

2018

Predictive Golf Analytics Versus the Daily Fantasy Sports Market

John O'Malley

Recommended Citation

O'Malley, John, "Predictive Golf Analytics Versus the Daily Fantasy Sports Market" (2018). *CMC Senior Theses*. 1969.
http://scholarship.claremont.edu/cmc_theses/1969

This Open Access Senior Thesis is brought to you by Scholarship@Claremont. It has been accepted for inclusion in this collection by an authorized administrator. For more information, please contact scholarship@cuc.claremont.edu.

Claremont McKenna College

Predictive Golf Analytics Versus the Daily Fantasy Sports Market

submitted to
Professor Eric Hughson

by
John H. O'Malley

for
Senior Thesis in Economics
Fall-Spring 2018
April 19, 2018

Acknowledgments

Thank you to the PGA Tour ShotLink Intelligence Program for their invaluable assistance, it was greatly appreciated.

Thank you to all of my professors who have helped me along the way, especially Professor Eric Hughson for his guidance throughout this project.

Thank you to my family and friends who have supported me throughout my life and academic career.

Table of Contents

Section I: Introduction	5
Section II: Literature Review	10
Section III: Data Review	14
Section IV: Empirical Process	23
Section V: Results	34
Section VI: Conclusion	72
Appendix A: References	74
Appendix B: Table Index	79

I. Introduction

Judging athletes by statistics has been common practice since the inception of professional sports leagues. In recent years, however, advanced statistics have become a part of the sports fan's vernacular. In the 1980's, Bill James created "sabermetrics," which is essentially a statistical analysis of baseball. The goal of sabermetrics is to identify which player attributes and baseball strategies contribute most directly to winning games as a team. The value of these statistics is being able to determine the market value of various skills a baseball player can demonstrate.¹ "Sabermetrics" came to the forefront through Michael Lewis's book *Moneyball*, which described how the low budget Oakland Athletics judged players to maximize expected wins and still fit their payroll.² Since *Moneyball* was published in 2003, advanced statistics have exploded onto the scene across all major U.S. sports, including golf. Indeed, in 1999, the PGA tour decided that advanced technology was needed to track statistical performance, so they began the ShotLink program.³ Until 2005, the system was only for PGA tour insiders use, when the ShotLink Intelligence program began to allow access to professors and Ph.D. students.⁴ The data supplied in the ShotLink program contains common statistics, as well as advanced statistics called Strokes Gained, which date back to the 2004 PGA season. Advanced statistics are generally used

¹ Rob Neyer, "Sabermetrics," *Encyclopedia Britannica.com*, accessed 17 April 2018, <https://www.britannica.com/sports/sabermetrics#ref1182350>.

² Benjamin Baumer and Andrew Zimbalist, *The Sabermetric Revolution: Assessing the Growth of Analytics in Baseball*, 2018, *University of Pennsylvania Press*, accessed 17 April 2018, <http://www.upenn.edu/pennpress/book/15168.html>.

³ "ShotLink Background," *ShotLink.com*, accessed 17 April 2018, <http://shotlink.com/about/background>.

⁴ "ShotLink Intelligence," *PGAtour.com*, accessed 17 April 2018, <https://www.pgatour.com/stats/shotlinkintelligence/overview.html>.

to judge a player's true level of performance; however, they are just as useful when attempting to predict a player's future outcomes.

In this study, advanced PGA statistics will be used to create a predictive model for a player's score at a certain course. This model will be used to attempt to see if the daily fantasy sports market for golf is efficient, specifically by testing if the players selected based on the model can return a profit on DraftKings PGA contests. The PGA events that will be analyzed are full field, 120 players or greater, stroke play events.

The PGA was established in 1916 by Rodman Wanamaker in New York. It was created to grow the game of golf, by hosting tournaments and employing professional golf instructors at clubs.⁵ In 1968, the PGA Tour began as a subsection of the PGA, which was for touring professionals instead of club professionals. The PGA is the largest professional golf tour in North America, as it runs most week-to-week professional golf tournaments. In 2018, there will be 47 tournaments hosted by the PGA Tour. Individual PGA tournaments are held annually, generally being played at the same course year-after-year. PGA tour events usually host 144 players, who compete for four days in what is called stroke play. Stroke play is simply a competition in which the player who takes the least total strokes wins. Each day, a player will play a round of eighteen holes, and after two days, rounds, about half of the players will be cut from the event. The top 70 players, including ties, will complete four rounds, and the individual who has taken the least strokes over the totality of the four days will be named the winner. Certain events have modified rules, for example the Career Builder Challenge has a cut after three days and is played on

⁵ "PGA Is Formed," *History.com*, 2009, accessed 17 April 2018, <https://www.history.com/this-day-in-history/pg-a-is-formed>.

multiple courses, while the WGC-Dell Match Play is a match play event instead of stroke play (meaning players compete head to head to move on in a seeded bracket). This paper will focus on the standard tournaments.

The nature of golf is that each individual player is only indirectly competing against one another, which is in stark contrast to most other professional sports. To win a tournament, a golfer must shoot a lower score than all of his competitors, but there is nothing a competitor can do to effect the play of any other individual. There is not an offensive and defensive side to a golf tournament, as there is in baseball, basketball, football, hockey, and soccer. Golf pits a collection of individuals against a course. This should make statistical golf predictions more accurate than those of other sports. For example, to predict the outcome of a baseball game there are many factors one must forecast: How will the starting pitchers pitch? How will the defense play behind them? Who will pitch after the starters? How well will whoever pitches after the starters pitch? How will each individual hitter hit? The amount of possibilities that must be taken into consideration is vast. In Golf, though, there is one question: Who will play best over the course of four days? The “defense” is the course, and the player is on the offensive always, as they attempt to shoot the lowest score possible. The “defense” in this case does not have to be forecasted for as it would be in the other sports mentioned above, since it is a known quantity. Most PGA Tour courses are played for multiple years, allowing for data to be collected and analyzed. By analyzing said data, a course profile can be created, which is entirely predictable from year-to-year. Since the courses are known, it should be possible to predict which players will most likely play well at a given tournament, which can be valuable information for someone trying to make money in the emerging Daily Fantasy

Sports industry. With all this said, the idea that golf could be more predictable because of less outside factors does not necessarily mean it is easier to win money in the fantasy sports market, as one can assume that this would be an advantage for all skilled competitors.

Fantasy Sports began, loosely, in 1962 with rules for how fantasy football could work being laid out. In 1963 the first draft occurred, with members of the Oakland Raiders organization picking players from the NFL to make their own “fantasy” teams and compete against each other based on how their drafted players perform on the field. By 1980, Fantasy Football Leagues had become public and the idea of fantasy sports had spread to baseball as well. With the internet boom in the 1990’s, fantasy sports went online and spread rapidly. In 2006, the Unlawful Internet Gambling Enforcement Act (UIGEA) became law, which took down the online poker industry, while allowing for fantasy sports contests as they were deemed a game of skill and not chance. The language of the UIGEA did not stipulate a difference between fantasy sports contests which lasted for the length of the season and those that were solely for a given day. This led to the rise of Daily Fantasy Sports. FanDuel was founded in 2009, which was a platform designed for fans to pick a roster of players competing on a given day in baseball, basketball, or football and wager money on their lineup against other users of the site. Shortly after FanDuel was founded, DraftKings was started in 2011, as their main competitor.⁶ Both companies are now valued at over a billion dollars.⁷ As the companies’ user bases grew so did their creativity, with

⁶ Nico Newman, “History of Fantasy Sports,” 4 April 2017, *Fantasy-Sport.net*, accessed 17 April 2018, <https://fantasy-sport.net/history-of-fantasy-sports/>.

⁷ Adam Kilgore, “Daily Fantasy Sports Websites Find Riches in Internet Gaming Law Loophole,” 27 March 2015, *The Washington Post*, accessed 17 April 2018, https://www.washingtonpost.com/sports/daily-fantasy-sports-web-sites-find-riches-in-internet-gaming-law-loophole/2015/03/27/92988444-d172-11e4-a62f-ee745911a4ff_story.html?noredirect=on&utm_term=.76dfb8653e45.

multiple different types of contests and additional sports added. DraftKings will be the focus of this paper, as they were the first to offer Daily Fantasy Golf contests.

DraftKings offers many contests, but they can be grouped into two main categories: Cash games and Guaranteed Prize Pool (GPP) games. Cash games are those with a greater chance of winning, but with smaller overall prizes. Guaranteed Prize Pool contests are large tournaments, with often thousands of players, in which only the top 20 or so percent make money.⁸ The payout scale is exponential, however, with the winner able to make many thousand's times their contest entry, for example the PGA Millionaire Maker Tournaments pay \$1,000,000 to the winner, with a lineup entry cost of only \$20. DraftKings establishes a limit to the amount of entries a single person can place in a contest, with some contests allowing up to 150 entries. This study is focused on GPP tournaments, specifically the \$3 entry fee, 150 lineups maximum, PGA Tournaments offered weekly.

Each \$3 entry is a ticket to construct a lineup of golfers under a given salary cap. DraftKings has a \$50,000 salary cap for players, who they price on a scale generally from around \$6,000 to \$14,000 based on DraftKings' ranking of their ability. Each lineup must consist of six golfers, and the total sum of the prices of the six must be \$50,000 or less.⁹

DraftKings Golf platform paired with the advanced statistics provided by the PGA Tour and the nature of the game of golf yields an opportunity to possibly beat the market in Daily Fantasy Sports and make a profit. This paper will look to analyze statistics of how each course on the PGA Tour plays in order to create a regression equation that will predict the player profile that should excel on the given course. The regression equation formed

⁸ "DFS Cash Games Versus Tournaments (GPP's)," 12 February 2016, *fantasysports.net*, accessed 17 April 2018, <https://www.fantasysports.net/dfs-cash-games-vs-tournaments-gpps>.

⁹ "Rules & Scoring," *DraftKings.com*, accessed 17 April 2018, <https://www.draftkings.com/help/rules/golf>.

by regressing past statistical results on a specific course to a player's score, coupled with the regression of binary variables about a player's history and form, should create a model that predicts which players will generally score the lowest at a given tournament. Using this information, players can be valued based on their DraftKings price and a group of players can be identified as good selections. By distributing these chosen players in different combinations throughout 150 lineups, hopefully, there will be a greater chance of placing highly in GPP contests and making a profit. Ultimately, this study does show a large positive profit, however, it is difficult to conclude success by the model with very limited observable results.

This study is organized as follows: Literature Review, Section II, which details past research on the predictability of golf through statistics, as well as how to value players and create optimal lineups to win DraftKings contests for sports other than golf; Data Review, Section III, which is an overview of the data used, detailing each variable and its importance; Empirical Process, Section IV, which explains step-by-step the general process for the creation of the predictive model for a certain tournament, as well as the process for valuing and selecting players to create 150 DraftKings lineups; Results, Section V, which gives a detailed explanation of the results of this study, looking at the most successful week individually, as well as the overall net gains/losses; and Conclusion, Section VI, which brings the results together and details the further research that could be conducted and the information that would be needed to improve this model.

II. Literature Review:

Success on the PGA tour is often defined by a player's overall earnings for the year, so there have been many empirical studies as to what a player's traits, statistics, yield the

greatest dollar value. Davidson and Templin (1986) was the first published research document that delved into the effect of different golf skills on a player's success, which they measured by earnings and season long scoring average. Their results showed that specific skill set differences had a greater effect on scoring average than earnings but that certain skills were clearly more beneficial.

In the years since Davidson and Templin (1986), many studies have been done to attempt to show the effect of certain skills on PGA performance, with putting and accuracy consistently being the skills most correlated to success. Some significant publications are Shmanske (1992), Finley and Halsey (2004), Alexander and Kern (2005), and Peters (2008), all of which are studies that show which statistics lead to the best year-long performance, meaning, in large part, which statistics yield consistency. There is significantly less published work on what may lead a golfer to be successful on any given week.

Shmanske (1992) observes strong putting to be the most significant characteristic of a successful golfer, in terms of earnings. As time goes on, the PGA tour and its courses evolve, with the main change in the past decades being increased length. Alexander and Kern (2005) attempted to review the previous publications claiming putting to be the key to earnings, with the thought that longer courses may put a higher premium on driving distance. Their results still showed putting to be the main contributing factor to earnings, despite it becoming marginally less so than in past years. Peters (2008) further corroborated the importance of putting on earnings, while also looking at the exterior factor of experience, which proved to have a positive impact as well.

Finley and Halsey (2004) looked into the effect of new stats, Bounce Back and Scrambling, on scoring average, while also looking at Simple Scoring Average versus Adjusted Scoring Average as a predictor of earnings. Simple Scoring Average is merely the average score of each round an individual plays over the course of the season, while Adjusted Scoring Average takes into account the average score of each player who played the round and adjusts it to see if an individual played better or worse than his competitors. Their finding that Simple Scoring Average was not highly correlated to earnings is significant, as in the past earnings and scoring had been used simultaneously as measures of success on tour. Adjusted Scoring Average is shown to be more important for earnings.¹⁰

More statistics evolved in the late 2000's to be used to determine a golfer's performance. Brodie (2008) and (2012) delved into new data being provided by the PGA tour, ShotLink data. The data was used to create a comparative metric for the relative value of a single putt and then extrapolated that number to accumulate the relative number of strokes gained or lost to the average player in a tournament. The idea of how many strokes could be gained or lost in relation to the average player in a tournament field, would become known as Strokes Gained statistics, which now are used for each shot on a golf course, broken into Off the Tee, Approaching the Green, Around the Green, and Putting.

Despite the array of work highlighting which golf skills most affect success, whether judged by scoring average or earnings, there is very little public research on which statistics yield success at any of the specific courses played annually on the PGA Tour.

¹⁰ David Scott Hunter, Juan Pablo Vielma, and Tauhid Zaman, "Picking Winners Using Integer Programming," *MIT.edu*, accessed 17 April 2018, <http://www.mit.edu/~jvielma/publications/Picking-Winners.pdf>.

With there being minimal work on course specific results for PGA tour players, there is no published work on how to predict performance for PGA Daily Fantasy sports. Daily Fantasy sports have exploded in the past decade behind leading companies, DraftKings and FanDuel. Both companies provide contests for PGA events, however, there is no work published on how to successfully profit off of said contests.

There is published work, however, on the merit of statistical modeling to profit off of DraftKings NBA contests. Barry, Canova, and Capiz recently completed a study to see if they could improve their chances of consistently winning money on DraftKings NBA by analyzing projected statistics relevant to DraftKings point scoring, as well as factors that could affect performance, such as rest and the opposing defenders. They managed to show improved accuracy for their projections when taking into account these factors.¹¹

Hunter, Vielma, and Zaman studied how to maximize the ability to win a contest with a top-heavy payout structure. They conducted this study using DraftKings Hockey and Baseball contests, in which a large percentage of the prize pool was paid out to the winner. Their hypothesis was that by putting in a large amount of entries, all of which having a large expected point value, a large volatility, and minimal correlation to each other, one would have the best chance of winning. Despite a small sample size, they yielded large enough winnings to not reject their hypothesis.¹²

My study will look to expand upon both research into PGA tour success and also Daily Fantasy Sports success. There are no published papers which focus on PGA Daily

¹¹ Christopher Barry, Nicholas Canova, and Kevin Capiz, "Beating DraftKings at Daily Fantasy Sports," *Stanford.edu*, accessed 17 April, 2018, <https://web.stanford.edu/class/stats50/files/BarryCanovaCapiz-paper.pdf>.

¹² Hunter, Vielma, and Zaman, "Picking Winners Using Integer Programming," *MIT.edu*, accessed 17 April 2018, <http://www.mit.edu/~jvielma/publications/Picking-Winners.pdf>.

Fantasy Sports. This study will provide new information as to which player statistics and other exterior factors affect a PGA player's success on a given course during a specific week, while also exploring how these results can be used to profit on DraftKings PGA contests.

III. Data Review

There were forty PGA Golf tournaments played during the 2017 Calendar season, however, the sample used for this project is much smaller. Only tournaments which hosted a full field of players, 144 or more, were considered. Furthermore, certain tournaments are not played on the same course each year, for example the three majors (US Open, Open Championship, and PGA Championship), which makes past years' statistics irrelevant to the coming year's event. Beyond changes in course and number of participants, DraftKings only provided the type of contest this model is created for (150 entry GPP) during the first portion of the season, before changing the number of entries allowed into their contests. Ultimately, there are nine tournaments, which fit the parameters necessary to test this hypothesis for conclusive results, and another six which were simulated and can be looked at to see general trends.

The data used was provided by the ShotLink® Intelligence Program, which began in 2005 and expanded in 2008 with a partnership with CDW. The program allows for professors and students to study advanced PGA statistics that are not made available to the public. The ShotLink database contains common statistics dating back many years, however, the highly advanced Strokes Gained statistics are only available since 2004. Strokes Gained statistics were developed by Mark Broadie of Columbia University, as a way of measuring a player's performance in specific skills against those of his competitors.

Strokes Gained Total takes a player's score for a round and compares it to the average score of the rest of the players in the competition during that round.¹³ The winner of a tournament will lead the field in Strokes Gained Total. Beginning in 2014, the PGA tour began to split Strokes Gained Total into two categories: Strokes Gained Tee-to-Green and Strokes Gained Putting.

For the purposes of this study, data before 2011 will not be examined. Individual tournament data from past years is only provided for those players who make the cut, with the cut being the top 70 players and ties, so for each year's hosting of the event there are 70+ data points. The previous three years' results at an event will be used to predict the current year, so for each regression there will be 210+ data points used.

There are eight independent variables used to predict the dependent variable, Scoring Average. Three binary variables: History, Form, and Weather, are regressed against the difference between historical projections of the individual's scoring averages from 2014-2016 and the true outcomes they achieved to further adjust the predicted dependent variable. These binary data points are collected from a review of the historical section of the ShotLink database, which shows individual player's finishing position results at events. The data ultimately input into the regression equation to predict an individual's score is the player's season long averages in the eight variables examined.

Dependent Variable – Adjusted Scoring Average: This is a weighted statistic of how an individual player scored with an adjustment for how the rest of the players in the field scored in the same round. The average score of each of the four rounds of the event

¹³ "Strokes Gained: How It Works," 30 May 2016, *PGATour.com*, accessed 17 April 2018, <https://www.pgatour.com/news/2016/05/31/strokes-gained-defined.html>.

will be subtracted from the course's par score, with the four resulting differences being added to the total strokes an individual took over the course of the tournament. The sum of the total strokes and these adjusting differences is then divided by the number of rounds played, four, yielding a weighted scoring average.¹⁴

Independent Variable 1 – Driving Distance: A distance measured in total average yards a player hits the ball off of the tee on all par 4 and par 5 holes, with the accuracy of the shot being ignored. The statistic attempts to show how far on average a player will hit the ball using a Driver. ShotLink uses GPS and laser measurement equipment to determine the total amount of yards a drive covers. In 2016, 85% of the shots used to determine a player's average driving distance were confirmed to be shots hit with a Driver, however, 15% were unconfirmed which club the player used to hit the ball off the tee. Not knowing what club was used by a player can skew the driving distance statistic, as a player who chooses to hit 3-Wood would have hit the ball farther had he chosen to use a Driver, yet the distance is attributed to his driving distance. Ultimately, this statistic is still the best measure of a player's ability to hit the ball a certain distance off of the tee.¹⁵

Independent Variable 2 – Driving Accuracy Percentage: A percentage of how many of a player's tee shots on par 4 and par 5 holes end up on the fairway. The statistic does not take into account the club hit off of the tee, so a player with a high percentage

¹⁴ RBC Heritage, "Statistics: Scoring Average," 15 April 2018, *PGATour.com*, accessed 17 April 2018, <https://www.pgatour.com/stats/stat.120.html>.

¹⁵ "A Review of Driving Distance – 2016," *USGA.org*, accessed 17 April 2018, <http://www.usga.org/content/dam/usga/pdf/Equipment/2016%20Distance%20Report.pdf>.

may not necessarily be hitting their Driver more accurately but may, in reality, just be using a different club.¹⁶

Independent Variable 3 – Strokes Gained Tee-to-Green: The sum of a player's Strokes Gained Off-the-Tee, Strokes Gained Approach-the-Green, and Strokes Gained Around-the-Green. Conversely, it is a player's Strokes Gained Total – Strokes Gained Putting. Strokes Gained Total on a hole is determined by a player's score minus the average score on all holes of the same distance. In other words, hypothetically, if a hole is 450 yards and the average score on holes of said length is 4.5, then a player who scores a 4 will have accumulated .5 Strokes Gained Total. The par of the hole has no affect. In general, a player's Strokes Gained Tee-to-Green represents how effectively a player is at getting the ball to the green at a distance closer than expected to the hole. The Strokes Gained Tee-to-Green for each hole is added up to determine Strokes Gained Tee-to-Green for a round. The season-long Tee-to-Green statistics are an average of each calculated round played. See Strokes Gained Putting below for more details on what is subtracted from Strokes Gained Total to find Strokes Gained Tee-to-Green.¹⁷

Independent Variable 4 – Strokes Gained Putting: A measure of the number of putts a player takes against the projected number of putts the average PGA tour player takes from a certain distance from the hole. For example, hypothetically, if a player is 20 feet from the hole and on average it takes a PGA Tour player 1.8 shots to get the ball in the hole from 20 feet, then a shot made from 20 feet would yield .8 Strokes Gained Putting.

¹⁶ RBC Heritage, "Statistics: Driving Accuracy Percentage," 15 April 2018, *PGATour.com*, accessed 17 April 2018, <https://www.pgatour.com/stats/stat.102.html>.

¹⁷ "Strokes Gained: How It Works," 30 May 2016, *PGATour.com*, accessed 17 April 2018, <https://www.pgatour.com/news/2016/05/31/strokes-gained-defined.html>.

The number of strokes gained or lost on each putt over the course of eighteen holes is accumulated to find a Strokes Gained Putting for the round. The season-long Strokes Gained Putting statistics are an average of a player's Strokes Gained Putting over the number of rounds played.¹⁸

Independent Variable 5 – Scrambling Percentage: A measurement of how likely a player is to make par or birdie after missing the green in regulation. To be on a green in regulation is being on the green in two strokes less than the par of the hole. To miss a green in regulation means that a player's third shot on a par 5, second shot on a par 4, or first shot on a par 3 lands off of the green. Scrambling percentage looks at every time a player is in such a position and finds the percentage of times the player still makes birdie, by making the shot from off of the green, or par, by making it in to the hole using just two shots from off of the green. The statistic emphasizes players who are good at chipping and pitching from around the green.¹⁹

Independent Variable 6 – Greens in Regulation Percentage: The total amount of times a player makes it onto a green in regulation divided by the number of holes played. As explained above under Scrambling Percentage, a green in regulation is a player being on the green in two strokes less than the par of the hole. This statistic highlights a player's ability to hit their irons or wedges onto the green.²⁰

Independent Variable 7 – Putts Per Round: The sum of the total number of putts a player hits divided by the number of rounds he has played. Does not take into account

¹⁸ Ibid.

¹⁹ Ibid.

²⁰ RBC Heritage, "Statistics: Greens in Regulation Percentage," 15 April 2018, *PGATour.com*, accessed 17 April 2018, <https://www.pgatour.com/stats/stat.103.html>.

distance of putt or how many strokes were hit before putting, both of which affect the total. Players who hit the green in regulation more frequently will likely take more putts, whether they are better putters or not, so the statistic can be skewed as to who the best putters truly are.²¹

Independent Variable 8 – Sand Save Percentage: A measurement of how likely a player is to make par or birdie after missing the green in regulation with the ball sitting in a sand bunker. Calculated in the same way as Scrambling Percentage, except with the condition of the ball being in the sand instead of just off the green in any location. This statistic highlights who is best at hitting the ball accurately out of a sand trap.²²

Binary Variable 1 – History: Certain players consistently play well at courses which do not fit their statistical profile. In order to take this into account, a player's past performances at a course must be taken into account. If enough data was collected, each individual who has played the event more than once could be projected for the given year and observed to either underperform or outperform their projection. Those who continually outperformed what was projected would be considered to have good history, which would be taken into account for the current year. However, not enough data is being analyzed in this study to be able to go back far enough to project each player's individual average over or under performance. With that being said, a player's history is still significant, so it must be incorporated in some other way. To do so, history is tracked for each past participant and then a standard linear regression is run on the sample of players with the independent

²¹ RBC Heritage, "Statistics: Putts Per Round," 15 April 2018, *PGATour.com*, accessed 17 April 2018, <https://www.pgatour.com/stats/stat.119.html>.

²² RBC Heritage, "Statistics: Sand Save Percentage," 15 April 2018, *PGATour.com*, accessed 17 April 2018, <https://www.pgatour.com/stats/stat.111.html>.

variable being a qualitative measure of a player's history and the dependent variable being the difference between the projection and real score. If the regression coefficient is statistically significant and negative, then it will be taken account into the final projection equation, as that means that players who were deemed to have good course history in the past have generally outperformed their projection by the given coefficient. The following chart dictates whether a player is given a 1 for good course history or a zero for poor history or none:

Table 1: Course History

Times Played	Made Cuts	Top 10	Top 5
1	1	1	1
2	2	1	0
3	2+	1	
4+	75%		

If a player has played the tournament four or more times, then they solely need to have made the cut in 75% or better of the time. A player who has played the tournament less times, though, needs to have performed better. Someone who has played only once must have finished in the top five to be labeled a 1 for good course history. One that has played the tournament twice must have made both cuts and finished in the top ten once. Lastly, someone who has played the tournament three times must have made the cut at least twice and must have a top ten finish.

Binary Variable 2 – Current Form: Certain players may not be statistically suited for a certain course nor have played the course well in the past, however, if a player is

playing extremely well in past weeks, they may continue their hot streak. The “hot hand fallacy” is often discussed by sports statisticians, as statistically players don’t get “hot,” since streaks of success are shown to be merely random occurrences that are to be expected from a large sample size.²³ In order to examine the “hot hand” in golf, current form will be used as a binary variable. A player’s form is still possibly significant, so it must be incorporated in some way. To do so, form is tracked for each past participant and then a standard linear regression is run on the sample of players with the independent variable being a qualitative measure of a player’s current form and the dependent variable being the difference between the projection and real score. If the regression coefficient is statistically significant and negative, then it will be taken into account in the final projection equation, as that means that players who were deemed to have good current form leading into the tournament in the past have generally outperformed their projection by the given coefficient.

Since most PGA Tour golfers do not play every week, multiple weeks will have to be looked at to determine who has been playing well recently. For the purpose of this study, the past six weeks will be viewed and a player must have played in three of them to achieve a 1. The following chart dictates the scenarios in which players will receive a 1 for good current form:

²³ Gary N. Curtis, “The Hot Hand Fallacy,” *fallacyfiles.org*, accessed 17 April 2018, <http://www.fallacyfiles.org/hothandf.html>.

Table 2: Current Form

Times Played Past 6 Weeks	Made Cut	Top 10
3	2	2
3	3	1+
4	3	1+
4	4	0+
5	3	2+
5	4	1+
6	4	2+
6	5	1+
6	6	0+

Binary Variable 3 – Weather: Rain and wind can have a serious impact on PGA Tour events. Certain players are more equipped to handle these challenges, so this needs to be taken into account if possible. The problem is that weather is hard to predict, with forecasts often being wrong. For the purpose of this study, tournaments that are expected to have wind over 20 mph on three or four of the days, and tournaments that are given a 75% chance or greater of rain on the first two days will take weather into account. The main issue with wind and rain is the affect it has on the ball when it is in the air. The longer and higher a ball is traveling through the air the larger the effect of the inclement weather, so players who naturally hit the ball lower are at an advantage.²⁴ When weather is deemed to be a factor, players who have an average apex height of their shots in the bottom 50% of the field will be given a 1 that indicates good play in bad weather. By looking at past tournaments that had weather as a factor, player's binary score for weather play based on apex height can be regressed on the difference between their score and their projection. If the result of this regression shows that apex height is significant to performance during the

²⁴ Butch Harmon, "Playing Great on Windy Days," 12 February 2012, *GolfDigest.com*, accessed 17 April 2018, <https://www.golfdigest.com/story/butch-harmon-windy-days>.

past tournaments with inclement weather, then it will be added into the projection for the coming week with forecasted bad weather.

IV. Empirical Process

Executive Summary:

This study will create a model for identifying a portfolio of players to select for weekly DraftKings PGA Golf tournaments. The process will first take historical predictions of the past three years, in order to see how similar the projected scores were to reality. The same standard linear regression equation that is used to predict the past three seasons will also be the beginning of our predictive equation for the 2017 event, before being adjusted to take into account binary variables. The binary variables will be found to be significant or not based on if they had significance when regressing them against the difference between projected and real past results. Once the binary variables have been incorporated into the base projection equation for 2017, the player's statistics for the current season will be input to find an expected score. Twenty-five players will be selected based on their projected scores and the value of them based on their given DraftKings' prices. Selected players will be randomized into 150 lineups to meet the DraftKings Salary Cap and the results will be judged based on the net returns.

1.1 - Historical Predictions

The first step in predicting a player's performance in a specific upcoming PGA tournament is to look at past results. Tournaments' names change frequently with the change of sponsors; however, the vast majority of PGA events are played at the same course for a substantial number of years. The fact that many tournaments are always played on the same course makes past statistics very significant to the event in any given year.

Even when the players participating in an event change, the skills that are most necessary for success on a certain course, at a specific tournament, should remain unchanged over the years. The first step for predicting the success of participants in a coming year's tournament is to attempt to predict the performance of players in the event in past years. For this model, the previous three year's tournament results will be used to predict any desired year. This test was done for the 2017 PGA Tour season, however, historical models were made for the 2014, 2015, and 2016 seasons as well, so 2011-2016 data is relevant.

The general model for the standard linear regression used to predict success will have the dependent variable of player n 's Scoring Average and the independent variables being player n 's performance in the statistics listed in the data section: Driving Distance (DD), Driving Accuracy (DA), Strokes Gained Tee-to-Green (SGTG), Strokes Gained Putting (SGP), Scrambling Percentage (SCRAM), Greens in Regulation (GIR), Putts/Round (PR), and Sand Save Percentage (SAND). Taking the data from 2011, 2012, and 2013, the linear regression should produce an equation that will predict how a player will score in the 2014 tournament:

$$\begin{aligned} SCR_{YEAR} = & \alpha + \beta_{DD}DD + \beta_{DA}DA + \beta_{STG}SGTG + \beta_{SGP}SGP + \beta_{SCRAM}SCRAM \\ & + \beta_{GIR}GIR + \beta_{PR}PR + \beta_{SAND}SAND + \varepsilon \end{aligned} \tag{1}$$

If certain variables are not shown to be significant, then they are removed from the data set, and the regression is run once more, but this time with fewer independent variables. Once an equation is found for 2014 with each statistic included being significant, the process will be repeated again using the data from 2012, 2013, and 2014 to predict the

2015 event, the data from 2013, 2014, and 2015 to predict the 2016 event, and finally from 2014, 2015, and 2016 to predict the upcoming 2017 event. These four regression equations (SCR_{2014} , SCR_{2015} , SCR_{2016} , SCR_{2017}) are the first step in predicting the 2017 results. The historical equations are necessary for finding how much of a difference there was between the projection and reality in 2014, 2015, and 2016, so, hopefully, other variables can be added that will take into account some of this error. The 2017 equation is the base equation for predicting the current year before being adjusted with binary variables.

1.2 – Independent Variable Inputs

Once equations are in place to predict a player's performance for a given year, inputs must be decided on. Players' season long statistics will be used from 2013, 2014, 2015, and 2016 to predict specific events in 2014, 2015, 2016, and 2017, when the event takes place during the first ten events of the calendar year, as players have not played enough leading into those first events for their season long averages to be indicative of their true skill set. By looking at the previous year, we are given a wider look at a player's skills, however, we do sacrifice the possibility that a player has improved or worsened significantly since the end of the past season, which is always possible in sports. For the 11th event of the season and on, season long averages for the year of the event in question will be used and input into the SCR equation. Accepting that players' season long averages are good predictors of their performance over multiple rounds still leaves a large margin for error, however, it is the best way to define their skills. The result of inputting season long average statistics for each individual should yield a projected score above or below the tournament average based on how well the individual tends to perform in the significant skills needed to optimize performance.

1.3 – Binary Variables and Adjusting Projection

Using past data becomes significant for checking how well the model fits to the actual results. The effectiveness of the linear regression equation can be judged by comparing the projected individual results for SCR_{2014} , SCR_{2015} , SCR_{2016} versus what the player's actual Scoring Average was. This comparison can be expected to yield a small positive or negative difference for each player depending on if they outperformed or underperformed on their season averages during the given week's tournament. Once these differences are found, the question is why an individual may have outperformed or underperformed from their season statistics. There are many factors that may play a role, which are not quantifiable. Some examples of factors that almost certainly affect a player are illness, travel/time change, sleep, family or personal problems, motivation, etc. There are however other factors, which despite the PGA not offering directly as statistics, can be used to try to diminish the error in the predicted Scoring Average and the player's actual one. The three that will be focused on are all binary variables, which are explained in the Data section above, Course History, Current Form, and Weather. To determine the effect, if any, that these variables may have on a player under or out performing his projected score, a simple linear regression will be run with the dependent variable being player n 's projection error (Real SCR – Projected SCR) and the independent binary variables Course History, Current Form, and possibly Weather. The equation is:

(2)

$$DIF = \alpha + \beta_{HIST}HIST + \beta_{FORM}FORM + \beta_{WEATHER}WEATHER + \varepsilon$$

The variables must be in the 95% confidence interval to be considered for the final equation. Along with being significant at the 95% level, the variables must also decrease the average error between projection and real score for the past three years. To check if this is the case, first the average error will be calculated based on our past results and then the past projections must be altered to account for the binary variables effect. To account for their effect, a player's projected score will have $\beta_{HIST}HIST + \beta_{FORM}FORM + \beta_{WEATHER}WEATHER$ added to it, with HIST, FORM, and WEATHER being 0 or 1 depending on n 's individual performances. The adjusted past projected scores are then subtracted from the player's real score from that year's event to find an individual's new projected difference, error. The adjusted errors are then averaged, and the average is compared to the average error before accounting for the binary variables. If the newly found adjusted average difference is less than the original average difference, then the binary variables effects will be included in the predictive equation for 2017. The equation to predict 2017 maintains the initial betas but now includes any significant binary variables as well:

(3)

$$\begin{aligned} SCRADJ_{2017} = & \alpha + \beta_{DD}DD + \beta_{DA}DA + \beta_{STG}SGTG + \beta_{SGP}SGP + \beta_{SCRAM}SCRAM \\ & + \beta_{GIR}GIR + \beta_{PR}PR + \beta_{SAND}SAND + \beta_{HIST}HIST + \beta_{FORM}FORM \\ & + \beta_{WEATHER}WEATHER + \varepsilon \end{aligned}$$

The 2016 or 2017 season long statistics will be input for the independent variables depending on when the event takes place, as stipulated above, to produce what should be a

more precise projected scoring average than the initial SCR_{2017} equation would have yielded.

1.4 – Player Selection

Once the predicted scoring averages are found based on the equation above, it is time to figure out how to select the best core, portfolio, of players to use on DraftKings. DraftKings typically posts its different contests, as well as player prices, on the Monday before a tournament, so this step must be completed shortly before the tournament tees off. As is discussed above, there are many different types of games on DraftKings, however, for the purpose of this experiment the large guaranteed prize pool contests will be the focus. In DraftKings largest weekly PGA contests, an individual player is allowed to submit up to 150 lineups. Since payouts decrease exponentially from the top, with the top ten finishing lineups securing a massive percentage of the overall winnings, the best way to return a profit is to maximize your chance of having one or more lineups fall in that top ten. In order to maximize potential winnings, this model will be used to create 150 distinct lineups each week.

Each lineup on DraftKings is made up of six golfers, with a total salary cap of 50,000, so when constructing 150 lineups, there must be 900 players selected at a maximum cost of \$7,500,000. Players are given a different price on DraftKings each week. The 900 players must be made up of a portfolio that takes into account the individual's projection, expected ownership, and price. One of the main difficulties is deciding how many players to pick and how many to fade entirely. For this model, 25 players will be selected each week. Due to price constraints and errors, certain players will not be picked or will be picked despite how they rank based on the SCRADJ equation.

To maintain balanced teams, which are not solely based around the top priced players and the highly volatile bottom priced ones, there are restrictions placed on how many expensive players will be owned on any given week. The rules are as follows:

1. Only 3 Players above \$10,000 can be owned;
2. 2 players over \$11,000 can be owned when players in the field are priced over \$12,000; and
3. If highest priced player in the field is <\$12,000 then only own one player above \$11,000.

These stipulations limit the amount of high priced players that will be picked. This eliminates the possibility of fully diversifying among the top players, who all could hypothetically win a tournament any week, however, this is accepted, as to fully benefit from a player's performance the player must be owned in a large percentage of lineups. What is meant by this is that to beat the field on average an individual must have a higher percentage of a player in his 150 lineups than is owned by the contest participants as a whole. For example, if Player A produces the most points on DraftKings and is included in ten of the 150 lineups you have entered, while Player A is owned in 50% of all lineups entered in the contest, then Player A's success is likely hurting your portfolio of lineups as a whole. By selecting only a few of the high priced options, an individual can have them highly concentrated throughout the 150 lineups, allowing for a massive advantage when the players selected play up to their projection.

Along with the stipulations that reduce the players allowed to be picked for this model, there are also qualitative measures for picking certain players, who for some reason are not projected to perform well at the event. These rules are purely based off of

DraftKings being a predictive model itself, and one that frequently has errors. Players who are priced under \$7,000 on DraftKings but that reside in the top 50 of the Official World Golf Rankings (OWGR) will be automatically selected to be one of the 25 players used. A low price on a top 50 player can generally be attributed to recent poor play or time off, however, said players are deemed by this model too talented to not be selected when near the minimum price.

Now the remaining 25 spots must be filled. The top 15 lowest projected scoring players based on SCRADJ will be selected, taking into consideration the rules stipulated above. Generally this yields fairly high priced players, as better players tend to be projected to score better on any given week. In order to pick cheaper players to fill out the rest of the player portfolio a simple value ratio is used, which reads:

(4)

$$\text{VALUE} = \text{DK PRICE} / (\text{PROJRANK} / \text{DKRANK})$$

DK PRICE, DraftKings price, is given. PROJRANK, projection rank, is a 1 to 144 ranking based on the projected scoring average from SCRADJ for each player in the field. DKRANK, DraftKings rank, is a rank from highest to lowest price, with the highest price being 1 and any players of equal price being tied for the same rank. For example, if the top three players were priced \$11,000; \$10,000; \$10,000, then the \$11,000 player would be DraftKings Rank 1, while the two \$10,000 players would each be DraftKings Rank 2. This value ratio will highlight players who are less expensive on DraftKings than the projected score would indicate. The remaining spots from the initial 25 are filled by the top ranked VALUE players. This equation could yield flaws based on scale if comparing players of

drastically different prices/ranks, however, because it is only selecting the final ten players who will be chosen, each is generally from the lowest price range, so the issue of scale is less of a concern.

1.5 – Weighting

Above touched upon the idea of needing to have a higher concentration of a player than the overall participants have in order to gain an advantage if said player performs well. The determination of how much of each player to own is tricky, since it cannot be known exactly how popular a pick a player will be on a certain week. For this reason, to simplify the process, instead of looking at individual player's average ownership, this study will focus on price tier ownership. Historically, players on DraftKings who are priced higher will also be owned at a higher percentage on average. This is in large part because there are less players in the top price tiers than in the lower ones. While there may only be four players above \$10,000, it is likely there are around ten players in the \$8,000's, more in the \$7,000's, and even more in the sub \$7,000 range. This unequal dispersion causes less differentiation among high priced players and more among low priced ones, so naturally high priced players are owned in a larger amount of lineups, while the low priced ones are more randomly strewn throughout. Once determining projected ownerships by tier, there must be an adjustment to have a higher percentage in the 150 lineups being created. In other words, to gain an advantage a higher percentage of each player will be held than is projected to be held by the total entries to the contest. To go massively overweight the goal will be to have twice the percentage amount of each player picked. The ownership percentages that will be used for the projection for the overall contest and also in the 150 lineups being created are as follows:

Table 3: Ownership Percentages

Player Price	Expected Ownership %	Portfolio Ownership Percent
>\$9,990	20%	40%
\$8,900<Price<\$10,000	15%	30%
\$7,900<Price<\$9,000	12%	24%
\$6,900<Price<\$8,000	10%	20%
Price<\$6,900	5%	10%

In an ideal scenario, these weights will allow for 150 lineups to be created within the \$7,500,000 total salary cap, however, far more frequently adjustments have to be made.

1.6 Adjusting Weights

To check for whether adjustments need to be made for player weighting, first it must be determined if the salary cap has been passed and by how much. First, multiply a player's percentage ownership by 150 and then, take the resulting number and multiply it by the player's DraftKings price. For example, a \$10,000 player would be in 40% of lineups, so $.40 * 150 = 60$ total and $60 * \$10,000 = \$600,000$. Once the total overall cost of owning each player is found, they can be summed, and if the resulting number exceeds \$7,500,000, then adjustments must be made to how much of each player is owned. The first step to do this adjustment is to move ownership from the most expensive players to the least expensive by price tier. Remove five from each of the top price tier players and disperse them to the next price tier. For example, if your top tier is above \$9,900, then remove five of the 60 lineups from each member of this tier and disperse the additional five or more slots to each player in the \$8,900 to \$10,000 tier. If the total salary is still higher than \$7,500,000, then take five off of each of the second-tier players and redistribute

them to the third tier. Continue this process of redistribution until the total salary is under \$7,500,000.

1.6 – Lineup Creation

Now that the portfolio of 25 players has been established, as well as how many times each player will be used in a lineup while fitting the salary cap (\$7,500,000) and total number of players (900), lineups must be created. In order to eliminate selection bias, lineups ought to be randomized. In order to randomize the lineups efficiently, they must be done in tiers, so that they do not grossly exceed the \$50,000 salary cap. Excel's `RANDBETWEEN(x,y)` function can be used to randomize once players are allotted a number. In order to create tiers, start from the highest owned players and work down. Tier 1 will be comprised of the highest owned player, followed by the next highest owned players until the total number of lineup slots allotted to tier one surpasses 150. For example, if the top two players are in 60 lineups each and the third player is in 45 lineups, then these three players would comprise tier 1, as $60+60+45=165$, which is greater than 150. Tier 2 and so on will be formed in the same way, with the remaining highest owned players being added together until their total amount of selections exceeds 150 or is the last remaining players. Once the tiers are created they will be randomized individually. Tier 1 is randomized throughout the first column of excel, with the spillover past 150 being randomized into the second column. Tier 2 will then be randomized into the second column with the spillover being randomized into column 3. This process will continue until six columns are filled with 150 players. The sum of the prices of each row are then summed to make sure the players fit in the \$50,000 salary cap. For the lineups which are too expensive, players will be switched with the lineups that are the farthest under the salary

cap, so that all 150 are playable on DraftKings. In the case of duplicate lineups, players can be switched from lineups that are under the salary cap to create greater randomness. The lineups would then be entered into the DraftKings system.

V. Results

Week 1 - Sony Open

The Sony Open is the first full field PGA event of the calendar year. The tournament has been played at Waialae Country Club in Honolulu, Hawaii since 1965.²⁵ Regressing the chosen statistics from the 2014-2016 gives the predictive base equation for the 2017 event:

$$\begin{aligned}
 & \text{(5)} \\
 SCR_{Sony2017} &= 63.2211 - (.6775)SGTG - (.1640)SGP - (.0611)GIR + (.3727)PR \\
 & + \varepsilon
 \end{aligned}$$

²⁵ Brent Kelley, "Sony Open in Hawaii Golf Tournament," 15 January 2018, *ThoughtCo.com*, accessed 17 April 2018, <https://www.thoughtco.com/sony-open-in-hawaii-golf-tournament-1565848>.

**Table 4:
Sony Regression Output**

<i>Regression Statistics</i>			
R Square		0.9371	
Adj R Square		0.9359	
Standard Error		0.2409	
Observations		214	

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
Intercept	63.2211	0.5368	117.7743
SGTG	-0.6775	0.0240	-28.2678
SGP	-0.1640	0.0087	-18.7857
GIR	-0.0611	0.0050	-12.1066
P/R	0.3727	0.0263	14.1463

Strokes Gained Tee-to-Green, Strokes Gained Putting, Greens in Regulation, and Putts/Round were significant statistics at the 95% level, while Driving Distance, Driving Accuracy, Scrambling, and Sand Save Percentage were insignificant. This equation emphasizes players who are very good at hitting approach shots to the green and putting, while devaluing off the tee driving and accuracy skills. Waialae Country club is 7044 yards, making it one of the ten shortest courses played annually on the PGA Tour.²⁶ Being such a short course, any player on the PGA tour can hit the ball far enough off the tee to reach the green easily with their second shot on par 4's, and because it is a par 70, there are only two par 5 holes which would require a long drive off the tee to reach the green in two shots. These aspects of the course make a player whose strength is hitting the ball onto the green

²⁶ "The 10 Longest and 10 Shortest Courses on the PGA Tour in 2015-16," 16 December 2016, *TheGolfNewsNet.com*, accessed 17 April 2018, <https://thegolfnewsnet.com/golfnewsnetteam/2016/12/16/longest-shortest-courses-pga-tour-2015-2016-101774/>.

close to the hole with their approaches and putting well once on the green the ideal player. A player's ability to hit a drive far off the tee or to be accurate off the tee with a drive is insignificant when the course is so short, as most players will be able to hit their more accurate irons off the tee and still reach the green with their next shot. Scrambling is also understandably insignificant, as with such an emphasis on reaching the green in regulation, those who are continually scrambling are likely not scoring well no matter how successful they are being. In other words, players need to be on the green in regulation for a chance to make birdies, and scrambling to make pars will not let a player compete.

In order to sharpen the equation to take into account a player's history at the course, the 2014, 2015, and 2016 tournaments must be back checked. Linear regressions of the key statistics from 2011-2013 to predict 2014, from 2012-2014 to predict 2015, and from 2013-2015 to predict 2016 are found to have the same four significant variables with slightly differing coefficients depending on the year. The player's season long averages from the year in question for the four statistical categories are then inserted as the independent variables to predict their score for the tournament. Then, subtracting the projections from their real average scores from the event differences are found, which are regressed on each player's individual binary course history. Current form and weather are not analyzed for this tournament, as it is the first tournament of the year so there is no current form data and the weather is not expected to be a factor. The regression yields:

(6)

$$DIF = -(.89184)HIST + \varepsilon$$

**Table 5:
Adjusted Sony Regression Output**

<i>Regression Statistics</i>	
R Square	0.2653
Adj R Square	0.2591
Standard Error	1.0161
Observations	161

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
HIST	-0.8918	0.1173	-7.6015

So now, adding this to the initial 2017 projection equation:

$$\begin{aligned}
 & \hspace{20em} (7) \\
 SCR_{Sony2017} &= 63.2211 - (.6775)SGTG - (.1640)SGP - (.0611)GIR + (.3727)PR \\
 & \quad - (.8918)HIST + \varepsilon
 \end{aligned}$$

The players playing the 2017 tournament are then projected by inputting their 2016 season long statistics into the equation, as well as a 0 or 1 for course history depending on whether they had met the qualifications detailed. Valuation based on the projections and lineup creation detailed above leads to the 150 lineups that were set for the DraftKings Sony Open \$3 GPP contest.

The net result of the \$450 investment was a loss of \$356. Eighteen of the 150 lineups placed in the money. Four of the lineups returned the \$3 investment along with \$3 of profit. Fourteen lineups returned the \$3 investment along with \$2 of profit. Fifteen of the twenty-five players selected made the cut for a 60% made-cut rate. This percent is too low to be profitable without extreme luck in the randomization of rosters, as rosters generally need at least five of six players to make the cut. Justin Rose who was the top

projected value by the model and the seventh highest priced player on DraftKings finished second in the event, however the cheaper value players did not perform to the projections, resulting in a losing week.²⁷

Week 2 – CareerBuilder Challenge

The CareerBuilder Challenge is the second full field PGA event of the calendar year. The tournament is played at three courses in the Coachella Valley in Southern California. Despite being played at three courses, each have similar layouts, as the desert golf courses are made in the same style. The PGA West Stadium course is played twice, so a player's ability on it is most significant.²⁸ Regressing the chosen statistics from the 2014-2016 gives the predictive base equation for the 2017 event:

(8)

$$SCR_{CBC2017} = 71.8419 - (.0382)DD - (.025)DA - (.23)SGTG - (.0184)SCRAM \\ - (.0834)GIR + (.6643)PR + \varepsilon$$

²⁷ "Sony Open in Hawaii," 12-15 January 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/sony-open-hawaii/>.

²⁸ "About the Tournament," *CareerBuilderChallenge.com*, accessed 17 April 2018, <https://www.careerbuilderchallenge.com/about-the-tournament>.

Table 6:
CareerBuilder Regression Output

<i>Regression Statistics</i>			
R Square		0.4401	
Adj R Square		0.4248	
Standard Error		1.1620	
Observations		226	

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
Intercept	71.8419	3.1219	23.0119
DD	-0.0382	0.0073	-5.2707
DA	-0.0250	0.0102	-2.4584
SGTG	-0.2300	0.0686	-3.3508
SCRAM	-0.0184	0.0084	-2.1831
GIR	-0.0834	0.0198	-4.2056
P/R	0.6643	0.0838	7.9296

Driving Distance, Driving Accuracy, Strokes Gained Tee-to-Green, Scrambling Percentage, Greens in Regulation, and Putts/Round were significant statistics at the 95% level, while Strokes Gained Putting and Sand Save Percentage were insignificant. The courses played for this tournament are all listed in the ten easiest courses played on tour, so in order to be successful a player must be well under par, so birdie making is key. The courses are all about the same length, around tour average, and all are par 72.²⁹ Par 5 scoring is key, as players will play sixteen par 5's over the four days. The emphasis on par 5 scoring makes Driving Distance and Driving Accuracy significant factors, since those who can potentially hit their drive far enough and accurate enough to be close enough to

²⁹ Patrick Mayo, "Fantasy Golf Picks: 2017 Careerbuilder Challenge Sleepers, Starts & Preview," 16 January 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/118134/2017-careerbuilder-challenge-picks-fantasy-golf-picks-sleepers-starts-course-preview-careerbuilder-picks/>.

the green to hit their second shot onto it will have the best chance to make birdie or even eagle. Scrambling Percentage would not seem to be highly valuable on a course where making birdies is so vital, however, the Bermuda grass rough at the Stadium Course is difficult to get out of and can lead to bogey and double bogey, so being able to get up and down from it is important. Strokes Gained Putting being insignificant is odd, however, it is likely correlated to the fact that in order to play well a player must make so many birdies that being close to the hole and having easy putts for birdie is more important than being able to make longer birdie putts. Even the best putters do not consistently make long birdie putts over four rounds, however, great ball strikers can hit it close to the hole frequently when their game is clicking, so when a massive number of birdies is necessary, the players giving themselves more close-range opportunities at birdie are generally more successful than those hoping to make many long putts. The players who have historically won this tournament in the past have likely hit it close enough to the hole repeatedly that they have not had to out putt their competitors on a Strokes Gained basis. Sand Save Percentage is likely insignificant, as those who win rarely hit it into the sand, whether they get up and down from the sand at a high rate or not has minimal impact on their score.

In order to sharpen the equation to take into account a player's history at the course, the 2015 and 2016 tournaments must be back checked. 2014 is not reviewed as there is not enough data from the previous three years to do so, since the tournament had a different format up until 2012. Linear regressions of the key stats from 2012-2014 to predict 2015 and from 2013-2015 to predict 2016 are found to have the same significant variables with slightly differing coefficients depending on the year. The players' season long averages from the year in question for the four statistical categories are then inserted as the

independent variables to predict their score for the tournament. Then, by subtracting the projections from their real average scores from the event differences are found, which are regressed on each player's individual binary course history. Current form and weather are not analyzed for this tournament, as it is only the second tournament of the year, so there is not enough significant current form data and the weather was not expected to be a factor. The regression yields:

(9)

$$DIF = -(1.779)HIST + \varepsilon$$

**Table 7:
Adjusted CareerBuilder Regression Output**

<i>Regression Statistics</i>			
R Square		0.2068	
Adj R Square		0.1973	
Standard Error		2.1507	
Observations		106	

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
HIST	-1.7791	0.3401	-5.2317

So now adding this to the initial 2017 projection equation:

(10)

$$SCR_{CBC2017} = 71.8419 - (.0382)DD - (.025)DA - (.23)SGTG - (.0184)SCRAM \\ - (.0834)GIR + (.6643)PR - (1.779)HIST + \varepsilon$$

The players playing the 2017 tournament are then projected by inputting their 2016 season long statistics into the equation, as well as a 0 or 1 for course history, depending on whether they had met the qualifications detailed. Valuation based on the projections and lineup creation detailed above leads to the 150 lineups that were set for the DraftKings CareerBuilder Challenge \$3 GPP contest.

The net result of the \$450 investment was a loss of \$22. Fifty-eight of the 150 lineups placed in the money. All 58 lineups returned the \$3 investment with profit dispersion as follows: two lineups yielded \$22, one \$12, three \$9, three \$7, eleven \$5, nineteen \$3, and nineteen \$2 of profit. Fifteen of the twenty-five players selected made the cut for a 60% made-cut rate, which is the same as the past week at the Sony Open. This percent is too low to be profitable without extreme luck in the randomization of rosters, as rosters generally need at least five of six players to make the cut. Three of the top four finishers in the tournament were selected, however, the winner of the event was not. Having these high finishers allowed the lineups to almost break even, with just a minimal loss, despite the high number of missed cuts. The majority of the missed cuts stem from the low DraftKings priced value plays selected. Of the eight players selected who were priced below \$7000 on DraftKings only one made the cut. Bud Cauley, who was \$6000, was ranked third by the projection model despite being priced above only eighteen of the 156 players in the larger field. Unfortunately, Cauley was the lone stand out from the low-price selections, and since most lineups contain at least one of the cheap players, there were very few which had enough players make the cut to return significant profit. Overall having

three players in the top four mitigated losses that were to be expected based on the 60% made-cut percentage.³⁰

Week 3 – Farmers Insurance Open

The Farmers Insurance Open is the third full field PGA event of the calendar year. The tournament has been played at Torrey Pines Country Club in La Jolla, California since 1968. Torrey Pines has a South and North course. Players will play one round at each course before the cut, however, both rounds after the cut will be played at the South Course.³¹ The three rounds of data from the South course will be analyzed. Regressing the chosen statistics from the 2014-2016 gives the predictive base equation for the 2017 event:

(11)

$$SCR_{Farmers2017} = 65.5192 - (.281)SGTG - (.0855)SGP - (.0121)SCRAM - (.1312)GIR + (.5281)PR + \varepsilon$$

³⁰ "CareerBuilder Challenge," 19-22 January 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/careerbuilder-challenge/>.

³¹ "The Farmers Insurance Open Returns to La Jolla January 25th," *LaJolla.com*, accessed 17 April 2018, <https://www.lajolla.com/article/regional-attractions/farmers-insurance-open-torrey-pines-san-diego/>.

Table 8:
Farmers Insurance Regression Output

<i>Regression Statistics</i>			
R Square		0.8742	
Adj R Square		0.8714	
Standard Error		0.4074	
Observations		233	

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
Intercept	65.5192	1.0539	62.1694
SGTG	-0.2810	0.0320	-8.7788
SGP	-0.0855	0.0131	-6.5315
SCRAM	-0.0121	0.0033	-3.6221
GIR	-0.1312	0.0085	-15.3473
PR	0.5281	0.0455	11.6100

Strokes Gained Tee-to-Green, Strokes Gained Putting, Greens in Regulation, Scrambling and Putts/Round were significant statistics at the 95% level, while Driving Distance, Driving Accuracy, and Sand Save Percentage were insignificant. Torrey Pines is consistently one of the toughest tournaments on tour to score low at, so bogey avoidance is at a premium. Players who can stay close to par and make a few birdies without hurting themselves with bogey or worse will be able to contend. This is due to the length of the South Course, at 7,698 yards it is one of the longest on tour, and the wind can wreak havoc, being situated near the ocean.³² The predictive equation emphasizes players who get to the green in regulation and putt well, while also being proficient at making par when they do

³² Patrick Mayo, "Fantasy Golf Picks: 2017 Farmers Insurance Open Sleepers, Starts & Preview," 23 January 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/118284/fantasy-golf-picks-2017-farmers-insurance-open-picks-sleepers-starts-preview-tiger-woods/>.

miss the green in regulation. Driving Distance and Accuracy are not significant, which at first seems odd based on the length of the course. Taking the wind into account, though, it is highly likely that the players who drive the ball the farthest might be most affected due to the higher apex and longer time in the air. Sand Save Percentage is yet again not significant, possibly due to players who win being on the green in regulation a high enough percentage that their few trips to the sand do not have a great effect on their scorecard.

History, Current Form, and Weather are all insignificant for this tournament. Despite having certain players who have dominated the course year after year, most notably Tiger Woods who has won there eight times, a player's history does not show up as significant when back testing. This is likely because of the difficulty of the course, as even players who have games which should yield success on the course have a very slim margin for error. Form is insignificant again as it is the third week of the season, and although some players will be teeing off for a third straight week, many are making their season debut. Weather has already been discussed as a factor on this course, with extreme wind always being possible when located on the water, however, the variability of wind is too difficult to predict for. The standard equation before adjusting for History, Current Form, or Weather is used to predict success.

The players playing the 2017 tournament are then projected by inputting their 2016 season long statistics into the equation. Valuation based on the projections and lineup creation detailed above leads to the 150 lineups that were set for the DraftKings Farmers Insurance Open \$3 GPP contest.

The net result of the \$450 investment was a loss of \$401. Eight of the 150 lineups placed in the money. One of the lineups returned the \$3 investment along with \$6 of profit.

Five lineups returned the \$3 investment along with \$3 of profit. Two of the lineups returned the \$3 investment along with \$2 of profit. Twelve of the twenty-five players selected made the cut for a 48% made-cut rate. This percent is far too low to be profitable. The model selected the winner of the tournament, Jon Rahm, as well as one of the players who tied for second, Charles Howell, but the lineups never stood a chance with such a poor made-cut percentage. Forty-Seven percent of players in the field made the cut, and the model only managed to have 48% make it, so it was incredibly unsuccessful for this week leading to a major loss.³³

Week 4 – Waste Management Phoenix Open

The Waste Management Phoenix Open is the fourth full field PGA event of the calendar year. The tournament has been played at TPC Scottsdale in Scottsdale, Arizona, since 1987.³⁴ Regressing the chosen statistics from the 2014-2016 gives the predictive base equation for the 2017 event:

(12)

$$SCR_{WMP2017} = 70.9218 - (.983)SGTG - (.2422)SGP + \varepsilon$$

³³ "Farmers Insurance Open," 26-29 January 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/farmers-insurance-open/>.

³⁴ "Tournament History," *wmPhoenixOpen.com*, accessed 17 April 2018, <https://wmphoenixopen.com/spectator-info/tournament-history/>.

**Table 9:
WM Phoenix Open Regression Output**

<i>Regression Statistics</i>			
R Square			0.9830
Adj R Square			0.9826
Standard Error			0.1672
Observations			220

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
Intercept	70.9218	0.4002	177.2298
SGTG	-0.9830	0.0108	-91.2049
SGP	-0.2422	0.0044	-55.6342

Only the Strokes Gained statistics, both Tee-to-Green and Putting were significant statistics at the 95% level, with all other statistics being insignificant. TPC Scottsdale generally ranks in the middle of PGA Tour courses in difficulty, with a score in the teens under par likely to win.³⁵ Driving Distance is insignificant, likely due to the dry Arizona climate, which allows for players to hit the ball farther than they generally would. A desert course, TPC Scottsdale is wide open, where missing the fairway could leave your ball at the base of a cactus or in a dry patch with a view of the green. Being on the fairway consistently always helps, but at this type of course, luck in relation to where a player misses the fairway could make driving accuracy ultimately insignificant. At a course where scoring significantly under par is necessary to win, scrambling to make par is less valuable, as hitting greens and making birdies is necessary. The greens at TPC Scottsdale are Bermuda grass, which are generally slower and thus easier to two-putt on. This is because a player's first putt is

³⁵ Patrick Mayo, "Fantasy Golf Picks: 2017 Waste Management Open Sleepers, Starts & Preview," 30 January 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/118466/fantasy-golf-picks-2017-waste-management-open-picks-sleepers-starts-preview/>.

unlikely to roll well past the hole on slow greens, so most players should be able to avoid three-putting. Putts/round is insignificant due to the ease of two-putting, however, Strokes Gained Putting is still significant as players who can make birdies will excel. Strokes Gained Tee-to-Green is the most significant variable, as those who are able to get close to the hole in regulation will be able to amass the most birdie opportunities.

In order to sharpen the equation to take into account a player's history at the course, the 2014, 2015, and 2016 tournaments must be back checked. Linear regressions of the key stats from 2011-2013 to predict 2014, 2012-2014 to predict 2015, and from 2013-2015 to predict 2016 are found to have the same significant variables with slightly differing coefficients depending on the year. The players' season long averages from the year in question for the four statistical categories are then inserted as the independent variables to predict their score for the tournament. By then subtracting the projections from their real average scores from the event, the differences are found, which are regressed on each player's individual binary course history. Current form and weather are not analyzed for this tournament, as it is only the fourth tournament of the year so there is not enough significant current form data and the weather in Arizona is expected to be warm and clear, which should benefit all. The regression yields:

(13)

$$DIF = -(.6946)HIST + \varepsilon$$

**Table 10:
Adjusted WM Phoenix Open Regression Output**

<i>Regression Statistics</i>			
R Square	0.1195		
Adj R Square	0.1142		
Std. Error	1.1986		
Observations	189		

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
HIST	-0.6946	0.1375	-5.0522

So now adding this to the initial 2017 projection equation:

(14)

$$SCR_{WMP2017} = 70.9218 - (.983)SGTG - (.2422)SGP - (.6946)HIST + \varepsilon$$

The players playing the 2017 tournament are then projected by inputting their 2016 season long statistics into the equation, as well as a 0 or 1 for course history, depending on whether they had met the qualifications detailed. Valuation based on the projections and lineup creation detailed above leads to the 150 lineups that were set for the DraftKings Waste Management Open \$3 GPP contest.

The net result of the \$450 investment was a gain of \$609. Ninety-three of the 150 lineups placed in the money. All 93 lineups returned the \$3 investments with profit dispersion as follows: one lineup yielded \$147, one \$97, two \$72, one \$47, one \$37, one \$24, four \$9, seven \$6, six \$5, five \$4, twenty-eight \$3, and thirty-six \$2 of profit. Nineteen of the twenty-five players selected made the cut for a 75% made-cut rate. This percent was high enough to be profitable. Ten of the top 23 finishers in the tournament were selected,

including the players who finished first, second, third, and one of two players who tied for fourth. Having these high finishers allowed for a large profit, however, unfortunately, the randomization did not create any lineups with all four of the selected top finishers, which would have exponentially increased the profit. Course history seemed to be the key to success this week, as each of the top players had past success at the course, with Winner Hideki Matsuyama having a particularly stellar past record at TPC Scottsdale.³⁶

Week 5 – AT&T Pebble Beach Pro-Am

The AT&T Pebble Beach Pro-Am is the fifth full field PGA event of the calendar year. The tournament is played at three courses around Pebble Beach, California. Despite being played at three courses, each are fairly similar, particularly in length. Pebble Beach is a par 72 measuring 6,816 yards, Spyglass Hill is a par 72 measuring 6,953 yards, and Monterey Peninsula is a par 71 measuring 6,914 yards.³⁷ Regressing the chosen statistics from the 2014-2016 gives the predictive base equation for the 2017 event:

(15)

$$SCR_{PB2017} = 66.5510 - (.3387)SGTG - (.1317)SGP - (.012)SCRAM \\ - (.0483)GIR + (.2855)PR + \varepsilon$$

³⁶ John Davis, "Waste Management Phoenix Open: Hideki Matsuyama Takes Aim at 3-Peat," 30 January 2018, *azCentral.com*, accessed 17 April 2018, <https://www.azcentral.com/story/sports/golf/phoenix-open/2018/01/30/waste-management-phoenix-open-hideki-matsuyama-takes-aim-3-peat/1081419001/>.

³⁷ "AT&T Pebble Beach Pro-Am: Tournament Information," *PebbleBeach.com*, accessed 17 April 2018, <https://www.pebblebeach.com/events/att-pebble-beach-pro-am/>.

Table 11:
Pebble Beach Regression Output

<i>Regression Statistics</i>			
R Square		0.6649	
Adjusted R Square		0.6561	
Standard Error		0.6239	
Observations		196	

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
Intercept	66.5510	1.4387	46.2568
SGTG	-0.3387	0.0315	-10.7669
SGP	-0.1317	0.0213	-6.1794
SCRAM	-0.0120	0.0050	-2.3914
GIR	-0.0483	0.0080	-6.0313
PR	0.2855	0.0499	5.7170

Strokes Gained Tee-to-Green, Strokes Gained Putting, Scrambling Percentage, Greens in Regulation, and Putts/Round were significant statistics at the 95% level, while Driving Distance, Driving Accuracy, and Sand Save Percentage were insignificant. With these three courses being extremely short by PGA standards and also the possibility of coastal winds, players can choose to hit very few drivers off the tee. Most players will be able to hit a long iron off the tee and still have a wedge or a short iron shot left to make the green in regulation. Due to players ability to not hit driver as frequently at this tournament, Driving Distance and Accuracy are both insignificant statistics. The challenge of these courses are the tiny greens, which are incredibly difficult to hit. The small greens make hitting the green in regulation, as well as being able to scramble to make par when a green

is missed, extremely significant.³⁸ Gaining strokes tee-to-green and putting are significant factors, as despite the small greens, the length of the courses allow for players to have many birdie opportunities in good conditions, which must be taken advantage of by gaining strokes against the field. Yet again, Sand Save Percentage is not significant, likely due to the small sample of times players who score well end up in the sand.

In order to sharpen the equation to take into account a player's history at the course, the 2014, 2015, and 2016 tournaments must be back checked. Linear regressions of the key stats from 2011-2013 to predict 2014, 2012-2014 to predict 2015, and from 2013-2015 to predict 2016 are found to have the same significant variables with slightly differing coefficients depending on the year. The players' season long averages from the year in question for the four statistical categories are then inserted as the independent variables to predict their score for the tournament. By then subtracting the projections from their real average scores from the event differences are found, which are regressed on each player's individual binary course history. Current form and weather are not analyzed for this tournament, as it is only the fifth tournament of the year so there is not enough significant current form data and it is not projected to rain and wind is too difficult to predict. The regression yields:

(16)

$$DIF = -(.586)HIST + \varepsilon$$

³⁸ Patrick Mayo, "Fantasy Golf Picks: 2017 AT&T Pebble Beach Pro-Am Sleepers, Starts & Preview," 6 February 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/118602/fantasy-golf-picks-2017-pebble-beach-picks-sleepers-starts-preview-pro-am/>.

**Table 12:
Adjusted Pebble Beach Regression Output**

<i>Regression Statistics</i>			
R Square	0.1060		
Adj R Square	0.0983		
Standard Error	1.0449		
Observations	131		

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
HIST	-0.5860	0.1493	-3.9257

So now adding this to the initial 2017 projection equation:

(17)

$$\begin{aligned}
 SCR_{PB2017} = & 66.5510 - (.3387)SGTG - (.1317)SGP - (.012)SCRAM \\
 & - (.0483)GIR + (.2855)PR - (.586)HIST + \varepsilon
 \end{aligned}$$

The players playing the 2017 tournament are then projected by inputting their 2016 season long statistics into the equation, as well as a 0 or 1 for course history depending on whether they had met the qualifications detailed. Valuation based on the projections and lineup creation detailed above leads to the 150 lineups that were set for the DraftKings Pebble Beach \$3 GPP contest.

The net result of the \$450 investment was a loss of \$221. Thirty-eight of the 150 lineups placed in the money. All 38 lineups returned the \$3 investments with profit dispersion as follows: one lineup yielded \$13, one \$9, three \$5, two \$4, eight \$3, and twenty-three \$2 of profit. 14 of the twenty-five players selected made the cut for a 56% made-cut rate, which is the same as the past week at the Sony Open. This percent is too

low to be profitable without extreme luck in the randomization of rosters, as rosters generally need at least five of six players to make the cut. Four of the top seven finishers in the tournament were selected, including Jordan Spieth, the winner of the event. Having these high finishers allowed the lineups to make back slightly more than 50% of the initial investment, however, the low percentage of made-cuts still left a relatively large net loss. The top three rated players by the model: Dustin Johnson, Jason Day, and Jordan Spieth finished 3rd, T-5th, and 1st respectively.³⁹ Only four of the top fifteen projected players missed the cut, however, seven of the next ten did make the cut. Of the ten players who were selected that were priced below \$7,000 on DraftKings only three made the cut. This shows a failure by the model to predict sleeper plays this week, with only the top-end players selected performing as expected.⁴⁰

Week 6 – Genesis Open

The Genesis Open is the sixth full field PGA event of the calendar year. The tournament has been played at Riviera Country Club in Pacific Palisades, California, consistently since 1973.⁴¹ Regressing the chosen statistics from the 2014-2016 gives the predictive base equation for the 2017 event:

(18)

$$SCR_{Genesis2017} = 71.5802 - (1.0147)SGTG - (.2565)SGP + \varepsilon$$

³⁹ "AT&T Pebble Beach Pro-Am," 9-12 February 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/att-pebble-beach-pro-am/>.

⁴⁰ *Ibid.*

⁴¹ "Tournament History," *GenesisOpen.com*, accessed 17 April 2018, <http://genesisopen.com/tournament-history>.

Table 13:
Genesis Open Regression Output

<i>Regression Statistics</i>	
R Square	0.9963
Adj R Square	0.9963
Standard Error	0.0767
Observations	225

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	71.5802	0.2091	342.3389	0.0000
SGTG	-1.0147	0.0083	-121.7951	0.0000
SGP	-0.2565	0.0029	-89.6179	0.0000

Only the Strokes Gained statistics, both Tee-to-Green and Putting were significant statistics at the 95% level, with all other statistics being insignificant. The equation for the Genesis is eerily similar to that of the Waste Management Phoenix Open, which despite being very different courses in terms of terrain makes some sense. Both courses generally have winners who score in the teens under par, and both can be conquered by players of very different skill sets. Riviera Country Club is 7,322 yards, which is around the average PGA tour yardage, however, the difference between holes' distances makes it more interesting. Six of the eleven par 4 holes are over 450 yards, which are long by tour standards, while one of the par 4 holes is 315 yards which is incredibly short. The course also features a short par 5, which is easily reachable in two strokes even for the tour's shorter hitters off the tee.⁴² Driving distance would appear to be important with six long par 4's, but distance

⁴² Patrick Mayo, "Fantasy Golf Picks: 2017 Genesis Open Sleepers, Starts & Preview," 13 February 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/118793/fantasy-golf-picks-2017-genesis-open-picks-sleepers-starts-preview/>.

can get players into trouble as well. On the short par 4 if long hitters decide to try to make the green in one stroke and fail, then they may wind up in the surrounding bunkers, which can be very penal. Players who do not drive the ball but are accurate can be just as effective as those driving the ball a long way. With that said, accuracy is not shown to be significant either, which can be attributed to the lack of accuracy needed by long hitters, as those who hit it a long way often will be able to reach the green even if they miss the fairway. In other words, there is more than one way to be successful driving the ball at Riviera. The Strokes Gained statistics are as usual the most indicative of success, with a player gaining strokes on his competitors in both facets of the game being vital.

In order to sharpen the equation to take into account a player's history at the course, the 2014, 2015, and 2016 tournaments must be back checked. Linear regressions of the key statistics from 2011-2013 to predict 2014, 2012-2014 to predict 2015, and from 2013-2015 to predict 2016 are found to have the same significant variables with slightly differing coefficients depending on the year. The players' season long averages from the year in question for the two statistical categories are then inserted as the independent variables to predict their score for the tournament. By then subtracting the projections from their real average scores from the event differences are found, which are regressed on each player's individual binary course history. Current form and weather are not analyzed for this tournament, as it is only the sixth tournament of the year so there is not enough significant current form data and the weather in the Pacific Palisades is expected to be sunny and clear, which should benefit all. The regression yields:

(19)

$$DIF = -(.612)HIST + \varepsilon$$

**Table 14:
Adjusted Genesis Open Regression Output**

<i>Regression Statistics</i>			
R Square		0.0773	
Adj R Square		0.0717	
Standard Error		1.4083	
Observations		179	

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
HIST	-0.6120	0.1584	-3.8625

So now adding this to the initial 2017 projection equation:

(20)

$$SCR_{Genesis2017} = 70.9218 - (.9828)SGTG - (.2421)SGP - (.612)HIST + \varepsilon$$

The players participating in the 2017 tournament are then projected by inputting their 2016 season long statistics into the equation, as well as a 0 or 1 for course history depending on whether they had met the qualifications detailed. Valuation based on the projections and lineup creation detailed above leads to the 150 lineups that were set for the DraftKings Genesis Open \$3 GPP contest.

The net result of the \$450 investment was a gain of \$45,623. Seventy-seven of the 150 lineups placed in the money. All 77 lineups returned the \$3 investments with profit dispersion as follows: one lineup yielded \$24,997, one \$14,997, one 2,497, one \$747, three \$497, one \$297, one \$147, one \$97, two \$67, two \$47, two \$27, four \$17, five \$9, four \$7, five \$5, fourteen \$4, ten \$3 and nineteen \$2 of profit. Twenty of the 25 players selected made the cut for an 80% made-cut rate. This percent was high enough to be profitable. Seven of the top 15 finishers in the tournament were selected, including the players who

finished first, one of two players tied for second, and two of four players tied for fourth. Having these high finishers allowed for a large profit, and with the benefit of some luck in the randomization process creating more than one lineup with all four of the top selected finishers, the Genesis Open provided the greatest win to-date. The top two entries would have won and finished second in the entire GPP tournament. This one week would be able to pay for at least 451 weeks of future contest entries, however, due to the small sample it is impossible to know if this was a result of luck or the skill of the model.⁴³

Week 7 – Honda Classic

The Honda Classic is the seventh full field PGA event of the calendar year. The tournament has been played at PGA National Golf Club Championship Course in Palm Beach Gardens, Florida, since 2007.⁴⁴ Regressing the chosen statistics from the 2014-2016 gives the predictive base equation for the 2017 event:

(21)

$$SCR_{Honda2017} = 69.9973 - (.9996)SGTG - (.9996)SGP - (.00002)SCRAM \\ - (.00006)GIR + (.0003)PR + \varepsilon$$

⁴³ "Genesis Open," 16-19 February 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/genesis-open/>.

⁴⁴ "Honda Classic Winners and History," 25 February 2018, *GolfBlogger.com*, accessed 17 April 2018, https://golfblogger.com/honda_classic_past_winners_and_history/.

**Table 15:
Honda Classic Regression Output**

<i>Regression Statistics</i>			
R Square		0.9999998	
Adjusted R Square		0.9999998	
Standard Error		0.0005171	
Observations		217	

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
Intercept	69.9973	0.0014	49802.7320
SGTG	-0.9996	0.0001	-16587.8787
SGP	-0.9996	0.0001	-12454.0378
SCRAM	0.0000	0.0000	-4.4921
GIR	-0.0001	0.0000	-5.8683
PR	0.0003	0.0001	4.5710

Strokes Gained Tee-to-Green, Strokes Gained Putting, Scrambling, Greens in Regulation, and Putts/Round were significant statistics at the 95% level, while Driving Distance, Driving Accuracy, and Sand Save Percentage were insignificant. The Championship Course has only been outside of the top ten most difficult courses on the PGA schedule in two seasons, solidifying it as one of the most challenging courses on tour with the average score being above par. The course is a par 70, which plays 7,140 yards and is littered with sand traps and water hazards.⁴⁵ Being in Florida, weather can also play a large factor at the tournament, however, for this year's installment, rain is not in the forecast and winds are only expected to be ferocious on Sunday afternoon. The Strokes Gained statistics are key

⁴⁵ Patrick Mayo, "Fantasy Golf Picks: 2017 Honda Classic Sleepers, Starts & Preview," 20 February 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/119146/fantasy-golf-picks-2017-honda-classic-picks-sleepers-starts-honda-classic-preview/>.

as usual, with players needing to beat their competitors in how they reach the green and how they get the ball in the hole in order to compete. Driving Distance and Accuracy are both insignificant, which can likely be attributed to the water hazards. With so many chances to lose a ball into the water off the tee, many players will resort to hitting irons instead of a driver off the tee in order to more likely keep the ball in play. Despite the many bunkers, Sand Save Percentage is still insignificant, however, scrambling is important and includes player's success out of the sand. The many bunkers make hitting the green in regulation a priority, as a missed green could result in a tough test to get up and down for par from the sand.

In order to sharpen the equation to take into account a player's history at the course, the 2014, 2015, and 2016 tournaments must be back checked. Linear regressions of the key statistics from 2011-2013 to predict 2014, from 2012-2014 to predict 2015, and from 2013-2015 to predict 2016 are found to have the same five significant variables with slightly differing coefficients depending on the year. The players' season long averages from the year in question for the four statistical categories are then inserted as the independent variables to predict their score for the tournament. By then subtracting the projections from their real average scores from the event differences are found, which are regressed on each player's individual binary course history. Weather is not analyzed for this tournament, as it is not expected to take effect until late Sunday. Current Form is analyzed for this week; however, results are insignificant, likely due to an influx of European players who come to Florida to begin to prepare for the Masters and have little trackable form. The regression yields:

(22)

$$DIF = -(.586)HIST + \varepsilon$$

**Table 16:
Adjusted Honda Classic Regression Output**

<i>Regression Statistics</i>			
R Square		0.1315	
Adj R Square		0.1253	
Standard Error		1.0887	
Observations		162	

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
HIST	-0.5899	0.1195	-4.9365

So now adding this to the initial 2017 projection equation:

(23)

$$SCR_{Honda2017} = 69.9973 - (.9996)SGTG - (.9996)SGP - (.00002)SCRAM \\ - (.00006)GIR + (.0003)PR - (.586)HIST + \varepsilon$$

The players playing the 2017 tournament are then projected by inputting their 2016 season long statistics into the equation, as well as a 0 or 1 for course history depending on whether they had met the qualifications detailed. Valuation based on the projections and lineup creation detailed above leads to the 150 lineups that were set for the DraftKings Honda Classic \$3 GPP contest.

The net result of the \$450 investment was a gain of \$403. Eight-four of the 150 lineups placed in the money. All 84 lineups returned the \$3 investments with profit

dispersion as follows: one lineup yielded \$97, two \$47, one \$37, two \$27, three \$17, three \$12, one \$9, one \$7, seven \$6, five \$5, seven \$4, nineteen \$3, and thirty-one \$2 of profit. Nineteen of the 25 players selected made the cut for a 75% made-cut rate. This percent was high enough to be profitable. Eight of the top 20 finishers in the tournament were selected, including the player who finished first, one of two players who tied for second, and two of six players tied for fourth. Having these high finishers allowed for a large profit, but the randomization process did not produce the best possible lineup which could have been formed out of the 25 players selected. Seven of the ten players below \$7,500 that were selected made the cut, with Billy Horschel, priced at \$7,200, placing in a tie for fourth. Furthermore, the number two ranked player by the model, Rickie Fowler, won the tournament, which shows that this week it was successful in picking the top players, as well as cheap value players.⁴⁶

Week 8 – Valspar Championship

The Valspar Championship is the eighth full field PGA event of the calendar year, which falls in week nine of the season. The WGC- Mexico is played during week seven, which only has a field of approximately sixty players, so was not analyzed. The Valspar Championship began in 2000 and is played at Copperhead Course at Innisbrook Resort in Palm Harbor, Florida.⁴⁷ Regressing the chosen statistics from the 2014-2016 gives the predictive base equation for the 2017 event:

⁴⁶ "The Honda Classic," 23-26 February 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/honda-classic/>.

⁴⁷ "Valspar Championship Winners and History," 11 March 2018, *GolfBlogger.com*, accessed 17 April 2018, https://golfblogger.com/valspar_championship_past_winners_and_history/.

(24)

$$SCR_{Valspar2017} = 70.9883 - (.987)SGTG - (.9849)SGP + \varepsilon$$

Table 17:
Valspar Regression Output

<i>Regression Statistics</i>			
R Square	0.988831986		
Adj R Square	0.988727122		
Standard Error	0.107841898		
Observations	216		

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
Intercept	70.9883	0.0095	7482.3466
SGTG	-0.9870	0.0081	-122.0771
SGP	-0.9849	0.0102	-96.8238

Only the Strokes Gained statistics, both Tee-to-Green and Putting were significant statistics at the 95% level, with all other statistics being insignificant. Similar to at the Honda Classic, the second Florida tournament is also very challenging.⁴⁸ The challenge of the Honda Classic was water hazards throughout, however, at the Valspar Championship it is tree lined fairways, which can keep a player from being able to reach the green in regulation if hit into. Driving accuracy would seem to be significant, however, so few players elect to hit their driver off the tee that driving statistics are actually inconsequential. Greens in Regulation and Scrambling are insignificant as well, despite seemingly being important when looking at the difficulty of the course. The key to success at Copperhead

⁴⁸ Patrick Mayo, "Fantasy Golf Picks: 2017 Valspar Championship Sleepers, Starts & Preview," 6 March 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/119604/fantasy-golf-picks-2017-valspar-championship-picks-sleepers-starts-valspar-preview-predictions/>.

is avoiding making double bogey or worse in order to keep a score close to par, as most players will be over par for the tournament. Gaining strokes on the field on the way to the green and also by making difficult putts, to either save par or make birdie, is the path to victory.

In order to sharpen the equation to take into account a player's history at the course, the 2014, 2015, and 2016 tournaments must be back checked. Linear regressions of the key statistics from 2011-2013 to predict 2014, from 2012-2014 to predict 2015, and from 2013-2015 to predict 2016 are found to have the same five significant variables with slightly differing coefficients depending on the year. The players' season long averages from the year in question for the two statistical categories are then inserted as the independent variables to predict their score for the tournament. By then subtracting the projections from their real average scores from the event differences are found, which are regressed on each player's individual binary course history. Weather is not analyzed for this tournament, as rain is not in the forecast and wins is not expected to be a constant. Current Form is analyzed for this week however results are insignificant. The regression yields:

(25)

$$DIF = -(.361)HIST + \varepsilon$$

**Table 18:
Adjusted Valspar Regression Output**

<i>Regression Statistics</i>			
R Square	0.0387		
Adj R Square	0.0327		
Standard Error	1.1830		
Observations	170		

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
HIST	-0.3610	0.1385	-2.6069

So now adding this to the initial 2017 projection equation:

(26)

$$SCR_{Valspar2017} = 70.9883 - (.987)SGTG - (.9849)SGP - (.361)HIST + \varepsilon$$

The players playing the 2017 tournament are then projected by inputting their 2016 season long statistics into the equation, as well as a 0 or 1 for course history, depending on whether they had met the qualifications detailed. Valuation based on the projections and lineup creation detailed above leads to the 150 lineups that were set for the DraftKings Valspar Championship \$3 GPP contest.

The net result of the \$450 investment was a loss of \$229. Thirty-nine of the 150 lineups placed in the money. All 39 lineups returned the \$3 investments with profit dispersion as follows: two yielded \$6, one \$5, two \$4, eleven \$3, and twenty-three \$2 of profit. Seventeen of the 25 players selected made the cut for a 68% made-cut rate. This percent could have been high enough to be profitable with high finishers and some luck in the randomization process, but unfortunately it was not this week. Only two of the top ten

players were selected this week, with none in the top five. Charl Schwartzel and Henrik Stenson finished sixth and seventh respectively in the tournament, with both being projected to be in the top five, however, mediocre results from the middle and lower tier players selected and failing to pick the winner, Adam Hadwin, resulted in a losing week.⁴⁹

Week 9 – Arnold Palmer

The Arnold Palmer Invitational is the ninth PGA event of the calendar year, and although not a full field event, the 120 players invited and the standard two day cut make this tournament similar enough to a full field event that it can be modeled. The invitational began in 1966 under the name, the Florida Citrus Open Invitational. The tournament has been played at Bay Hill Club and Lodge since 1979 and took on the name the Arnold Palmer Invitational in 2007.⁵⁰ Regressing the chosen statistics from the 2014-2016 gives the predictive base equation for the 2017 event:

(27)

$$SCR_{API2017} = 71.9976 - (.00001)DA - (.9997)SGTG - (.25)SGP \\ - (.00001)SCRAM - (.00004)GIR + (.0002)PR + \varepsilon$$

⁴⁹ "Valspar Championship," 9-12 March 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/valspar-championship/>.

⁵⁰ "History: Since 1979," *ArnoldPalmerInvitational.com*, accessed 17 April 2018, <https://arnoldpalmerinvitational.com/history>.

Table 19:
Arnold Palmer Invitational Regression Output

<i>Regression Statistics</i>			
R Square		0.9999	
Adj R Square		0.9999	
Standard Error		0.0006	
Observations		226	

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
Intercept	71.99760	0.00154	46870.36936
DA	-0.00001	0.00001	-2.38129
			-
SGTG	-0.99973	0.00006	16962.29791
			-
SGP	-0.24995	0.00002	12272.96259
SCRAM	-0.00001	0.00000	-3.02344
GIR	-0.00004	0.00001	-2.58358
PR	0.00022	0.00007	3.27540

Driving Accuracy, Strokes Gained Tee-to-Green, Strokes Gained Putting, Scrambling, Greens in Regulation, and Putts per Round were all significant, while Driving Distance and Sand Save Percentage were not. Bay Hill is a par 72 course that measures 7,419 yards.⁵¹ The four par 5's provide many scoring opportunities, but the course has plenty of water and sand to ruin a player's round. Similar to the Honda Classic there is water that must be avoided, however, unlike a few weeks back, Bay Hill is generally too long to not hit a driver off the tee on most holes. Instead of hitting irons to avoid the water like they did at

⁵¹ Patrick Mayo, "Fantasy Golf Picks: 2017 Arnold Palmer Invitational Sleepers, Starts & Preview," 13 March 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/119817/fantasy-golf-picks-2017-arnold-palmer-invitational-picks-sleepers-starts-bay-hill-preview-predictions/>.

the Honda Classic, here players must be accurate with their drivers. The Strokes Gained statistics are significant as always, as if a player hopes to win, they must make enough birdies by gaining strokes against their competitors. Scrambling and Greens in Regulation are likely both significant for the same reason, which is the multitude of bunkers surrounding the greens. Being able to hit the green in regulation and avoid scrambling out of the bunkers or thick rough is a huge advantage. If a player is to miss the green, though, he had better be scrambling at a high level given the difficult conditions. Driving Distance is insignificant, as this course is more reliant on accuracy.

In order to sharpen the equation to take into account a player's history at the course, the 2014, 2015, and 2016 tournaments must be back checked. Linear regressions of the key statistics from 2011-2013 to predict 2014, from 2012-2014 to predict 2015, and from 2013-2015 to predict 2016 are found to have the same five significant variables with slightly differing coefficients depending on the year. The players' season long averages from the year in question for the two statistical categories are then inserted as the independent variables to predict their score for the tournament. By then subtracting the projections from their real average scores from the event differences are found, which are regressed on each player's individual binary course history. Weather is not analyzed for this tournament, as rain is not in the forecast and wind is not expected to be present throughout. Current Form is analyzed for this week; however, results prove insignificant. The regression yields:

(28)

$$DIF = -(0.4099)HIST + \varepsilon$$

Table 20:
Adjusted Arnold Palmer Invitational Regression Output

<i>Regression Statistics</i>			
R Square	0.0456		
Adj R Square	0.0397		
Standard Error	1.2456		
Observations	171		

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>
HIST	-0.4099	0.1438	-2.8500

So now adding this to the initial 2017 projection equation:

(29)

$$SCR_{API2017} = 71.9976 - (.00001)DA - (.9997)SGTG - (.25)SGP \\ - (.00001)SCRAM - (.00004)GIR + (.0002)PR - (.4099)HIST + \varepsilon$$

The players playing the 2017 tournament are then projected by inputting their 2016 season long statistics into the equation, as well as a 0 or 1 for course history, depending on whether they had met the qualifications detailed. Valuation based on the projections and lineup creation detailed above leads to the 150 lineups that were set for the DraftKings Arnold Palmer Invitational \$3 GPP contest.

The net result of the \$450 investment was a loss of \$336. Sixteen of the 150 lineups placed in the money. All 16 lineups returned the \$3 investments with profit dispersion as follows: Four yielded \$3 and twelve yielded \$2. Eighteen of the 25 players selected made the cut for a 72% made-cut rate. This percent could have been high enough to be profitable with high finishers and some luck in the randomization process, but unfortunately it was

not this week. Henrik Stenson was the model's number two ranked player and was owned in 40% of lineups, however, he missed the cut. Along with Stenson who was the second highest priced player picked, the low-priced players also struggled, as four of the five players priced under \$7,000 missed the cut. The combination of the most owned player and almost the entire bottom tier of players missing the cut left not a single lineup with all six players having made the cut, which resulted in a large loss.⁵²

DraftKings Change – Week 10 and Beyond

After week 9, DraftKings altered their contest structure, so the 150-lineup model for the \$3 GPP was no longer relevant. The model was run for six more weeks to track missed cut percentage and winners. Here are the results for those six weeks:

Table 21: Week 10 and Beyond

Tournament	Made Cut Ratio	Winner Selected
Shell Houston Open	15/25	0
RBC Heritage Invitational	18/25	0
Valero Texas Open	14/25	0
The PLAYERS	19/25	0
Dean and Deluca Inv.	18/25	1
The Memorial Tournament	18/25	0

Based on the results above, it seems likely that the Dean and Deluca Invitational would have been profitable, while the Shell Houston Open and Valero Texas Open certainly would not have been. The RBC Heritage, PLAYERS Championship, and Memorial

⁵² "Arnold Palmer Invitational Presented by MasterCard," 16-19 March 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pgs-tour/2017/arnold-palmer-invitational-presented-mastercard/>.

Tournament would likely have been minor losses to minor wins depending on conditions that cannot be fully predicted.

Ultimately, over the course of the nine weeks tested on DraftKings there was a net profit of \$45,070. Nine weeks is far too small of a sample size to determine if this profit was made because of the skill of the model or if the Genesis Open was just a massive outlier. The one massive win, \$45,623, would cover 101 weeks of playing 150 lineups and still be in the positive, even with 101 weeks of losing. In order to see if the model truly works, it would have to be carried out even beyond 101 weeks, which is multiple PGA seasons. Here are the overall results for all 15 weeks tested:

Table 22: Overall Results

Tournament	Made Cut Ratio	Winer Selected	Return (USD)
Sony Open	15/25	0	-356
Career Builder Challenge	15/25	0	-22
Farmers Insurance Open	12/25	1	-401
Waste Management Phoenix Open	19/25	1	609
AT&T Pebble Beach Pro-Am	14/25	1	-221
Genesis Open	20/25	1	45,623
Honda Classic	19/25	1	403
Valspar Championship	17/25	0	-229
Arnold Palmer Invitational	18/25	0	-336
Shell Houston Open	15/25	0	N/A
RBC Heritage Invitational	18/25	0	N/A
Valero Texas Open	14/25	0	N/A
The PLAYERS Championship	19/25	0	N/A
Dean and DeLuca Invitational	18/25	1	N/A
The Memorial Tournament	18/25	0	N/A
Totals	67%	40%	\$45,070

VI. Conclusion:

This study ultimately shows that each course played on the PGA Tour is distinctly different, which lends to different skills needing greater or less emphasis at certain venues. The predictive equations used for each week highlight which skills have historically led a player to success at the specific course. The second part of the study, which was to create a portfolio of players for each week to invest in and make a profit on DraftKings has a less clear result. The final net gain is quite positive; however, it can be clearly attributed to one week's results, which could be a massive outlier. In order to see if the winnings are a result of the skill of the model or luck, the process would need to be completed many more times over before any type of conclusive result could be found. One thing that is demonstrated is that DraftKings is a legal platform that can lead to large profit, which could be a profitable investment endeavor if a model is found to be significantly positive over the long run.

In order to better the results of this study, first one would have to continue to employ the same exact process for many more weeks until a significant result was found. Beyond the number of weeks of observations needed, there are other things that could potentially be tested to improve the model. There are statistics that were not used for this model that could have been incorporated to potentially improve its predictability, for example the Strokes Gained Tee-to-Green statistic could have been broken down into separate components (Strokes Gained Off-the-Tee, Strokes Gained Approach, Strokes Gained Around-the-Green) rather than used as one number, which could be tested to see which tactic is more predictive. Along the same line, many decisions were qualitatively made for this model, which may or may not be optimal, with testing of multiple different options being needed to decide. For example picking twenty-five players per week was a decision

that structurally defined which players would be in the portfolio, and without further testing, it is impossible to know whether that is more or less than the optimal amount.

Going forward, hopefully there will be more people attempting to use predictive statistics to beat the DraftKings community in order to profit off of weekly PGA golf tournaments. The general process for the model used in this study, using historical data from the specific event to predict who should play well in the current year based on their current statistics and valuing the players based on their DraftKings price could be repeated in many slightly different ways. Hopefully, over time there are enough separate models created that an optimal one will be discovered, which will have the same fundamental core of this study.

Appendix A: References

- "A Review of Driving Distance – 2016," *USGA.org*, accessed 17 April 2018, <http://www.usga.org/content/dam/usga/pdf/Equipment/2016%20Distance%20Report.pdf>.
- "AT&T Pebble Beach Pro-Am," 9-12 February 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/att-pebble-beach-pro-am/>.
- "AT&T Pebble Beach Pro-Am: Tournament Information," *PebbleBeach.com*, accessed 17 April 2018, <https://www.pebblebeach.com/events/att-pebble-beach-pro-am/>.
- "About the Tournament," *CareerBuilderChallenge.com*, accessed 17 April 2018, <https://www.careerbuilderchallenge.com/about-the-tournament>.
- "Arnold Palmer Invitational Presented by MasterCard," 16-19 March 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/arnold-palmer-invitational-presented-mastercard/>.
- Barry, Christopher, Nicholas Canova, and Kevin Capiz, "Beating DraftKings at Daily Fantasy Sports," *Stanford.edu*, accessed 17 April, 2018, <https://web.stanford.edu/class/stats50/files/BarryCanovaCapiz-paper.pdf>.
- Baumer, Benjamin and Andrew Zimbalist, *The Sabermetric Revolution: Assessing the Growth of Analytics in Baseball*, University of Pennsylvania Press, accessed 17 April 2018, <http://www.upenn.edu/pennpress/book/15168.html>.
- "CareerBuilder Challenge," 19-22 January 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/careerbuilder-challenge/>.
- Curtis, Gary N., "The Hot Hand Fallacy," *fallacyfiles.org*, accessed 17 April 2018, <http://www.fallacyfiles.org/hothandf.html>.
- "DFS Cash Games Versus Tournaments (GPP's)," 12 February 2016, *fantasysports.net*, accessed 17 April 2018, <https://www.fantasysports.net/dfs-cash-games-vs-tournaments-gpps>.
- Davis, John, "Waste Management Phoenix Open: Hideki Matsuyama Takes Aim at 3-Peat," 30 January 2018, *azCentral.com*, accessed 17 April 2018, <https://www.azcentral.com/story/sports/golf/phoenix-open/2018/01/30/waste-management-phoenix-open-hideki-matsuyama-takes-aim-3-peat/1081419001/>.

- "Farmers Insurance Open," 26-29 January 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/farmers-insurance-open/>.
- "Genesis Open," 16-19 February 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/genesis-open/>.
- Harmon, Butch, "Playing Great on Windy Days," 12 February 2012, *GolfDigest.com*, accessed 17 April 2018, <https://www.golfdigest.com/story/butch-harmon-windy-days>.
- "History: Since 1979," *ArnoldPalmerInvitational.com*, accessed 17 April 2018, <https://arnoldpalmerinvitational.com/history>.
- "Honda Classic Winners and History," 25 February 2018, *GolfBlogger.com*, accessed 17 April 2018, https://golfblogger.com/honda_classic_past_winners_and_history/.
- Hunter, David Scott, Juan Pablo Vielma, and Tauhid Zaman, "Picking Winners Using Integer Programming," *MIT.edu*, accessed 17 April 2018, <http://www.mit.edu/~jvielma/publications/Picking-Winners.pdf>.
- Kelley, Brent, "Sony Open in Hawaii Golf Tournament," 15 January 2018, *ThoughtCo.com*, accessed 17 April 2018, <https://www.thoughtco.com/sony-open-in-hawaii-golf-tournament-1565848>.
- Kilgore, Adam, "Daily Fantasy Sports Websites Find Riches in Internet Gaming Law Loophole," 27 March 2015, *The Washington Post*, accessed 17 April 2018, https://www.washingtonpost.com/sports/daily-fantasy-sports-web-sites-find-riches-in-internet-gaming-law-loophole/2015/03/27/92988444-d172-11e4-a62f-ee745911a4ff_story.html?noredirect=on&utm_term=.76dfb8653e45.
- Mayo, Patrick, "Fantasy Golf Picks: 2017 AT&T Pebble Beach Pro-Am Sleepers, Starts & Preview," 6 February 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/118602/fantasy-golf-picks-2017-pebble-beach-picks-sleepers-starts-preview-pro-am/>.
- Mayo, Patrick, "Fantasy Golf Picks: 2017 Arnold Palmer Invitational Sleepers, Starts & Preview," 13 March 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/119817/fantasy-golf-picks-2017-arnold-palmer-invitational-picks-sleepers-starts-bay-hill-preview-predictions/>.
- Mayo, Patrick, "Fantasy Golf Picks: 2017 CareerBuilder Challenge Sleepers, Starts & Preview," 16 January 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/118134/2017-careerbuilder-challenge-picks-fantasy-golf-picks-sleepers-starts-course-preview-careerbuilder-picks/>.

- Mayo, Patrick, "Fantasy Golf Picks: 2017 Farmers Insurance Open Sleepers, Starts & Preview," 23 January 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/118284/fantasy-golf-picks-2017-farmers-insurance-open-picks-sleepers-starts-preview-tiger-woods/>.
- Mayo, Patrick, "Fantasy Golf Picks: 2017 Genesis Open Sleepers, Starts & Preview," 13 February 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/118793/fantasy-golf-picks-2017-genesis-open-picks-sleepers-starts-preview/>.
- Mayo, Patrick, "Fantasy Golf Picks: 2017 Honda Classic Sleepers, Starts & Preview," 20 February 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/119146/fantasy-golf-picks-2017-honda-classic-picks-sleepers-starts-honda-classic-preview/>.
- Mayo, Patrick, "Fantasy Golf Picks: 2017 Valspar Championship Sleepers, Starts & Preview," 6 March 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/119604/fantasy-golf-picks-2017-valspar-championship-picks-sleepers-starts-valspar-preview-predictions/>.
- Mayo, Patrick, "Fantasy Golf Picks: 2017 Waste Management Open Sleepers, Starts & Preview," 30 January 2017, *RotoExperts.com*, accessed 17 April 2018, <http://rotoexperts.com/118466/fantasy-golf-picks-2017-waste-management-open-picks-sleepers-starts-preview/>.
- Newman, Nico, "History of Fantasy Sports," 4 April 2017, *Fantasy-Sport.net*, accessed 17 April 2018, <https://fantasy-sport.net/history-of-fantasy-sports/>.
- Neyer, Rob, "Sabermetrics," *Encyclopedia Britannica.com*, accessed 17 April 2018, <https://www.britannica.com/sports/sabermetrics#ref1182350>.
- "PGA Is Formed," *History.com*, 2009, accessed 17 April 2018, <https://www.history.com/this-day-in-history/pga-is-formed>.
- RBC Heritage, "Statistics: Driving Accuracy Percentage," 15 April 2018, *PGATour.com*, Accessed 17 April 2018, <https://www.pgatour.com/stats/stat.102.html>.
- RBC Heritage, "Statistics: Greens in Regulation Percentage," 15 April 2018, *PGATour.com*, accessed 17 April 2018, <https://www.pgatour.com/stats/stat.103.html>.
- RBC Heritage, "Statistics: Putts Per Round," 15 April 2018, *PGATour.com*, accessed 17 April 2018, <https://www.pgatour.com/stats/stat.119.html>.

RBC Heritage, "Statistics: Sand Save Percentage," 15 April 2018, *PGATour.com*, accessed 17 April 2018, <https://www.pgatour.com/stats/stat.111.html>.

RBC Heritage, "Statistics: Scoring Average," 15 April 2018, *PGATour.com*, accessed 17 April 2018, <https://www.pgatour.com/stats/stat.120.html>.

"Rules & Scoring," *DraftKings.com*, accessed 17 April 2018, <https://www.draftkings.com/help/rules/golf>.

"ShotLink Background," *ShotLink.com*, accessed 17 April 2018, <http://shotlink.com/about/background>.

"ShotLink Intelligence," *PGAtour.com*, accessed 17 April 2018, <https://www.pgatour.com/stats/shotlinkintelligence/overview.html>.

"Sony Open in Hawaii," 12-15 January 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/sony-open-hawaii/>.

"Strokes Gained: How It Works," 30 May 2016, *PGATour.com*, accessed 17 April 2018, <https://www.pgatour.com/news/2016/05/31/strokes-gained-defined.html>.

"The Farmers Insurance Open Returns to La Jolla January 25th," *LaJolla.com*, accessed 17 April 2018, <https://www.lajolla.com/article/regional-attractions/farmers-insurance-open-torrey-pines-san-diego/>.

"The Honda Classic," 23-26 February 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/honda-classic/>.

"The 10 Longest and 10 Shortest Courses on the PGA Tour in 2015-16," 16 December 2016, *TheGolfNewsNet.com*, accessed 17 April 2018, <https://thegolfnewsnet.com/golfnewsnetteam/2016/12/16/longest-shortest-courses-pga-tour-2015-2016-101774/>.

"Tournament History," *GenesisOpen.com*, accessed 17 April 2018, <http://genesisopen.com/tournament-history>.

"Tournament History," *wmPhoenixOpen.com*, accessed 17 April 2018, <https://wmphoenixopen.com/spectator-info/tournament-history/>.

"Valspar Championship," 9-12 March 2017, *GolfChannel.com*, accessed 17 April 2018, <https://www.golfchannel.com/tours/pga-tour/2017/valspar-championship/>.

"Valspar Championship Winners and History," 11 March 2018, *GolfBlogger.com*,
accessed 17 April 2018,
https://golfblogger.com/valspar_championship_past_winners_and_history/.

Appendix B: Table Index

Table 1: Course History	20
Table 2: Current Form	22
Table 3: Ownership Percentage	32
Table 4: Sony Regression Output	35
Table 5: Adjusted Sony Regression Output	37
Table 6: CareerBuilder Regression Output	39
Table 7: Adjusted CareerBuilder Regression Output	41
Table 8: Farmers Insurance Regression Output	44
Table 9: WM Phoenix Open Regression Output	47
Table 10: Adjusted WM Phoenix Open Regression Output	49
Table 11: Pebble Beach Regression Output	51
Table 12: Adjusted Pebble Beach Regression Output	53
Table 13: Genesis Open Regression Output	55
Table 14: Adjusted Genesis Open Regression Output	57
Table 15: Honda Classic Regression Output	59
Table 16: Adjusted Honda Classic Regression Output	61
Table 17: Valspar Regression Output	63
Table 18: Adjusted Valspar Regression Output	65
Table 19: Arnold Palmer Invitational Regression Output	67
Table 20: Adjusted Arnold Palmer Invitational Regression Output	69
Table 21: Week 10 and Beyond	70
Table 22: Overall Results	71