

PERFORMANCE AND COMPENSATION ON THE EUROPEAN PGA TOUR: A STATISTICAL ANALYSIS

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

ABSTRACT

This paper investigates the determinants of player performance as measured by scoring average and earnings on the European Professional Golf Association (EPGA) tour for the 2006 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 both scoring average and earnings on the EPGA. We further find that driving distance and driving accuracy are approximately equally important in determining scoring average.

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 money winnings as measures of success on the European 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 success on the EPGA tour.

LITERATURE REVIEW

There are several strains of research on professional golf performance based on the statistics compiled by the PGA and LPGA and EPGA tours. One of the first studies of the statistical determinants of success in professional golf was by Davidson and Templin [2]. 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 [11], 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. [1] 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 [4] 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 [8] 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 [9] 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 [3] 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.

METHODOLOGY

The primary research method for this paper is multiple regression analysis with scoring average and money won per event as the dependent variables (the money winnings equation is estimated in log form due to heteroskedastic residuals in the untransformed version), and the general set of performance statistics as the explanatory variables.

The general models may be represented as:

$$SA_i = \beta_0 + \beta_1 GIR_i + \beta_2 DD_i + \beta_3 DA_i + \beta_4 PPR_i + \beta_5 SS_i + \varepsilon_i, \quad (1)$$

and

$$\ln(M_i / E_i) = \beta_0 + \beta_1 GIR_i + \beta_2 \ln DD_i + \beta_3 DA_i + \beta_4 \ln PPR_i + \beta_5 SS_i + \varepsilon_i \quad (2)$$

Where,

SA = Scoring average (strokes per round)

M/E = money winnings per event

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,

and the *i* subscript refers to the *i*th observation (here the individual player), and *ln* is symbol for the natural log.

SUMMARY STATISTICS ON THE EPGA TOUR

Table I represents the summary statistics for the 2006 EPGA tour. For 2006, the EPGA tour reported full statistics on 178 players. Unlike the PGA tour in the United States, the EPGA tour in 2006 was not dominated by a single player. Paul Casey was the leading money winner (2.5 million euro), Sergio Garcia had the lowest stroke average (70.04), but the gaps between those leaders and the other top players were relatively small.

Table I: Summary Statistics for the 2006 EPGA Tour

Variable	Mean	Standard Deviation	Minimum	Maximum
Scoring Average (SA)	71.90	0.755	70.04	74.03
Greens in Regulation (GIR)	66.08%	3.63%	57.8%	75.6%
Driving Distance (DD)	284.28	9.97	260.2	314.1
Driving Accuracy (DA)	59.51%	5.89%	42.0%	75.6%
Putts per Round (PPR)	29.69	0.767	28.0	32.6
Sand Save Percentage (SS)	51.92%	8.63%	24.6%	75.9%
Putts per GIR (PPG)	1.79	0.032	1.72	1.90
Money Winnings (M)	€390,929.00	€443,790.23	€1,842.00	€2,454,084.00

(n = 178)

SOME REGRESSION RESULTS

In this section we present and discuss the regression results for scoring average and money winnings.

Scoring Average

Table II presents the results of the regressions with scoring average (*SA*) as the dependent variable. The column titled Regression 1 contains the empirical estimates for equation 1 from above. The coefficients for each variable have the anticipated signs, however only *GIR* and *PPR* are strongly significant in the statistical sense (*DD* is weakly significant). The equation explains over 85% of the variation in scoring average and the standard error is slightly over one-quarter of a stroke. It is clear from regression 1 in Table II that *GIR* and *PPR* are dominant variables in the equation, and *PPR* is fairly near its theoretical value of 1.

In order to assess the effects of variables such as driving distance and accuracy, we offer Regression 2. Since *DD* and *DA* are dominated by *GIR* as explanatory variables in equation 1, at least in part because the percentage of greens that a golfer hits in regulation is in fact determined by the distance and accuracy of his drives, we eliminate *GIR* from Regression 2. In this way we believe a clearer assessment of those skills, and some evidence of the value of each can be estimated. Notice that all of the variables remaining in Regression 2 are signed in accord to theory and statistically significant. However, without *GIR* in the regression, the explanatory power of the regression is significantly reduced—the adjusted R-squared value is only .33. It is often considered a controversy as to whether driving distance (*DD*) or driving accuracy (*DA*) is more important for professional golfers. Much of the recent commentary (and some research) tends to emphasize distance over accuracy. We offer Regression 2 as contradictory evidence. First, each of those two variables is statistically significant at approximately the same levels. Second, assuming regression 2 is properly specified, some indication of the relative importance of the two variables may be offered. Suppose other things equal, a player improved one standard deviation in each

of *DD* and *DA*. Driving distance would increase by 9.97 yards and driving accuracy would rise by 5.89%. The product of the coefficient estimate and the respective one standard deviation change would represent the change in scoring average. For *DD* this procedure would lower scoring average by .361 strokes per round ($-0.0362 \cdot 9.97$) and for *DA*, the corresponding effect would lower scoring by .352 strokes per round ($-0.0597 \cdot 5.89$). While it is clear that each of these variables is important in scoring average, it is not at all clear which is the more important influence on scoring average for the EPGA tour. Based on this analysis, their respective contributions appear to be approximately the same.

Table II: Regression Results: Scoring Average = Dependent Variable

Variable/ Summary Statistics	Regression 1	Regression 2	Regression 3	Regression 4
Constant	60.63	72.40	79.07	59.30
<i>GIR</i>	-0.2038* (-25.13)		-0.1087* (-8.14)	-0.2092* (-29.86)
<i>PPR</i>	0.8881* (26.03)	0.4711* (7.33)		0.8900* (28.82)
<i>DD</i>	-0.0047*** (-1.62)	-0.0362* (-6.40)		
<i>DA</i>	-0.0038 (-0.78)	-0.0597* (-6.28)		
<i>SS</i>	-0.0011 (-.415)	-0.0125** (-2.28)		
$\bar{R}^2$	.8561	.3315	.2695	.8563
<i>SEE</i>	.2862	.6170	.6450	.2861
$F_{k, n-k-1}$	211.62*	22.94*	66.30*	528.31*

(notes: $n = 178$; k = number of regressors; t-statistics in parentheses; *, **, *** indicate significance at $\alpha < .01$, .05, and .10, respectively)

Regressions 3 and 4 are examples of more parsimonious estimations. In Regression 3, greens in regulation (*GIR*) alone explain 26% of the variance in scoring average. It is interesting to note that the corresponding statistic is similar for the US PGA tour, but approximately 75% for the LPGA tour (see [10]). Finally, Regression 4 indicates that an equation including only *GIR* and *PPR* explains as much of the variance in scoring average as does Regression 1. Put differently, only 15% of the variance of scoring average is explained by factors other than *GIR* and *PPR*.

We also experimented with putts per green in regulation (*PPG*) rather than *PPR*. It turns out that when *GIR* is in the equation, *PPR* is a stronger explanatory variable, and when *GIR* is not in the equation, *PPG* is the better explanatory variable. The reason for this, we believe, is that players who miss a larger proportion of greens (lower *GIR*) have fewer total putts (because they are able to get the ball closer to the hole when they eventually get the ball on the green). Indeed the simple correlation between *GIR* and *PPR* is 0.535, clear evidence of the effect we describe.

Tournament Winnings

Since different tournaments have different purses, tournament winnings per event is a more difficult variable to explain. Table III depicts the results of the regressions aimed at predicting tournament winnings. In the specifications presented here, the form of the dependent variable is the natural log of

tournament money winning per event ($\ln(M/E)$). Since it is necessary to control for the number of events that a particular player enters, winning per event is a natural transformation.¹ The log form is chosen for two reasons: first, the residuals are heteroskedastic unless the log transformation is used and, second, the predicted winning per event is bounded by zero in log form, but predicted money per event can be negative without the log transformation.

Table III: Regression Results: Log of Money per Event = Dependent Variable

Variable/ Summary Statistics	Regression 1	Regression 2	Regression 3	Regression 4
Constant	90.96	-16.33	-0.4438	98.25
<i>GIR</i>	0.2613* (12.89)		0.1454* (7.12)	0.2634* (15.09)
<i>Ln(PPR)</i>	-30.92* (-12.13)	-14.96* (-4.81)		-31.41* (-12.71)
<i>Ln(DD)</i>	1.027 (0.495)	12.602* (4.82)		
<i>DA</i>	-0.0056 (-0.452)	0.0664* (4.31)		
<i>SS</i>	0.0062 (0.957)	.0210* (2.37)		
$\bar{R}^2$	.5893	.1975	.2192	.5916
<i>SEE</i>	.7158	1.001	.9871	.7139
$F_{k, n-k-1}$	51.80*	11.89*	50.67*	129.20*

(notes: n = 178; k = number of regressors; t-statistics in parentheses; * indicates significance at $\alpha < .01$, .05, and .10, respectively)

The explanatory variables *PPR* and *DD* are in log form so that their coefficients are interpreted as estimates of elasticities. Since the other variables are already in percentages, they are not in log form. Their coefficients (*GIR*, *DA*, *SS*) when multiplied by 100) are sometimes called semi-elasticities, i.e., the percentage change in the (untransformed) dependent variable due to a one unit change in the independent variable.

Regression 1 of Table III explains approximately 59% of the variation in the log of money per event across players on the EPGA tour. Again, the coefficients on sand save percentage (*SS*), driving distance (*DD*) and driving accuracy (*DA*), suggest little independent predictive power for any of those variables in regression 1. The coefficient for *GIR* suggests that a one-unit change in *GIR* implies a 26.13% increase in money winnings per tournament. The coefficient for *PPR* suggests that a 1% decrease in putts per round leads to a 30.92% rise in money winnings per event. Since *DD* and *DA* operate through *GIR*, their interpretation is left for regression 2. In regression 2, the *GIR* variable is dropped. Notice that the coefficients for *PPR*, *DD*, *DA*, and *SS* are all correctly signed and statistically significant. Regression 2 is offered to assess (as in Table II) the effects of *DD* and *DA* on the dependent variable without *GIR* absorbing most of their effects. In that regression, the elasticity of money winnings per event with respect to *PPR* is estimated to be 15.0%, that is, a one percent decrease in *PPR* implies a 15% increase in money per event. In regression 2, the implied elasticity of *DD* is 12.6%, meaning that if *DD* increases by 1%,

¹ It is also (of course) possible to control for events entered by entering the number of events as an explanatory variable.

winnings per event are estimated to increase by 12.6%. The semi-elasticity for *DA* is estimated to be 6.64%.²

Regression 3 is offered to assess the effect of *GIR* alone on money winnings per tournament. *GIR* in this estimation has an implied semi-elasticity of 14.5%, and explains almost 22% of the variance in tournament winnings. Finally Regression 4 suggests that *GIR* and *PPR* explain approximately 60% of the variation in money winnings per event across players—the same proportion as the less parsimonious model in Regression 1. Further the *GIR* and *PPR* coefficient estimates for Regression 4 are nearly identical to those of Regression 1.

CONCLUSIONS

We present evidence of the determinants of success on the European Professional Golf Association (EPGA) 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 assess the effects of driving accuracy and driving distance, it is necessary to remove *GIR* from the estimating equations. Those formulations suggest that driving accuracy and driving distance are approximately equally important in determining scoring average.

A comparison of the United States PGA tour, the European PGA tour, and the Ladies PGA tour is a potential future project.

REFERENCES

- [1] 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.
- [2] Davidson, J. D. & Templin, T. J. (1986) "Determinants of Success Among Professional Golfers." *Research Quarterly for Exercise and Sport*, 57, 60-67.
- [3] 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.
- [4] Engelhardt, G. M. (1995) "'It's Not How You Drive, It's How You Arrive': The Myth." *Perceptual and Motor Skills*, 80, 1135-1138.
- [5] 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.
- [6] Englehardt, G. M. (2002) "Driving Distance and Driving Accuracy Equals Total Driving: Reply to Dorsel and Rotunda." *Perceptual and Motor Skills*, 95, 423-424.

² One must be careful with this interpretation. Since the explanatory variable is in percentage form, a one unit change is one percentage point, but that is a larger percentage of the mean of the variable, since all mean percentages for these variables are considerably less than 100%.

- [7] 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.
- [8] Moy, R. L. and Liaw, T. (1998) "Determinants of Professional Golf Tournament Earnings." *The American Economist*, 42, 65-70.
- [9] Nero, P. (2001) "Relative Salary Efficiency of PGA Tour Golfers." *The American Economist*, 45, 51-56.
- [10] 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.
- [11] Shmanske, S. (1992) "Human Capital Formation in Professional Sports: Evidence from the PGA Tour." *Atlantic Economic Journal*, 20, 66-80.
- [12] 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*.
- [13] *Proceedings of the World Scientific Congress of Golf*. London: E & FN Spon. 199-204.
- [14] Woolridge, Jeffrey M., *Introductory Econometrics: A Modern Approach*, South-Western College Publishing, 2000.