PERFORMANCE ON THE PGA TOUR: A STATISTICAL ANALYSIS

C. Barry Pfitzner and James Robinson, Department of Economics/Business,
Randolph-Macon College, Ashland, VA, bpfitzne@rmc.edu

ABSTRACT

This paper investigates the determinants of player performance as measured by scoring average on the Professional Golf Association of America (PGA) tour for the 2012 season. Among other findings, this research shows that the percentage of greens reached in regulation (*GIR*) and putts per round (*PPR*) are by far the most important determinants of scoring average on the PGA tour. We choose in this paper to focus on direct golf skills and among other results, we find that driving distance and driving accuracy are not only important, but are approximately equally important in determining scoring average. These results contrast with recent findings from a Harvard group [9].

INTRODUCTION

Professional golf tours keep a variety of performance statistics presumed to measure important skills related to success. One dominant statistic is greens in regulation—the percentage of golf holes for which the player reaches the surface of the green in at least two fewer strokes than the par score for that hole. Other major statistics include driving distance (*DD*), driving accuracy (*DA*) which measures the percentage of drives in the fairway of the hole being played, sand saves (*SS*) which measures the percentage with which a player takes two or fewer strokes to hole the ball from greenside bunkers, putts per round (*PPR*), and putts per green reached in regulation (*PPG*). Each of those measures, *GIR*, *DD*, *DA*, *SS*, *PPR* (or *PPG*), are related in theory to scoring and scoring is clearly related to monetary success.

The purpose of this paper is to provide empirical estimates to aid in determining if and how those statistics are related to scoring average and, therefore, success on the PGA golf tour. This paper will employ regression techniques to capture the influence of the measures of the skills enumerated in the previous paragraph on performance on the PGA tour.

LITERATURE REVIEW

There are several strains of research on professional golf performance based on the statistics compiled by the PGA, the LPGA (Ladies Professional Golf Association) and European PGA tours. One of the first studies of the statistical determinants of success in professional golf was by Davidson and Templin [4]. Utilizing data from the 1983 PGA (119 of the top 125 money winners) in a multiple regression framework, Davidson and Templin found that greens in regulation (*GIR*), putting (*PPR*), and a combined driving efficiency measure were capable of explaining 86% of the variation in scoring average for the PGA tour, with *GIR* the most important single variable. When the dependent variable was earnings, putting was slightly more important statistically than the other explanatory variables, based on standardized beta coefficients. Shmanske [14], also using a multiple regression framework for data from the 1986 PGA tour (the top 60 money winners), finds that putting and driving distance are the two most important skills in determining success on the PGA tour. When player money winnings per event are the dependent variable, he finds no significant role for *GIR* as an explanatory variable. Shmanske also attempts to estimate the greatest payoff for practice, and finds the greatest payoff is for putting practice. Belkin et al. [2] utilize PGA statistics for three years (1986-88) in correlation and step-wise regression frameworks. Their research confirms the importance of *GIR* and putts per round (*PPR*) as dominant variables in determining scoring average, with lesser, but statistically important roles for driving distance, driving accuracy and sand saves. They conclude that their research confirms the importance of tour statistics in predicting scoring average.

A 1995 paper by Englehardt [6] concludes that the rankings of the top 10 money winners are *not* significantly correlated with *GIR* for 1993 and 1994 PGA seasons, and cites an increasingly important role for "total driving," which is the sum of the ranks in driving distance and driving accuracy. This study utilizes, however, a sample size of only 10. Moy and Liaw [11] find evidence that conflicts with that from Englehardt for the same PGA year. They find statistically important roles for driving distance, driving accuracy, *GIR*, and putting in determining earnings on the PGA tour for the 1993 season. The latter study utilizes a multiple regression framework and a much larger sample size than Englehardt. Moy and Liaw's work also includes analysis of the LPGA and the Senior PGA tours and they offer the general conclusion that a well rounded game is necessary for success in professional golf. Nero [12] using data from the 1996 PGA tour finds statistically important roles for driving distance, driving accuracy, putting, and sand saves in determining money won. Interestingly, Nero does not include *GIR* in his analysis. Nero also estimates a frontier earnings function in an attempt to identify the most efficient golfers—that is those golfers who earn more than that predicted by the regression equation.

Dorsal and Rotunda [5] using data from the top 42 players on 1990 PGA tour found that *GIR* was the most important variable determining scoring average, and that driving accuracy was more important than driving distance. Their analysis included simple correlation analysis and multiple regression techniques. They also used scoring average, top 10 finishes, and money winnings as dependent variables.

More recently, Alexander and Kern [1] offer some evidence that driving distance has become more important over time as a determinant of success on the PGA tour. Callan and Thomas [3] us a multi-equation approach wherein scoring average is modeled as a function of the normal skill set and earnings are then modeled as a function of scoring average. Finally, Ezekowwitz [9] finds no role for driving distance or driving accuracy as determinants of scoring average. Because of the *ad hoc* nature of the method employed (step-down regression) and implicit assumptions made in [9], we don't think the conclusions drawn from that study should be taken seriously.

METHODOLOGY

The primary research method for this paper is multiple regression analysis with scoring average as the dependent variable a general set of performance statistics as the explanatory variables.

The general model may be represented as:

$$SA_i = \beta_o + \beta_1 GIR_i + \beta_2 DD_i + \beta_3 DA_i + \beta_4 PPR_i + \beta_5 SS_i + \varepsilon_i, \quad (1)$$

Where,

SA = Scoring average (strokes per round)
GIR = greens in regulation (percentage of greens reached in regulation or fewer strokes)
DD = driving distance in yards
DA = driving accuracy (the percentage of drives in the fairway)
PPR = putts per round
SS = percentage of sand saves
i = references the i^{th} observation—here the individual player

SUMMARY STATISTICS ON THE PGA TOUR

Table I represents the summary statistics for the 2012 PGA tour. For 2012, the PGA tour reported full statistics on 190 players. These statistics can be described succinctly: the mean scoring average on tour is approximately 71; players hit 65% of the greens in regulation; they drive the ball an average of 290 yards; hit 61% of the fairways; they average 29 putts per round; and they save par (or better) from green side bunkers 49% of the time.

Table I: Summary Statistics for the 2012 PGA Tour

Variable	Mean	Standard Deviation	Minimum	Maximum
Scoring Average per (SA)	70.89	0.710	68.87	73.00
Greens in Regulation (GIR)	64.91%	2.68%	57.7%	70.34%
Driving Distance (DD)	290.05	8.41	268.9	315.5
Driving Accuracy (DA)	61.07%	4.71%	47.3%	73.0%
Putts per Round (PPR)	29.2	0.49	27.89	30.50
Sand Save Percentage (SS)	48.56%	6.22%	29.7%	65.4%

(n = 190)

SOME REGRESSION RESULTS

In this section we present and discuss the regression results for scoring average. Table II contains the results for several regressions with scoring average as the dependent variable. In regression 1, every variable coefficient in the general model specified in the previous section is correctly signed and statistically significant at less than $\alpha = .01$. The equation explains almost 80% ($\overline{R}^2 = .78$) of the variation in scoring average across players on the PGA tour for the 2012 season. Note that the estimated coefficient for *PPR* in regression 1 is equal to its theoretical value of one. That is, if a player averages one more putt per round, his score would be one stroke higher, other things equal.

Given the results, it is possible to estimate the practical importance of each of the explanatory variables on scoring average. Here we choose to estimate the effect on scoring average of a one-standard deviation improvement in each of the explanatory variables. This is accomplished by computing the product of each estimated coefficient and the standard deviation of that variable. For example, if a player were to improve on *GIR* by one standard deviation (2.68%), scoring average would be estimated to fall by approximately ½ stroke per round (-0.1828*2.68 = -0.49), or approximately 2 strokes over a four round tournament. Using the same method, *PPR* has a nearly identical effect. If a player could improve on *PPR* by one standard deviation (one-half *fewer* putts per round), *SA* would be predicted to fall by 0.49 strokes. The effect improving sand save percentage by one standard deviation would amount to an estimated improvement of about .11 in scoring average per round.

Much has been written and spoken regarding driving distance and driving accuracy and their relative importance. Based on regression 1, the effect of a player increasing driving distance by one standard deviation (8.41 yards) is estimated to lower scoring average by .294 strokes per round, and if a player could improve driving accuracy by one standard deviation (hit 4.71% more fairways), scoring average would fall by .217 strokes per round. These effects we would judge as similar, with a slight edge going to driving distance. However, recognize that these two measures of performance are, in general, inversely related. Across players, the correlation between *DD* and *DA* is -0.52 and that negative correlation is highly significant statistically. What this means in practice is that even if a player were to find a way (exercise, technique, equipment, etc.) to improve driving distance, that player would likely hit fewer fairways. Using simple trigonometry, if a fairway were 30 yards wide, a 260 yard drive with a ± 3.3º error in direction finds the edge of the fairway, and at 300 yards, the same error lands the ball in the rough.

Though estimations similar to regression 1 are common in the literature and often are the basis for the argument that *GIR* represent the most important statistical determinant of scoring average, we are interested in golf *skills* as they relate to scoring average. Greens-in-regulation, we argue, is not directly a golf skill, but is, rather, the effect of other golf skills. A player who drives the ball both long and accurately will on average hit more greens in regulation, assuming some degree of proficiency with iron shots to the green. (We also recognize that professional golfers are at times willing to miss greens versus hitting greens in places that are likely to results in higher scores than a missed green in a better position.)

We offer regressions 2 and 3, in which the explanatory variables are, in our opinion, direct measures of skills on which professional golfers continually seek to improve. In regression 3 with *GIR* omitted, the coefficients are again correctly signed and meet strict tests of statistical significance. The regression equation explains less of the variation in scoring average across players, that is, $\bar{R}^2$ is now .60. The calculations for the effects of the distance versus accuracy in driving question again yield similar effects. A one standard deviation improvement would lower scores by 0.53 strokes per round for *DD* and 0.50 for *DA*. Note that these estimated effects are now larger with *GIR* no longer in the equation.

Table II: Regression Results: Scoring Average = Dependent Variable

Variable/ Summary Statistics	Regression 1 Dependent Variable = SA	Regression 2 Dependent Variable = SA	Regression 3 Dependent Variable = SA	Regression 4 Dependent Variable = SA
Constant	67.46	72.40	78.01	52.13
GIR	-0.1829* (-12.17)			-0.2589* (-19.63)
PPR	0.9971* (13.28)	0.4499* (5.58)	0.7567* (10.43)	1.217* (16.93)
DD	-0.0349* (-8.34)	-0.0630* (-13.46)	-0.0448* (-10.59)	
DA	-0.0460* (-5.83)	-0.1058* (-12.78	-0.0975* (-14.39)	
SS	-0.0183* (-3.81)	-0.0332* (-5.36)	-0.0242* (-4.72)	
IRONS			-0.1203* (-9.82)	
$\bar{R}^2$	.78	.60	.74	.69
SEE	.3335	.4468	.3629	.3931
$F_{k, n-k-1}$	134.75*	73.21*	108.04*	215.07*

(notes: n = 190; k = number of regressors; t-statistics in parentheses;
* indicates significance at $\alpha < .01$)

The decline in explanatory power of regression 2 versus regression 1 is likely due, at least in part, to *GIR* acting as a proxy in regression 1 for another golf skill—namely iron play. The PGA tour currently keeps numerous statistics that can be utilized to measure skill with irons. We choose to use greens-in-regulation from the fairway, which we denote as *IRONS* in regression 3 from Table II. Since the regression includes driving distance, the equation controls for the fact that some players are closer to the greens than others. Note that the explanatory power of regression 3 ($\bar{R}^2 = .74$) is similar to that of regression 1, and all of the estimated coefficients of the variables are again correctly signed and meet strict tests of statistical significance. For our purposes, we strongly prefer regression 3 to regression 1 on the theoretical grounds that the explanatory variables in regression 3 are direct skills. Interestingly, on the driving distance versus accuracy question, regression 3 suggests that, other things equal, *DA* is slightly more important than *DD*, again measured as before. A one standard deviation improvement lowers *SA* by .377 for *DD* and .46 for *DA*. Also a one standard deviation improvement in the measure of iron play (2.75% more greens hit from the fairway) would lower *SA* by 0.33.

Since regressions 2 and 3 result in estimates for *PPR* that differ significantly from the theoretical value of 1, we also estimated equations with other measures of putting efficiency (strokes gained and others). The value of explanatory power of the equation was higher with *PPR* than with any of the other measures of putting skills.

Finally for reference only, regression 4 suggests that *GIR* and *PPR* explain as much as 70% of the variation in scoring average across players—a result consistent with much of the prior literature. Regression 4, however, is not directly indicative of the skills we wish to measure.

CONCLUSIONS

We present evidence of the determinants of success on the US Professional Golf Association (PGA) tour. Whether the measure of success in scoring average or money winnings, the percentage of greens reached in regulation (*GIR*) and a measure of putting success (here, putts per round) are dominant explanatory variables in regression formats. To asses the effects of driving accuracy and driving distance, it is desirable to remove *GIR* from the estimating equations. Those formulations suggest that driving accuracy and driving distance are approximately equally important in determining scoring average on the PGA tour.

REFERENCES

[1] Alexander, D.L., and W. Kern (2005) "Drive for Show and Putt for Dough?: An Analysis of the Earnings of PGA Tour Golfers." *Journal of Sports Economics*, volume 6, number 1, 46-60.

[2] Belkin, D.S., Gansneder, B., Pickens, M., Rotella, R.J., & Striegel, D. (1994) "Predictability and Stability of Professional Golf Association Tour Statistics." *Perceptual and Motor Skills*, 78, 1275-1280.

[3] Callan, SJ., and J.M. Thomas (2007) "Modeling the Determinants of a Professional Golfer's Tournament Earnings: A Multiequation Approach." *Journal of Sports Economics*, volume 8, number 4, 394-411.

[4] Davidson, J. D. & Templin, T. J. (1986) "Determinants of Success Among Professional Golfers." *Research Quarterly for Exercise and Sport*, 57, 60-67.

[5] Dorsel, T. N. & Rotunda, R. J. (2001) "Low Scores, Top 10 Finishes, and Big Money: An Analysis of Professional Golf Association Tour Statistics and How These Relate to Overall Performance." *Perceptual and Motor Skills*, 92, 575-585.

[6] Engelhardt, G. M. (1995) "'It's Not How You Drive, It's How You Arrive': The Myth." *Perceptual and Motor Skills*, 80, 1135-1138.

[7] Englehardt, G. M. (1997) "Differences in Shot-Making Skills among High and Low Money Winners on the PGA Tour." *Perceptual and Motor Skills*, 84, 1314.

[8] Englehardt, G. M. (2002) "Driving Distance and Driving Accuracy Equals Total Driving: Reply to Dorsel and Rotunda." *Perceptual and Motor Skills*, 95, 423-424.

[9] Ezekowitz, J. (2009) "Predictors of PGA Tour Scoring Average: Does the Driver Matter?," http://harvardsportsanalysis. wordpress.com/2009/11/19/predictors-of-pga-tour-scoring-average-does-the-driver-matter/

[10] Jones, R.E. (1990) "A Correlation Analysis of the Professional Golf Association (PGA) Statistical Ranking for 1988." In A.J. Cochran (Ed.), *Science and Golf: Proceedings of the First World Scientific Conference of Golf*. London: E & FN Spon. 165-167.

[11] Moy, R. L. and Liaw, T. (1998) "Determinants of Professional Golf Tournament Earnings." *The American Economist*, 42, 65-70.

[12] Nero, P. (2001) "Relative Salary Efficiency of PGA Tour Golfers." *The American Economist*, 45, 51-56.

[13] Pfitzner, C.B. and Rishel, T. (2005) "Performance and Compensation on the LPGA Tour: A Statistical Analysis," *International Journal of Performance Analysis in Sport,* University of Wales Institute, Cardiff, Volume 5, Number 3.

[14] Shmanske, S. (1992) "Human Capital Formation in Professional Sports: Evidence from the PGA Tour." *Atlantic Economic Journal*, 20, 66-80.

[15] Wiseman, F., Chatterjee, S. Wiseman, D. and Chatterjee, N. (1994) "An Analysis of 1992 Performance Statistics for Players of the U.S. PGA, Senior PGA, and LPGA Tours." In A.J. Cochran and M. R. Farrally (Eds.), *Science and Golf: II*.

[16] *Proceedings of the World Scientific Congress of Golf*. London: E & FN Spon. 199-204.

[17] Woolridge, Jeffrey M., *Introductory Econometrics: A Modern Approach,* South-Western College Publishing, 2000.