Have Advanced Analytics Influenced Golf Betting Odds?

Reko Laaksonen
Master's Thesis in Information Systems
Supervisor: Anna Sell, Xiaolu Wang
Faculty of Social Sciences, Business and Economics, and Law.
Åbo Akademi University
Turku 2023

Abstract

Subject: Information Systems	
Writer: Reko Laaksonen	
Title: Have Advanced Analytics Influenced Golf Betting Odds?	
Supervisor: Anna Sell, Xiaolu Wang	
Abstract:	
The study tries to find an answer to whether betting lines for major championships of golf are following the Strokes Gained model or The Official World Golf Ranking (OWGR). The literature review focuses on how the Strokes Gained model, the Official World Golf Ranking and the betting market function and what influences those three factors. The study focuses on To Win odds of the 19 major championships played in 2018 – 2022. The method of the study was to conduct a Pearson’s correlation test between the given odds and a player’s position in the OWGR and his Strokes Gained Total and Strokes Gained Approach statistics. Afterwards a Granger Causality test was conducted for the 23 players who were ranked inside the top 75 of the OWGR throughout the studied time period. The study concludes that the player’s position in the OWGR is the strongest correlating factor with odds with some exceptions. The Granger causality test gave various answers: Most importantly that a player’s position in the OWGR can be seen as starkest Granger causer of odds when a player is ranked highly in the OWGR, whereas a lower ranked player’s odds tend to be Granger caused by his Strokes Gained metrics.	
Keywords: Granger causality, golf, strokes gained, correlation.	
Date: 7.11.2023	**Number of pages**: 94

Contents

1 Introduction

This chapter will briefly present the subject, motivates why the subject is chosen, and describes the structure and objective of this thesis. This chapter also presents research questions and the methods used for answering them.

1.1 Introduction

Golf is quite new to advanced analytics. Until recently, the only available metrics was scoring average, lengths of shots, number of putts taken and other trivial, somewhat non-descriptive numbers. The problem in question is that the two metrics studied in this thesis function in a completely different manner; The other one makes it easy to understand performance in any given time frame whereas the other one is an accumulation of a player's long-term performance. The only reasonable way to measure performance was the Official World Golf Rankings (OWGR), still used as an official measure on who the best player in the world is. OWGR is based on an algorithm that considers players' performance in tournaments from the last two years.

However, not every tournament is the same since players participating are not always the same. When a tournament starts, each player contributes to what is called the Strength of Field, which is used to determine how hard it is to win that tournament. Strength of Field is formed by calculating a total from numbers that every player brings into the mix, but with starkly non-linear effects; Players ranked number 10 and 11 in OWGR, together contribute as much to the Strength of field as player ranked number five in the OWGR. Players ranked worse than 50^{th} in the OWGR, are irrelevant for the Strength of Field.

Then came along Mark Broadie, a professor at Columbia University who created the Strokes Gained model, nowadays used in every possible way in golfing world. Broadcasts use it to compare players and give the viewer better information of the ongoing tournament, players use it in training to get better on areas they are not relatively good at, and even avid amateur golfers have taken up the Strokes Gained model to get better in playing the game they play for fun.

Strokes Gained, in short, is used in same ways as Expected Goals (xG) - number is used in football and other team sports to predict a team's performance in a statistical manner to forecast results. In football's case, the result being goals, in golf's case the result being strokes taken to finish a round of golf.

Strokes gained (SG) consists of four different measured numbers: SG Off-The-Tee or Driving, SG Approach, SG Around-The-Green and SG Putting. These four numbers form a sum, called Strokes Gained or Strokes Gained Total.

The betting industry for long has relied on all possible data when giving out odds for games and tournaments. When betting on football and comparing the odds with given xG-numbers, it is quite clear to see how the odds are formed and what influences those. For example, American football's odds are heavily influenced by the presence or absence of the starting quarterback. In golf, on the other hand, some oddities show up, in a form of over-appreciating former performances or depreciating recent performances. Broadie's study from 2014 shows that approach-play, SG Approach, is the most important number in performing well, thus trumping a long-lived myth in the golfing world, that putting is the most important part of the game if one wants to win.

Because the best players, ranked 1 to 50 in OWGR measurement, gather almost exclusively only four times a year, for major tournaments which are the equivalent of Grand Slams in tennis, the study is going to focus on odds given for those four tournaments from year 2018 to the end of year 2022.

1.2 Objective

The objective of this study is to provide new information on whether Mark Broadie's Strokes Gained model, and a sub-part of it, is correlated with odds given by betting companies, when giving To Win odds before major golf tournaments and whether the given odds have some type of trend to them. The study also investigates the Official World Golf Ranking's influence on odds.

Golf's nature as opposed to many other sports is that the game is not played by two opposing sides; a player plays against many other competitors, henceforth making predicting a winner much harder. Also, a tournament of golf is played on four consecutive days which brings in more variables due to weather, and the tournament organizers, as the course is set up differently each day of the tournament.

OWGR and Strokes Gained as measurements go somewhat hand in hand, but in such a way that OWGR indirectly depends on Strokes Gained whereas Strokes Gained does not have any dependence to the OWGR. Strokes Gained gives a much more realistic picture of a player's performance for a time period as long as is desired to inspect, whereas the OWGR is focused on a player's performance during a longer period. The OWGR as a statistic relies heavily on whether a player has managed to win tournaments whereas the Strokes Gained statistic gives insight into how a player has performed in comparison to other players. The Strokes Gained statistic does not take wins into account but is a measurement of true performance on the golf course which is important as a win is often decided by one shot or even a play off.

The study tries to find which of the chosen metrics correlates the most with the given odds. The study tries to seek answers for whether a different metric correlates differently with a different tournament and are there trends to be found for each of the four major tournaments.

To address the objective of this research, the thesis will seek the answers to the following research questions:

1. *Do the given odds correlate more with the OWGR or Strokes Gained statistics?*
2. *Can Strokes Gained Approach metric be used to predict future odds?*

1.3 Method

The study is going to be a quantitative analysis with the use of Pearson's correlation and Granger Correlation.

The study's research design is exploratory. Exploratory design can be used when the problem in hand is not defined clearly and there are no clearly defined hypotheses (Cooper & Schindler, 2011). The purpose of an exploratory design is to study that which has not been previously studied. An exploratory study often results from an examination of the literature in which the researcher cannot find the answer to the question. Henceforth, Granger Causality was chosen as the used method because of its nature of giving a hypothetical answer whether a time series can be used to predict another time series.

The thesis will conduct a literature review of the theory behind how betting odds are set and which factors in general influence them, and on theory behind Mark Broadie's Strokes Gained model and its application in analyzing performance. The literature review will give reasons for choosing the studied metrics. The thesis will also explain how the Official World Golf Ranking system works and what is relevant to know about it regarding this study.

Data used for the study is collected from the PGA Tour and OWGR. Additional data was sought from the United States Golf Association, The Professional Golfers Association of America, Augusta National Golf Club and The Royal and Ancient Golf Club of St. Andrews. The reason for choosing these four institutions is that they run the four major championships and therefore each of those four institutions collect and store data of scoring on their championships that are The Masters, The PGA Championship, The United States Open Championship and The Open Championship.

1.4 The structure of the thesis

The upcoming thesis will be structured as follows: Chapter 2 will be a literature review focusing on the theory behind the Strokes Gained model and its applications, and on theory how betting odds are set. In Chapter 3 methods will be explained, the data collected and used in the study described and results will be discussed. After presenting the results in Chapter 4, in Chapter 5 findings and limitations of the study and future research opportunities will be discussed.

2 Evaluation Metrics for the Performance of Professional Golfers

This chapter will present a theoretical frame for Strokes Gained statistics and how to analyze it, as well as how to analyze the Official World Golf Rankings and what influences the betting market. The chapter begins by presenting how golf performance has been understood for long, and further continues to present Strokes Gained and its parts. The chapter concludes by presenting how the betting market works, and which factors have an influence on it.

2.1 The Official World Golf Ranking

The mission of the Official World Golf Ranking (OWGR) is to administer and publish, on a weekly basis, a transparent, credible, and accurate Ranking based on the relative performances of players participating in male Eligible Golf Tours worldwide (The Official World Golf Ranking, 2023). The board of the OWGR consists of people representing the different golf tours around the world. The most prominent golfing tours of the world are the PGA Tour which is played in the United States, the DP World Tour which is mostly played in Europe and Asian Tour which is played in Asia. These tours generally have the fields that yield the most ranking points for the players (Broadie & Rendleman, 2013). Field is the term used when talking about the players who are playing in a tournament, therefore forming a Field of Players.

2.1.1 How the Official World Golf Ranking works

The OWGR System is run over rolling Ranking Periods. Ranking points are maintained at full value for a 13-week period from the relevant Ranking Date on which they were awarded to place additional emphasis on recent performances. Ranking Points are then reduced in equal decrements for the remaining 91 weeks of the relevant Ranking Period (The Official World Golf Ranking, 2023).

Each player is then ranked according to their average points during the Relevant Ranking Period, which is determined by dividing a player's Total Points, which is the total number of Ranking Points awarded to that player in the relevant Ranking Period, as adjusted to give effect to the principle above by multiplying by the Weight to give the player's Adjusted Points, by the number of Eligible Tournaments they have played during that Ranking Period, subject to the minimum and maximum divisors set out below.

There is a minimum divisor of 40 Eligible Tournaments over the Ranking Period, with no more than the most recent 52 Eligible Tournaments that the player concerned has played in during the relevant Ranking Period counting towards a player's position in the Official World Golf Ranking (The Official World Golf Ranking, 2023).

An effect this system has is that a player might have recently not performed at his best, but due to the divisor, the worst performances are left out. This causes that a player might be ranked incorrectly, meaning that the recent performances are not representative of the position in the OWGR.

2.1.2 The Strength of Field

When entering a tournament, each individual player brings his contribution to the Strength of the Field. Therefore, tournaments elected for the study are the four major tournaments in which the best possible fields are. To demonstrate why so-called regular events are not elected for the study, a random week in 2022 and tournaments played on the PGA Tour and DP World Tour on that week were chosen from the OWGR archive.

The field of a regular PGA Tour or DP World Tour event might consist of very varying skill levels of players, therefore making studying those betting odds highly irrelevant due to difference in skill. To illustrate, a look into the field of the Spanish Open, a tournament of the DP World Tour, in 2022 will explain the problem. The highest ranked player of that tournament was ranked 6th in the world at that time. The second-best player in the field was ranked 30th. The 3rd, 4th and 5th best players of the field were ranked 71st, 72nd and 73rd in the world (The Official World Golf Ranking, 2023).

Studying such an event and its odds will automatically have an inbuilt inequality between players' odds, because of one player's superiority to other players in the field. Such field is also not strong since the contributing players are of low-ranking places in the Official World Golf Ranking.

The winner of Spanish Open earned 14,196 OWGR points. The same week the Spanish Open was being played in Madrid, on the PGA Tour a tournament called Shriners Childrens' Open was being played on the other side of the Atlantic Ocean. The winner of that event earned 42,186 OWGR points (The Official World Golf Ranking, 2023). The 2nd and 3rd place finishers of that tournament earned more ranking points than the winner of the Spanish Open.

The winners of 2022 major tournaments earned 100 OWGR points respectively (The Official World Golf Ranking, 2023). Those tournaments have what is known as maxed-out Strength of Field, meaning that all the players whose ranking can contribute to the Strength of Field are playing in that tournament.

To summarize, a random PGA Tour event's Strength of Field is 2,37 times weaker than the field of a major tournament. Respectively, a random DP World Tour event's field is 7,04 times weaker than the fields of the major tournaments (The Official World Golf Ranking, 2023).

2.1.3 Problems regarding the Official World Golf Ranking system

OWGR rankings are, in a sense, subjective to which tournaments each player has played. A player might, due to his category, not be able to play on the higher ranked regular tournaments (PGA Tour, 2023; DP World Tour, 2023). Because of that, the given player might have to play against weaker players on a weekly basis, without encountering the players who are ranked higher than him but for major tournament weeks. Playing successfully against weaker fields has propelled a player high in the OWGR granting him eligibility to play in major tournament, without really being challenged by better players.

Also, a player playing mostly on the PGA Tour, will most likely have a more correct ranking than a player playing on the other tours. The concurrency on that tour tends to yield the highest Strength of Field every week, thus giving a player, regardless of his ranking, ranking points more correctly due to the better players mostly playing on that tour (The Official World Golf Ranking, 2023).

As explained in the previous chapter, playing on the PGA Tour will cause a player to be most correctly ranked in the OWGR. Using data from 2002 to 2010 and comparing the results from the OWGR and score-based methods, a study found that PGA Tour golfers are penalized by an average of 26–37 OWGR ranking positions compared to non-PGA Tour golfers (Broadie & Rendleman, 2013).

Henceforth, as explained shortly, Strokes Gained gives a much more realistic view into how that player has performed in comparison to all players. One must keep in mind that great playing most likely yields positive Strokes Gained numbers. Due to the not-so-good playing from the rest of the players in those tournaments, finishing in a good position in a tournament with a weaker field, comes easier.

Winning a tournament always yields a relatively large number of ranking points which can manipulate the OWGR if a player manages to win, or finish in high position, on many occasions on a weaker golf tour (Broadie & Rendleman, 2013).

After the summer of 2022 and the emergence of LIV tour, the use of OWGR data becomes even more irrelevant when assessing odds for major tournaments, because LIV Tour tournaments are not granted OWGR points, henceforth making the use of Strokes Gained the only feasible method when comparing the performance of different players playing different tours.

LIV Tour is a breakaway tour funded by the Saudi Arabian government. The tour, as of now, is not granted Official World Golf Ranking points. Henceforth, players who joined that tour, have had a declining position in the OWGR since joining the tour. After the formation of that tour, the OWGR does not measure the performance of players playing exclusively on that tour.

2.2 Strokes Gained model

The reason for inventing Strokes Gained was the lack of information traditional golf statistics provided. Those statistics were not able, in depth, to explain what the separating factor between the best professional golfers and the average professional golfer is (Broadie, M. 2014). In co-operation with the PGA Tour's ShotLink system, an immense amount of data was collected for Broadie to study.

2.2.1 The problems with traditional golf statistics

The statistics of old gave easily understandable numbers, but with little to no context. Golf was for a long time analyzed through number of putts taken, average length of drives, number or fairways hit, up-and-down percentage, sand-saves and such statistics that give information but don't consider the context (Broadie, M. 2014).

A statistic often shown in golf broadcasts is the number of putts a player has taken during his round. An example to illustrate the problem of such statistic could be the following: For a round of golf, player A has carded a score of 67 strokes and player B has carded a score of 72 strokes. Player A took 31 putts and player B took 27 putts respectively.

Player A clearly carded a better score whilst taking more putts. Also, this type of statistics is not too informative on, for example, the length of the putts each player has taken, which is a big factor in how many putts a player takes (Broadie, M. 2014). One can easily understand that player A has hit other shots than putts much better than player B, because his score is better. Still, this provides no insight whatsoever into how player A or B managed to card these scores and what the separating factors between those two rounds were (Broadie, M. 2014).

2.2.2 Inventing Strokes Gained

To combat the problems of the older golf statistics Mark Broadie wanted to invent a system that makes it possible to compare and assess each part of the game individually between players as well as a whole (Broadie, M. 2014). Strokes Gained is derived from Dynamic Programming, a method developed by Richard Bellman in the 1950's.

Dynamic Programming is, effectively, breaking a problem into sub-parts and finding the most effective way to execute each part of the problem to maximize the outcome of the whole problem. Bellman-Ford's Shortest Path graphic is usually shown to visualize the problem visualized in Figure 1 (Bellman, R. 1966; Broadie, M. 2014). It is a systematic approach for thinking ahead, a skill needed in playing golf on a high level.

Dynamic Programming has also been used in other sports to gain competitive advantage. Most known sports use context of Dynamic Programming is baseball. Batting Average, On-base Plus Slugging Percentage and Ultimate Zone Rating are statistics used in baseball that have all been derived from Dynamic Programming (Broadie, M. 2014; Hirotsu & Wright, 2004).

The use of Dynamic Programming was made famous in popular culture by the 2011 movie, Moneyball, which is based on real life events. It starred Brad Pitt acting as the general manager of the Oakland Athletics who is trying to gain competitive advantage against richer clubs with more effective statistical analyses in hiring players.

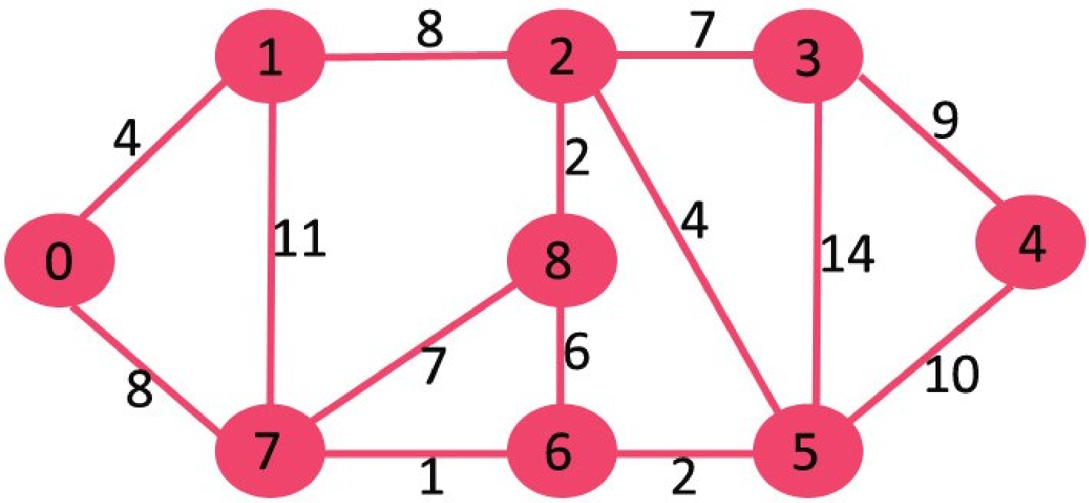


Figure 1. Bellman-Ford Shortest Path

Playing a hole of golf, effectively, is solving a problem as shown in the Bellman-Ford Shortest Path graphic. Hitting a tee-shot into a position from which the following, approach-shot, is as effectively and easily executable as possible, to make the putt on the green as easily executable as possible. A continuous deterministic process is a continuum of decisions to be made at different points of it (Bellman, R., & Dreyfus, S. 2010).

From Strokes Gained perspective, this leads to a performance measure in which the progress to the hole is measured not in yards but in the decrease in the average strokes to hole out (Broadie, M. 2014). The Strokes Gained value of a single golf shot is the decrease in the average number of strokes to hole out, minus one to account for the stroke the player has taken (Broadie, M. 2014).

To exemplify, each year on the 3rd hole of the Masters tournament, players face choice whether to lay up or go for the green. The hole is 350 yards long, which, in ideal conditions is drivable. Drivable means that a player only needs to hit one shot in order to reach the green on a par 4 hole.

The decision players are facing is whether to try and hit it close or hit it short of the four fairway bunkers. The former alternative giving the player a short pitch shot to the uphill green and the latter alternative leaving the player with a longer approach shot but taking the pressure off hitting it into the fairway bunkers.

From Strokes Gained perspective, the former alternative will more likely yield relative gains on Strokes Gained Off the Tee statistics and cause a relative loss to Strokes Gained Approach statistics, whereas the latter alternative is more likely to yield relative gains on Strokes Gained Approach and cause a relative loss to Strokes Gained Off the Tee. The two alternatives are illustrated in Figure 2.

Scoring-wise, the longer tee shot though, will on average always yield a better result due to professional golfers' average amount of shots to hole out from a shorter distance and therefore also a better Strokes Gained yield (Broadie, M. 2014).

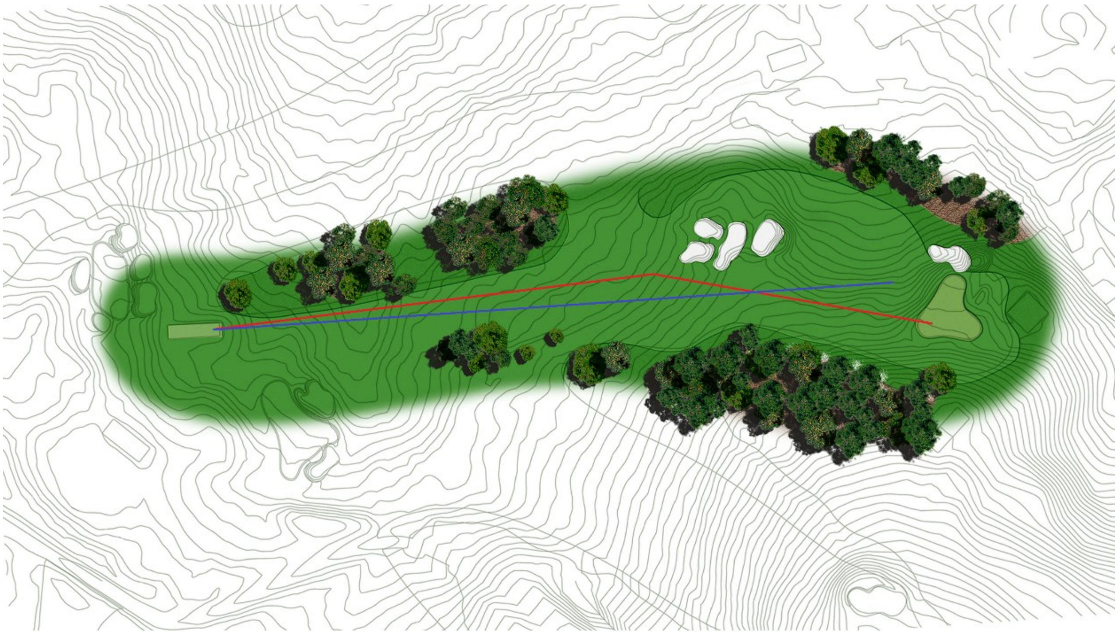


Figure 2. Augusta National's 3rd hole (The Masters, 2023).

Why a player chooses to lay up or go for the green can be explained by statistics (Broadie, M. 2014), but it is also a psychological mystery of many underlying factors which derive from players' current position in tournament play and whether a risk is worth taking (World Scientific Congress of Golf & Cochran, A. J.1992).

Statistically, the hole shown in Figure 2, yields almost double the number of birdies when the hole is cut in the center of the green versus when it is located on the left side of the green, near the green-side bunker, on the small peninsula. A birdie is a score that is one shot better than the Par of the hole. Par is an abbreviation of Professional Average Result, used to describe what professional golfers would on average score on a given hole.

In occasions when the hole is cut to the center of the green, only a small percentage of the field hit a lay-up shot from the teeing ground. On the other hand, when the hole is cut to the small peninsula, only a small percentage of the field intent to drive the ball close to the green. Strokes Gained statistics will explain why scoring from further away from the hole becomes considerably harder (Broadie, M. 2014).

The statistics of the 3rd hole from the 2023 Masters Tournament are shown in Table 1. The hole was cut to the center of the green for rounds one and three and to the left side peninsula for rounds two and four. As explained earlier, the decisions players made off the tee, depending on the location of the hole, had a stark influence on the number of birdies made (The Masters, 2023).

Table 1 Statistics from the 2023 Masters Tournament (The Masters, 2023).

Round	Par	Average	Rank	Eagles -	Birdies	Pars	Bogeys	Double Bogeys +
R1	4	4.0233	12		16	54	15	1
R2	4	4.1977	7		9	53	22	2
R3	4	3.8302	14		14	35	3	1
R4	4	4.0000	14		8	38	6	1

Decision making in golf is by no means easily explainable and varies between players, depending on their position in the tournament, their strengths, and psychological factors, how the hole in front of them presents to each individual player's mind thus making some decisions the players make not understandable through what the data and statistics inform to be the best possible decision (Davies et al., 2014), (Broadie, M. 2014).

2.2.3 Why use Strokes Gained as a measure of performance

Strokes Gained is the accumulation of fractional gains and losses of each individual shot and therefore makes comparing players' total performance easy (Broadie, M. 2014). With this system different shot types can be compared with each other. Strokes Gained is count with the formula below.

$$(Starting\ position\ Strokes\ Gained - Finishing\ position\ Strokes\ Gained) - 1 = Strokes\ Gained\ (or\ lost)\ for\ a\ shot$$

One can easily assess a given player's strengths and weaknesses if needed to. The system always uses number zero as a baseline, or a benchmark for the average of a PGA Tour player, and therefore analyzing performance by comparing a player’s statistics with the baseline, or against other player’s statistics, gives real comparable numbers for assessment.

In the 2021-2022 PGA Tour season, the Scoring Average of all players was 71,092 strokes (PGA Tour, 2023). Scoring Average is the only existing statistic Broadie (2014) considers reliable as it correlates strongly with Strokes Gained Total and its importance is discussed further along in the thesis.

2.3 Strokes Gained Putting

The definition of Strokes Gained Putting is the PGA Tour average number of putts to hole out from a given distance, minus the number of putts taken. Strokes Gained of an individual putt is the decrease in the average putts to hole out minus one, to account for the stroke taken. Strokes Gained Putting utilizes a Dynamic Programming approach, and analyzes putting as a multistage problem (Broadie, M. 2014). On average, a professional golfer takes 29 putts per round (PGA Tour, 2023). That is the most of a single type of shot a player takes on a single round.

For a long period of time, putting was considered as the most important factor in one’s success on the golf course. Players’ success on the greens was measured through two statistics: Greens-in-Regulation (GIR) and number of putts taken. These two statistics have little or nothing to do with one-another.

$$GIR = PAR\ of\ the\ Hole - 2\ strokes$$

If a golfer is regularly on the green in regulation, he is most likely to score well. GIR therefore is a measure of ball-striking, which means all other shots but putting (Broadie, M. 2014).

Broadie’s (2014) study concludes that putting, on average, plays a part of 15% of a golfer’s scoring advantage with 85% of the scoring advantage coming from the rest of the shots. On average the contribution of putting of the winner of a tournament is 35 % while the off green play accounts for 65% (Broadie, M. 2014). Therefore, putting well contributes to winning greatly, but still only a bit over a third of the total gains a player has gotten.

Experiments in golf putting demonstrate that skilled performance is streaky. The tendency for outcome sequences to form streaks was greatest when the task difficulty was such that about half the trials were successful (Gilden & Wilson, 1995). This means, that a player's putting performance is highly volatile.

On tournaments Broadie (2014) studied from 2004 to 2012, winning a tournament with negative Strokes Gained Putting occurred 14 times over 315 tournaments.

On the other end of the spectrum, winning a tournament with negative Strokes Gained off-green play, only occurred on 2 instances over 315 tournaments. This goes to show, that winning in spite of putting, or with worse putting, is possible, but a player is highly unlikely to win with below-average off-green play (Broadie, M. 2014).

2.3.1 Volatility of putting and its influence on short term statistics

On average, the top 40 players gain a bit over 1 stroke per round, of which 0,15 strokes come from putting, whereas on winning occasions, solely on putting, the winner gains a bit over 1 stroke. On the other hand, the overall average Strokes Gained number of a tournament winner being 3,5 strokes per round. It means that percentwise, putting's volatility plays a role in winning, but does not alone, or at least is highly unlikely alone, to guarantee a good week, much less a win (Broadie, M. 2014).

Another notion of putting is that Strokes Gained gives an insight in how the player generally performs on the greens but can be very volatile during small sample size. As an example, Tiger Woods win at Bay Hill in 2008. He gained 1 stroke per round on putting during the 4 days of competition. On the last hole he faced a 24 feet putt that he needed to hole to win the tournament outright. He managed to hole that putt to win. On that one successful putt he gained 0,9 shots (Broadie, M. 2014), (PGA Tour, 2023).

This means that the successful putt on the last hole steered his Strokes Gained Putting number from 0,1 to 1,0. During that tournament, his off green play accounted for 2,4 strokes gained per round. Had he not holed the putt, the Strokes Gained number would have been significantly lower, only because of putting and more importantly, because of just one putt, which at a correct time also secured him winning the tournament.

This also goes to show, that margins of winning a golf tournament are small and a well-timed great shot will affect the results, and Strokes Gained numbers, starkly.

Clearly, the PGA Tour statistics do not explain why some players, when in a position to win can make crucial up and downs and hole vital putts. Other psychological factors, besides the player's ability to handle competitive pressure, may also differentiate between successful and unsuccessful golfers (World Scientific Congress of Golf & Cochran, A. J.1992).

Henceforth, one would assume that betting companies, when giving lines, are looking at the average total Strokes Gained number of a player, or the total number minus putting, which Broadie (2014) has studied to be of lesser importance.

2.4 Strokes Gained Off the Green

The problem with assessing a player's performance using traditional golfing statistics, is illustrated exceptionally with the 2012 PGA Tour's Player of the Year Award recipient's statistics.

He won four tournaments during that season, one of which was a major, he won the money list, he had the lowest scoring average and yet somehow, he was not ranked inside the top 50 in the three main golfing statistics used at that time: Total Driving, Greens-in-Regulation and Strokes Gained Putting (Broadie, M. 2014).

As explained before, putting is to be disconnected from other shots. Total Driving and Greens-in-Regulation, though, give a very unclear picture of how the player has really performed because they mix putting and off green playing.

When that same successful player's performance is studied through Strokes Gained statistics, it gives a much clearer understanding for why that player had been as successful as he was during that season.

His Strokes Gained Total ranking was first, which is consistent with the lowest scoring average he had during that time span. He was ranked second in Strokes Gained Off the Tee, second in Strokes Gained Approach, 35th in Strokes Gained Around the Green and 73rd in Strokes Gained Putting (Broadie, M. 2014), (PGA Tour, 2023).

The influence of Strokes Gained Off the Green on a player's performance will be explained in the following chapters.

2.5 Strokes Gained Driving

Strokes Gained Off the Tee measures players' performance on tee shots on par-4 and par-5 holes. A player on average takes 14 of these shots during a round of golf (PGA Tour, 2023). It assesses player's abilities in hitting it long and straight. The straightness aspect of it assesses the player's ability to hit a fairway, which is a crucial part of player's ability to hit better shots to the green (Broadie & Ko, 2009). Hitting a fairway increases the likelihood of a player to hit it closer to the hole drastically (Broadie, M. 2014).

2.5.1 Strokes Gained Driving's influence on scoring.

Ability to score is a combination of hitting it long and straight but Broadie and Ko (2009) illustrate that Directional Accuracy contributes to scoring more than hitting it long. From 2004 to 2012, the ten best players in Strokes Gained Driving averaged Directional Accuracy of 3,27 degrees, while the average of the 40 best players in the same category averaged a Directional Accuracy of 3,3 degrees.

The Strokes Gained Driving average respectively for the ten best players being 0,75, and 0,54 for the 40 best players (Broadie, M. 2014). The average driving distance of the top 10 players was 294 yards, while the same average for the top 40 players was 289.

A small difference in Directional Accuracy can be seen yielding a gain of 0,21 strokes per round meaning an average difference of 0,84 strokes during four rounds of tournament play between a top 10 and a top 40 player. Keeping in mind, the average of the PGA Tour being 0,0 strokes gained per round, means the top 10 players gain 3 strokes on the field per tournament on average, and the top 40 players on average gain 2,16 strokes on the field respectively. Small gains cumulating for a long period of time will yield better results in the long run.

Broadie (2014) illustrates the importance of driving it long by comparing a player who steadily drives the ball the same distance and a player who has the ability to drive it longer when needed. Broadie (2014) shows that the player gains more shots by driving it longer than he loses by driving it shorter. This sort of effect is called nonlinearity.

As illustrated also in Figure 2, how to play the 3rd hole at the Masters tournament, the average driving distance and the percentage of fairways hit is a bad assessment of a player's performance. Therefore, Strokes Gained Driving trumps these older and long used statistics.

2.5.2 Nonlinear effect in driving.

Jensen's inequality states that averages are flawed when the effects are nonlinear (Jensen, 1906). Broadie (2014) uses a joke to describe this phenomenon: If one sticks his head to oven and feet to a refrigerator, on average that person will feel comfortably warm.

What this means in golfing terms, is that each drive must be measured individually to be able to use it as a metric of true performance. A longer, straighter or a shorter, crooked drive yields drastically larger gains or losses than an average drive. The further away from the average, the bigger the gains or losses will be (Jensen, 1906). Jensen's inequality plays an even larger role on Strokes Gained Approach.

Broadie's (2014) simulation shows that a 20-yard gain in driving distance for a player already longer than PGA Tour average, will yield a gain of 0,75 strokes per each round. That might not sound much, but during four tournament rounds, it will cumulate to 3 strokes gained against the average player. This same 20-yard gain further explains the example shown in how to play the 3rd hole at the Masters Tournament.

On average, Strokes Gained Driving contributes to 28% of the top 40 players' scoring advantage. As discussed earlier, putting, on average contributes to 15% of the player's scoring advantage. Therefore, Strokes Gained Driving can be seen twice as important to scoring well compared to Strokes Gained Putting. (Broadie, M. 2014)

2.6 Strokes Gained Around the Green

Strokes Gained Around the Green is used to measure shots taken from under 100 yards to the hole except for putts (Broadie, M. 2014). Traditionally, the statistics used to perform analysis on players' performance near the green were Scrambling, sometimes known as up-and-down percentage and Sand-save percentage.

As Broadie (2014) explains, these statistics do not tell whether a player manages to get an up-and-down or a sand-save because of his good playing from off the green, or because of good putting. Strokes Gained Around the Green is not only used when a player misses a green in regulation but also on shorter shots on par 5 holes after the player has not reached the green in two shots.

2.6.1 Strokes Gained Around the Green's influence on scoring advantage.

On average during a round, professional golfers hit 10 shots belonging to Strokes Gained Around the Green category. According to Broadie (2014), this statistic accounts for 17% of a top 40 player's scoring, 0,19 strokes gained on average between those players. Strokes Gained Around the Green is slightly more important for a player's scoring advantage than Strokes Gained Putting.

These types of shots occur most often when a player has not hit an optimal shot belonging to Strokes Gained Approach category.

2.7 Strokes Gained Approach

Strokes Gained Approach measures all shots taken from over 100 yards to the hole except the tee-shots on par 4 and par 5 holes (Broadie, M. 2014). A professional golfer, on average, hits 18 shots belonging to Strokes Gained Approach category per round, which is understandable due to the fact that there are 18 holes on a golf course.

Strokes Gained Approach combats the problematics the older statistics that are used to measure approach performance had. The problems are mainly related to the linear approach Greens Hit and Proximity to Hole statistics had in evaluating performance. The nonlinear approach and design that Strokes Gained has, gives a clearer picture how well a player hits his approach shots to the green.

Broadie's (2014) study concludes that of the four Strokes Gained measurements, Strokes Gained Approach is the single most important factor in a player's ability to score well. The reason for this derives partly from putting statistics and more specifically, the probability to hole out from a given distance.

The data collected for Broadie's study states that Strokes Gained Approach contributes 40% of a player's ability to score well. That number trumps the importance of the influence on scoring advantage of the other Strokes Gained categories clearly.

Through 2004-2012, the player ranked 120th in Strokes Gained Approach from 150 to 200 yards from the hole on average hits the shot to 30 feet from the hole. During that time span, that same player won three times on the PGA Tour. The player who is first in the same category, on average hit his shot to 27 feet from the hole from that same distance and won 39 times from 2004 to 2012 (PGA Tour, 2023).

To give the reader even better perspective on importance of small margins, the PGA Tour top 40 players' average from that distance is 29,4 feet from the hole and the average of all PGA Tour players is 31 feet from the hole (Broadie, M. 2014). Relatively the players are close to each other, but stark differences start to show when time periods get longer.

The player with 39 wins between 2004 and 2012, is Tiger Woods, who also is the leader of Strokes Gained Average and Strokes Gained Approach. Approach shots from any given distance can be picked and he will be the first on that list. Woods' success in real life gives support to Broadie's (2014) theory of the importance of Strokes Gained Approach.

From 2004 to 2012 Tiger Woods, on average, gained 0,4 strokes more than the second-best player of Strokes Gained Approach metric. On average, the influence of Strokes Gained Approach on a player's performance was said to be 40%. On Tiger Woods' instance, that number is 46%, which means that he gained even more on his good approach play (Broadie, M. 2014), (PGA Tour, 2023).

Furthermore, no player in the top 10 of Strokes Gained Average is ranked worse than 30th in Strokes Gained Approach. In contrast, the same number for Strokes Gained Driving is 198th, in Strokes Gained Around the Green the worst ranked is 124th and in Strokes Gained Putting 193rd.

To elaborate even further, of the top 10 players in Strokes Gained Average metric, eight are also ranked in the top 10 of Strokes Gained Approach which can also be seen when counting the average position of players.

The top 10 of Strokes Gained Average metric are on average ranked 54th in Strokes Gained Driving, 9,54th in Strokes Gained Approach, 37,90th in Strokes Gained Around the Green and 89,27th in Strokes Gained Putting.

These averages give a slightly different picture of importance because the order of Strokes Gained Driving and Strokes Gained Around the Green has turned around. One must remember, though, that a player hits, on average, more shots that fall into Strokes Gained Driving and henceforth its importance is greater than that of Strokes Gained Around the Green (Broadie, M. 2014).

2.7.1 Why a player needs to hit it close to the hole

The hole-out percentage for a professional golfer decreases quite harshly the further away from the hole the player's approach shot ends.

Professional golfers are expected to hole 50% of their putts from 8 feet (2,43 meters). From 10 feet (3,04 meters) the percentage drops down to 40, and respectively from 15 feet (4,57 meters) the percentage drops down to 23.

Probability to hole out a putt increases hugely from inside 8 feet. Those percentages are:

58% from 7 feet (2,1 meters)

66% from 6 feet (1,8 meters)

77% from 5 feet (1,5 meters)

88% from 4 feet (1,2 meters)

96% from 3 feet (0,91 meters).

From Strokes Gained perspective, a missed put from 3 feet (0,91 meters) will make a player lose 0,9 strokes, whereas a missed putt from 10 feet (3,04 meters) costs a player 0,4 strokes. This, on the other hand, means that there are not many strokes to be gained on short putts, but the player already has gained strokes from his approach shot which has ended up close to the hole (Broadie, M. 2014).

2.8 Summarizing Strokes Gained and its parts.

Strokes Gained Approach clearly is the single most important factor with its 40% contribution to scoring. Henceforth it is chosen for the metric which correlation to odds is to be studied along with Strokes Gained Average. While Broadie (2014) shows that putting is only slightly of lesser importance in winning occasions with its 35% influence in scoring advantage in comparison to Strokes Gained Approach, it will not be chosen as a metric for this study because the stark difference in influence it has on average performance and winning performance.

Jensen's Inequality's importance as a phenomenon shows clearly when illustrated by four players' yearly earnings on the PGA Tour (Table 2). In 2012, Player A was 10th on the season long money list having earned 4 million US Dollars, his scoring average was 70,00. Player B respectively had earned 1,6 million and had a scoring average of 70,75. Player C earned 360000 US Dollars and had a scoring average of 71,50. Player D earned 335000 Dollars and had a scoring average of 72,20 (PGA Tour, 2023). On average, the players earned just over one million US dollars during the 2012 PGA Tour season (PGA Tour, 2023).

Table 2 Importance of small gains in longer time periods.

2012 Season	Scoring Average	Earnings	Money list position
Player A	70	4.000.000	10th
Player B	70,75	1.600.000	50th
Player C	71,5	335.000	170th
Winner	69,63	8.000.000	1st

To contextualize furthermore, the player who won the 2012 money list, earned 8 million US Dollars (PGA Tour, 2023). He had a scoring average of 69,63, only 0,37 strokes better than player A, but earning double the money and winning three tournaments more. Small gains in Strokes Gained Statistics in long run, prove themselves to be important.

2.8.1 Delimitations Strokes Gained has when used for betting purposes.

As Broadie (2014) has calculated, winning a golf tournament approximately requires that the player gains 3,7 strokes per round against the field. Therefore, a player with a higher Strokes Gained Average baseline, needs to perform only slightly better to win than an average player, who needs to outperform himself to have a chance to win. Putting has proven to be highly volatile (Gilden & Wilson, 1995), which is also shown by Broadie's (2014) study.

Tiger Woods needs to elevate his game in order to win by gaining 1,3 more strokes due to his baseline of Strokes Gained being 2,4, whereas the player ranked 40th in Strokes Gained Average from 2004 to 2012, on average needs to elevate his game by 2,9 strokes. A player is no going to hit it 20 yards further overnight, due to physical limitations humans have, but can with help of good putting win a tournament. However, it is highly unlikely to be able to successfully estimate when a player is going to putt better, or worse, than normally.

2.8.1.1 Geographical influences on performance.

Streakiness in performance, especially in putting, has proven to be the key in winning a tournament (Broadie, M. 2014), (Gilden & Wilson, 1995). Studying a player's previous performance on greens on a given course can be of some help, but that is highly influenced by weather conditions, which vary yearly. Strokes Gained still manages to give picture on who is likelier to win because he does not need to elevate his performance from his average drastically to win.

Another problem in predicting on-green performance in major tournaments is geographical. But for Masters Tournament, the courses change yearly at The Open Championship, PGA Championship and US Open. Furthermore, Broadie (2014) adds that different players have been shown to perform better on different grass types. Rana's (2016) study supports Broadie's findings on how putting on Poa Annua greens is different to putting on Bentgrass greens or Bermuda greens and that different players perform better in different conditions.

Henceforth, Strokes Gained will give information on how players have performed recently, and it can be scaled to various time periods to study performance, but it is not all-knowing and does not account for humane errors and psychological mishaps. Psychological factors cannot be left out of the equation when trying to estimate performance (Finn, 2008). As Jensen's Inequality states that averages are flawed when the outcomes of two different acts vary harshly (Jensen, 1906), one must keep in mind that a player might have his worst performance or his best performance of the season at any given date, and at that point, averages are unimportant.

2.9 Sports betting

In general, when we talk about sport or football betting, we mean the process of forecasting the result of a game and place bets according to a prediction. Specifically, in order for a bet to exist there should exist two sides: the bookmakers who determine the prices and the players who bet in various sports based on the given odds. Depending on the result of the game, one of the two – the bookmaker or the player – wins and the other loses (Ioulianou et al, 2011).

Sports betting is a one-time occurrence where an outcome which is the subject of a wager either occurs or does not occur (Malaska & Virtanen, 2007).

2.9.1 The betting market

The betting market comprises of many bookmakers who determine odds, and of the players who are setting bets and wagers.

According to Xu (2011) and Makropoulou & Markellos (2011), a bookmaker has one single mission; To make profit on the odds they provide for the public to play. This drives them to set the odds high enough to be competitive, but low enough that betting on them is not profitable. Therefore, though the posted odds may not reflect the bookmakers' true probabilistic beliefs, they can still be viewed as probabilistic assessments of a sporting event's outcome, or, in other words, as forecasts (Štrumbelj & Šikonja, 2010). They continue by declaring that the actual value of the bet is known once the uncertainty has been resolved (i.e., the actual outcome of an event is known).

As long as the bettor preferences and perceptions are unbiased, bookmakers do best by setting informationally efficient odds that reflect the true winning probability of the underlying event. Otherwise, bookmakers can sustain large losses if bettors are able to recognize and exploit the biased odds (Levitt, 2004), (Flepp, Nüesch & Franck, 2016). In the presence of sentimental bettors who prefer bets with particular characteristics and who do not necessarily choose the bets with the highest expected return, optimal bookmaker pricing becomes more complex.

Popular examples of bettor sentiment include the optimistic/perception bias (Levitt, 2004). That phenomenon causes bettors to overrate the winning probability of certain teams, and the loyalty bias which prevents bettors from betting against the team, or player, they support. Bettor sentiment leads to an asymmetric volume demand even when the bookmaker odds reflect the true winning probability of the underlying event (Flepp, Nüesch & Franck, 2016).

2.9.2 Odds

Odds are a number supplied by a bookmaker. By that number, a player will decide on whether to bet on that wager or not (Virtanen & Vänni, 2005). For example, if a player is offered odds 2,5 and he sets a wager of 10 euros, the player has the possibility to get 25 euros of which his winnings are 15 euros.

In a classic home win, draw or away win wager, with given odds, the bookmaker tries to sell his likelihood for these three different events happening, with his own commission subsidized from it. The bigger the commission of the bookmaker, the smaller the given odds. Commission of the bookmaker is count through 1 – the return of interest of the bookmaker.

2.9.2.1 Return percentage.

Return percentage for a player, is the percentual proportion of a player's gross win of wagers he has made.

The theoretical return percentage for bookmaker can be count through the reciprocal number of all the different odds given for a wager, a counting a sum of them (Virtanen & Vänni, 2005).

For example, a match of golf is being played between Tiger Woods and Luke Donald. Woods' odds are 1,3 and Donald's odds are 3,2. Reciprocal numbers of the odds are 0,769 and 0,3125. Therefore, the sum of those is 0,769 + 0,3125 = 1,0815. The return percentage of a bookmaker is the reciprocal number of 1,0815, which is 1 / 1,0815 = 0,924, which means 92,4 percents.

2.9.3 Determining probability

Determining probability, or likelihood, can be done from a subjective or an objective point of view (Ali, 1977). The selection between these methods depends on the time frame from which the likelihood for a certain event is surveyed from.

2.9.3.1 Subjective probability

Subjective probability according to Ali (1977) is what the people, the crowd, think about the likelihood of different outcomes. In his study, Ali (1977), uses a methodology based on subjective and objective probability for horses to win a race. Subjective probability in his study therefore means which horse the crowd favors the best. The subjective winning probability π for a horse *h* can be determined by the share wagers *X* of the total turnover bet on horse *h*. (Ali, 1977). Malaska & Virtanen (2007) add that subjective probability could be translated to a result of a person's thought process which cannot be measured objectively.

$$\pi_h = x_h / w$$

Malaska & Virtanen (2007) add that subjective probability can be explained as a result of a person's thought process which cannot be measured objectively. They add that subjective thought process does not mean irrationality but rather a different point of view. Galavotti (2001) states that all probabilities are subjective and there are various ways to evaluate probabilities. He adds that estimating probabilities is a complex process which includes various simultaneous variables.

Because there is no correct means to estimate subjective probabilities it is likely that a bookmaker and a bettor have different estimates of the subject in matter. When making a wager, a bettor might put more weight on a piece of information, and that same piece of information might be totally ignored by a bookmaker when giving out odds.

2.9.3.2 Objective probability

Defined by Ali (1977), the objective winning probability of a horse is defined to be the proportion of times the horse wins, when the race is repeated an infinitely large number of times. Thus, a race can be taken as a binomial trial with this objective probability for the winning outcome. In a race the total of the objective probabilities is one.

A horse race according to Ali (1977) can be thought to be a binomial probability test, where the objective probability estimates a horse's likelihood to win a race. The sum of each competitor's objective probability is one. Both subjective and objective probabilities vary according to horses in a single start. In a single start, the horses racing, have different probabilities of winning. Completely identical starts are a very rare occurrence. Henceforth, estimating objective probability is impossible, because the same start is only raced once (Ali, 1977).

The same problem applies to golf tournaments thoroughly because every field is different.

Objective probability can be defined for games of chance, such as roulette or coin toss, because every variable is known beforehand. Therefore, for sports betting it is impossible. The nature of sports betting is different to games of chance which do not possess natural variability (Figlewski, 1979).

2.9.4 The different types of bettors

Makropoulou & Markellos (2011) consider the players making wagers to be of three different types, based on how much information they possess. There is casual or uninformed, so-called noise bettors, expert or informed bettors and privately informed bettors or insiders.

Noise bettors do not bet on the basis of information, but randomly. They could include those that bet in favour of a team they support, or, against a team they dislike. The betting process of those players is random.

The informed bettors observe the public information flow and make their decisions based on the news they receive by the market. Compared with noise bettors, informed bettors may be viewed as individuals that are skillful in processing information.

Insiders bet on the basis of private information they possess, unknown to the bookmaker, whereas informed bettors do not hold any private information and place their bets according to public information (e.g., forecasts by analysts, news about players' injuries, weather changes, etc). The bookmaker is assumed to be an informed agent, possessing the same level of skill as informed bettors (Makropoulou & Markellos, 2011). Moreover, a change in odds during the betting period can drive change in players' betting behavior.

The founder of European Sports Betting Consultants, Jorma Vuoksenmaa, in his Guide to Sports Betting (1999), recognizes four different types of bettors on the market.

Bettors who make wagers for fun and to gain a rush of excitement and adrenaline. They play many different types of wagers without a systematic approach in their betting process.

Bettors who consider making wagers a hobby. They estimate being capable for profitable betting through simple methods and strategy. A driving factor is the sensation of making a correct wager and more prominently, getting a wager correct through the sense of knowing how to make correct wagers. Normally, a hobbyist does not make profit in long run. Hobbyists and for-fun players comprise the biggest percentage of players in the market.

Semi-professionals get close to breaking even. Some make profit. They have managed to get rid of the mistakes the hobbyists and those who bet for fun make. Their hobby pays for itself.

Professionals, who comprise the smallest percentage of players in the market. They make profit and living out of making wagers. They have a systematic approach and methodology in their behavior.

	A rational player	An emotional player
The style of play	Thoughtful, against subjective thinking	Impulsive
Gathering information	Gathering information from many sources	Fast, feeling based decisions
Methods	Evaluates probabilities	Goes for winners
The most important aspect of betting	Gained value	Reliability
Goals of betting	Profit	Rush of excitement
Amount of wagers	Single wagers	Parlays of many kinds
Betting	In relation to funds and probabilities	All or nothing
Time frame of wagers	Long term	Short term
Relationship towards winning	Knows it will come at some point	"I told you so"
Relationship towards losing	Losing is an essential part of betting	"I told you so"
Control of betting	Continuous bookkeeping	No bookkeeping

Figure 3. The two types of bettors according to Suuraho (2002).

In figure 5, Suuraho (2002) presents the two types of bettors: Rational players, and emotional players. The criteria are based on how those two types of players act on the betting market. Suuraho adds that the behavior of the rational player is comparable to a person acquiring durable goods whereas the behavior of the emotional player is resemblant of a person acquiring daily goods.

Suuraho's (2002) study also exhibits the purchase processes of the two player types. Noteworthy is that the rational player's process starts from gathering information and evaluating alternatives. That player's process is not necessarily moving forward to identifying a need but can be cancelled if the wager and odds are not found profitable. In that case, the process starts from the beginning.

The emotional player's process starts from identifying a need, which might be that the player, for example, has decided to go to a pub to watch a live football game, henceforth necessitating a need to make a wager to make the watching the match more interesting. Gathering information and evaluating alternatives play a relatively small part in the betting process of the emotional player.

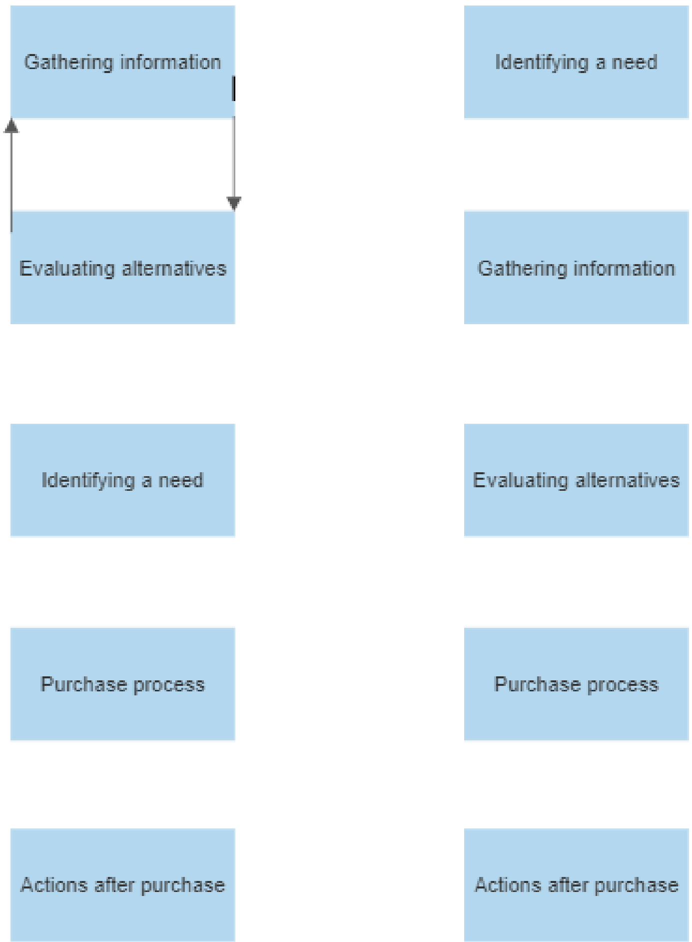


Figure 4 The betting processes of the two player types (Suuraho, 2002).

Assessing information available from Strokes Gained data in conjunction with the data from the Official World Golf Ranking ought to be a vital part in case a rational player intends to bet on golf wagers. The emotional player, according to Suuraho and Vuoksenmaa, will likely bet on a player he wishes to win while seeking a rush on adrenaline while watching the tournament play.

2.9.5 Favourite-Longshot Bias

The most tested bias in betting is the Favourite-Longshot bias, first observed in 1949 by Richard Griffith, a psychologist, while at horse race in Kentucky.

Griffith (1949) inspected thousands of wagers for horse races in the United Stated. He noticed that the smaller the odds of a horse, the more play volume that wager gets. Those who systematically played the favorite horse, the one with the lowest odds, won more and lost less on long term than those who bet on the other horses of the starts. Ali (1979) observed the effectivity of the horse race betting market and noticed that playing the favorite has a better expected return of interest than the other horses. That means, the favorite horses are underrated in relation to the other horses in that start. Ali observed 20247 starts on three different racetracks between 1970 and 1974 and concluded that the favorite horse was always less bet on in relation to the other horses. That implies that the Favourite-Longshot bias is apparent. Harville (1973) and Snyder (1978) have also observed the bias to be real, meaning the favorite was underplayed. They observed that the underdogs are too heavily bet on.

There are exceptions to Favourite-Longshot bias's truthfulness. Busche & Hall (1988) performed a study on horse race betting in Hong Kong from 1981 to 1986. They used same methods as Ali (1979) and noticed that the bettors in Hong Kong did not act as Favourite-Longshot bias would suggest but rather acted accordingly which meant the underdogs got less wagers and the favorites got most of the betting volume.

3 Methodology

3.1 Pearson's correlation

Correlations are conducted to measure how the betting odds move in relation to the chosen indicators, which are Official World Golf Ranking, Strokes Gained Total and Strokes Gained Approach.

Pearson's Correlation presents the strength of the association between two variables, but it does not show the strength of causation nor if the variables are affected by some other factors. In this study, the Pearson correlation method is used. The Pearson correlation method measures the correlation between the values of -1 and 1 (Nettleton, 2014).

$$r_{xy} = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum_i x_i^2 - n\bar{x}^2}\,\sqrt{\sum_i y_i^2 - n\bar{y}^2}}.$$

The value of 0 informs of no correlation.

The value of -1 informs a perfect negative correlation in which the two variables move in opposite directions.

The value of 1 informs a perfect positive correlation in which two variables move in the same direction.

Correlation informs of the linear relationship between the variables (Nettleton, 2014). The relationship between the betting odds and the three indicators shall be observed using Pearson's correlation coefficient.

Correlation and the correlation significance will not be tested between the three indicators. They are not completely independent of each other, but Broadie's study shows that Strokes Gained Approach is already known to be an important part of Strokes Gained Total having a 40% influence on it on average, and therefore a further analysis is not needed. The position of a player in the Official World Golf Ranking is known to be subjective and related to his performance in tournaments he has played, having no straightforward relation to Strokes Gained statistics. Correlation between a player's Official World Golf Ranking and Strokes Gained numbers is known not to be absolute as seen and confirmed when studying the data because the player's position in the OWGR depends on other factors, such as which tournament the player has played and how the competitors have fared in tournaments they have played.

Pearson correlation coefficient is a commonly used method in various sciences.

Smith et., al (2016) used Pearson correlation to study how golfers of different skill levels use their feet to generate clubhead speed and whether a stable, balanced position of the body correlates with better clubhead speed. Their study concluded that golfers displaying early movement of medial-lateral center of pressure to the front foot in the backswing or greater range and rate of movement to the front foot in the downswing were more likely to have higher clubhead velocity. Golfers that moved their center of gravity towards the front foot earlier in the backswing, however, were more likely to have lower clubhead velocity (Smith, et.al, 2016).

Fedorick et,.al (2012) used Pearson correlation to determine whether a golfer with a lower handicap is more prone to wrist injuries than a golfer with a higher handicap. Their study population consisted of NCAA Division 1 golfers with low handicaps and of golfers whose handicap was higher than 10. The study focuses on the differences in swing mechanics between players of different skill levels and more specifically the position of the lead wrist at impact with the ball. Golfers with more ulnar deviation incur greater loading forces on the ulnar side of the wrist that may predispose them to certain injuries (Fedorick, et.,al, 2012). The study concludes that the higher handicap golfer is more prone to "cocking or casting" henceforth diminishing the pressure the ball is hit with whereas the lower handicap golfer does not "cock or cast” the club. In the low handicap golfers, the decreased radial deviation in the lead arm during the swing and at ball contact, as well as the greater angle of descent when the forearm was parallel to the ground may increase the stress placed on the ulnar side of the lead wrist during the downswing and at ball contact (Fedorick, et.,al, 2012).

Brown et.,al (2006) used Pearson correlation to determine whether a golfing tourist is prone to treating all costs of a golf trip as bundled costs instead of sunk costs. They found a strong correlation between the distance traveled for the trip and the cost of greens fees. The data from the study indicates that most golfers, especially golf tourists, do bundle the quality costs with the intermediate costs of transportation, lodging, and food. Therefore, visitors play relatively more high-quality rounds of golf in relation to lower quality rounds of golf than do locals (Brown, et.al, 2006). They found that various factors have effect on whether a traveling golfer considers a bundle to be feasible but in general, bundle deals are preferred if structured properly, yielding financial gains to both the seller and the buyer. The study suggests that companies selling golf travel packages may use the results to gain information on how a traveling golfer is prone to behave.

Deheshti, Firouzjah & Alimohammadi (2016) in their study used Pearson correlation to research correlation between a foreign brand’s image and trust among buyers of foreign sports brands. The results of their Pearson’s correlation test showed a significant relationship between image and brand trust (r=0.68). The components of brand image (brand, services, suitability, variety, quality, and atmosphere) with the coefficients of 0.39, 0.53, 0.45, 0.55, 0.51, and 0.54, respectively, had significant correlation with brand trust (Dehesti, Firouzjah & Alimohammadi, 2016). The results of the correlation tests in their research showed that, generally, brand image has a significantly positive relationship with brand trust.

Pearson Correlation has been used in a variety of different industries and arts to study problems. It is an informative way to describe how two things might or might not correlate because of each other.

3.2 Granger causality

Granger Causality is a statistical hypothesis test invented by a British mathematician Clive Granger in 1969. It is used to determine whether a time series is useful in forecasting another time series. The notion of Granger Causality is: If lagged values of Y help predict current values of X in a forecast formed from lagged values of both Y and X, then Y is said to Granger cause X (Thurman & Fisher, 1988). Thurman and Fisher in their study used Granger Causality to estimate whether the amount of chicken is relevant when forecasting the number of eggs or vice versa.

Granger Causality studies the causality of time series in a short time interval. The null hypothesis of a Granger Causality test is that there is no causality to be found between two time series. The contra hypothesis is that the time series Y causes the movement of the time series X (Malkamäki, 1992). However, Y causing movement in X does not signify that X is a consequence or a result of Y. Granger Causality test will also depict the amount of lag, time delay, is needed for Y to cause movement in X or vice versa.

The amount of lag chosen for a Granger Causality test depends on the dataset that is being studied (Woolridge, 2015). Woolridge adds that Granger Causality only reveals the direction causality but does not indicate whether the causality is codirectional or vice versa.

Granger Causality is an established testing method in statistical research. The p-value of the test should be under 0.05 to be seen Granger caused and the higher the f-value, the better. An F-test is any statistical test in which the test statistic has an F-distribution under the null hypothesis. It is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled. Exact F-tests mainly arise when the models have been fitted to the data using least squares (Adepoju, K. A., Chukwu, A. U., & Shittu, O. I. (2016).

Due to odds being a number given after having the available statistics, this study does not research whether odds Granger cause the player's position in the OWGR or his Strokes Gained metrics. Therefore, the study intents to find out whether one of those three metrics Granger cause odds.

Narayan & Smyth (2003) used Granger causality to test whether the price of admission to spectate Melbourne Cup, a premiere horse racing event in Australia, is Granger causing the attendance to be higher or whether the income a person Granger causes admission. They observed that in the long-run both income and the price of admission are equally important Granger causers for a person to attend. They add that the authorities running the races play a big part of setting the prices and henceforth can impact attendance starkly. They note that the existing studies of attendance at performance events, and in particular professional team sports, which assume that causality runs from admission price to attendance tend to be based on a simplistic demand and supply model (Narayan & Smyth, 2003). They conclude that Few studies in the literature on professional team sport even attempt to address the simultaneity issue using instrumental variable techniques.

Hall, Szymanski & Zimbalist (2002) have used Granger causality test to determine whether a bigger payroll of a sports team tends to lead the team to perform better between 1980 and 2000. They found that while there is no evidence that causality runs from payroll to performance over the entire sample period, the data shows that the cross-section correlation between payroll and performance increased significantly in the 1990s (Hall et., al, 2002). The study focused on the MLB which is the premiere league of baseball and on the English Premier League football. Their findings show that a team's performance cannot be seen Granger caused directly in the MLB, due to it having restrictions for player trades and not having a free market, which is the case of the professional football leagues, such as the English Premier League. The performance of an English football club can be seen directly Granger caused by the bigger payroll.

Jagielski et, al. (2002) used Granger Causality test to study whether the transfer spending of the top European football clubs caused them to perform better in their respective leagues. They conclude that it has little importance in the teams' performance. The study only focused on the bigger clubs and the study does not account for the performance of the so called weaker or lesser clubs.

Dobson & Goddard (1998) used Granger causality to test whether the lagged revenue Granger causes better performance of an English football club in the future. They studied the revenues and performance of all the football clubs which maintained their League status from 1946 to 1994, after which the English Football Association altered their league system, and the Premier league became. They conclude that unless the wealth of ticket admissions is shared through all teams in the league, the bigger clubs will gain more success through lagged revenues.

Jane, San & Ou (2009) used Granger causality to determine whether "Yankee Paradox" exist and whether the relationship between salary structure and team performance by taking the dispersion of salaries as well as total payroll of professional baseball teams in Taiwan into consideration when examining the possibility that a causal link between salary structure and team performance exists (Jane, San & Ou, 2009). The study presents that a team with many average players performs better than a team with a mix of super-star and less-talented players. The results indicate that Granger causality runs only from the dispersion of salaries to team performance, but not vice versa (Jane, San & Ou, 2009).

As well as Pearson Correlation, Granger Causality is used in a variety of different problems and industries. It will give answers to even more complicated problems which require a longer scope of studying.

4 Empirical study

This chapter will introduce the exploratory collection process of the data and explain why the three different metrics were chosen for studying correlation between them. The collection criteria of the picked odds will also be discussed and explained.

4.1 Collecting the data

The data from the Official World Golf Ranking is freely accessible at their website. Strokes Gained data is freely accessible at the PGA Tour's website. Betting odds were collected from bookmakers' websites which also are freely accessible.

Data was collected by hand picking from three sources. The Official World Golf Ranking stores historical positions of players on any given week of the year from its start in 1986. Strokes Gained data was collected from PGA Tour's statistics. Betting odds were collected from sports media outlets, who store articles in their archive.

The Official World Golf Ranking and Strokes Gained metrics are not exact reflections of each other, but rather a high ranking in either of those statistics is an indicator of a high ranking in the other as well.

Odds were picked from different bookmakers with the criteria that the odds must have been given in the beginning of the tournament week before tournament play had commenced on Thursday. Earliest betting odds for major tournaments are nowadays given months in advance but those are nearly impossible to be found afterwards, and unnecessary, for this study. Odds given at the start of each major tournament week were the ones needed for the study. Those odds are the likeliest to reflect the Official World Golf Ranking that was updated the Sunday before.

Odds collected for the study were converted to decimal numbers from what are called American and British odds for the purpose to be easier to understand.

American odds are a number with a plus sign or a minus sign before it, such as +110 or -110. A plus signed number can be converted to a decimal number with the following formula.

$$+110 = 1 + \frac{110}{100} = 2{,}10$$

A minus signed number can be converted to a decimal number with the following formula.

$$-110 = 1 + \frac{100}{110} = 1{,}91$$

American odds tell directly what the percentual net profit of a set wager is but are not used in Europe, henceforth the conversion was done.

British odds are given in form of fractions. For example, a decimal number 1,27, would be given in form of 3/11 in the United Kingdom. Those are read in the following manner: Betting 11 units can yield a net profit of 3 units. These were converted for the same reason as given for the conversion of the American units.

Data pre-processing was done when singling out which Strokes Gained metrics give relevant information in studying correlation between them and odds. The other part was done when singling out odds that fulfilled the criteria above and converting them to decimal numbers.

4.1.1 Picking relevant Strokes Gained statistics.

Choosing to study Strokes Gained Total and Strokes Gained Approach statistics is because as explained in chapter two, Broadie (2014), has studied Strokes Gained Approach to be the single most important factor of Strokes Gained metrics for a player's scoring.

Strokes Gained Total on the other hand gives an overall picture of how a player performs. A player might have a starkly positive Strokes Gained Approach but in case a player otherwise, in Strokes Gained Total metric, is not performing well, the positive Strokes Gained Approach metric is of little importance.

Strokes Gained data was handpicked. The starting point of the Strokes Gained values is the first tournament of the PGA Tour season and it cumulates throughout the given season, painting a picture of a player's overall performance that season instead of focusing on shorter time periods before each of the four major tournaments.

At minimum, before the first major tournament of each season, 25 tournaments were played giving sufficient information on the players' performance. Players have not played the same number of events but setting the starting point of Strokes Gained data to the start of each season, will have given enough opportunities for each player to play a sufficient number of tournaments for the data to be reliable.

Notable is that the PGA Tour season is not in unison with the calendar year, but rather commences in September and ends in August the next year.

Coronavirus pandemic in 2020 caused an abnormal schedule for the 2019-2020 PGA Tour season. The pandemic caused the season to be paused from March 8th to June 11th. Because of that, players had already played more tournaments until the first major tournament, which was played from August 6th to 9th.

Therefore, an argument can be made whether Strokes Gained metrics should have been picked from June 11th forward because of a change of playing form the pause could have caused to some players. After looking at that alternative, it became clear that most of the players hadn't played enough tournaments after June 11th for the data to be reliable due to its scarcity and it would have caused recency-bias as well as the lack of it depending on the player.

The ranking of each player picked for the data was that to which the player had moved to after the play of the tournament preceding a major tournament had come to an end.

4.1.2 Description of the dataset

The handpicked dataset's size is 7125 data points.

It comprises of 19 major tournaments from 2018 to 2022. Each tournament's data includes the top 75 players in the Official World Golf Ranking from the start of the tournament week with respective odds, Strokes Gained Total and Strokes Gained Approach statistics.

The data includes five Masters tournaments, five PGA Championships, five United States Open Championships and four Open Championships. The Open Championship was not played in 2020 due to restrictions related to Coronavirus pandemic in the United Kingdom at that time. Henceforth only four of those are included in the data.

The lowest odd in the dataset given for a player is 8.00 and the highest is 501.00.

4.2 Studying correlation

The first correlation study was done with the whole dataset, meaning all 75 players in each tournament, to research what the overall picture is like. Afterwards the same study was performed for the top 50 players and thereafter for the top 25 players.

4.2.1 Top 75 correlation coefficients

The correlation between the player's OWGR position and odds is the only positively correlating metric, which it should be. Using Pearson's correlation, the correlation coefficient between the OWGR position and odds ranges from 0.127 to 0.777 between tournaments and is 0.504 on average.

Correlations between odds and Strokes Gained Total and Strokes Gained Approach respectively, are negative, which they should be since the higher the Strokes Gained value, the better the player has performed which should cause the odds to be lower. On the other hand, the lower the Strokes Gained value, the higher the odds should be and vice versa.

Strokes Gained Total's correlation coefficient with odds ranges between -0.685 and -0.334 and is -0.466 on average.

Strokes Gained Approach is the least correlating metric with odds amongst the top 75 players. The coefficients of it range between -0.501 and -0.179 and is -0.313 on average.

This means, that Strokes Gained Approach is the metric that correlates the least with the odds.

4.2.2 Top 25 correlation coefficients

When performing the same test to the top 25 players, the correlation relationships of the three chosen metrics with the given odds become even clearer.

The correlation coefficient between the OWGR and odds ranges between 0.383 and 0.887 with the average is 0.674.

The correlation coefficient between Strokes Gained Total and odds' ranges between -0.015 and -0.772 with the average is -0.440.

The correlation coefficient of Strokes Gained Approach and odds ranges between -0.028 and -0.669 and the average is -0.284.

A player's position on the Official World Golf Ranking clearly is the most important factor in giving odds. The correlation coefficient between the OWGR and odds rose by 0.170 compared to the correlation coefficient of the same metrics for the top 75 players.

4.2.3 Correlation coefficients for the four major tournaments

The correlation test was performed to study how the odds of each of the four major tournaments correlate with the three metrics. The test was run with the top 75 players, the top 50 players and the top 25 players.

Through all player categories, the odds of every four tournaments correlate the least with Strokes Gained Approach metric. The coefficient varies between -0.15 and -0.42 and has an average of -0.34.

Strokes Gained Total has the second strongest correlation coefficient with the odds. It varies between -0.40 and 0.63 and has an average of -0.50.

The Official World Golf Ranking has the strongest correlation with the odds. The coefficient ranges between 0.48 and 0.75 and has an average of 0.68.

4.2.4 Correlation coefficients of different players

The dataset included 23 players that appeared in the top 75 of the Official World Golf Ranking through all studied tournaments. The correlation test showed that not all players' odds correlate accordingly with the Official World Golf Ranking and Strokes Gained data.

When studying the single players and disregarding two exceptions, the Official World Golf Ranking correlates the strongest with the given odds.

Strokes Gained Total correlates strongly with the odds when the player is also ranked high in the Official World Golf Ranking.

The figures of these 23 players' statistics with respective correlation numbers and Granger Causality test results will be presented shortly.

4.3 Results

This section presents the results of the Granger Causality test performed on the 23 players as well as their respective Strokes Gained Total, Strokes Gained Approach and OWGR statistics and the results of Pearson Correlation test. The results are presented first individually by player and afterwards a summary of the Granger Causality test is presented.

4.3.1 The test results of the 23 players

The results of the players are listed in alphabetical order. The upper figure for each player represents a time series of the player's Strokes Gained Statistics and odds. The right side of the upper figure represents the odds, and the left side represents the Strokes Gained values. The lower figure presents a time series of the players positions in the Official World Golf Ranking and odds. The values on the left represent the odds and the player's position in the Official World Golf Ranking.

The results were of four different types: Either none of the metrics Granger-cause the odds, one of the metrics Granger-cause the odds, two of the metrics Granger-cause odds or all three metrics Granger-cause odds. One player had all the three metrics Granger-causing the odds whereas seven players had no metrics Granger-causing the odds. OWGR was the metric that most often Granger-caused the odds. This is because most of the 23 players who appeared in the top 75 on all 19 tournaments, were also ranked high in the OWGR, for example Patrick Cantlay.

Players such as Tyrrell Hatton, who had relatively high Strokes Gained numbers, had either of those two statistics as the Granger-causer of odds. Players who had no metrics Granger-causing the odds were a diverse group with different positions in the OWGR through the data. No specific single type of player can be singled out to be seen belonging to this group.

The values under the graphs report the players’ Pearson correlation coefficient between the odds to the OWGR, Strokes Gained Average and Strokes Gained Approach. The correlation coefficient between odds and the OWGR should be correlating positively, whereas the odds and the two Strokes Gained metrics should correlate negatively.

4.3.1.1 Patrick Cantlay

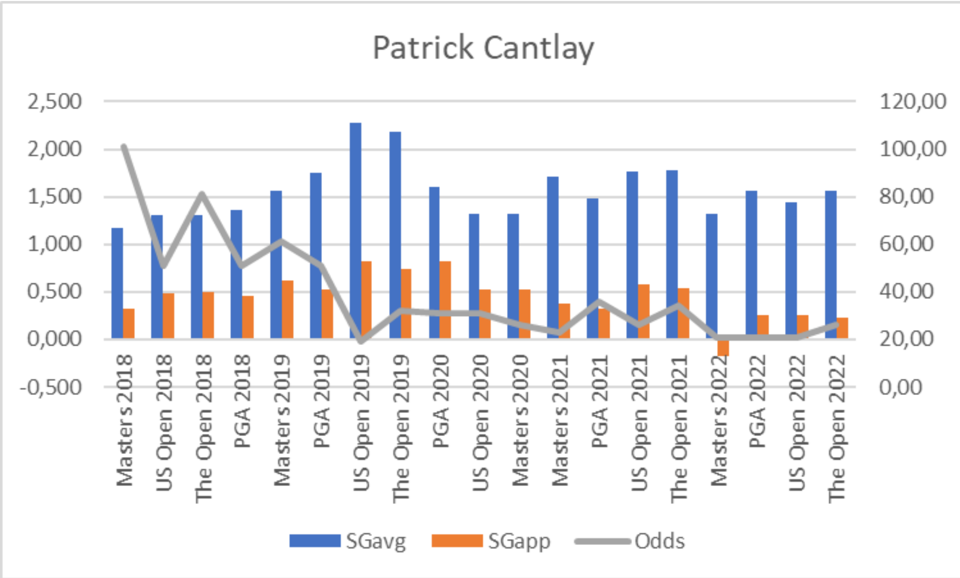


Figure 5 Patrick Cantlay's Strokes Gained statistics (Left column) and odds (Right column).

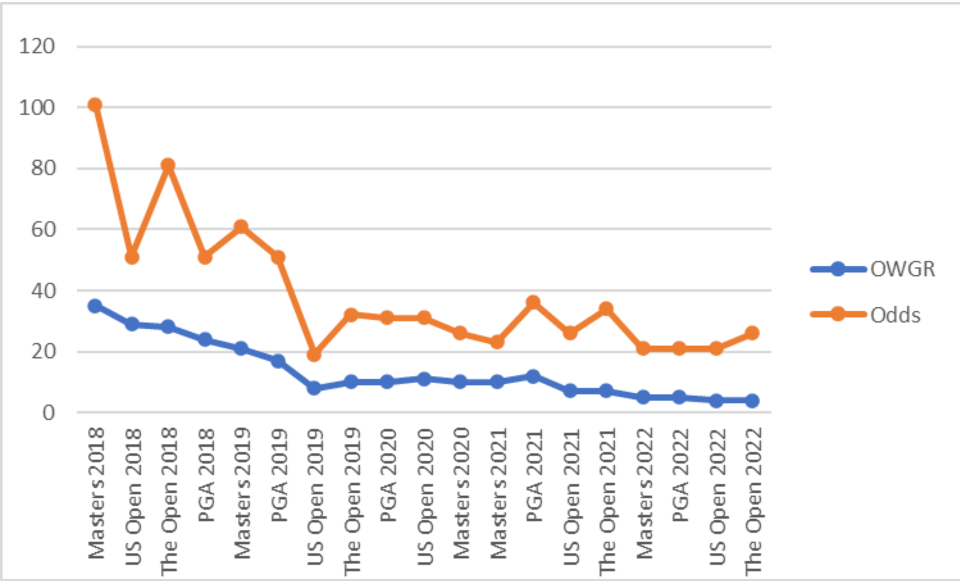


Figure 6 Patrick Cantlay's position in the OWGR and odds.

Patrick Cantlay's OWGR correlation is 0,923559 while his Strokes Gained Average correlation is -0,43565 and his Strokes Gained Approach correlation is 0,061473. His correlation statistics show that his odds heavily correlate due to his position in the OWGR and not on the Strokes Gained data.

4.3.1.2 Paul Casey

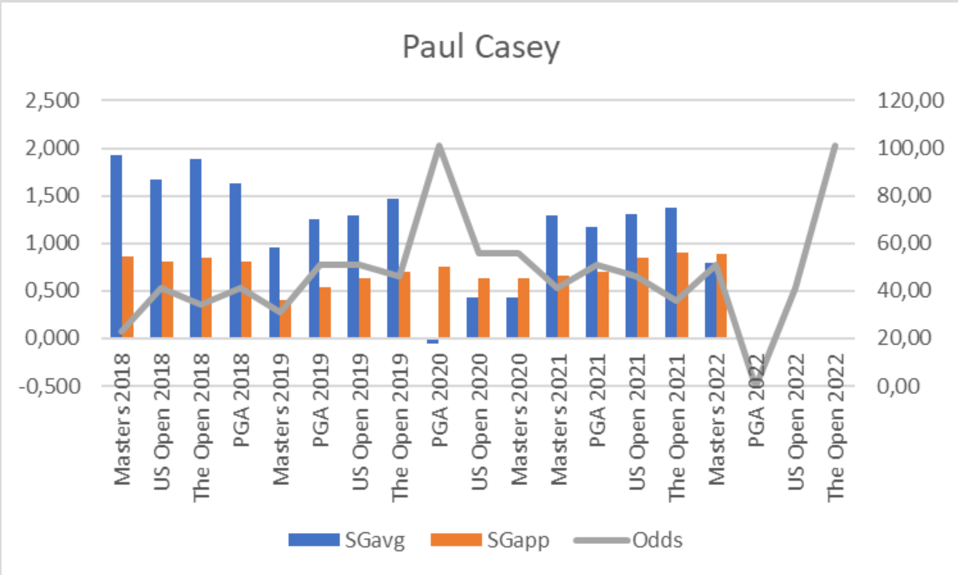


Figure 7 Paul Casey's Strokes Gained statistics (Left column) and odds (Right column).

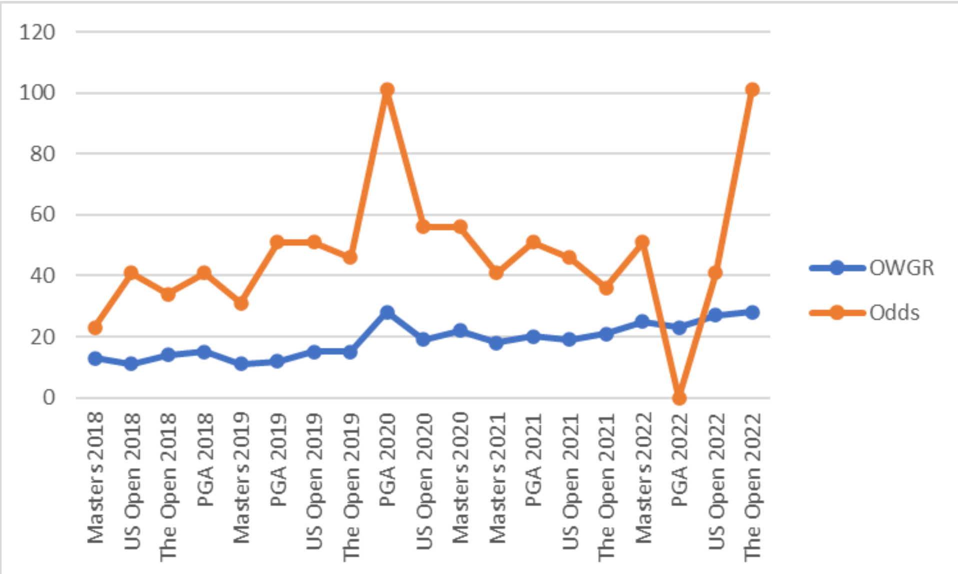


Figure 8 Paul Casey's position in the OWGR and odds.

Paul Casey's OWGR correlation is 0,488033 while his Strokes Gained Average correlation is -0,41098 and his Strokes Gained Approach correlation is -0,02583. His correlation data is not too informative but rather gives an insight to how an 'average' player is rated by bookkeepers.

4.3.1.3 Bryson DeChambeau

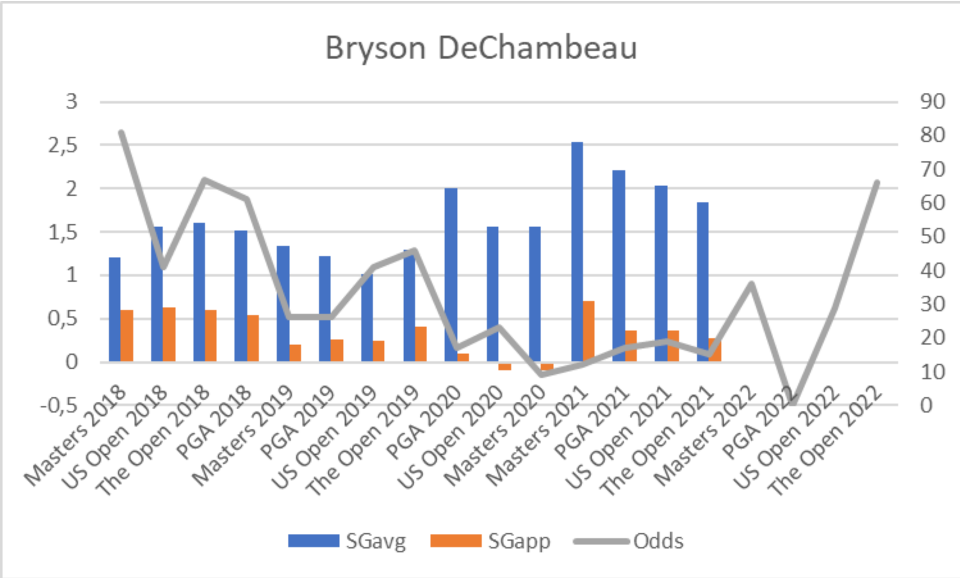


Figure 9 Bryson DeChambeau's Strokes Gained statistics (Left column) and odds (Right column).

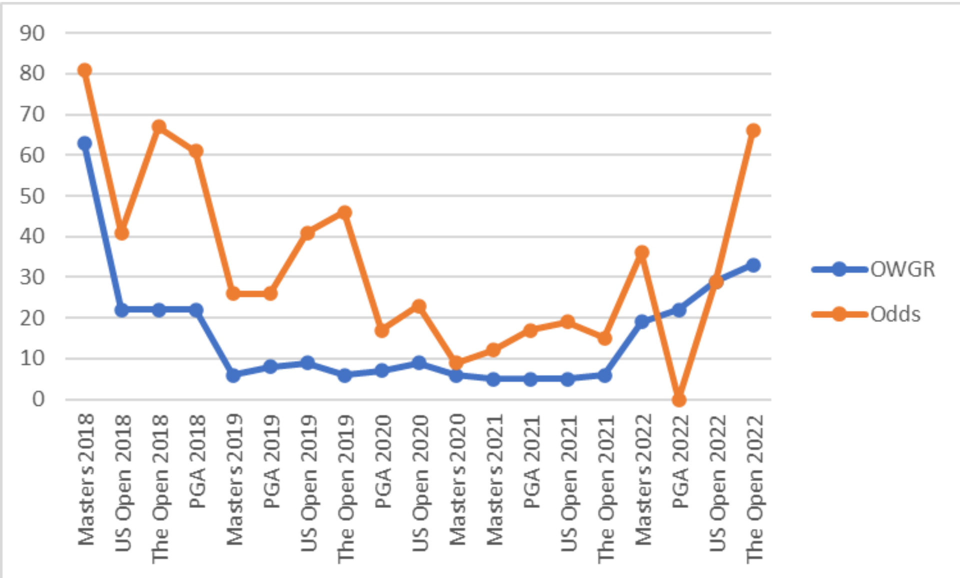


Figure 10 Bryson DeChambeau's position in the OWGR and odds

Bryson DeChambeau's OWGR correlation is 0,704386 while his Strokes Gained Average correlation is -0,21723 and his Strokes Gained Approach correlation is 0,413013. His correlation statistics show that his odds heavily correlate due to his position in the OWGR and not due to Strokes Gained data.

4.3.1.4 Tony Finau

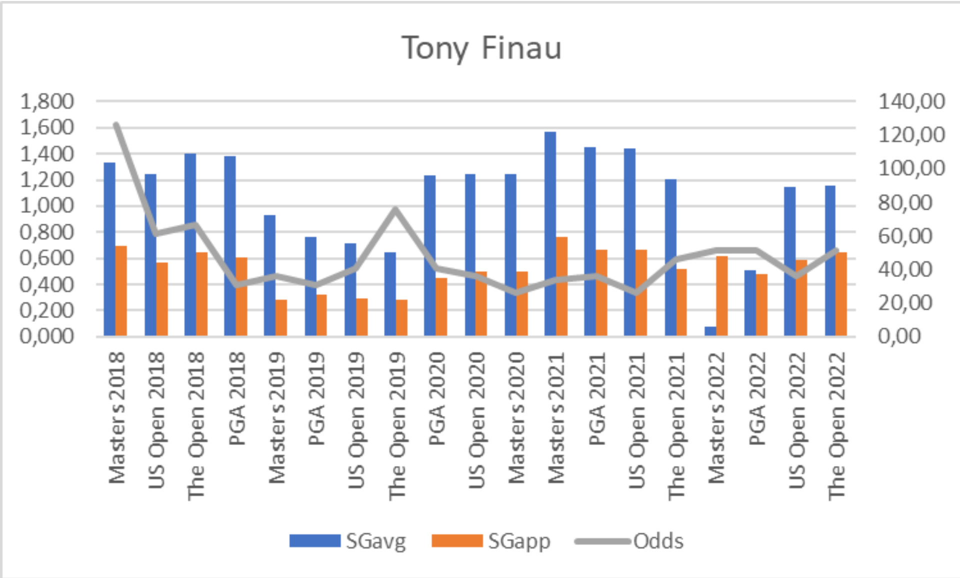


Figure 11 Tony FInau's Strokes Gained statistics (Left column) and odds (Right column).

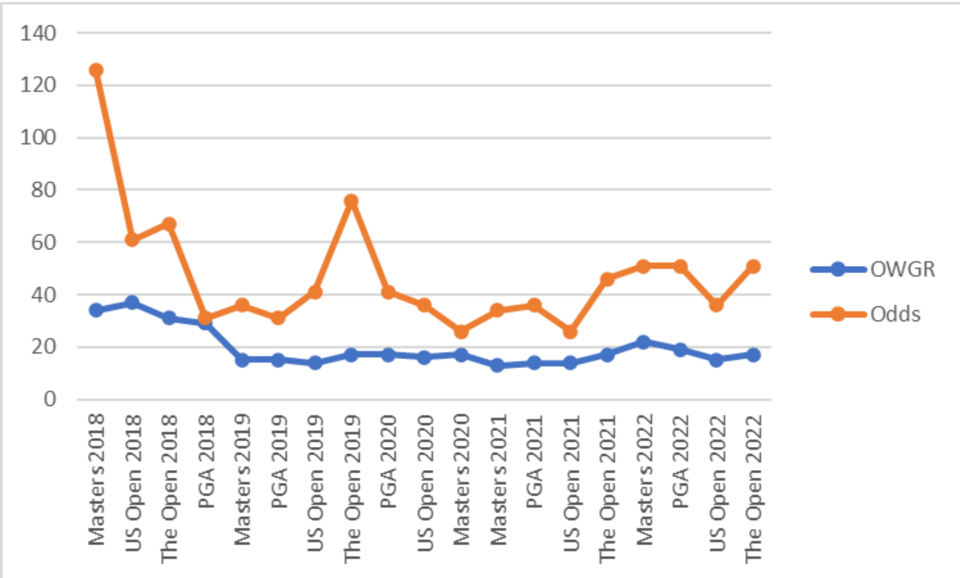


Figure 12 Tony Finau's position in the OWGR and odds.

Tony Finau’s OWGR correlation is 0,632195 while his Strokes Gained Average correlation is -0,06619 and his Strokes Gained Approach correlation is 0,141052. His correlation statistics show that his odds heavily correlate due to his position in the OWGR and not due the Strokes Gained data.

4.3.1.5 Matthew Fitzpatrick

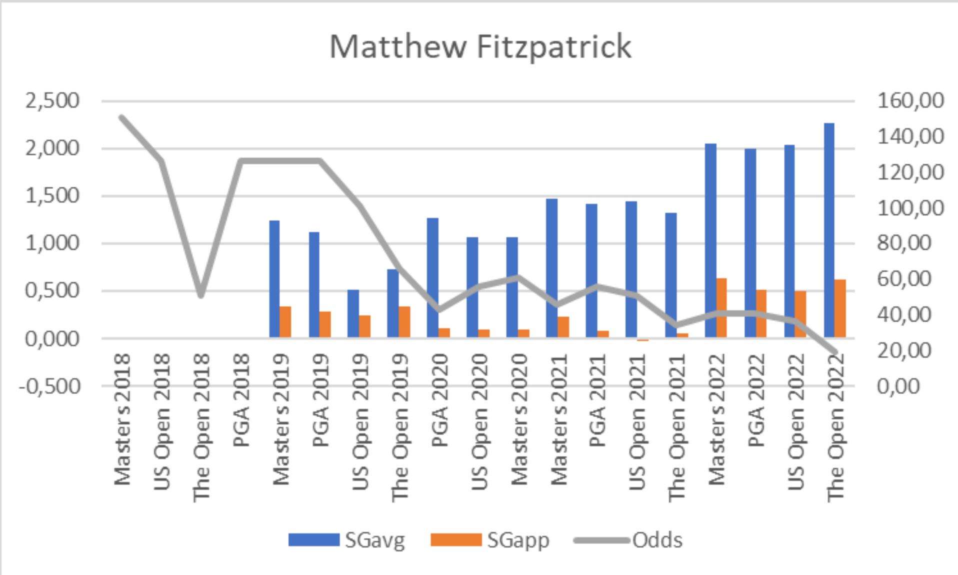


Figure 13 Matthew Fitzpatrick's Strokes Gained statistics (Left column) and odds (Right column).

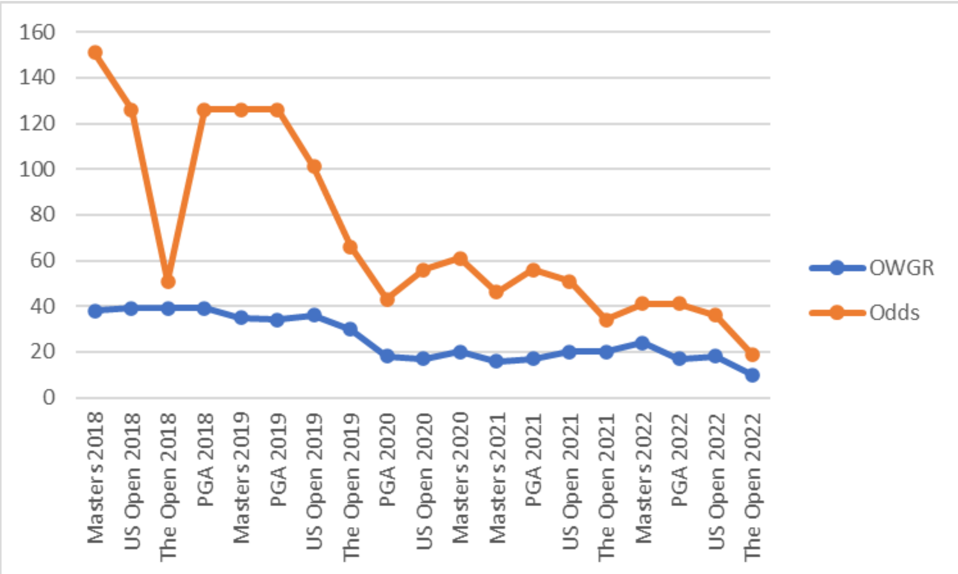


Figure 14 Matthew Fitzpatrick's position in the OWGR and odds.

Matthew Fitzpatrick's OWGR correlation is 0,829644 while his Strokes Gained Average correlation is -0,70466 and his Strokes Gained Approach correlation is -0,37526. His correlation statistics show that his odds heavily correlate due to his position in the OWGR but also on his Strokes Gained average data. His Strokes Gained approach even has a relative high correlation coefficient. One can clearly see a pattern in good performance and getting lower odds onwards from the start of the study period.

4.3.1.6 Tommy Fleetwood

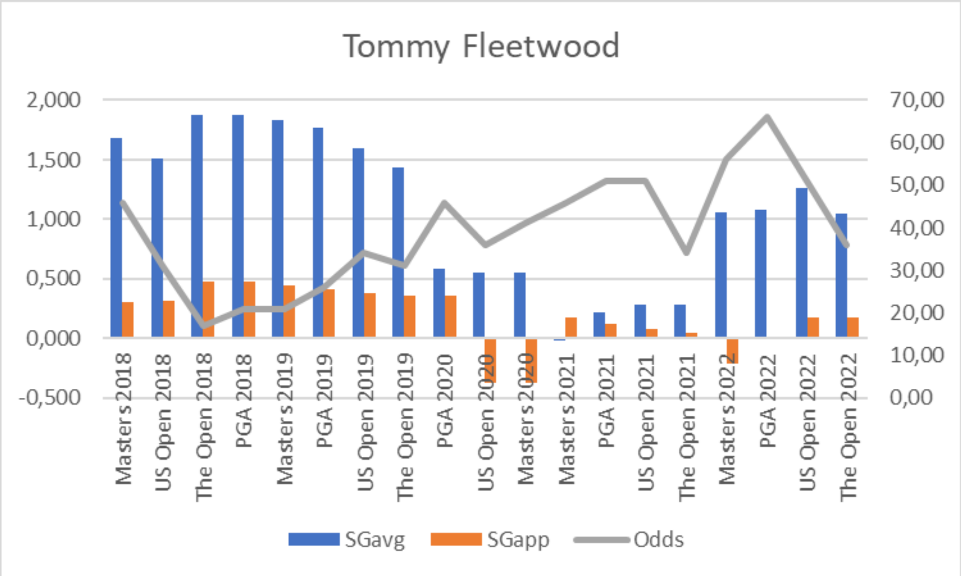


Figure 15 Tommy Fleetwood's Strokes Gained statistics (Left column) and odds (Right column).

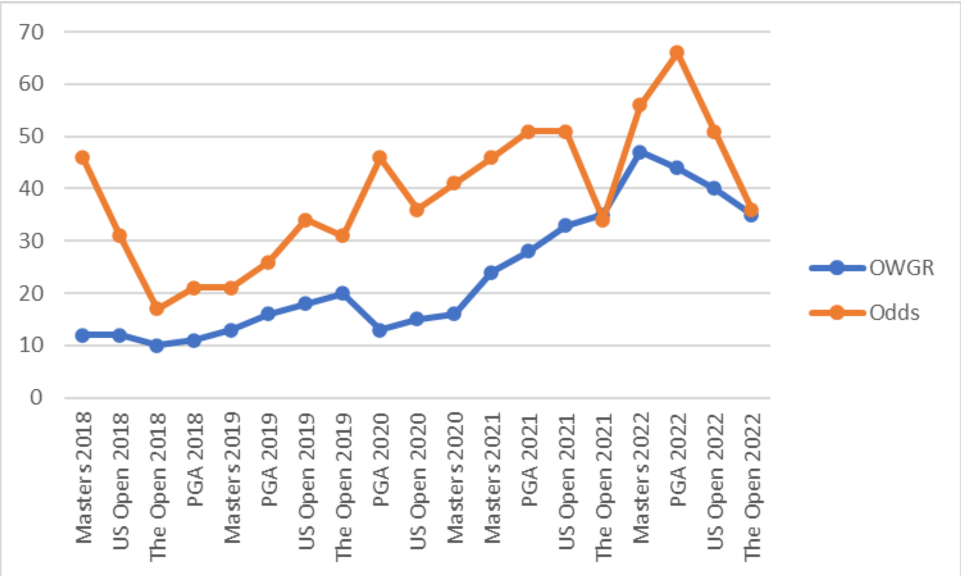


Figure 16 Tommy Fleetwood's position in the OWGR and odds.

Tommy Fleetwood's OWGR correlation is 0,710391 while his Strokes Gained Average correlation is -0,55715 and his Strokes Gained Approach correlation is -0,53481. His correlation statistics show that his odds heavily correlate due to all three metrics. His metrics also show that his decrease in performance during the study period was being considered by the bookmakers.

4.3.1.7 Sergio Garcia

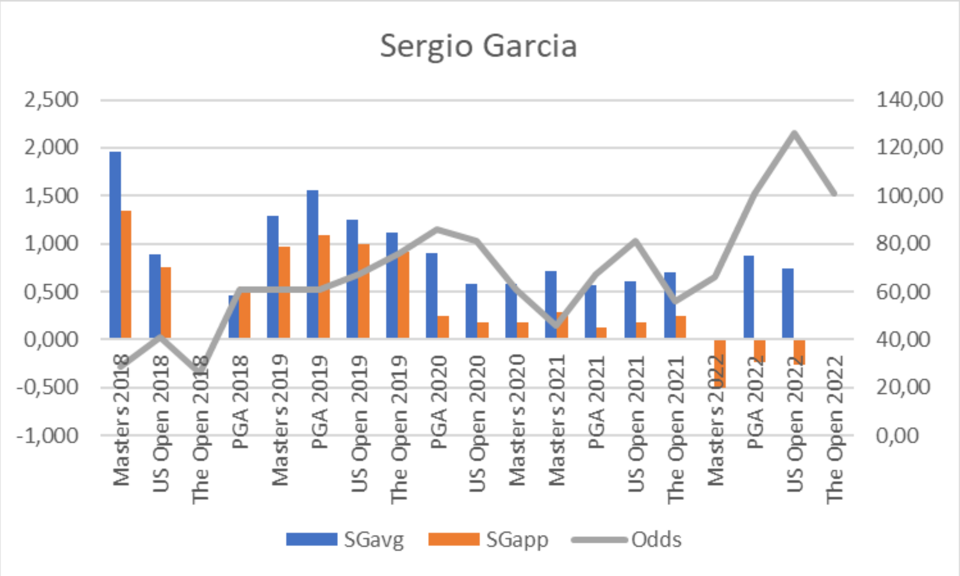


Figure 17 Sergio Garcia's Strokes Gained statistics (Left column) and odds (Right column).

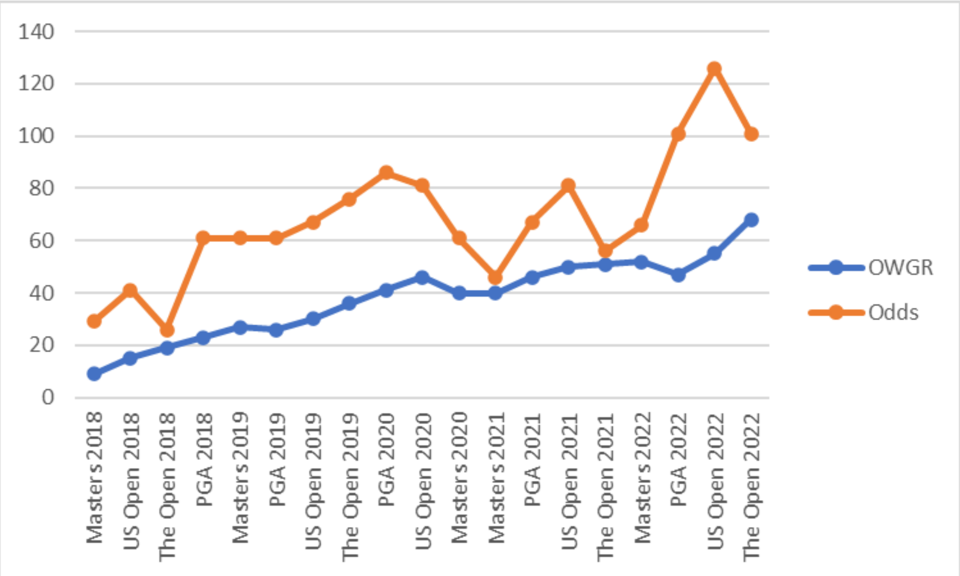


Figure 18 Sergio Garcia's position in the OWGR and odds.

Sergio Garcia's OWGR correlation is 0,751985 while his Strokes Gained Average correlation is -0,18681 and his Strokes Gained Approach correlation is -0,47988. His correlation statistics show that his odds heavily correlate due to his position in the OWGR and to some degree with his Strokes Gained Approach data.

4.3.1.8 Tyrrell Hatton

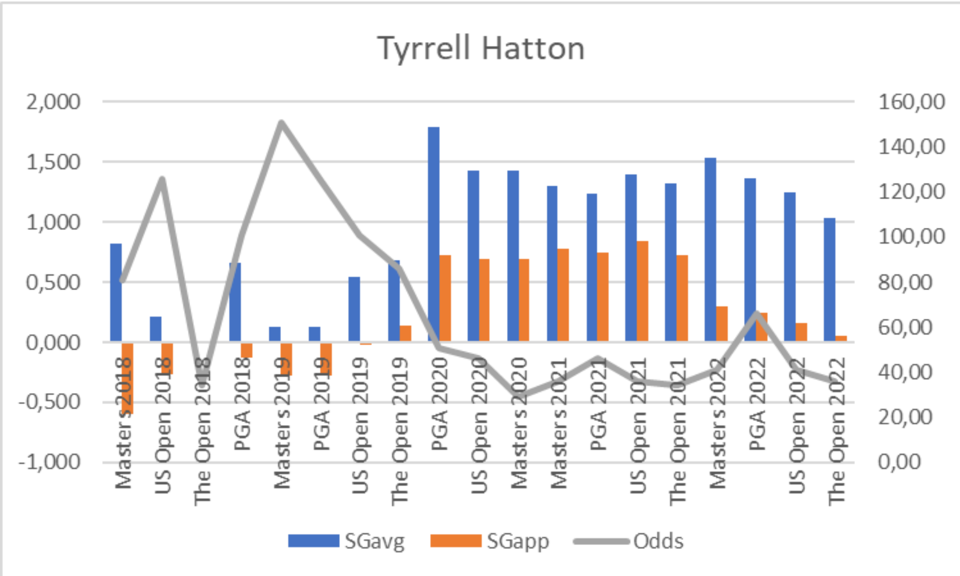


Figure 19 Tyrrell Hatton's Strokes Gained statistics (Left column) and odds (Right column).

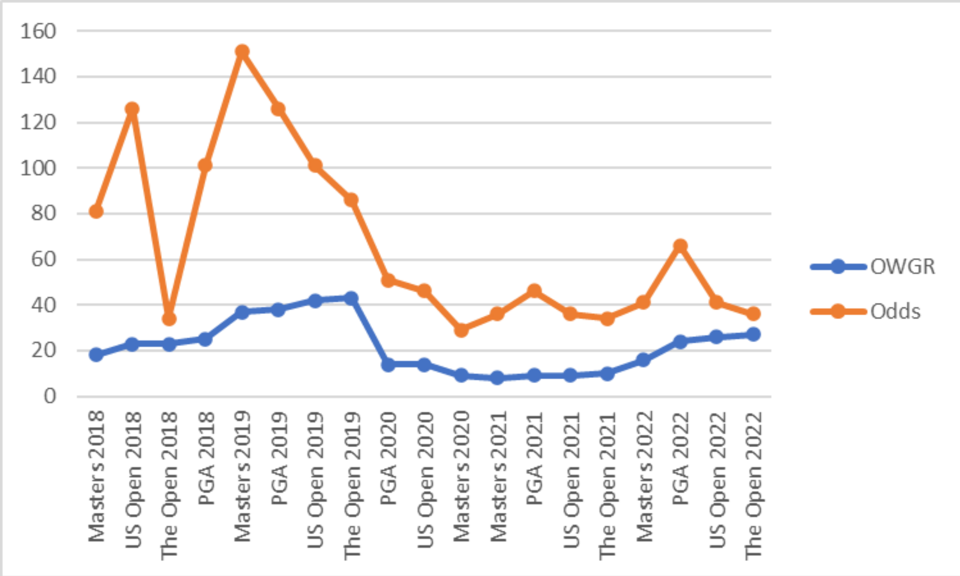


Figure 20 Tyrrell Hatton's position in the OWGR and odds.

Tyrrell Hatton's OWGR correlation is 0,702922 while his Strokes Gained Average correlation is -0,72247 and his Strokes Gained Approach correlation is -0,72425. His correlation statistics show that his odds heavily correlate due to all three metrics. The sudden lowering of his odds in 2019 can be explained through winning a few tournaments in the start of the year.

4.3.1.9 Dustin Johnson

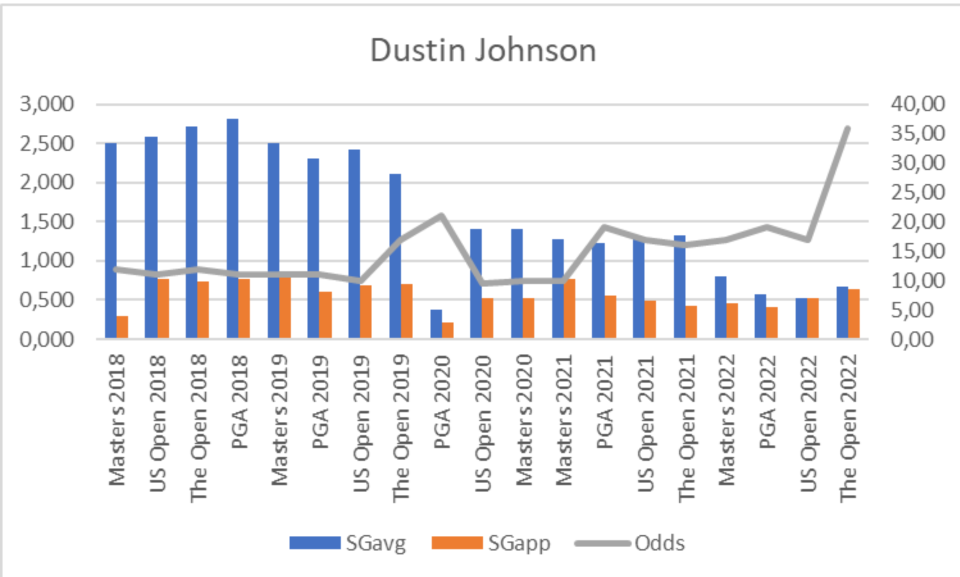


Figure 21 Dustin Johnson's Strokes Gained statistics (Left column) and odds (Right column).

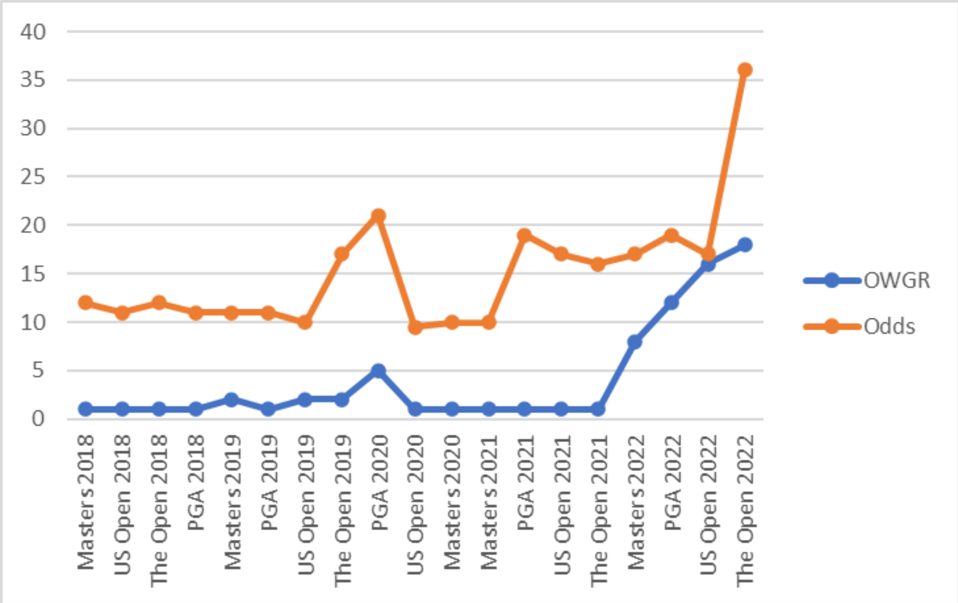


Figure 22 Dustin Johnson's position in the OWGR and odds.

Dustin Johnson's OWGR correlation is 0,742763 while his Strokes Gained Average correlation is -0,61383 and his Strokes Gained Approach correlation is -0,028431. His correlation statistics show that his odds heavily correlate due to his position in the OWGR and due to the Strokes Gained average data. His odds adjusted accordingly after his slump in performance after 2021.

4.3.1.10 Kevin Kisner

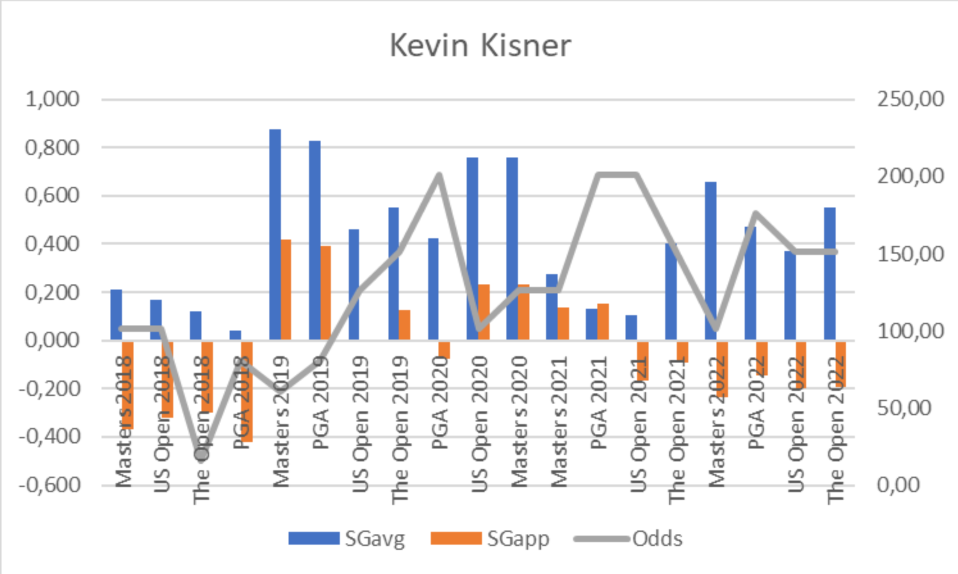


Figure 23 Kevin Kisner's Strokes Gained statistics (Left column) and odds (Right column).

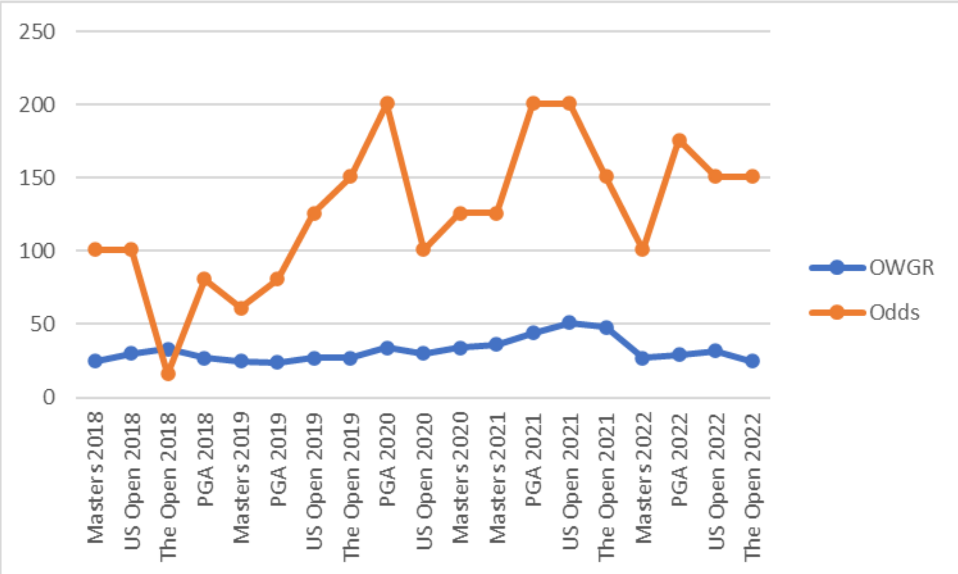


Figure 24 Kevin Kisner's position in the OWGR and odds.

Tony Finau's OWGR correlation is 0,531859 while his Strokes Gained Average correlation is -0,1667 and his Strokes Gained Approach correlation is 0,009325. His correlation statistics show that his odds are somewhat reflective of his position the OWGR but otherwise he is an unsteady performer on the golf course and the odds might reflect his recent performance before major tournaments.

4.3.1.11 Brooks Koepka

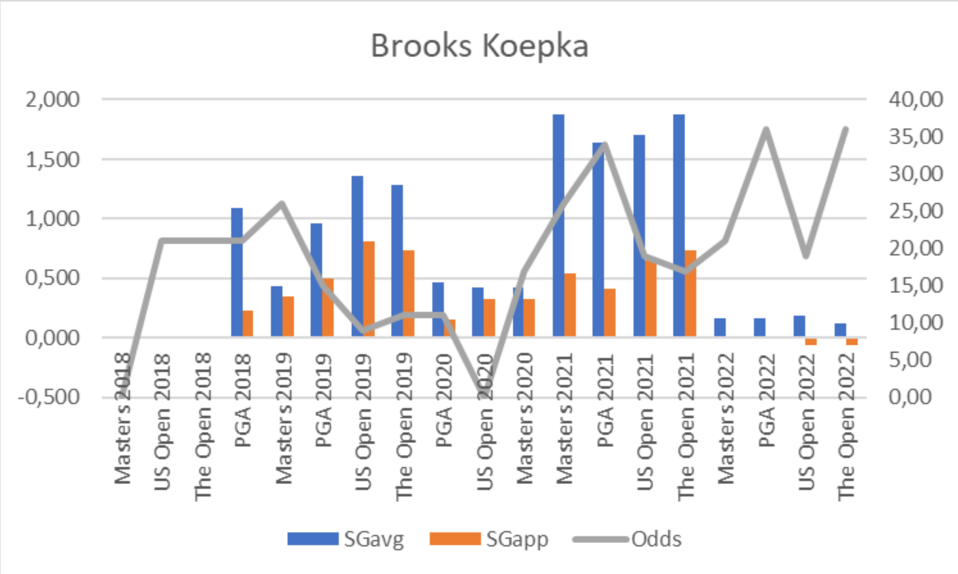
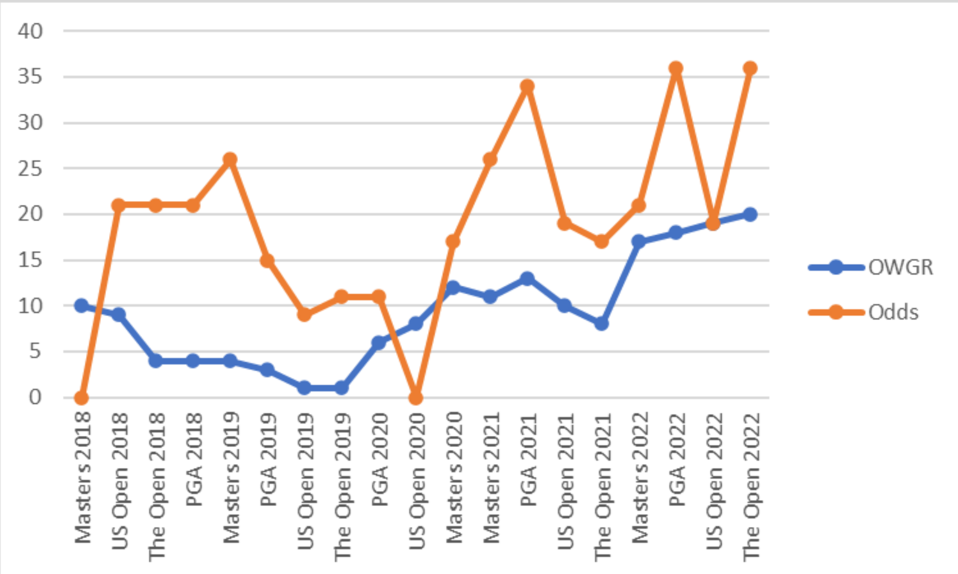


Figure 25 Brooks Koepka's Strokes Gained statistics (Left column) and odds (Right column).



Brooks Koepka’s OWGR correlation is 0,51035 while his Strokes Gained Average correlation is 0,039252 and his Strokes Gained Approach correlation is -0,24333. His correlation statistics show that his odds do not heavily corelate to any direction. He is the best performer in major tournaments from 2017 to 2023 by a country mile and his odds get depreciated because of his performance in regular tournaments.

4.3.1.12 Marc Leishman

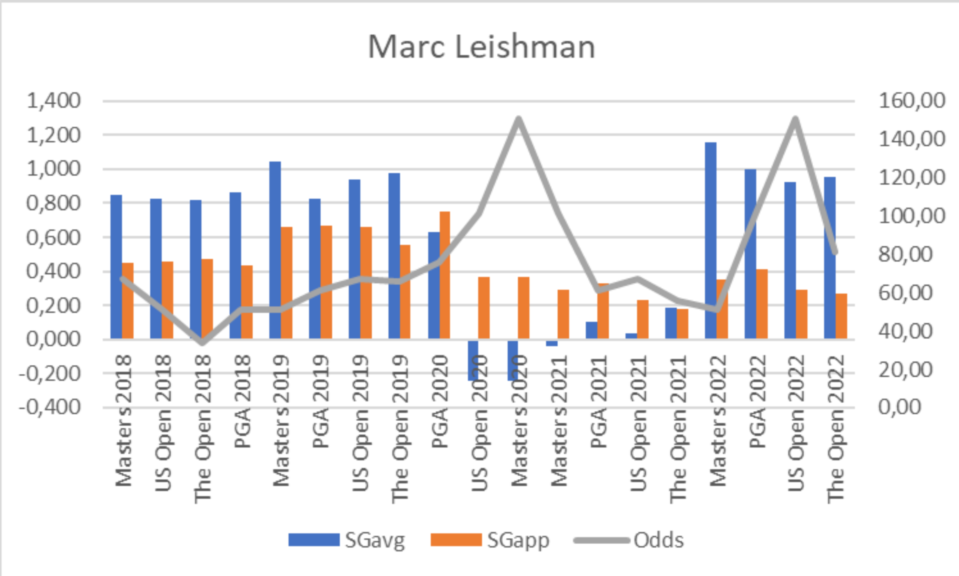


Figure 26 Marc Leishman's Strokes Gained statistics (Left column) and odds (Right column).

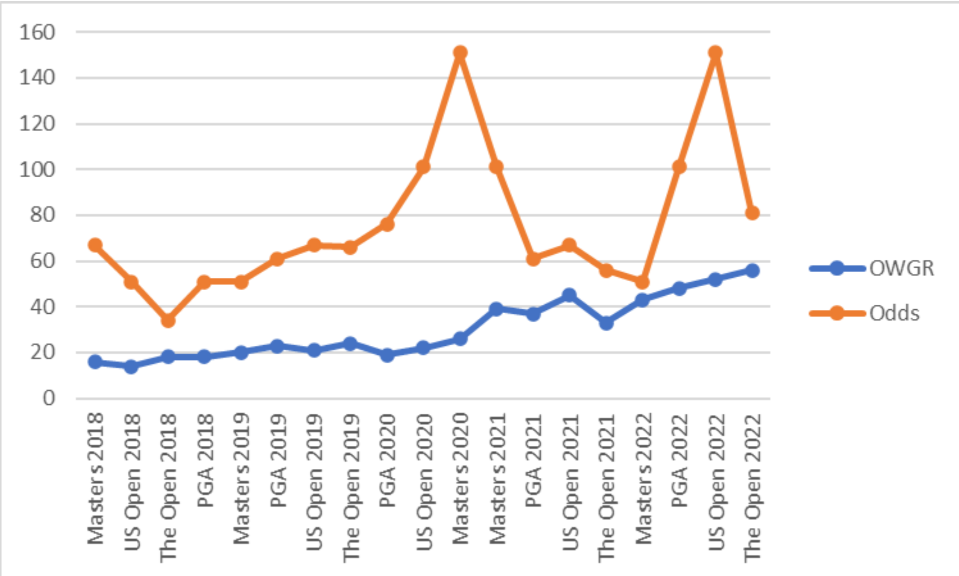


Figure 27 Marc Leishman's position in the OWGR and odds.

Marc Leishman’s OWGR correlation is 0,407502 while his Strokes Gained Average correlation is -0,37489 and his Strokes Gained Approach correlation is -0,28397. His correlation statistics show that his odds do not correlate strongly with any metric. An unsteady performer with declining results under the study period.

4.3.1.13 Hideki Matsuyama

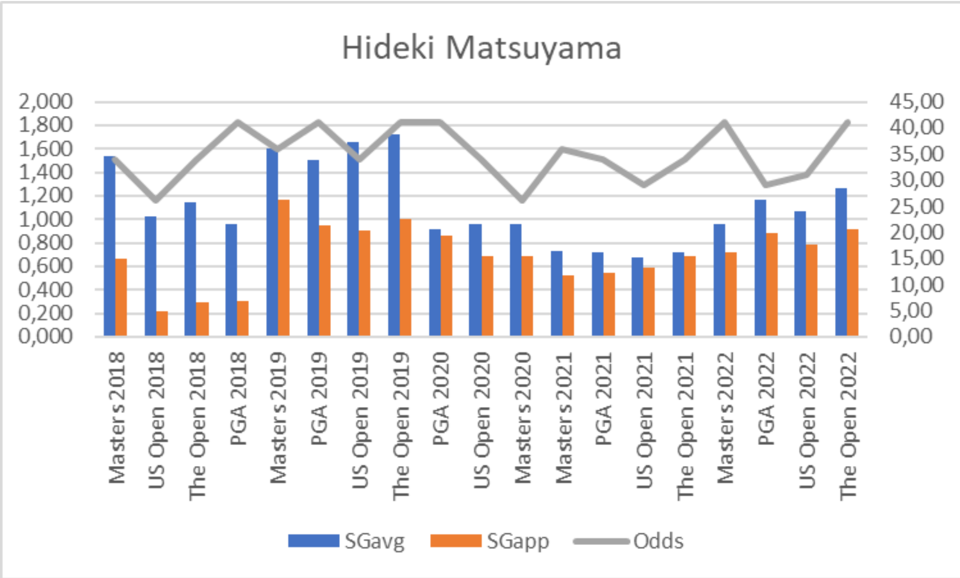


Figure 28 Hideki Matsuyama's Strokes Gained statistics (Left column) and odds (Right column).

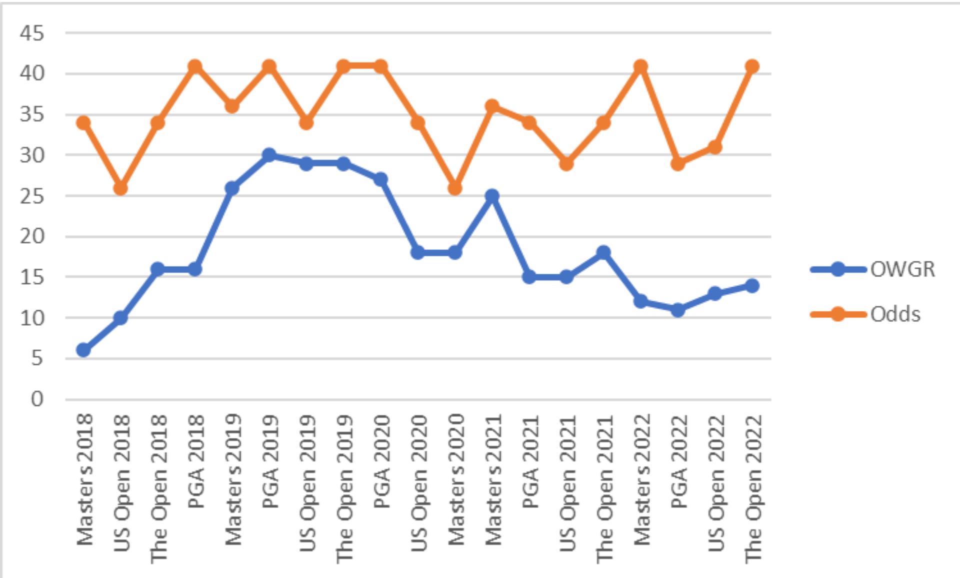


Figure 29 Hideki Matsuyama's position in the OWGR and odds.

Hideki Matsuyama’s OWGR correlation is 0,440437 while his Strokes Gained Average correlation is 0,275017 and his Strokes Gained Approach correlation is 0,328773. His correlation statistics show that his odds can only be explained through his position in the OWGR although it does not correlate starkly. He has before the study period been a top 5 player in the world but experience a slump thereafter.

4.3.1.14 Rory McIlroy

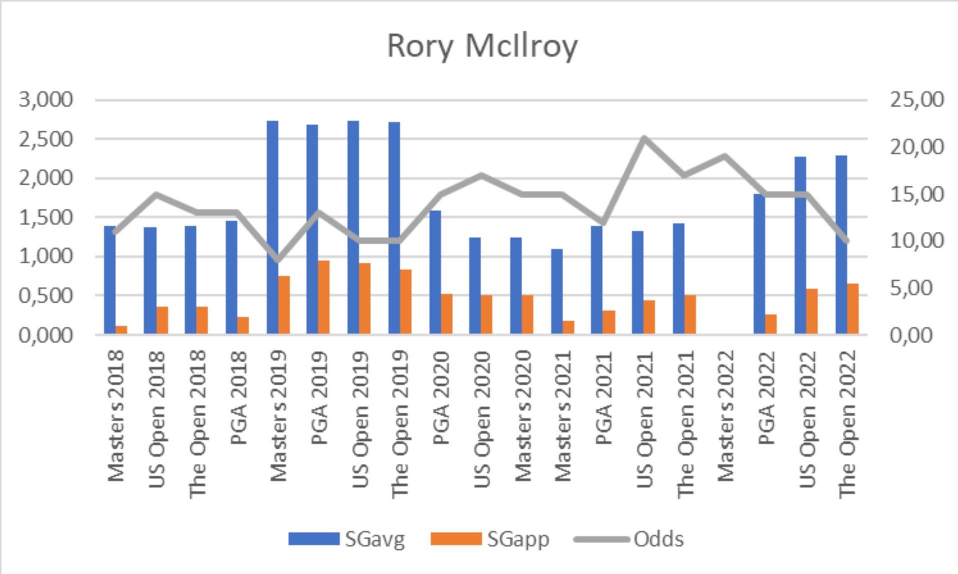
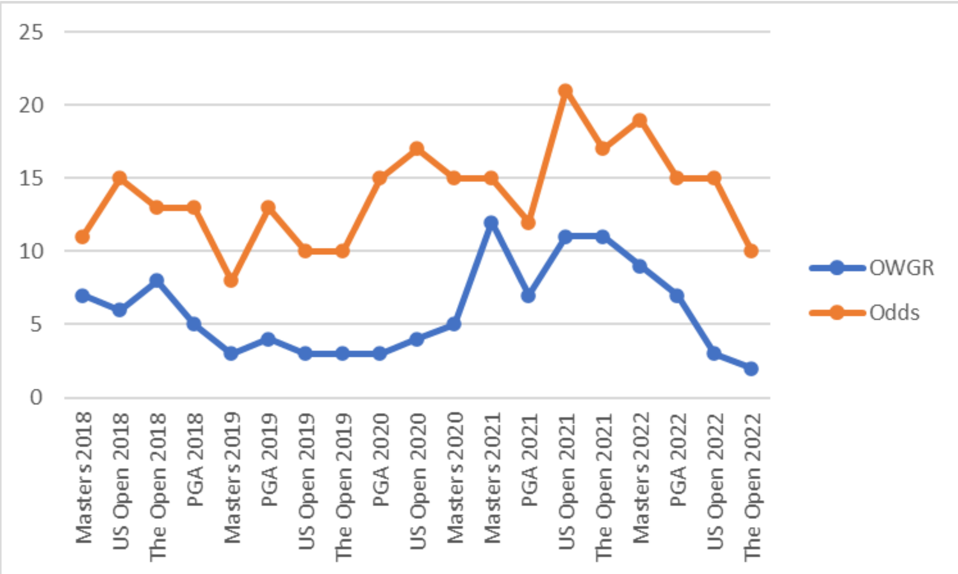


Figure 30 Rory McIlroy's Strokes Gained statistics (Left column) and odds (Right column).



Rory McIlroy's OWGR correlation is 0,610401 while his Strokes Gained Average correlation is -0,70003 and his Strokes Gained Approach correlation is -0,44848. His correlation statistics show that his odds heavily correlate due to his position in the OWGR and due to the Strokes Gained average data. He did not win many tournaments during the study period, therefore the odds were at times higher than the odds of the players with similar position in the OWGR.

4.3.1.15 Kevin Na

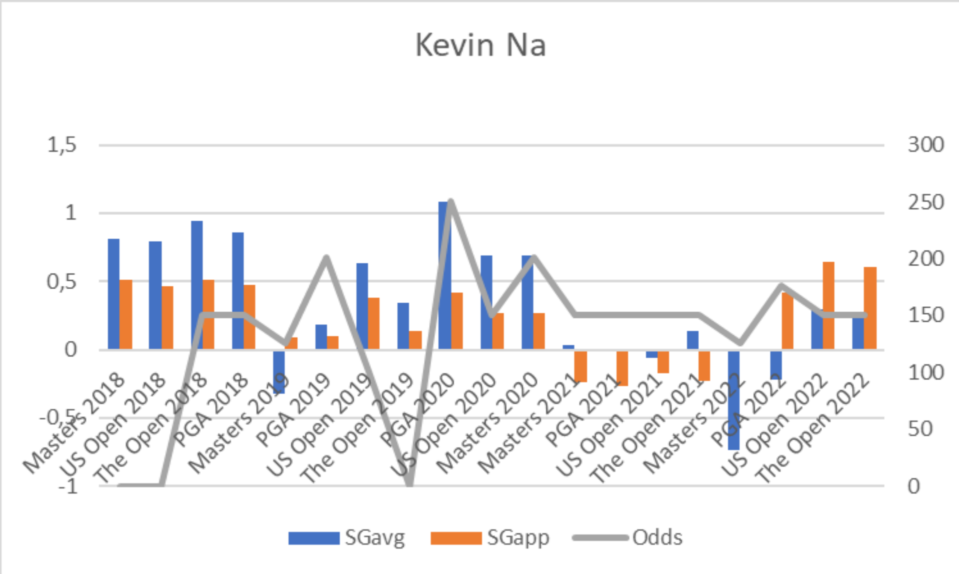


Figure 31 Kevin Na's Strokes Gained statistics (Left column) and odds (Right column).

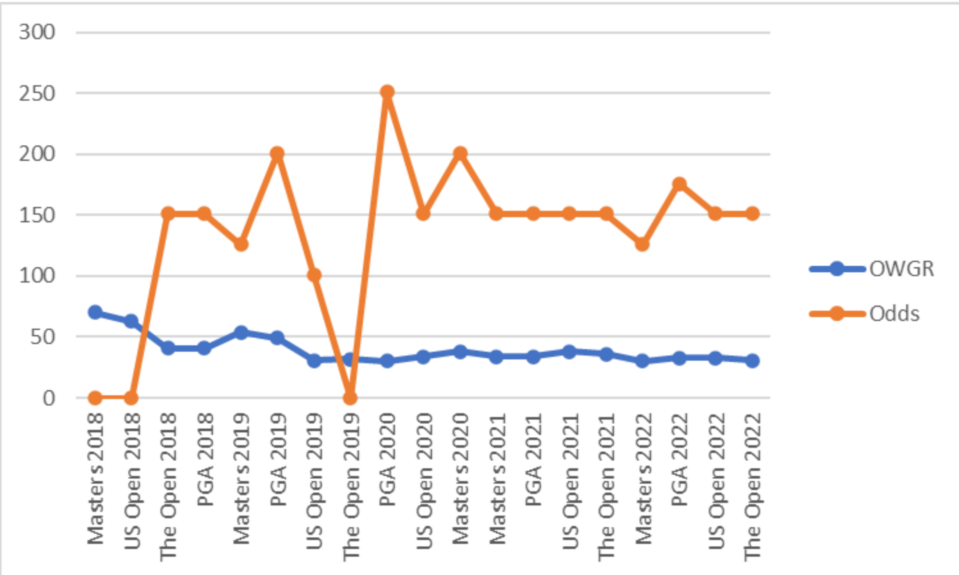


Figure 32 Kevin Na's position in the OWGR and odds.

Kevin Na's OWGR correlation is -0,53208 while his Strokes Gained Average correlation is -0,07084 and his Strokes Gained Approach correlation is -0,1161. His correlation statistics show that his odds do not correlate at all with the data, rather the opposite.

4.3.1.16 Louis Oosthuizen

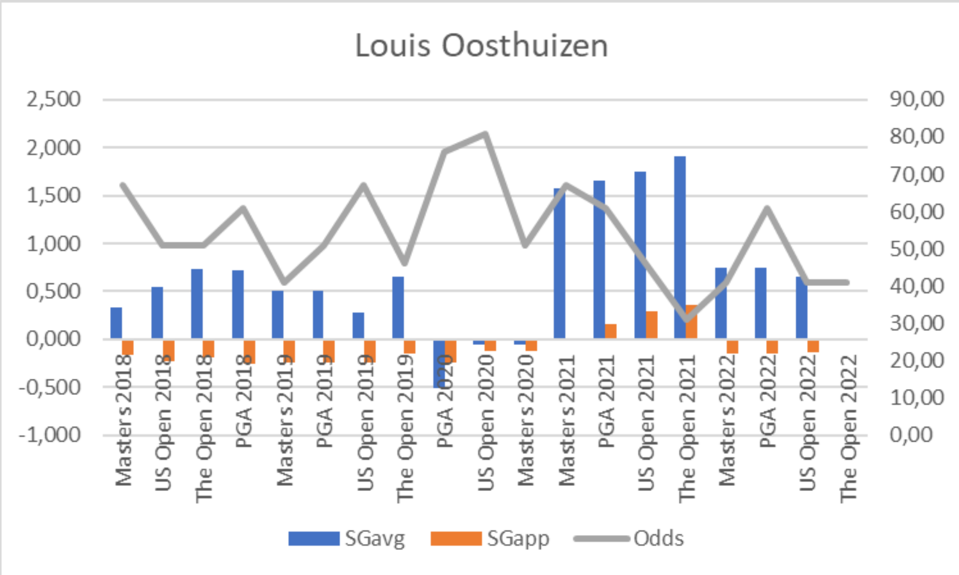


Figure 33 Louis Oosthuizen's Strokes Gained statistics (Left column) and odds (Right column).

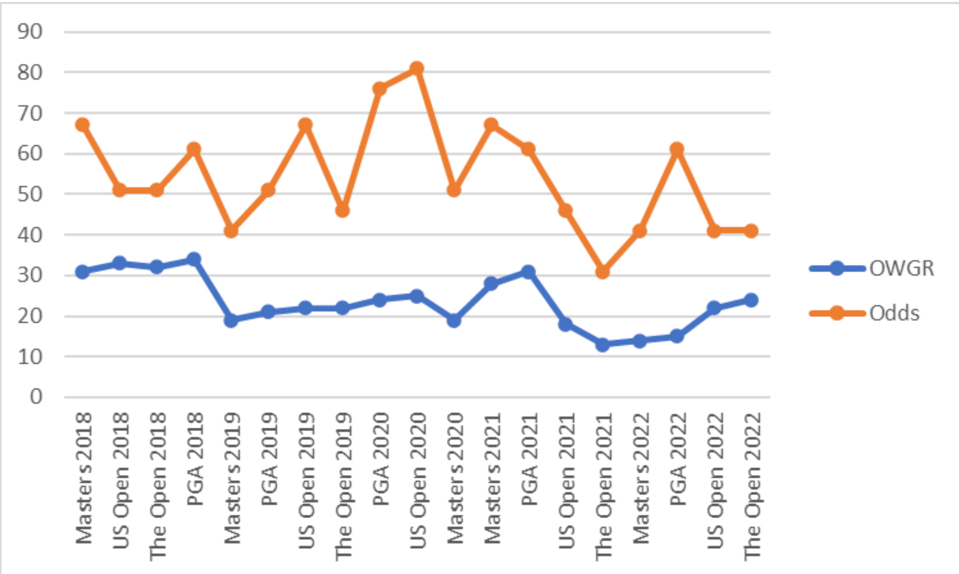


Figure 34 Louis Oosthuizen's position in the OWGR and odds.

Louis Oosthuizen's OWGR correlation is 0,453909 while his Strokes Gained Average correlation is -0,37628 and his Strokes Gained Approach correlation is -0,35306. His correlation statistics show that his odds do not heavily correlate with any statistic. He gets his lowest odds for The Open because he has previously won that and apparently the bookmakers take it into account. He has been a steadily good performer in the Masters which will also influence his odds for that tournament.

4.3.1.17 Jon Rahm

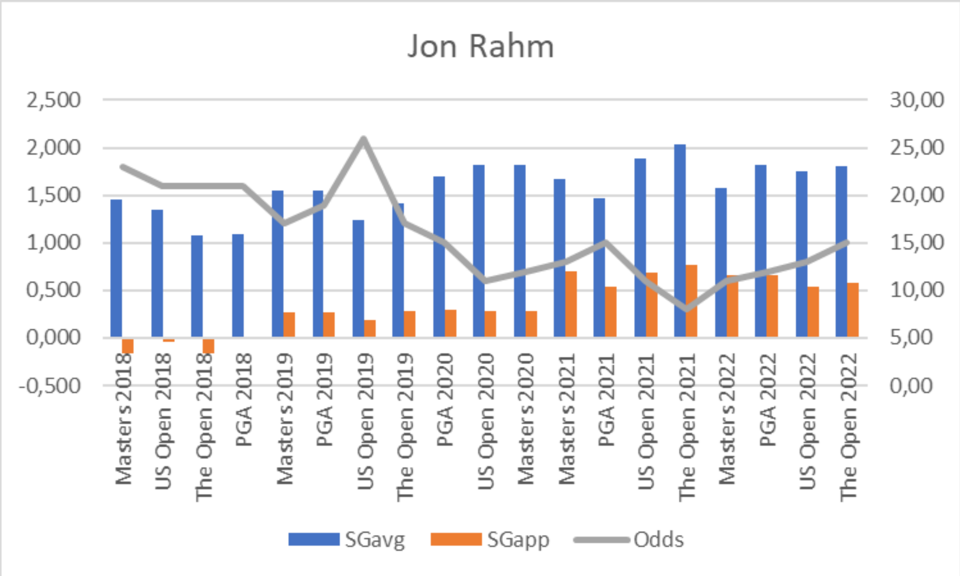


Figure 35 Jon Rahm's Strokes Gained statistics (Left column) and odds (Right column).

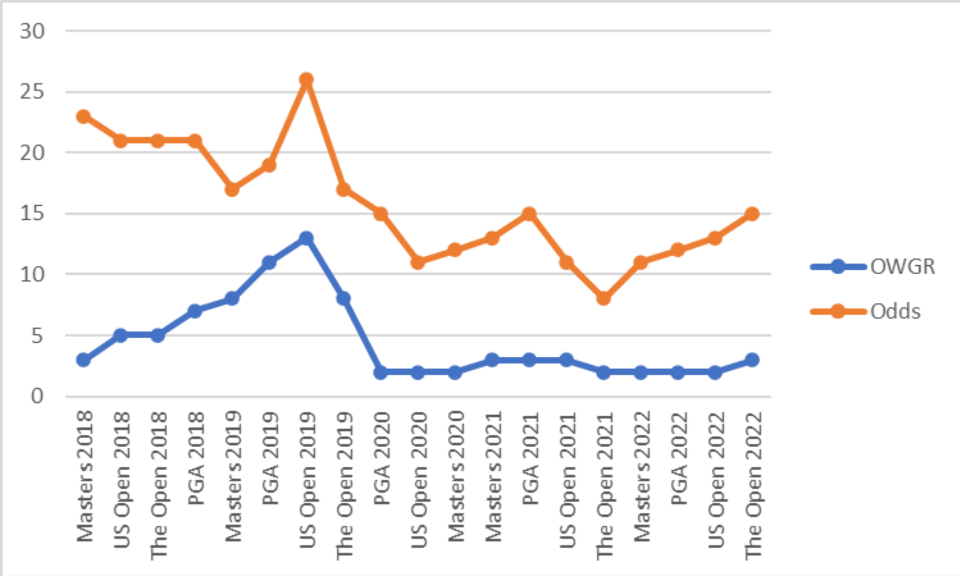


Figure 36 Jon Rahm's position in the OWGR and odds.

Jon Rahm's OWGR correlation is 0,712827 while his Strokes Gained Average correlation is -0,85216 and his Strokes Gained Approach correlation is -0,81244. His correlation statistics show that his odds heavily correlate with all the statistics and react accordingly.

4.3.1.18 Patrick Reed

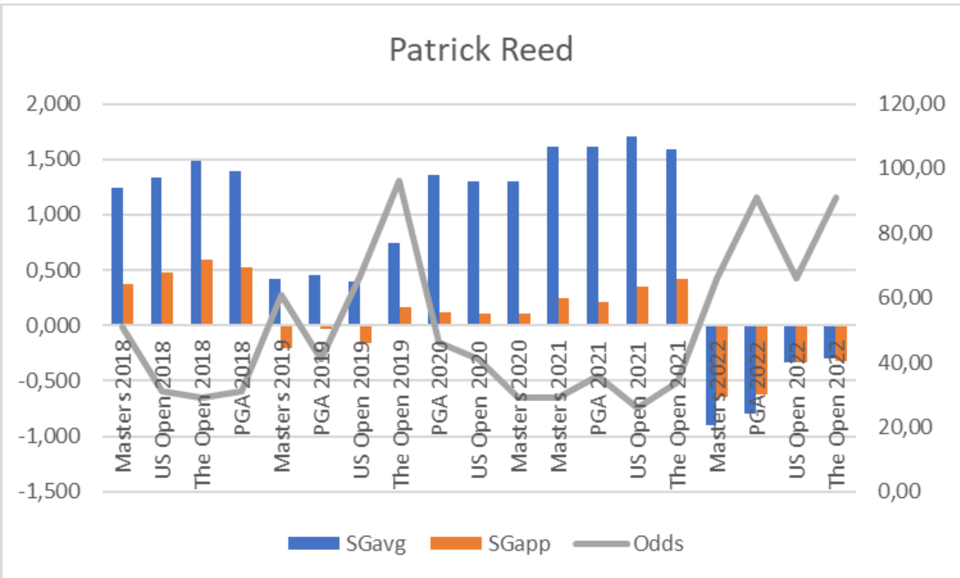


Figure 37 Patrick Reed's Strokes Gained statistics (Left column) and odds (Right column).

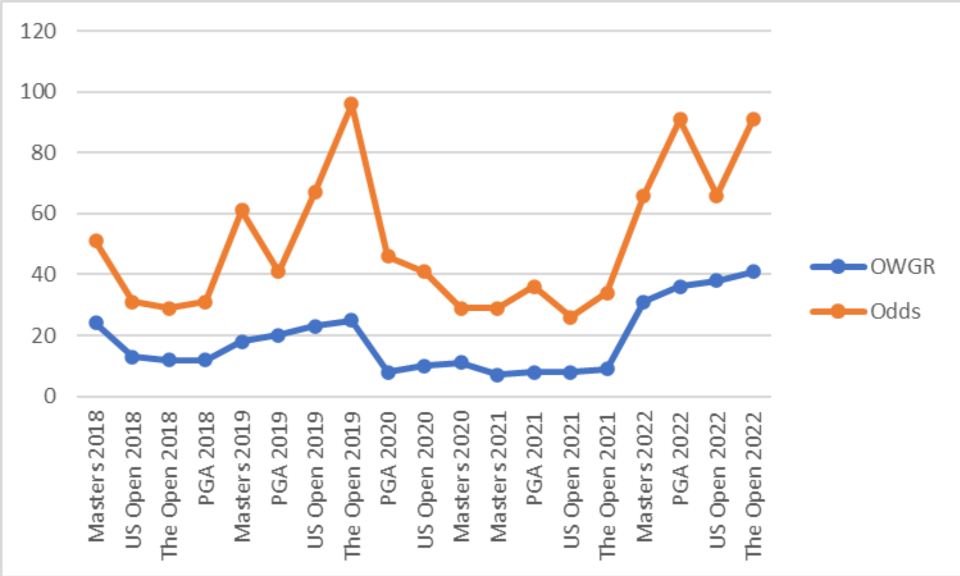


Figure 38 Patrick Reed's position in the OWGR and odds.

Patrick Reed's OWGR correlation is 0,848771 while his Strokes Gained Average correlation is -0,79031 and his Strokes Gained Approach correlation is -0,71409. His correlation statistics show that his odds heavily correlate with all the statistics and react accordingly.

4.3.1.19 Justin Rose

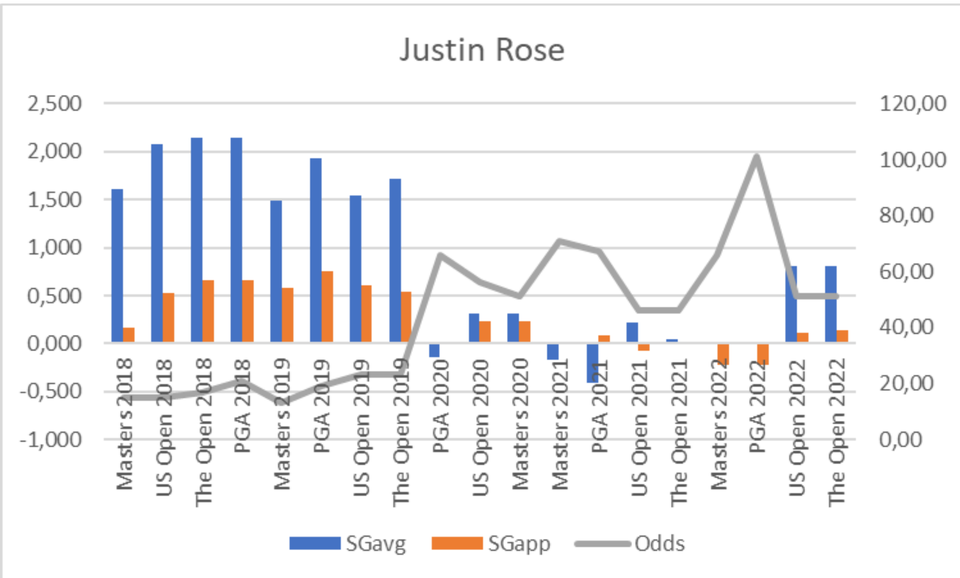


Figure 39 Justin Rose's Strokes Gained statistics (Left column) and odds (Right column).

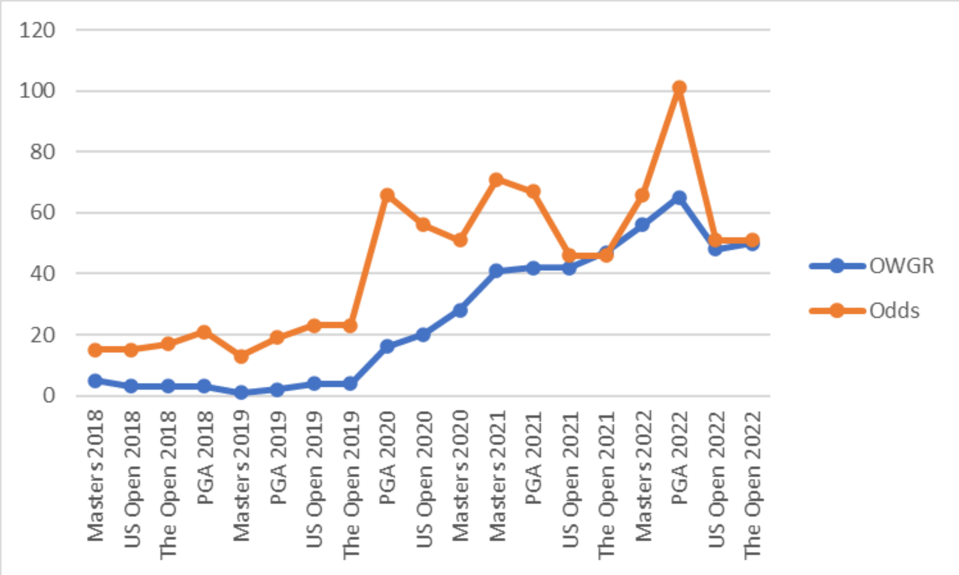


Figure 40 Justin Rose's position in the OWGR and odds.

Justin Rose’s OWGR correlation is 0,847199 while his Strokes Gained Average correlation is -0,88229 and his Strokes Gained Approach correlation is -0,84336. His correlation statistics show that his odds heavily correlate with all the statistics and react accordingly.

4.3.1.20 Xander Schauffele

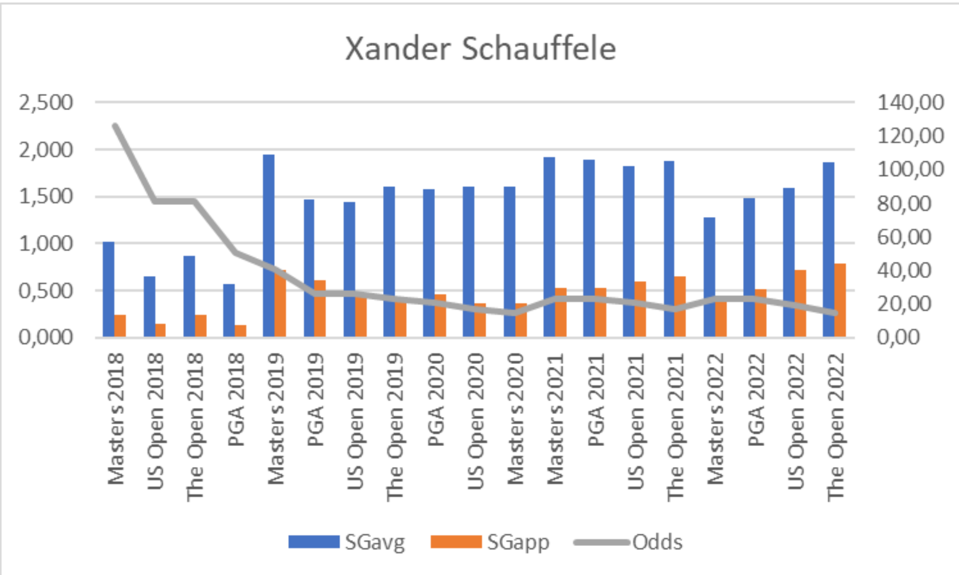


Figure 41 Xander Schauffele's Strokes Gained statistics (Left column) and odds (Right column).

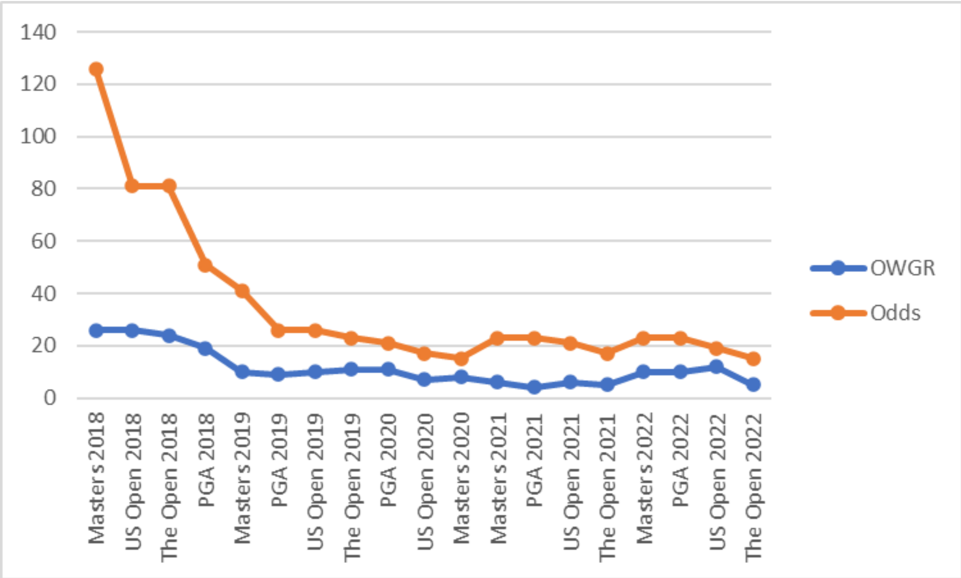


Figure 42 Xander Schauffele's position in the OWGR and odds.

Xander Schauffele's OWGR correlation is 0,905696 while his Strokes Gained Average correlation is -0,69626 and his Strokes Gained Approach correlation is -0,62213. His correlation statistics show that his odds heavily correlate with all the statistics and react accordingly.

4.3.1.21 Webb Simpson

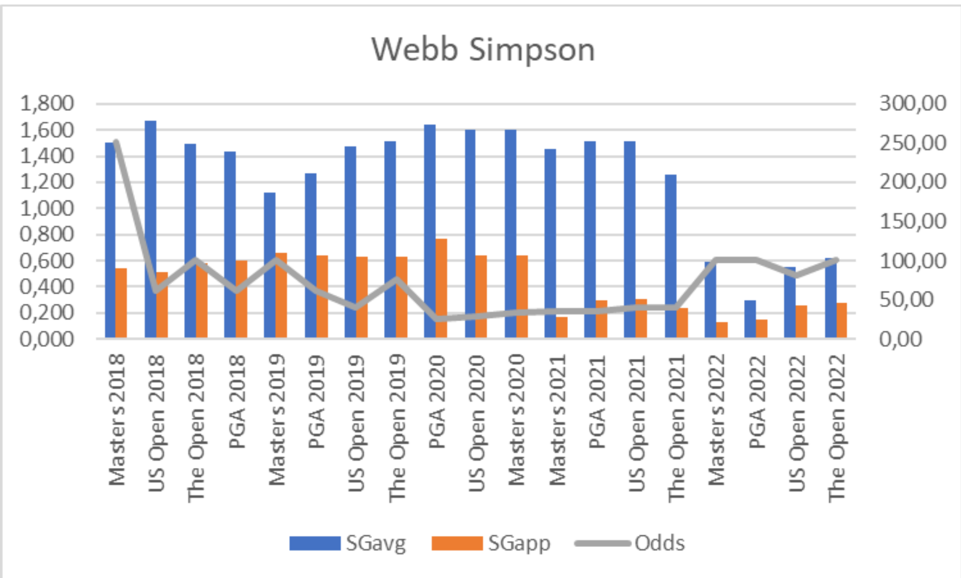


Figure 43 Webb Simpson's Strokes Gained statistics (Left column) and odds (Right column).

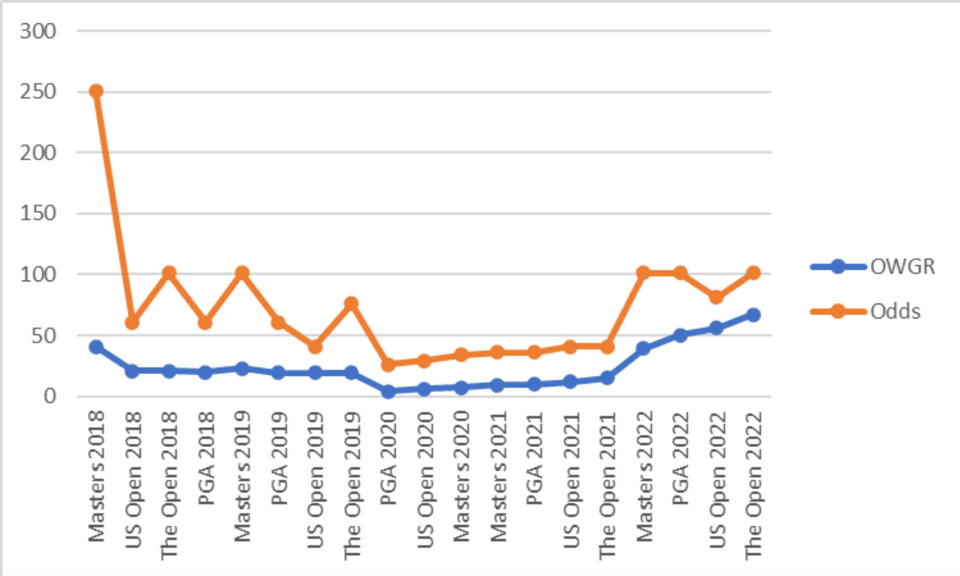


Figure 44 Webb Simpson's position in the OWGR and odds.

Webb Simpson's OWGR correlation is 0,597413 while his Strokes Gained Average correlation is -0,27332 and his Strokes Gained Approach correlation is -0,05502. His correlation statistics show that his odds correlate heavily with his position in the OWGR.

4.3.1.22 Cameron Smith

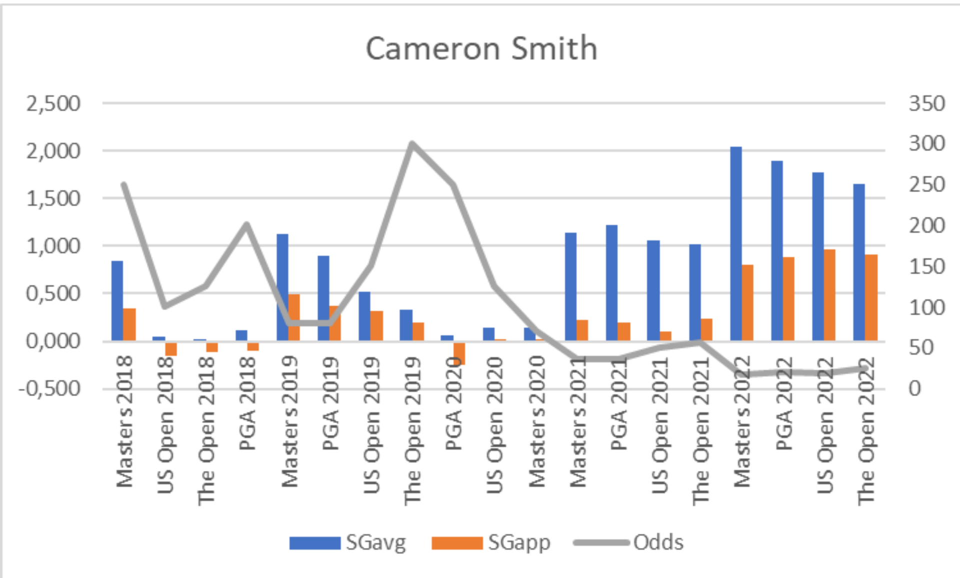


Figure 45 Cameron Smith's Strokes Gained statistics (Left column) and odds (Right column).

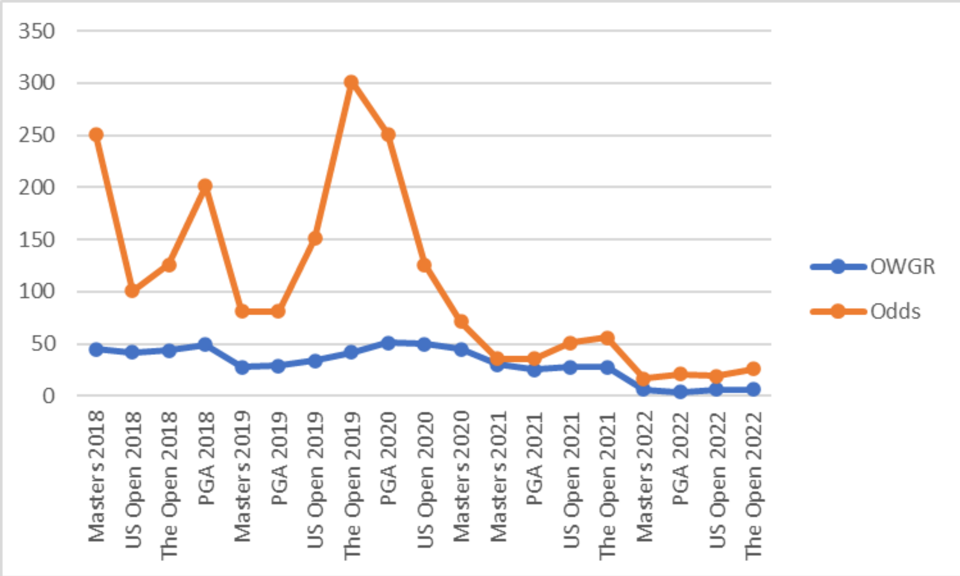


Figure 46 Cameron Smith's position in the OWGR and odds.

OWGR correlation: 0,740137.

Strokes Gained Average correlation: -0,67068.

Strokes Gained Approach correlation: -0,52773.

Cameron Smith's OWGR correlation is 0,740137 while his Strokes Gained Average correlation is -0,67068 and his Strokes Gained Approach correlation is -0,52773. His correlation statistics show that his odds heavily correlate with all the statistics.

4.3.1.23 Justin Thomas

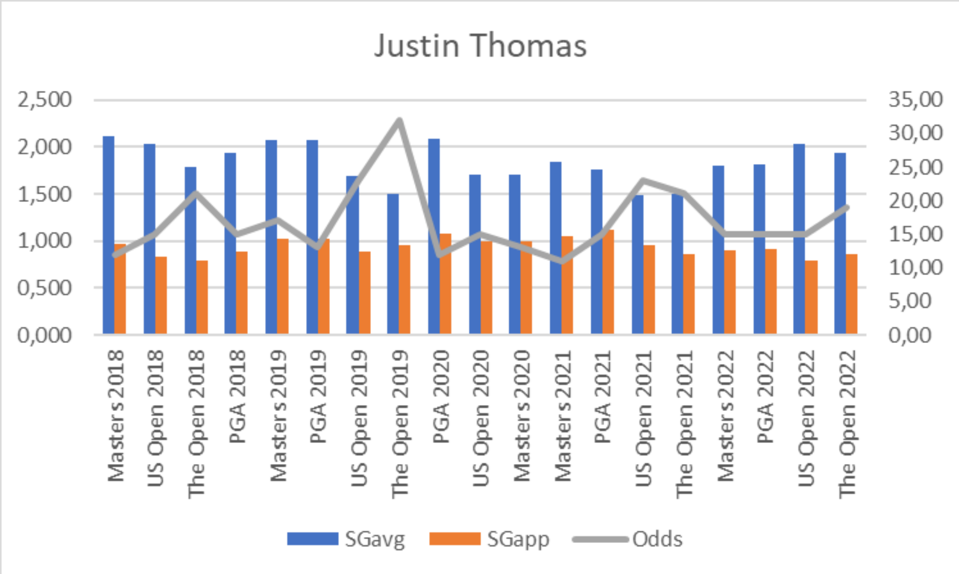


Figure 47 Justin Thomas' Strokes Gained statistics (Left column) and odds (right column).

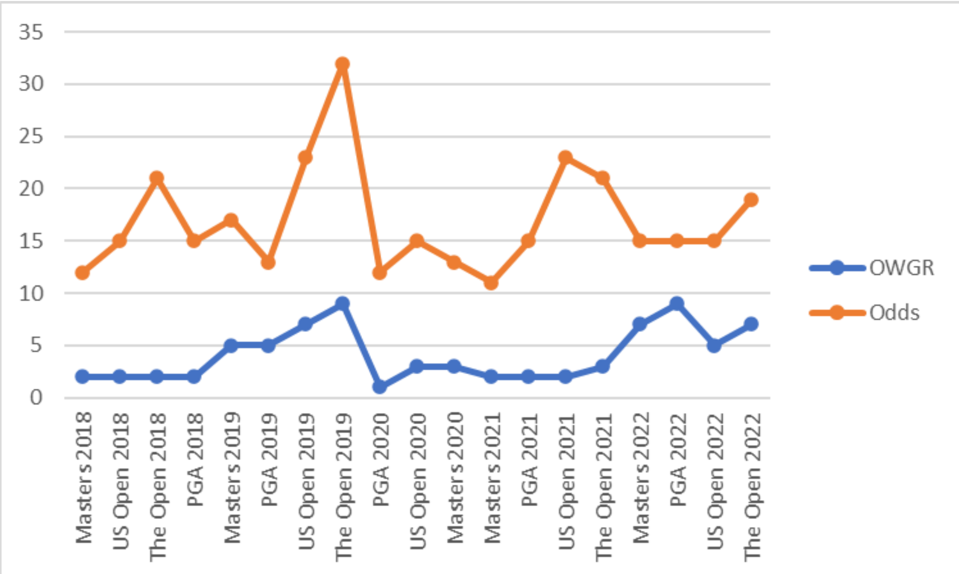


Figure 48 Justin Thomas' position in the OWGR and odds.

Justin Thomas’ OWGR correlation is 0,451329 while his Strokes Gained Average correlation is -0,67708 and his Strokes Gained Approach correlation is -0,32972. His correlation statistics show that his odds heavily correlate with the Strokes Gained average data. He did not win many regular tournaments during the study period.

4.3.1.24 Summary of players’ results

The statistics form a blend of many different results which are quite varying. Most of the players constantly ranked in the top 10 of the OWGR tend to have it as the steering factor of the odds. The players ranked here and there during the study period tend to have Strokes Gained data as the nominating factor off odds. The odds of players trending upwards seem to be a combination of all the metrics.

4.3.2 Granger Causality test results

Table 3 Granger Causality test results of the 23 players who appeared in the top 75 of the OWGR in all studied tournaments.

Player	The metric	The number of observations	F-test value	P-value	The number of lags
Patrick Cantlay	OWGR	23	42,2160	0,0000	1
	SGavg	23	0,7690	0,6290	5
	SGapp	23	0,3434	0,5666	1
Paul Casey	OWGR	23	0,5852	0,7201	5
	SGavg	23	1,7963	0,2078	2
	SGapp	23	5,9428	0,0161	2
Bryson DeChambeau	OWGR	23	4,9358	0,0273	2
	SGavg	23	3,1362	0,0802	2
	SGapp	23	0,4308	0,5215	1
Tony Finau	OWGR	23	1,5642	0,3783	5
	SGavg	23	0,8371	0,3734	1
	SGapp	23	1,1890	0,4038	4
Matthew Fitzpatrick	OWGR	23	2,6662	0,1101	2
	SGavg	23	10,5406	0,0027	3
	SGapp	23	2,1792	0,1602	3

Tommy Fleetwood	OWGR	23	4,6002	0,0488	1
	SGavg	23	4,3049	0,0556	1
	SGapp	23	4,3934	0,0535	1
Sergio Garcia	OWGR	23	1,6595	0,2310	2
	SGavg	23	2,0139	0,2995	5
	SGapp	23	2,2886	0,1513	1
Tyrrell Hatton	OWGR	23	1,6567	0,2447	3
	SGavg	23	9,7976	0,0034	3
	SGapp	23	7,0837	0,0187	1
Dustin Johnson	OWGR	23	10,2925	0,0059	1
	SGavg	23	1,2192	0,2869	1
	SGapp	23	5,1891	0,1029	5
Kevin Kisner	OWGR	23	1,5976	0,3714	5
	SGavg	23	1,7617	0,2134	2
	SGapp	23	2,6873	0,1219	1
Brooks Koepka	OWGR	23	8,7658	0,0097	1
	SGavg	23	3,2031	0,1834	5
	SGapp	23	1,9012	0,1882	1
Marc Leishman	OWGR	23	2,1441	0,2816	5
	SGavg	23	1,0443	0,4572	4
	SGapp	23	1,5182	0,3882	5
Hideki Matsuyama	OWGR	23	5,2580	0,0229	2
	SGavg	23	5,5536	0,0196	2
	SGapp	23	1,7431	0,2585	4
Rory McIlroy	OWGR	23	13,9775	0,0273	5
	SGavg	23	1,5043	0,3913	5
	SGapp	23	0,9184	0,4702	3
Kevin Na	OWGR	23	7,3742	0,0082	2
	SGavg	23	2,5544	0,1464	4
	SGapp	23	3,7163	0,1545	5
Louis Oosthuizen	OWGR	23	0,6513	0,6468	4
	SGavg	23	5,2967	0,0361	1
	SGapp	23	3,1661	0,0954	1

Jon Rahm	OWGR	23	2,2236	0,1508	2
	SGavg	23	1,5819	0,2928	4
	SGapp	23	0,8603	0,3683	1
Patrick Reed	OWGR	23	6,0531	0,0265	1
	SGavg	23	10,0181	0,0064	1
	SGapp	23	3,8236	0,0694	1
Justin Rose	OWGR	23	2,8755	0,0954	2
	SGavg	23	2,9494	0,1065	1
	SGapp	23	2,2658	0,1463	2
Xander Schauffele	OWGR	23	5,1124	0,0248	2
	SGavg	23	2,8749	0,0955	2
	SGapp	23	3,0951	0,0989	1
Webb Simpson	OWGR	23	9,7504	0,0070	1
	SGavg	23	5,2564	0,0365	4
	SGapp	23	10,7792	0,0066	4
Cameron Smith	OWGR	23	10,1976	0,0423	5
	SGavg	23	3,4740	0,0849	4
	SGapp	23	14,8368	0,0251	5
Justin Thomas	OWGR	23	0,6085	0,5601	2
	SGavg	23	2,5544	0,1464	4
	SGapp	23	1,3293	0,3245	3

The table informs each player’s test results. To gain a player the most optimal results required different lags and the number of lags varied quite starkly between a player’s three different statistics. For example, Patrick Cantlay’s statistics clearly show how starkly his OWGR statistics steers his odds. On the contrary, Justin Thomas’ OWGR statistic is not important at all when the odds are being given. Tyrrell Hatton can be seen as a player whose Strokes Gained statistics are the ones steering his odds instead of his position in the OWGR.

5 Conclusions and Discussion

The thesis conducted a Granger causality test and a Pearson correlation test to seek answers whether the odds for major golf tournaments are prone to be able to be explained by the three different chosen metrics. The results show that each individual player's odds can be explained by a different metric. Although, the players in the top 10 of the OWGR are prone to have their position in the OWGR as the nominal factor of odds. In that case, the Strokes Gained data can be used as the determining factor between separating those players, but also players in other categories of the OWGR. Some players' odds, though, are not easily explainable by the conducted tests and feel a coin-toss or an educated guess by the bookmakers. Player's odds, whose position in the OWGR rose steadily during the studied period, were often prone to depend on their Strokes Gained data. This partly, because their true potential can not be totally measured by their position in the OWGR, and their so-called ceiling and roof are unknowns. Even the bookmakers need to make educated guesses and have unknowns affecting their decision making.

The results might vary depending on the chosen time frame. Different players have been in different positions in the OWGR during different times and their performance, Strokes Gained statistics, vary through times. Golf media often presents the reader with the statistics from two to three months prior a tournament. In some cases, shorter or longer periods are used depending on what serves the purpose with each player when trying to pump him up. Crowd favorites are often given lower odds due to the high volume of bets placed on them and as a counter, the bookmakers must make sure to minimize the losses if that player was to win.

Double the amount of bets on Tiger Woods to win the Masters than any other golfer.

The Masters Outright Winner Splits

Masters Winner Splits				
Rank	Golfer	Odds	% of Handle	% of Bets
1	Tiger Woods	+4000	16%	14%
2	Justin Thomas	+1200	8%	7%
3	Brooks Koepka	+2000	7%	5%
4	Cameron Smith	+1400	5%	5%
5	Patrick Cantlay	+2500	5%	4%
6	Collin Morikawa	+2000	5%	4%
7	Jordan Spieth	+2200	4%	3%
8	Dustin Johnson	+1600	4%	3%
9	Jon Rahm	+1000	4%	4%
10	Scottie Scheffler	+1200	3%	4%
Provided by DraftKings Sportsbook				

6:15 PM · Apr 7, 2022

Figure 49 Betting statistics of DraftKings Sportsbook published on their Twitter account after closing the books for the 2022 Masters Tournament.

On the other hand, dumb money odds are present as well. As presented in Figure 52, Tiger Woods got the most of betting volume in the 2022 Masters Tournament, although he could barely sustain walking the course due to a broken leg.

Players who have previously won a major tournament get lower odds than a player who is in a similar position in the OWGR or who has comparable Strokes Gained numbers, who has not previously won one the four major tournaments. During the time period researched in this thesis, 2018 – 2022, 11 new players won a major tournament. It means that 58% of the time, major tournaments are won by a player who previously has not won one. On the other hand, 42% of the time a player who previously as won, will win again.

Historically, from the first Open Championship played in 1860, 462 major tournaments have been played and there have been 230 different winners which percentwise means that 50% of the played tournaments, a new winner has emerged. During modern times, after the Second World War, 152 new winners have emerged through 306 tournaments, meaning that for 49% of tournaments played, a new winner has emerged.

After 1997 the Masters Tournament, another watershed moment in golf with Tiger Woods winning his first major tournament, a total of 100 major tournaments have been played with 57 new players winning. These historical statistics partly explain the reasons as for why a player who previously has won, will get lower odds. Even a player who has been ridden by injuries, but is a former winner, will get low odds, as can be seen through the data of this study in the case of Brook Koepka. Furthermore, a player ranked outside of the top 25 of the OWGR is highly unlikely to win a major tournament but a former winner is seen a possible candidate to win. The defending champion is also favored slightly in odds. This is probably because the four major tournaments are being played on different styles of courses and each championship normally requires different aspects of a player's game to be on point. Historically, The Open Championship has produced older winners than the other three major tournaments. Hitting it long is not a requirement on the courses that tournament is played on which is an equalizing factor and brings in more candidates for winning. On the contrary, the US Open often requires a player to hit it long and straight and is punishing for shorter hitters. The Masters Tournament is often producing winners who are great at hitting approach shots. The course has wide fairways that are forgiving to players who might not be the most accurate off the tee, but the slopy greens, that require precise approach shots, are a separating factor between players. Picking a winner out of a hundred and odd players is difficult, but the historical data and Strokes Gained statistics, collected by Mark Broadie, prove that in truth, but for rare occasions, the number of possible winners is much smaller than the whole field. Recently, paysites providing deeper information of players' performance have emerged. Data Golf, one of the websites, provides people with information such as "course fit", which is model studying a player's Strokes Gained data and comparing it to what a course of a given week has historically required, statistically, of a player to win. How companies really set the odds, is a black box.

The OWGR's integrity after the emergence of the LIV tour, became even more questionable after the 2023 Masters Tournament and the PGA Championship. Brooks Koepka finished second in the former and won the latter, while ranked 78th in the world before the Masters Tournament. Strokes Gained data backed his good performance but since LIV tour's tournaments are not being granted OWGR points, his position in it has been slowly declining.

According to this study, each player's odds are formed based on different level of importance put on the three studied metrics. This can be used as an advantage by a conscious bettor with the right know-how. Choosing a winner out of a whole field of players is a difficult task but with the help of Broadie's studies, as well as this thesis, the chance of picking correctly will be much more likely than without using the help of statistics.

Strokes Gained as a concept is not widely known amongst golfers, except for the ones interested in data and how to use it. Most golfers might have heard the term but have little to no idea what it is. Although it has been for long used by professional golfers, it is only now emerging into the knowledge of the weekend-golfers who enjoy the game as a hobby. The study shows that odds given for betting can not be easily explained but rather require studying. The type of data used in this study is free to access and will most likely give a good picture of the players' performance before a tournament one is willing to bet on.

The results of the study also show that the odds are sometimes nothing but an educated guess. Some players' data is such that no trends can be found which in part makes it difficult for the bookmakers to set the odds. Strokes Gained remains a good tool when assessing a player's performance and in liaison with the OWGR will paint a good picture on which players are trending upwards and which players are trending downwards. Golf analytics are evolving all the time and players have developed their own versions of Strokes Gained model. The models only they have access to include information such as where did they aim their shot and from which direction the wind was blowing. Those to factors will heavily influence the income of a shot and therefore influence the Strokes Gained statistic available for us.

5.1 Research Questions

1. *Do the given odds correlate more with the OWGR or Strokes Gained statistics?*

The correlation coefficient of the players ranked high in the OWGR is larger than the correlation coefficient of the players ranked lower in the OWGR. Strokes Gained statistics are a more correlating factor when separating lower ranked players. Granger causality test's results can be seen backing up this. The players ranked high in the OWGR had that as the Granger-causer of odds whereas the lower ranked players' odds were Granger-caused more often by Strokes Gained statistics.

2. *Can Strokes Gained Approach metrics be used to predict future odds?*

When a player has a largely positive Strokes Gained Approach statistic in conjunction with otherwise positive Strokes Gained Total statistics, it can be seen as a factor. But for one player, Strokes Gained Approach does not singlehandedly Granger cause odds.

5.2 Future Research

Trying to guess a correct winner out of a field of some 150 players is tough and analytics can help to identify hot players. What is much is easier than picking a winner, is to correctly guess that a player will finish inside the top 20 or even top 10 of a golf tournament. A study could be conducted to study how the players who have finished inside the top 20 of past major tournaments and using that data to find out how often players ranked high in the Strokes Gained statistics will finish inside top 20.

Also, a study could be conducted on how often in the past, the favorites before the tournament, have performed well. Was it worth betting on the favorites or whether it would have been more feasible to bet on a player ranked high in the Strokes Gained statistics? Using that data, guidelines could be drawn what is normally needed Strokes Gained wise in order for a player to finish in the top 20.

The study conducted in this thesis could be optimized by trying to find out which time frame of the Strokes Gained data correlates the strongest with the given odds. By doing that, a bettor could more easily find out whether a bet is feasible and give better answers to whether betting companies in fact always seem to use a specific time frame when giving odds.

6 Svensk sammanfattning

6.1 Introduktion

Avancerad statistisk analys är måttligt nytt i golf. Fram till bara nyligen var de enda tillgängliga statistiker rätt så icke-beskrivande siffror. Det enda rimliga sättet att mäta spelarnas prestation var den Officiella Världsgolfrankingen (OWGR), som fortfarande används som ett officiellt mått på vem som är den bästa spelaren i världen. OWGR är baserad på en algoritm som tar hänsyn till spelarnas prestationer i turneringar från de senaste två åren. Alla turneringar är inte lika med varandra eftersom spelare som deltar i turneringarna inte alltid är de samma. När en turnering startar, bidrar varje spelare till det som kallas för Fältstyrka, som används för att avgöra hur svårt det är att vinna den turneringen och det orsakar delvis partiskhet i rankingen.

Mark Broadie, en professor vid Columbia University skapade Strokes Gained (Förtjänade slag) -modellen för att bättre kunna analysera professionella spelares prestationer. Numera används modellen på alla möjliga sätt i golfvärlden. Televisionssändningar använder den för att jämföra spelarnas kunskaper och ge tittaren bättre information om den pågående turneringen, spelare använder den i deras träning för att bli bättre på områden de inte är relativt bra på, och även ivriga amatörgolfare har tagit upp Strokes Gained-modellen för att bli bättre. Strokes Gained, kort sagt, används på samma sätt som Expected Goals (Förvantade mål) används i fotboll och andra lagsporter för att förutsäga ett matchresultat genom att studera lagens prestation statistiskt. Då resultaten förutsägs, tas det i fotboll i beaktande antalet mål ett lag förväntas göra, medan det som tas i beaktande i golf är antalet golfslag som spelarna förväntas slå för att avsluta en golfrunda. Strokes Gained (SG) består av fyra olika mått: SG Off-The-Tee eller Driving (Från tee), SG Approach (Inspel), SG Around-The-Green (Närspel) och SG Putting (Puttning). Dessa fyra siffror bildar en summa, som kallas för Strokes Gained eller Strokes Gained Total. Syftet med denna studie är att ge ny information om huruvida Mark Broadies Strokes Gained-modell, och en del av den, korrelerar med odds som ges av spelbolag när de ger Vinnare-odds före majorgolfturneringarna, och ifall det kan hittas någon trend. Studien undersöker också den Officiella Världsgolfrankningens inverkan på odds. För att nå målet med denna forskning kommer avhandlingen att söka svar på följande forskningsfrågor:

1. *Korrelerar de givna oddsen mer med OWGR eller med Strokes Gained Total-statistiken?*
2. *Påverkar statistiken Strokes Gained Approach (inspel) de givna oddsen?*

6.2 Litteraturöversikt

Det här kapitlet kommer att presentera en teoretisk ram för Strokes Gained-statistik, och vad den Officiella Världsgolfrankingen är samt vad som påverkar vadslagningsmarknaden.

Uppdraget för den Officiella världsgolfrankingen (OWGR) är att på veckobasis administrera och publicera en transparent, trovärdig och korrekt ranking baserad på de relativa prestationerna för spelare som deltar i kvalificerade golfturer för män över hela världen (The Official World Golf Ranking, 2023).

Anledningen till att uppfinna Strokes Gained var bristen på information från traditionella golfstatistiker. De statistikerna kunde inte på djupet förklara vad skiljefaktorn mellan de bästa professionella golfarna och den genomsnittliga professionella golfaren var (Broadie, M. 2014). För att bekämpa problematiken med den äldre golfstatistiken ville Mark Broadie uppfinna ett system som gjorde det möjligt att jämföra och bedöma varje del av spelet individuellt mellan spelare såväl som en helhet (Broadie, M. 2014). Strokes Gained kommer från Dynamic Programming (Dynamisk programmering), en metod som utvecklades av Richard Bellman på 1950-talet. Dynamisk programmering innebär i praktiken att man delar upp ett problem i underdelar och hittar det mest effektiva sättet att lösa varje del av problemet för att maximera resultatet. Dynamisk programmering har även använts i andra sporter för att få konkurrensfördelar. Den mest kända sporten som använder sig av dynamisk programmering är baseball.

Ur Strokes Gained-perspektiv leder Dynamisk programmering till ett prestationsmått där framstegen till hålet inte mäts i yards utan i minskningen av det genomsnittliga antalet slag till hålet (Broadie, M. 2014). Värdet för Strokes Gained för ett enstaka golfslag är minskningen av det genomsnittliga antalet slag till hål, minus ett för att ta hänsyn till det slaget som spelaren har tagit (Broadie, M. 2014). Strokes Gained är ackumuleringen av vinst- och förlustbråkdelar av varje enskilt skott och gör det enkelt att jämföra spelarnas totala prestation (Broadie, M. 2014). Man kan enkelt bedöma en spelares styrkor och svagheter om det behövs. Modellen använder alltid numret noll som en baslinje.

När vi i allmänhet pratar om sportvadslagning, menar vi processen att förutsäga resultatet av ett spel och slå vad enligt en förutsägelse. Specifikt, för att en satsning ska existera bör det finnas två sidor: spelbolagen som bestämmer priserna och spelarna som slår vad i olika sporter baserat på de givna oddsen. Beroende på spelresultatet vinner en av de två – spelbolaget eller spelaren – och den andra förlorar (Ioulianou et al, 2011). Sportspel är en engångsföreteelse där ett resultat som är föremål för en vadslagning antingen inträffar eller inte inträffar (Malaska & Virtanen, 2007). Enligt Xu (2011) och Makropoulou & Markellos (2011) har ett spelbolag ett enda uppdrag; att tjäna pengar på oddsen som de tillhandahåller. Detta driver spelbolagen att sätta oddsen tillräckligt högt för att vara konkurrenskraftiga, men tillräckligt lågt för att det inte ska vara lönsamt att satsa på dem. Även om de publicerade oddsen kanske inte återspeglar spelbolagens sanna sannolikhetsuppfattningar, kan de fortfarande ses som sannolikhetsbedömningar av ett sportevenemangs slutresultat, eller med andra ord, som prognoser (Štrumbelj & Šikonja, 2010). Makropoulou & Markellos (2011) anser att spelare som slår vad kan delas i tre olika typer baserat på hur mycket information de har. Det finns tillfälliga spelare som inte är så insatta, experter och privatinformerade spelare som kallas för insiders.

6.3 Empirisk studie

Som studiens metod valdes Grangers kausalitetstest samt Pearsons korrelationstest. Korrelationstestet utförs för att mäta hur oddsen rör sig i förhållande till de valda indikatorerna, som är den Officiella världsgolfrankingen, Strokes Gained Total och Strokes Gained Approach. Pearsons korrelation visar styrkan i sambandet mellan två variabler, men den visar inte styrkan av orsakssamband eller huruvida variablerna påverkas av några andra faktorer. Pearson-korrelationsmetoden mäter korrelationen mellan värdena -1 och 1 (Nettleton, 2014). Korrelation informerar om det linjära sambandet mellan variablerna (Nettleton, 2014). Förhållandet mellan speloddsen och de tre indikatorerna observerades med Pearsons korrelationskoefficient.

Grangers kausalitet används för att avgöra om en tidsserie är användbar för att förutsäga en annan tidsserie. Grangers kausalitet kunde förklaras på följande sätt: om fördröjda värden på X hjälper till att förutsäga nuvarande värden på Y i en prognos som bildas från fördröjda värden på både X och Y, så sägs X att Granger-orsaka Y (Thurman & Fisher, 1988). Grangers kausalitet studerar kausaliteten av tidsserier i ett kort tidsintervall. Nollhypotesen för ett Grangers kausalitetstest är att det inte finns någon kausalitet att hitta mellan två tidsserier. Kontrahypotesen är att tidsserien Y orsakar rörelsen av tidsserien X (Malkamäki, 1992).

6.4 Resultat

Pearsons korrelationstestresultat bevisar att spelarens position på världslistan är det måttet som oftast korrelerar starkast med de givna oddsen. Somliga spelare med höga Strokes Gained-värden har Strokes Gained Total som det mest korrelerande måttet med de givna oddsen. Strokes Gained Approach var aldrig det mest korrelerande måttet med oddsen men dess korrelationskoefficient har ständigt blivit starkare under årens lopp. Det samma fenomenet har inträffat med Strokes Gained Totals korrelationskoefficient.

Granger kausalitetstestet gav varierande svar på alla de 23 spelarna som hittades i datats 19 tidpunkter. De spelarnas testresultat var av fyra olika slag; att de givna oddsen Granger-orsakades av inget mått, ett mått, två mått eller alla tre måtten. Oddsen av spelarna inom världslistans top 10 Granger-orsakades nästan alltid av måttet OWGR. Oddsen av de spelarna som var rankade utanför världslistans top 10 var Granger-orsakade av Strokes Gained måtten.

7 References

Adepoju, K. A., Chukwu, A. U., & Shittu, O. I. (2016). On the Kumaraswamy Fisher Snedecor Distribution. *Mathematics and Statistics*, *4*(1), 1-14.

Ali, M. M. (1977). Probability and utility estimates for racetrack bettors. Journal of political Economy, 85(4), 803-815.

BELLMAN, R., & Dreyfus, S. (2010). *Dynamic Programming* (Vol. 33). Princeton University Press. https://doi.org/10.2307/j.ctv1nxcw0f

Broadie, M., & Ko, S. (2009, December). A simulation model to analyze the impact of distance and direction on golf scores. In Proceedings of the 2009 Winter Simulation Conference (WSC) (pp. 3109-3120). IEEE.

Broadie, M., & Rendleman Jr, R. J. (2013). Are the official world golf rankings biased?. Journal of Quantitative Analysis in Sports, 9(2), 127-140.

Brown, M. T., Rascher, D. A., McEvoy, C. D., & Nagel, M. (2006). Treatment of travel expenses by golf course patrons: Sunk or bundled costs and the first and third laws of demand. *Available at SSRN 1690896*.

Busche, K., & Hall, C. D. (1988). An exception to the risk preference anomaly. Journal of business, 337-346.

Cooper, D. R., & Schindler, P. S. (2011). Chapter 6 Research Design: An Overview.

Business Research Methods, pp 121-141, Twelfth edition. Published by: McGraw Hill / Irwin ISBN 978-0-07-352150-3

Davies, T. C., Collins, D., & Cruickshank, A. (2014). So what do we do with the rest of the day? Going beyond the pre-shot routine in professional golf. International Journal of Golf Science, 3(2), 163-175.

Deheshti, M., Adabi Firouzjah, J., & Alimohammadi, H. (2016). The relationship between brand image and brand trust in sporting goods consumers. *Annals of Applied Sport Science*, *4*(3), 27-34.

Dobson, S. M., & Goddard, J. A. (1998). Performance and revenue in professional league football: evidence from Granger causality tests. *Applied Economics*, *30*(12), 1641-1651.

DP World Tour. (1.4.2023). *Exemption categories.* https://www.europeantour.com/dpworld-tour/players/exemption-categories/

Eviews 5 User´s Guide. 2004. USA: Quantitative Micro Software.

Fedorcik, G. G., Queen, R. M., Abbey, A. N., Moorman III, C. T., & Ruch, D. S. (2012). Differences in wrist mechanics during the golf swing based on golf handicap. *Journal of Science and Medicine in Sport*, *15*(3), 250-254.

Figlewski, S. (1979). Subjective information and market efficiency in a betting market. Journal of Political Economy, 87(1), 75-88.

Finn, J. (2008). An introduction to using mental skills to enhance performance in golf: Beyond the bounds of positive and negative thinking. International Journal of Sports Science & Coaching, 3(1_suppl), 255-269.

Flepp, R., Nüesch, S., & Franck, E. (2016). Does bettor sentiment affect bookmaker pricing?. Journal of Sports Economics, 17(1), 3-11.

Galavotti, M. C. (2001). Subjectivism, objectivism and objectivity in Bruno de Finetti's Bayesianism. In Foundations of Bayesianism (pp. 161-174). Springer Netherlands.

Gilden, D. L., & Wilson, S. G. (1995). Streaks in skilled performance. Psychonomic Bulletin & Review, 2, 260-265.

Griffith, R. M. (1949). Odds adjustments by American horse-race bettors. The American Journal of Psychology, 62(2), 290-294.

Hall, S., Szymanski, S., & Zimbalist, A. S. (2002). Testing causality between team performance and payroll: The cases of Major League Baseball and English soccer. *Journal of sports economics*, *3*(2), 149-168.

Harville, D. A. (1973). Assigning probabilities to the outcomes of multi-entry competitions. Journal of the American Statistical Association, 68(342), 312-316.

Hirotsu, N., & Wright, M. (2004). Modeling a baseball game to optimize pitcher substitution strategies using dynamic programming. Economics, Management and Optimization in Sports, 131-161.

Ioulianou, M., Christofi, I., Talattinis, K., Chalkias, K., & Stephanides, G. (2011). Automatic identification of low-risk trading opportunities in sport betting exchange markets. BALCOR 2011, 108.

JAGIELSKI, M., OGASAWARA, Y., YONEYAMA, T., IKUJI, T., TSUCHIYA, H., & SUGO, T. (2022). Transfer spending and sports performance on the example of selected European football clubs. The Granger causality approach. *Journal of Physical Education & Sport*, *22*(9).

Jane, W. J., San, G., & Ou, Y. P. (2009). The causality between salary structures and team performance: A panel analysis in a professional baseball league. *International Journal of Sport Finance*, *4*(2), 136-150.

Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta mathematica, 30(1), 175-193.

Levitt, S. D. (2004). Why are gambling markets organised so differently from financial markets?. The Economic Journal, 114(495), 223-246.

The Masters. 1.3.2023. *Hole no. 3.* https://www.masters.com/en_US/course/hole3.html

Makropoulou, V., & Markellos, R. N. (2011). Optimal price setting in fixed-odds betting markets under information uncertainty. Scottish Journal of Political Economy, 58(4), 519-536.

Malaska, P., & Virtanen, I. (2007). Uskottavaan päättelyyn perustuvasta todennäköisyysteoriasta. Tutkijasta johtajaksi, 33-59.

Malkamäki, M. 1992. ”Cointegration and Causality of Stock Markets in Two Small Open Economies and Their Major Trading Partner Nations”. Bank of Finland discussion papers, 16/92.

McMillen, J. (Ed.). (2005). Gambling cultures: Studies in history and interpretation. Routledge.

Narayan, P. K., & Smyth, R. (2003). Attendance and pricing at sporting events: empirical results from Granger causality tests for the Melbourne Cup. *Applied Economics*, *35*(15), 1649-1657.

Nettleton, D. (2014). Commercial data mining: processing, analysis and modeling for predictive analytics projects. Elsevier.

The Official World Golf Ranking. (3.4.2023). *About us*. https://www.owgr.com/about-owgr

The Official World Golf Ranking. (5.4.2023). *Archive*. https://www.owgr.com/archive

The Official World Golf Ranking. (3.4.2023). How the ranking system works. https://www.owgr.com/how-the-ranking-works

PGA Tour. (3.4.2023). *Scoring Average.* https://www.pgatour.com/stats/detail/120

PGA Tour. (3.4.2023). *Putts Per Round*. https://www.pgatour.com/stats/detail/119

PGA Tour. (5.4.2023). *Official Money*. https://www.pgatour.com/stats/detail/109

PGA Tour. (10.4.2023). *Priority Rankings*. https://www.pgatour.com/players/priority-rankings

Rana, S. S. (2016). Influence of Annual Bluegrass on Putting Green Trueness and Control of Weedy Poa Species in Kentucky Bluegrass and Creeping Bentgrass Turf (Doctoral dissertation, Virginia Tech).

Smith, A. C., Roberts, J. R., Kong, P. W., & Forrester, S. E. (2017). Comparison of centre of gravity and centre of pressure patterns in the golf swing. *European journal of sport science*, *17*(2), 168-178.

Snyder, W. W. (1978). Horse racing: Testing the efficient markets model. The Journal of finance, 33(4), 1109-1118.

Štrumbelj, E., & Šikonja, M. R. (2010). Online bookmakers' odds as forecasts: The case of European soccer leagues. International Journal of Forecasting, 26(3), 482-488.

Suuraho, V. (2002). Kioskilta kotitietokoneelle - Urheiluvedonlyönnin muuttuvat palvelukanavat. Markkinoinnin laitos. Pro Gradu -tutkielma. Helsingin Kauppakorkeakoulu.

Thurman, W. N., & Fisher, M. E. (1988). Chickens, eggs, and causality, or which came first. American journal of agricultural economics, 70(2), 237-238.

Vuoksenmaa, J. (1999). Urheiluvedonlyönti – Voittajan opas. Gummerus Kirjapaino Oy. Jyväskylä.

VÄNNI, M., & VIRTANEN, M. (2005). Suomalaisten vedonlyöjien käyttäytyminen-sijoittajia vai pelureita? (Master's thesis).

Wooldridge, J. M. (2015). Introductory econometrics: A modern approach. Cengage learning.

World Scientific Congress of Golf & Cochran, A. J. (1992). *Science and golf*. E & FN Spon : Chapman & Hall.

Xu, J. S. (2011). ONLINE SPORTS GAMBLING: A LOOK INTO THE EFFICIENCY OF BOOKMAKERS'ODDS AS FORECASTS IN THE CASE OF ENGLISH PREMIER LEAGUE. (Unpublished undergraduate dissertation). University of California, Berkeley.