

A Study of the Performance and the Determinants of Winning for Professional Female Golfers

Ming-Ly Hsu, Lecturer
LEE-MING Institute of Technology
New Taipei City, Taiwan

Kuei-Pin Kuo, Assistant Professor
National Pingtung University of Science and Technology
Ping-tung, Taiwan

Abstract—The purpose of this paper was to evaluate the performance and the determinants of winning in golfers of the Ladies Professional Golf Association (LPGA) in 2010. This study used a two-stage data envelopment analysis (DEA) approach: the first stage of this study involved measuring the efficiency scores of female golfers, and the second stage of the study used Tobit regression analysis to explain the variations in these efficiencies. Eight of the examined golfers achieved an efficient approach, whereas 103 golfers were inefficient. There were significantly positive effects on efficiency scores from the following items: “driving distance”, “driving accuracy percentage”, “putts per round”, and “money leaders”. Thus, both driving and putting were determinants of success in LPGA tournaments in 2010. This study marked the first attempted application of either a censored data model or a limited dependent variables model to the context of Taiwanese athletes, and the results of the Tobit regression analysis explained how golfers improve with respect to efficiency and competitiveness.

Keywords-component; *Tobit Regression Analysis, Censored Data Model, Putts, Data Envelopment Analysis, Determinants*

I. INTRODUCTION

The analysis of sports data provides a useful application for econometric and managerial techniques (Nero, 2001). At the highest level of competition, athletes must possess not only overall skill but also sufficient logistical support; in other words, coordination of a wide variety of factors is required to achieve perfect scores in competitions. To focus on the training perspective, the first goal of modern golf training is to win the tournament; thus, for a golf coach to understand the technical situation and demands of each athlete, the analysis of all types of competition skills and statistics after the game has become very important.

Sport-related applications of analytical approaches are interesting and entertaining; moreover, sports provide large quantities of data for analytical purposes. In particular, there are applications of data envelopment analysis (DEA) to baseball. Anderson and Sharpe (1997) use DEA to evaluate the performance of baseball hitters and calculate a composite batter index (CBI). Studies have increasingly addressed and discussed a growing number of baseball-related issues (Sexton & Lewis, 2003; Lewis & Sexton, 2004; Lewis, Sexton, & Lock, 2007; Lewis, Lock, & Sexton, 2009). Fried, Lambrinos,

and Tyner (2004) utilized DEA to assess golfer performance but were not aware of any applications of DEA to golf.

Nero (2001) used regression analysis to estimate golfer earnings through an examination of four performance measures: putting average, driving distance, driving accuracy and number of sand saves. Because the assessment of golf techniques involves multi-level considerations and numerous inputs and outputs, an increasing number of studies have been employing DEA to investigate the performance of golfers. Fried et al. (2004) examined the performance under pressure and the personal competence of professional golfers of the PGA (Professional Golf Association), the LPGA (Ladies Professional Golf Association) and the SPGA (Senior Professional Golf Association) tours of the United States in 1998.

Regression analysis is typically used to predict or judge the directions and magnitudes of correlations between two variables, whereas DEA is more frequently employed to determine the relative efficiency among decision making units (DMUs); thus, these two research methods should be applied to appropriately address distinct investigative topics. Coelli, Rao and Battese (1998) observed that scholars in the management field have generally adopted a two-stage method to assess the influence of different management factors on efficiency. Lovell, Walters and Wood (1995) summarised related literature that used the traditional ordinary least squares (OLS) method of calculating regressions and found that the efficiency values are largely estimated from corrections to the DEA model. However, related literature revealed that if the least squares method is adopted to estimate the value of the parameter, bias and inconsistency are typically produced (Cooper, Seiford & Tone, 1999). Therefore, a two-stage search for efficiency values can be adopted that generally produces technical efficiency values that are smaller than 1; these values differ from the unlimited range that is found for the dependent variables of the general regression model. Thus, with the Tobit regression model (Tobin, 1958), which is a type of censored data model (also known as a limited dependent variables model), in the regression analysis, the aforementioned issues regarding the generation of bias and inconsistency can be solved.

In summary, a two-stage DEA is adopted in this study to investigate the issue of the game-participation efficiency of 2010 LPGA golfers. In the first stage of the study, DEA is applied to assess the game-participation efficiency of the athletes and create efficiency rankings; in the second stage of the study, Tobit regression is used to analyse and inspect determinants (which are also referred to as key factors or winning factors) that affect the game-participation efficiencies of the athletes. Through the analysis of the annual game-participation efficiencies and determinants of winning for the examined athletes, this study seeks to produce results that can be used as an important reference for facilitating superior future performance.

II. METHODOLOGY

A. Research Targets

This study sought to analyse the complete annual performance of golfers who participated in the 2010 LPGA Tour. In particular, this investigation focused on the top 120 golfers in the money rankings for 2010. Among these athletes, several golfers did not participate for entire quarters of the year or retired during the course of the year (for example, Lorena Ochoa retired in 2010 after only attending six tournaments); these golfers and certain athletes with incomplete data were removed from the examined pool of golfers, reducing the size of the pool to 111 golfers.

B. Data Envelopment Analysis

The methods that economists employ to evaluate production efficiency and issues can be broadly classified into the two categories: parametric approaches and non-parametric approaches. DEA is a non-parametric method that was first proposed by Charnes, Cooper and Rhodes (1978). Initially, this method assumed a situation involving constant returns to scale (CRS); subsequently, Banker, Charnes and Cooper (1984) expanded this approach into a variable returns to scale (VRS) model. DEA does not require assumptions to be made regarding the production function, and its only requirement is that the set of production possibilities must be convex. Further, DEA employs a linear planning technique to obtain the efficiency envelopment curve that is used as the basis for evaluating the efficiency of each DMU, and it is applicable to the efficiency assessment of multiple inputs and outputs.

Before DEA is used to assess each DMU, the correlation between input and output is not determined; hence, the output efficiency values reflect relative efficiency instead of absolute efficiency. DEA can be utilized not only to calculate the efficiency value of each DMU, but also to adjust the input and output variables of each DMU in a way that maximises efficiency. Therefore, to achieve an optimal model for assessing the game-participation efficiency of the golfers who participated in the US LPGA in 2010, this study has chosen to employ DEA and has adopted a VRS output-oriented model to obtain the DEA results.

C. Tobit Regression

Tobit regression analysis was proposed by Tobin in 1958. It is applicable to the analysis of whether the dependent variable of a dataset is segmented or sectioned. If this approach is used to assess efficiency values that are between 0 and 1, this type of regression analysis may be regarded as a limited dependent variables model or a censored data model (Coelli et al., 1998).

To further investigate the determinants of winning with respect to the game-participation efficiency of LPGA athletes in 2010, the game-participation efficiency value of LPGA athletes is used as the dependent variable. As mentioned above, because the value of this dependent variable is between 0 and 1, the regression model of this study can be considered a censored data model or a limited dependent variables model. This characterisation holds because, although the independent variable can correspond to any observation value, the dependent variable will nonetheless only incorporate a subset of the corresponding observation values; thus, even if the limited observation value has a higher proportion, the expectation value of the error item will not necessarily be zero (Powell, 1986). Traditionally, certain studies in the extant literature have adopted the least squares method for estimating parameter values; however, OLS will generate biases and inconsistency (Cooper et al., 1999). Therefore, to avoid these issues, this investigation has adopted the Tobit regression as the empirical model for its analysis.

In the Tobit regression empirical model of this study, the dependent variable is the technical efficiency (TE) of each examined golfer. The independent variables include events, money leaders (moneys), driving distance (dd), driving accuracy percentage (dap), greens in regulation percentage (girp), sand save percentage (ssp) and putts per round (ppr). Hence, the Tobit regression model of this investigation may be expressed as follows:

$$TE = \beta_0 + \beta_1 \text{events} + \beta_2 \text{moneys} + \beta_3 \text{dd} + \beta_4 \text{dap} + \beta_5 \text{girp} + \beta_6 \text{ssp} + \beta_7 \text{ppr} + \varepsilon_i$$

D. Definition of Terms

Building on the work of Fried et al. (2004), this study created seven input and output evaluation indices. In particular, the variable of events is the output index for this investigation, whereas the six items of money leaders, driving distance, driving accuracy percentage, greens in regulation percentage, sand save percentage and putts per round are the input indices for this study. In the design of this research, the nature of the indices was considered; for example, putts per round were converted to its reciprocal value to achieve calculation consistency. Each input or output index is defined in accordance with LPGA standards as follows:

1. Events: the annual number of events that were played in the 2010 LPGA Tour by an athlete.
2. Money leaders: the sum (in US dollars) of the annual monetary prizes that were earned by an athlete during the 2010 LPGA Tour.

3. Driving distance: an athlete's average driving distance, in yards.
4. Driving accuracy percentage: the percentage of occasions on which an athlete drove her golf ball from the tee to the fairway.
5. Greens in regulation percentage: the percentage of greens that an athlete reached in regulation.
6. Sand save percentage: the percentage of occasions during which an athlete successfully saved her golf ball from the sand after failing to reach the green in regulation.
7. Putts per round: the average number of putts that an athlete required on the green during the course of each round.

III. RESULTS

A. *The Efficiency and Ranking of LPGA Golfers in 2010*

This study investigated the game-participation performance and determinants (factors that significantly affect performance and winnings) for LPGA Tour golfers in 2010. The original study target and evaluation index data were the top 120 highest earning athletes of 2010; after removing athletes who did not complete a tournament for an entire quarter, those who retired and those who presented insufficient data, 111 golfers remained in this study target. In total, seven input/output evaluation indices were examined. The descriptive statistical results are presented in Table 1.

First stage: This study sought to assess the monetary rankings, game-participation efficiencies and efficiency rankings of the examined LPGA athletes in 2010. Moreover, the input and output data for each athlete were introduced into the DEA Solver (version 5.0) software package for calculations, and the variable scale return output model was applied to generate technical efficiency performance (Table 2).

There were eight game-participation efficient athletes on the LPGA Tour in 2010: Na Yeon Choi was first, Jiyai Shin was second, Cristie Kerr was third, Yani Tseng was fourth, Suzann Pettersen was fifth, Ai Miyazato was sixth, Michelle Wie was ninth and Vicky Hurst was 28th in the money rankings for the LPGA for this year. The efficiency values for these eight athletes are all equal to 1, indicating that these athletes are relatively efficient; that is, their input/output ratio is relatively maximised. The efficient values for these game-participation efficient athletes are all located at the same equal-output curved surface and form a reference set; this set may be used for comparisons involving athletes who have not achieved game-participation efficiency.

The technical skills of athletes who have not achieved efficient game-participation merit further reinforcement and promotion.

Except for Yani Tseng, who has achieved game-participation efficiency, none of the other Taiwanese golfers on the LPGA Tour in 2010, including Candie Kung (money ranking: 36th place; efficiency value: 0.382; efficiency ranking: 60), Amy Hung (money ranking: 48th place; efficiency value:

0.233; efficiency ranking: 83) and Teresa Lu (money ranking: 63rd place; efficiency value: 0.330; efficiency ranking: 67) demonstrated efficient game-participation. In addition, Yu-Ping Lin did not participate in tournaments during 2010.

B. *Determinants that Affected the Efficiency of LPGA Golfers in 2010*

In the second stage of this study, a Tobit regression model was applied to identify and assess the influential factors (Table 3) that affected the monetary performance of LPGA golfers in 2010. In this process, the technical efficiency value of each of the 111 examined golfers from the first stage of the analysis was used as the dependent variable. Accumulated years of experience, height and nationality of the athletes were examined as environmental variables for predicting game-participation performance, but no significant results were obtained for these factors.

Therefore, in the Tobit regression model of this study, the input-output index that was obtained in the first stage of the analysis is selected as an independent variable to determine the influential factors (determinants) of the game-participation performance by the athletes. From Table 3, it can be observed that at a significance level of 1%, positive correlations may be observed between efficiency and each of the following factors: driving accuracy percentage, driving distance, money leaders and putts per round. The importance of driving distance with respect to efficiency was in accordance with the modern trend of power-oriented play, whereas the correlation between the game-participation efficiency of LPGA golfers and the putts per round factor was in accordance with the results of Fried et al. (2004). In both the current study and this earlier investigation, this finding demonstrated that putting technique is a highly important and influential aspect of the LPGA game.

The events factor demonstrates a significant negative correlation to the game-participation performance of LPGA athletes, implying that this index was not able to enhance the game-participation efficiency of the examined golfers.

Moreover, an insignificant correlation is observed between golfer efficiency and either of the two indices of greens in regulation percentage and sand save percentage.

IV. DISCUSSION AND CONCLUSIONS

A. *The Efficiency and Ranking of LPGA Golfers in 2010*

In the first stage of this study, there were eight golfers who reached efficiency on the LPGA Tour in 2010: Na Yeon Choi was first, Jiyai Shin was second, Cristie Kerr was third, Yani Tseng was fourth, Suzann Pettersen was fifth, Ai Miyazato was sixth, Michelle Wie was ninth and Vicky Hurst was 28th in the money rankings for the LPGA in 2010. For athletes with an efficiency value of 1, the input resource is efficiently located in the LPGA context for the entire year; that is, the input events and all of the technical performances have reached an efficient output. The Taiwanese golfer Yani Tseng was one of the golfers who led the rankings of efficient game-

participation; this fact helps explain the origins of her outstanding scores and performance. In addition, in the future, if clear rankings of the efficient athletes must be determined, various research methods, such as assessments of super efficiency, can be used to obtain these rankings.

There were 103 golfers who were relatively inefficient, and their technical performance level must still be enhanced and promoted. Therefore, this study focused on investigating the efficiency of these golfers. In fact, suggestions regarding the extent of improvement and the reference learning targets that can be achieved for these inefficient golfers may be acquired through slack variable analysis, scale return analysis and assessments of the reference set; this topic is a future research direction that follows from the current investigation.

Except for Yani Tseng, who has been efficient, other Taiwanese golfers did not achieve game-participation efficiency. In particular, these other athletes include Candie Kung (money ranking: 36th place; efficiency value: 0.382; efficiency ranking: 60), Amy Hung (money ranking: 48th place; efficiency value: 0.233; efficiency ranking: 83) and Teresa Lu (money ranking: 63rd place; efficiency value: 0.330; efficiency ranking: 67). Although Yu-Ping Lin did not participate in the LPGA Tour in 2010, the number of Taiwanese golfers who participated in the Tour did not change because Yani Tseng was a new participant. All of the examined Taiwanese athletes demonstrated significant enhancements in both money ranking and efficiency, indicating the enhancement of the golf technique and competitiveness of Taiwanese LPGA golfers. This development augurs well for the future environment of women's professional golf in Taiwan.

B. Determinants that Affected the Efficiency of LPGA Golfers in 2010

This study initially established the three variables of accumulated years of experience, athlete height and athlete nationality as the environmental parameters for predicting game-participation performance; however, none of these variables proved to be significantly related to performance. Therefore, the second stage of this investigation utilized the input/output index from the DEA of the first stage of the study as the independent variable, and the Tobit regression model was used to obtain key factors that influenced the monetary performance of LPGA golfers in 2010.

Driving accuracy percentage and driving distance were both significantly correlated with golfer efficiency, demonstrating the prominence of driving distance for LPGA athletes and emphasising the importance of improvements in golf equipment, scientific training and weight training for facilitating the development of greater power for professional golfers. Driving distance and accuracy considerations provide the primary explanations for why the strength of the competition in women's professional golf is rapidly approaching the level of competition for male golfers. Moreover, the importance of these factors disproves the

assertion, promulgated by Fried et al. (2004), that driving distance is merely a publicity stunt.

Money leaders are positively correlated with golfer efficiency; that is, the examined golfers who obtain greater earnings demonstrate superior performance. This finding is in accordance with prior expectations and verified the hypotheses of this study regarding the game-participation efficiency and monetary awards of the examined athletes.

Putts per round are positively correlated to the efficiency of LPGA golfers. This result was not only consistent with the findings of Fried et al. (2004), who also employed the DEA research method, but also supported the viewpoint "drive for show, putt for money", implying that putting technique has a role and importance that will affect the results in distinct golfing associations, such as the PGA and LPGA, and in various tournaments.

The events factor is significantly negatively correlated with golfer efficiency, indicating that this factor does not enhance the competitive performance of the examined golfers. In other words, the viewpoint of many athletes, that enhanced competitive performance should result from more events, was not supported, and strategic efficiency appears to be a more important consideration with respect to athletic performance.

Further, an insignificant correlation was observed between efficiency and the examined indices of greens in regulation percentage and sand save percentage. This finding shows that shot-making skills and sand save percentages were not the primary influences for efficient golfing performance; thus, these skills were not the key techniques for determining the probabilities of winning on the LPGA Tour in 2010. In a prior study that compared different research methodologies, sand save percentage was found to yield a result that completely opposed initial predictions.

V. CONCLUSIONS

1. The efficiency rankings of LPGA golfers in 2010:
 - (1) Eight of the studied LPGA athletes reached maximum relative efficiency in 2010 (including Taiwanese golfer Yani Tseng, who was in fourth place in the money rankings for the year).
 - (2) One hundred and three golfers did not attain game-participation efficiency (including domestic athletes Candie Kung, who was in 36th place in the money rankings for the year; Amy Hung, who was in 48th place in the money rankings for the year; and Teresa Lu, who was in 63rd place in the money rankings for the year).
2. The determinants that affected the efficiency of LPGA golfers in 2010 were:
 - (1) Significant positive correlations were found among driving accuracy percentage, driving distance, monetary awards and putts per round. This result demonstrates that trends towards increased power and putting skill are not only the key factors that affected the efficiency of LPGA golfers in 2010, but also important factors in determining monetary rewards (as both driving and

putting considerations were correlated with monetary awards). The two-stage approach of this study has thus verified an important hypothesis regarding monetary leaders.

- (2) There was a significant negative correlation between the events factor and the game-participation performance of the athletes, indicating that this technical index cannot enhance athlete efficiency. Instead, the game-participation strategy of the athlete throughout the entire year is a more critical concern.

The greens in regulation percentage and sand save percentage did not appear to exert any significant influence on golfer efficiency.

VI. SUGGESTIONS

This study examines the performance of LPGA golfers in 2010. In the first stage of this investigation, DEA is adopted to calculate the efficiency value of each examined athlete; in the second stage, a Tobit regression approach that uses a censored data model is adopted to investigate the external environmental factors (determinants) that affect the monetary performance of 2010 LPGA golfers. This approach differs from the method adopted by Fried et al., (2004), who used only monetary winnings as a dependent variable for predicting golfer performance. However, this study further analysed the efficiency of the golfers using a model with restricted data and limited dependent variables, and this investigation constitutes the first example of this approach in the field of sports research. Thus, the results of this study may be used as a reference for future research and or for various concepts and applications.

Various research methods may be used in the future to rank the performance of efficient athletes, such as assessments of super efficiency or fuzzy measurements. The extent of improvement that is possible for golfers and the appropriate reference and learning targets for these athletes can be estimated using slack variable analysis, scale return analysis and examination of the reference set. These approaches and methods warrant further investigation.

The findings of the current study indicate that, at the present time, driving skill and putting skill are key determinants of the efficiency of LPGA golfers. Due to the rapid advancement of modern technology, the distance and difficulty of golf courses are continuously being enhanced; thus, the skills required of golfers are also continually evolving, with respect to driving distance, driving accuracy, and shot-making skills. As a result, the continual tracking of LPGA-related technical factors, including the analysis and investigation of the influence of different environmental variables, should occur. (For instance, this study has investigated items such as accumulated years of experience, height and nationality.) In addition, professional golf for male athletes at different competitive levels, such as the Asian Tour, the European Tour and the US PGA Tour, may also merit future study.

In 2011, the performance of Taiwanese athlete Yani Tseng, who was champion of the LPGA Tour, was very impressive. Her efficiency and performance with respect to the various determinants examined in this study were markedly different to those of her competitors. Such outstanding performances should be analysed and the data stored in a database that includes information over the long term; in other words, the performance of golfers should be continuously monitored, to harness information for strategy and analysis. The findings from this database can also be utilized as an important reference for conducting training and developing competitive skills. The eventual goal of these suggested measures is to enhance the competitiveness and visibility of Taiwanese athletes in international golf.

Figures and Tables

Table 1: Descriptive statistics of lpga tour golfers in 2010

	events (I ₁)	moneys (O ₁)	dd (O ₂)	dap (O ₃)	grip (O ₄)	ssp (O ₅)	ppr (O ₆)
Total	111	111	111	111	111	111	111
Minimum	9.00	28.00	225.10	.52	.52	.18	31.35
Maximum	24.00	1871.00	274.50	.77	.75	.67	28.11
Mean	17.90	325.01	247.97	.67	.66	.39	29.94
Std. dev.	3.79	402.02	8.91	.06	.04	.07	1538.46

Table 2: Money, efficiency and ranking of lpga golfers in 2010

Money ranking	Golfers	Efficiency	Efficiency ranking
1	Na Yeon Choi	1.000	1
2	Jiyai Shin	1.000	1
3	Cristie Kerr	1.000	1
4	Yani Tseng*	1.000	1
5	Suzann Pettersen	1.000	1
6	Ai Miyazato	1.000	1
7	In-Kyung Kim	0.999	9
8	Song-Hee Kim	0.829	35
9	Michelle Wie	1.000	1
10	Paula Creamer	0.973	33
11	Inbee Park	0.774	38
12	Katherine Hull	0.715	40
13	Morgan Pressel	0.513	44
14	Amy Yang	0.513	45
15	Brittany Lincicome	0.924	34
16	Sun Young Yoo	0.783	36
17	Mika Miyazato	0.484	50
18	Angela Stanford	0.497	46
19	Jee Young Lee	0.486	48
20	Maria Hjorth	0.999	13
21	Stacy Lewis	0.442	54
22	Jimin Kang	0.475	51
23	Karrie Webb	0.416	57
24	Anna Nordqvist	0.346	64
25	Karine Icher	0.517	43
26	Christina Kim	0.471	52
27	Kristy McPherson	0.487	47
28	Vicky Hurst	1.000	1
29	Juli Inkster	0.433	56
30	Azahara Munoz	0.309	71

Table 3: Influential factors affecting efficiency of 2010 lpga golfers

Independent variables	Coefficient	Std. error	t-statistic
Constant	-0.0499	1.8538	-0.0269
events (I ₁)	-2.8865***	0.6299	-4.5823
moneys (O ₁)	-0.0278***	0.0046	-6.0402
dd (O ₂)	-0.0095	0.0076	-1.2499
dap (O ₃)	-3.8193***	1.2195	-3.1320
grip (O ₄)	0.0014***	0.0002	9.0370
ssp (O ₅)	0.3947***	0.0625	6.3109
ppr (O ₆)	-0.3718	0.4003	-0.9288

Log likelihood function=1.3018.

*** Significance at 1% level.

** Significance at 5% level.

* Significance at 10% level.

REFERENCES

- ANDERSON, T. R., & SHARP, G. P. (1997). A new measure of baseball batters using DEA. *Annals of Operations Research*, 73: 141-155.
- BANKER, R. D., CHARNES, A., & COOPER, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(7): 1078-1092.
- CHARNES, A., COOPER, W. W., & RHODES, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6): 429-444.
- COELLI, T., RAO, D. S. P., & BATTESE, G. E. (1998). *An Introduction to Efficiency and Productivity Analysis* (2nd ed.). New York, NY: Springer.
- COOPER, W. W., SEIFORD, L. M., & TONE, K. (1999). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Boston, MA: Kluwer Academic Publishers.
- FRIED, H. O., LAMBRINOS, J., & TYNER, J. (2004). Evaluating the performance of professional golfers on the PGA, LPGA and SPGA tours. *European Journal of Operational Research*, 154: 548-561.
- LEWIS, H. F., LOCK, K., & SEXTON, T. R. (2009). Organizational capability, efficiency, and effectiveness in Major League Baseball: 1901-2002. *European Journal of Operational Research*, 197: 731-740.

AUTHORS PROFILE

Authors' names & institutions:

First author: Ming-Ly Hsu

LEE-MING Institute of Technology

Second Author: Kuei-Pin Kuo

National Pingtung University of Science and Technology

Corresponding author:

Name: Kuei-Pin Kuo

Email: kweibin@mail.npust.edu.tw

Mailing address: 1, Shuefu Road, Neipu, Pingtung 912, Taiwan

Tel: +886-8-7703202#7228

Fax: +886-8-774-0471