore, have had profound impact in statistical analysis of PGA tour players.

**Literature Review**

Little has been done in terms of research directly connected to determining earnings on the PGA tour. My research will be one of the first to incorporate the variables Strokes Gained Tee to Green and Strokes Gained Putting, placing valuable emphasis on the addition of these two variables making other performance measures obsolete.

Donald Alexander and William Kern, in their paper "Driver for Show and Putt for Dough?" examines the determinants on PGA Tour golfers earnings from 1992-2001. They however take a time series approach and look at how stats have affected earnings over their time period. Furthermore, they seek to challenge the idea of "Drive for show and putt for dough" claiming that driving distance is in fact a strong determinant in annual earnings. Their essay discusses similar variables and topics as my essay will discuss, however I am seeking to emphasize the importance of driving accuracy over distance.

The final article I examined, "Relative Salary Efficiency of PGA Tour Golfers" was written by Peter Nero and focuses on four variables with regards to salary efficiency: putting average, driving distance, driving accuracy, and sand saves. My model accounts for all of those, with the exception of putting average being replaced by a more accurate statistic, Strokes Gained Putting. My paper seeks to expand on the findings of Mr. Nero. He mentions in his essay that the best players were ones who "were able to put themselves in situations that utilize their strengths and avoid their weaknesses." This paper will examine driving accuracy and recommend players consistently put themselves in better situations by hitting more fairways, sacrificing a small amount of distance.

**Data and Methods**

I utilized data given at PGAtour.com. I first identified the top 70 earnings players of 2015 and traced their stats for the past 4 years, from 2012-2015. Because the website is still in the process of updating its statistical bank, some data on players was only able to be collected for the past 3, or even two years. Because of the lack of some of the data, I am only limited to 251 observations, which will have some effect on my model.

Using the data I collected, after testing several models I came up with two OLS regression models, both using the log of earnings as the outcome variable in order to produce a more linear affect in my explanatory variables. The OLS equation and variables used in those equations will be listed below:

*Earnings*

The earnings in US dollars ($), that the player accumulated over the course of that year, starting on January 10$^{th}$ and ending with October 10$^{th}$. The earnings in this model do not account for events played and earnings won on other tours, like the European or Canadian tours. The earnings are specific to PGA tour play only. The variable will be logged to produce a more linear outcome in the coefficients to take away from huge random increases in earnings (i.e. Jordan Spieth in 2015).

*Years on Tour and Years on Tour squared*

These two variables will be interpreted together. In my results I will use the turning point to identify when age starts to negatively affect earnings. The variable accounts for the number of years a PGA player has spent on the PGA tour (excluding Web.com tours and other mini tours). I decided to use this over an age variable because often players go off and play for a variety of

other leagues like the Web.com tour or European tours, and I specifically wanted data for years on the PGA tour, age would have accounted for those other tours.

*Events*

The number of Events played within a year is easily calculated; I just went through and look to see how many tournaments each player played on the PGA Tour for that season. This stat is only for the PGA Tour events it does not account for some of the players going overseas to play on other Tours such as the European Tour. Furthermore, this stat includes events where players may have missed a cut, or withdrew from the tournament.

*Average Strokes Gained Tee to Green (SGT2G)*

The per-round average of the number of strokes the player was better or worse than the field average on the same course at the same event minus the players strokes gained putting. This statistic was introduced in 2011 and has widely been accepted as the standard measure of one's ability to drive the ball and replace overall ball striking.

*Average Strokes Gained Putting (SGPUTT)*

The measure of how well someone putts compared to the field average, taking into account the length of the first putt. You then compare the number of actual putts taken to the average number of putts the field would take from those given distances. I chose to use this because it is widely accepted as the measure of one's ability to putt. Golf is a single man sport, so comparative statistics will produce a more accurate representation of your ability. Furthermore, the stat accounts for putts from all distances, which is a broader category than specific putting distances.

*Average Driving Distance*

The average number of yards per measured drive (the first shot you take on a par 4 or par 5). These drives are measured on two holes per round. The PGA tour will often set up a device called a Trackman, a little white box that measures ball flight and spin and accurately predicts the distance of that drive. Care is taken to select two holes which face opposite directions to counteract the effects of wind. Drives are measured at the point where the ball comes to rest, even if it is not in the fairway.

*Average Driving Accuracy Percentage*

The percentage of time a tee shot lands in the fairway (regardless of club choice). On most courses a player is likely to have 14 chances to hit the fairway from the tee box. This does not account for par 3 holes, only par 4 and 5 holes. A player is most likely going to use a driver when hitting off the tee box, however as this paper will attempt to explain, may not always be the best option. As a note to the readers, and something we will see in the summary of the variables, drivers are the clubs with the highest spread in terms of location of impact, making them the most inaccurate club in a standard bag of 14 for the average golfer.

*Green In Regulation Percentage (GIR)*

The percentage of time a player is able to hit and land a ball on a green in regulation. A ball is considered “in regulation”, if any part of the ball is touching the putting surface after the GIR stroke has been taken. The GIR stroke is determined by subtracting 2 from par (1$^{st}$ stroke on a par 3, 2$^{nd}$ stroke on a par 4, 3$^{rd}$ stroke on a par 5).

*Scrambling Percentage*

The percent of time a player misses the green but still makes a par or better. This is often a measure of each player’s ability around the green with his wedge.

*Sand Save Percentage*

Calculated as the number of times a player is able to get up-and-down (out of the bunker followed by a single putt in the hole) from a greenside bunker, regardless of score. Up and down in golf refers to when a player missed the green, was able to successfully get the ball on the green with his next shot, then putt the ball in the hole.

**Table 1. Descriptive statistics for PGA players 2012-2015**

| | | Min | Max |
|---|---|---|---|
| YEARSONTOUR | 7.854 | 1 | 22 |
| | [4.96] | | |
| EVENTS | 22.787 | 4 | 36 |
| | [5.3] | | |
| EARNINGS | 2443748 | 0 | 1.20E+07 |
| | [1683262] | | |
| SGT2G | 0.435 | -2.384 | 2.318 |
| | [0.652] | | |
| SGPUTTING | 0.122 | -1.154 | 1.421 |
| | [0.314] | | |
| DRIVINGDIS~E | 292.984 | 272.5 | 317.7 |
| | [8.47] | | |
| DRIVINGACC~Y | 0.614 | 0.375 | 0.7368 |
| | [0.049] | | |
| GIR | 0.660 | 0.5389 | 0.7352 |
| | [0.027] | | |
| SCRAMBLING | 0.588 | 0.438 | 0.7727 |
| | [0.035] | | |
| SANDSAVE | 0.506 | 0.3548 | 0.6515 |
| | [0.06] | | |
| Observations | 251 | | |

standard deviations in brackets

It is important to remind the reader here that this data set is a set of the top 70 golfers on the PGA tour, in terms of earnings, so the level of skill is considered above average. It is also imperative to remind the reader that this data set was collected from the years 2012-2015. That knowledge is important because technological advances have leveled off, and the comparative advantage of having improved technology has worn off essentially. Players are competing at high levels using various sponsors (Titlist, Ping, TaylorMade, Wilson, etc.). This will lend itself to my thesis of promoting driving accuracy over distance off the tee.

The summary statistics in **Table 1** however, are still useful in explaining my results and the findings of this paper. The first two stats to look at are the Strokes Gained Tee to Green and Strokes Gained Putting. Strokes Gained Tee to Green has an average of .435 strokes gained for the top 70 players, which makes sense, because they are the elite players and should have an advantage based on pure skill over the other members of the tour. Strokes Gained Putting's average is a little weaker, presumably because you take less putts overall on a green, so there are less strokes to be gained; that average is .122 strokes gained. I am more interested in the Standard Deviations however. The standard deviations are both relatively high for these two stats, .652 strokes for sgt2g and .314 strokes for sgputt. This signifies a wide variance in the play of the players at even the elite level, which will help make my model more accurate.

I wish to discuss two more summary variables in this section; Driving Accuracy and Driving Distance. If we examine the maximum and minimum values for Driving Distance we can see that the range is roughly 47 yards which is equivalent to about two club lengths, which is not much of a range change. If we examine the Driving Accuracy range however, we can see a much wider gap of accuracy percentage off the tee. The range on that statistic is 36 percent, a huge range considering the data set is for the top 70 golfers on the PGA Tour, meaning even the best players in the world struggle with accuracy off the tee. If I were to expand this data set to include the top 100 or even 200 players on the PGA Tour, we would see that driving distance plateaus and stays consistent in that range of 47 yards, whereas accuracy's range is actually expanded.

As stated above, I will be using two OLS models, one with the introduction of Strokes Gained Tee to Green and Strokes Gained Putting and one without them, plus the variables I omitted when adding those in. I did this to show how those two variables have made models

more accurate. I regressed Log earnings to get a more linear affect in my variables and to show the percent changes, rather than large complex numbers. The equations can be found below:

1.) $Log(Earnings)=B_o + B_1(normDRIVACC) + B_2(normSANDSAVE) + B_3(EVENTS) + B_4(Yearsontour) - B_5(Yearsontoursq) + B_6(normGIR) + B_7(DRIVDIST) + B_8(normSGPUTT) + E$

2.) $Log(Earnings)=B_o + B_1(NormSGT2G) + B_2(NormSGPUTT) + B_3(DRIVACC) + B_4(SANDSAVE) + B_5(EVENTS) + B_6(Yearsontour) - B_7(Yearsontoursq) + B_8(DRIVDIST) + E$

3.) $Log(Earnings)=B_o + B_1(normDRIVACC) + B_2(normSANDSAVE) + B_3(EVENTS) + B_4(Yearsontour) - B_5(Yearsontoursq) + B_6(normGIR) + B_7(normSCRAMB) + B_8(DRIVDIST) + B_9(normSGPUTT) + E$

All of these variables are important statistics in assessing the performance of a PGA Tour player. I chose these variables because they are distinct categories that a player can isolate during practice in order to improve the skill and thus performance. When running the first model, the only two variables that were not significant at the 90% level were yearsontour & yearsontoursq and Driving Accuracy in model 1. After running another regression without the sgt2g variable, Driving Accuracy became significant at every level. This means that sgt2g and Driving Accuracy explain similar things, which holds with theory. I decided to keep it in my model however, because I wanted to show how important the magnitude of the coefficient is. I decided to keep yearsontour and yearsontoursq in my model because theoretically they provide an age variable for performance on the PGA Tour. As you age, theoretically you are more prone to injuries and thus your performance will be hindered and your earnings will decrease. My model does show a decrease when we assess the turning point value in the results section of this paper. Sand Save Percentage had a slightly higher P-value than all the other statistics, mainly because the two strokes gained statistics are taking some of the explanatory power away from that variable,

however I wanted a measure of a players ability from the sand, to isolate how much time a player should focus on practice in the bunkers.

In terms of theory, the signs on the coefficients seem to make sense. Both Strokes Gained Tee to Green and Strokes Gained Putting should be positive because as you increase those statistics, you are performing better than the rest of the field and thus are able to capture a higher amount of the "purse" for that tournament. Driving Accuracy, which I have already discussed in immense detail, should be positive because as you increase the amount of fairways you hit, the more likely you are to score better on a hole, thus decreasing your overall score, leading to a lower overall round. Driving Distance, similar to Driving Accuracy, should be positive because the farther you hit the ball, the shorter your approach shot becomes and the more control you have over the spin and landing spot of the ball, making it easier to par or birdie holes, generating a lower overall score. Events should be positive because even if you don't win an event, you still get a percentage of the overall purse amount, meaning the more tournaments you play, the higher your earnings will be. According to this data, it would appear that PGA players are using the aforementioned point and playing in more tournaments. Sand Save Percentage, GIR, and Scrambling should all be positive because they are measures of specific performance variables and as you increase these, your overall performance increases, leading to lower scores and higher earnings. These three variables are especially important, as around 70% of a golfers shots occur on or around the green.

**Table 2. OLS Results PGA Tour Data 2012-2015**

| VARIABLES | (1)<br>lnearn | (2)<br>lnearn | (3)<br>lnearn |
|---|---|---|---|
| normsgputt | 0.0922*** | 0.0885*** | 0.0844*** |
| | [0.0134] | [0.0106] | [0.0138] |
| EVENTS | 0.0533*** | 0.0567*** | 0.0513*** |
| | [0.00755] | [0.00591] | [0.00756] |
| YEARSONTOUR | 0.0459* | 0.0328* | 0.0485* |
| | [0.0251] | [0.0198] | [0.0250] |
| yearsontoursq | -0.000699 | -0.000887 | -0.000846 |
| | [0.00125] | [0.000984] | [0.00124] |
| normdrivacc | 0.0645*** | 0.00958 | 0.0594*** |
| | [0.0129] | [0.0102] | [0.0131] |
| normsandsave | 0.0341*** | 0.00922 | 0.0257*** |
| | [0.00673] | [0.00559] | [0.00776] |
| DRIVINGDISTANCE | 0.0514*** | 0.0138** | 0.0498*** |
| | [0.00675] | [0.00594] | [0.00675] |
| normgir | 0.0682*** | | 0.0652*** |
| | [0.0187] | | [0.0186] |
| SGT2G | | 0.833*** | |
| | | [0.0640] | |
| normscramb | | | 0.0301** |
| | | | [0.0143] |
| Constant | -12.45*** | 7.421*** | -12.75*** |
| | [2.196] | [2.356] | [2.185] |
| Observations | 251 | 251 | 251 |
| R-squared | 0.526 | 0.706 | 0.535 |

Standard errors in brackets

The first piece of my results I would like to discuss is the difference between the three models, referencing the importance of Strokes Gained Tee to Green and Strokes Gained Putting with regards to their impacts on the R-squared value. The R-squared value for my second model is nearly 20% points greater than my first and third models, meaning the addition of the SGT2G contains more explanatory power. The introduction of Strokes Gained Tee to Green and Strokes Gained Putting by the PGA tour has made my model more accurate than previous models in the literature (Nero 2001), and has allowed for a more encompassing measure of skill on the PGA related to earnings. As mentioned above the addition of the two strokes gained variables is

starting to make other variables less significant in determining earnings, because the other variables are an overall measure of ball-striking, driving, iron play, wedge play, and putting; the statistical categories used in my model to determine earnings and performance.

When analyzing the coefficients I will be looking at regression 1, according to **Table 2.** I am using this model because even though it has the lowest R-squared value, all of the variables except for Yearsontoursq are significant. Furthermore, I am able to explain why the R-squared value is lower than the other two; the absence of SGT2G, which I took out of my main model because I wanted to show the changes amongst the variables that fall under the explanatory variable SGT2G.

The first two stats I wish to discuss, and arguably the most important to my theory, are Driving Accuracy and Driving Distance. With a 1 unit increase in driving accuracy (percentage point) a player can expect a 6.45% increase in his earnings, all else equal. With a 1 yard increase in driving distance a player can expect a 5% increase in earnings, all else equal. This means that in terms of earnings and overall performance, driving accuracy is more important that driving distance. This is relevant because it disproves the idea that longer hitters are more likely to perform better than shorter hitters. Furthermore, it lends itself to my hypothesis: PGA Tour players and any golfer for that matter, should rely more heavily on a lower club choice for their first shot, like a 3 wood, that has similar distance to a driver and is often hit consistently straight. This can be further backed up when looking at the correlation coefficients below between driving accuracy and distance.

**Table 3. Correlation Coefficients for Model 1.**

| | lnearn | norms~2g | normsg~t | DRIVIN~E | DRIVIN~Y | SANDSAVE | EVENTS | YEARSO~R | yearso~q |
|---|---|---|---|---|---|---|---|---|---|
| lnearn | 1 | | | | | | | | |
| normsgt2g | 0.6802 | 1 | | | | | | | |
| normsgputt | 0.2984 | -0.0862 | 1 | | | | | | |
| DRIVINGDIS~E | 0.1556 | 0.3039 | -0.2016 | 1 | | | | | |
| DRIVINGACC~Y | 0.2 | 0.2853 | -0.0309 | -0.5808 | 1 | | | | |
| SANDSAVE | 0.2651 | 0.1601 | 0.3501 | -0.1973 | 0.0439 | 1 | | | |
| EVENTS | 0.2713 | -0.1181 | 0.1303 | -0.2172 | 0.1297 | -0.0114 | 1 | | |
| YEARSONTOUR | 0.1591 | 0.2155 | -0.0402 | -0.0798 | 0.0743 | 0.1217 | -0.186 | 1 | |
| yearsontou~q | 0.1527 | 0.2043 | -0.013 | -0.1386 | 0.1194 | 0.1337 | -0.1536 | 0.9537 | 1 |

The marginal effect of taking a 3 wood over a driver is about 40 yards of pure distance, however when you take into account the favorable roll assuming a player hits the fairway and lower ball flight of the 3 wood, the distance loss is mitigated.

Another staggering result from my regression analysis is the coefficient on sand save percentage. This was the smallest coefficient I had, meaning that sand play is rather irrelevant when looking at earnings. On average, a player can expect a 3.4% increase in their earnings, which is small relative to other variables. I originally thought sand save percentage was going to have a larger impact on earnings than is commonly talked about, however as the model shows, it is not that significant in determining earnings.

Yearsontour and Yearsontoursq have to be interpreted together in order to find the turning point; the point where yearsontour starts to negatively affect earnings. When attempting to find the turning point however, because the statistic was the most insignificant from zero, I was getting answers that theoretically made little sense. Because of this, I can conclude that yearsontoursq does not belong in the model and realistically does not negatively affect earnings. As a player increases his years on tour by one year, he can expect a 4.5% increase in his earnings, all else equal. This makes sense because this variable acts as an experience variable,

the more you age, the more experience you gain and the better you are. Furthermore, the PGA only allows a certain number of people to carry PGA player cards, so if you are amongst the older members of the tour, you have to be able to compete with the younger generations, meaning you must be highly skilled to retain your card. This is one explanation for why the yearsontoursq variable is irrelevant and insignificant. Another essay may look at a pure age variable rather than an experience variable, accounting for deteriorating physical condition as one ages.

The final variable I wish to talk about in my model is the only variable that accounts for putting, Strokes Gained Putting. I left this Strokes Gained variable in my model because there were no other variables that were correlated with it, unlike the Strokes Gained Tee to Green statistic. According to my model, with a .1 stroke increase in strokes gained putting a player can expect a 9.2% increase in their annual earnings, all else equal. This was my highest coefficient and was the most significant of all the variables. This means that putting is the highest determinant in annual earnings, consistent with pervious literature (Alexander and Kern 2005).

**Conclusion**

There are several conclusions that can be drawn from this paper and these findings. The first is the overall importance of the statistics Strokes Gained Putting and Strokes Gained Tee to Green. With the addition of these two variables into a model, most other explanatory variables became insignificant. This means that these two variables are accounting for similar effects on logged earnings as the other variables. The two variables are beginning to make the other variables unnecessary when running models and measuring overall skill. This is evident when we examine the R-squared values, as mentioned above.

The next conclusion we can draw from the results is that driving accuracy is more important than driving distance, which directly contradicts previous literature (Alexander and Kern 2005). Using the correlation coefficient we can directly see that driving distance is negatively correlated with driving accuracy, making a driver the most inaccurate club in a golfer's bag. This can translate into in round behavior; a player should lean more heavily on his 3 wood off the tee box, especially on par 4 holes. This leaves the golfer more accuracy off the tee and sacrifices little distance, as most players are able to hit their three wood only 30-40 less than their drivers. The decrease in club gives the player more opportunity to hit the fairway and have a better lie into the green, which means more opportunities to hit the green in regulation. As a player averages one more percentage point increase in GIR he can expect a 6.8% increase in his annual earnings. All around, a 3 wood shot in the middle of a fairway is worth more in pure earnings than a driver in the hazard or on a tough lie in the thick grass.

The third conclusion we can draw from these results is the fact that players should play on the PGA Tour as long as they can, most likely until they physically cannot compete. The payoffs are higher on the PGA Tour versus the Champions tour, because the overall purse amounts per tournament are higher. Even if an older golfer is toward the bottom of the rankings, he is still making more money than a player on the Champions tour.

In terms of practicing routines to maximize efficient earnings, a player should focus on two things; ball striking and putting. If a player can hit the ball pure and straight with irons and woods, he will maximize his earnings. Furthermore, more time should be spent away from the driving range and on the putting surface. I am recommending that a player spend twice as much time practicing putting as he does practicing drives on the range. This recommendation comes from the results, citing putting as the most important in annual earnings and performance.

References

Alexander, and Kern.” Drive for Show and Putt for Dough”; Sage Journal. 46-60., 2014. Web.

Nero, Peter. “Relative Salary Efficiency of PGA Tour Golfers”; The American Economist. 51-56., 2001. Web.