

MIT Open Access Articles

*How to Catch a Tiger: Understanding
Putting Performance on the PGA TOUR*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Fearing, Douglas, Jason Acimovic, and Stephen C Graves. "How to Catch a Tiger: Understanding Putting Performance on the PGA TOUR." *Journal of Quantitative Analysis in Sports* 7.1 (2011): n. pag. © 2011 American Statistical Association

As Published: <http://dx.doi.org/10.2202/1559-0410.1268>

Publisher: Walter de Gruyter

Persistent URL: <http://hdl.handle.net/1721.1/107940>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



Journal of Quantitative Analysis in Sports

Volume 7, Issue 1

2011

Article 5

How to Catch a Tiger: Understanding Putting Performance on the PGA TOUR

Douglas Fearing, *Harvard Business School*
Jason Acimovic, *MIT Operations Research Center*
Stephen C. Graves, *MIT Sloan School of Management*

Recommended Citation:

Fearing, Douglas; Acimovic, Jason; and Graves, Stephen C. (2011) "How to Catch a Tiger: Understanding Putting Performance on the PGA TOUR," *Journal of Quantitative Analysis in Sports*: Vol. 7: Iss. 1, Article 5.

DOI: 10.2202/1559-0410.1268

©2011 American Statistical Association. All rights reserved.

Brought to you by | Massachusetts Institute of Technology - MIT Libraries

Authenticated | 18.101.8.220

Download Date | 3/9/14 5:50 AM

How to Catch a Tiger: Understanding Putting Performance on the PGA TOUR

Douglas Fearing, Jason Acimovic, and Stephen C. Graves

Abstract

Existing performance metrics utilized by the PGA TOUR have biases towards specific styles of play, which make relative player comparisons challenging. Our goal is to evaluate golfers in a way that eliminates these biases and to better understand how the best players maintain their advantage.

Through a working agreement with the PGA TOUR, we have obtained access to proprietary “ShotLink” data that pinpoints the location of every shot taken on the PGA TOUR. Using these data, we develop distance-based models for two components of putting performance: the probability of making the putt and the remaining distance to the pin conditioned on missing. The first is modeled through a logistic regression, the second through a gamma regression. Both models fit the data well and provide interesting insights into the game. Additionally, by describing the act of putting using a simple Markov chain, we are able to combine these two models to characterize the putts-to-go for the field from any distance on the green for the PGA TOUR. The results of this Markov model match both the empirical expectation and variance of putts-to-go.

We use our models to evaluate putting performance in terms of the strokes or putts gained per round relative to the field. Using this metric, we can determine what portion of a player’s overall performance is due to advantage (or loss) gained through putting, and conversely, what portion of the player’s performance is derived off the green. We demonstrate with examples how our metric eliminates significant biases that exist in the PGA TOUR’s Putting Average statistic. Lastly, extending the concept of putts gained to evaluate player-specific performance, we show how our models can be used to quickly test situational hypotheses, such as differences between putting for par and birdie and performance under pressure.

KEYWORDS: golf, putting, logistic regression, gamma regression, Markov model

Author Notes: We would like to thank both the editor and the anonymous referee for their useful comments. We would also like to thank Lee Bushkell of the PGA TOUR for reaching out to MIT and providing us access to the ShotLink data. Without Mike Vitti, this project would have never begun, and we are especially grateful for his and Kin Lo’s helpful feedback and advice throughout this research. Last, we would like to thank Mary Lou Bohn, Joe Gomes, and Megan Morgan of Titleist for continually challenging us to enhance our model and investigate interesting and relevant scenarios.

1 Introduction

The sport of golf has an extremely rich tradition, with origins dating back to at least the 15th century (Cornish, 1993). Other popular sports, such as baseball, basketball, and American football, have developed loyal followings of fans who pore over statistics on a regular basis. But, statistical analysis of the game of golf has lagged.

The game of golf is well-suited to such analysis. Much like baseball, the game of golf breaks down into a sequence of well-defined discrete events, that is the shots taken. Unfortunately, golf analysis currently suffers from a few significant drawbacks. The first is that data officially reported by the PGA TOUR is limited to a small number of aggregate statistics including Drive Distance, Drive Accuracy, Greens in Regulation, and Putting Average. The second issue is that the statistics that are reported do not do a particularly good job of differentiating golfer performance. For example, there is no way to tell whether a golfer's low Putting Average is due to exceptional putting performance or equally impressive performance on approach shots. The third issue is that the reported individual statistics are heavily biased by the difficulty of the courses the golfer has played, and professional golfers play in different sets of tournaments over the course of the year.

Our work is focused on putting, which accounts for approximately 40% of the strokes taken. In order to develop a truer performance metric for putting, we attempt to address the latter two concerns mentioned above: distinguishing putting performance and controlling for green difficulty. Based on a working agreement with the PGA TOUR, we have analyzed six years of ShotLinkTM shot tracking data for thirty of the top courses on the PGA TOUR. From this data, we have developed two statistical models for analyzing the key aspects of successful putting: 1) making the putt, and 2) leaving it close when missing. We demonstrate that when combined using a simple Markov model, these two independently trained models explain the vast majority of expected putting performance for the field of PGA TOUR golfers.

Additionally, we use this Markov model to define a putting performance metric, *putts gained per round*, based on the concept of *shot value* described in Section 1.3. By applying adjustments to our two statistical sub-models, we are able to account for the fact that PGA TOUR tournaments vary significantly in both the challenge of the putting greens and the quality of the field of golfers competing. These adjustments allow us to perform a direct comparison between golfers who have played in different sets of tournaments. Conveniently, calculating putts gained per round also allows us to determine each player's

performance off the green. That is, of the difference between a player's round score and the round scoring average for the field, we can determine what portion corresponds to putting performance, with the remainder coming off the green.

Last, we demonstrate how we can use our combined, adjusted Markov model to analyze various putting situations. First, we evaluate and compare golfer performance on par versus birdie putts. The benefit of using our model for this question is that it allows us to directly assess the difference in terms of scoring impact. Next, we use our model to analyze golfer performance under pressure during the last round of each tournament. Based on the underlying statistical model, we evaluate and analyze the confidence intervals around these results.

1.1 The Game of Golf

A PGA TOUR golf tournament consists of four rounds played over four days (usually Thursday through Sunday). Each day, the hole is moved to a different location on each of 18 greens. After the second round, typically, the players with the 70 best scores make "the cut," and advance onto rounds 3 and 4. The winner of the tournament is the player with the lowest cumulative score over the four days of the tournament.

Until now, golfers have been evaluated by a limited number of metrics. One metric of interest to golfers and fans is *greens in regulation* (GIR). To reach the green in regulation, the golfer must reach the green within "par minus 2" strokes. For example, on a par 3 hole, the golfer must reach the green in one stroke from the tee to have it considered a *green in regulation*. The metric most commonly used to evaluate putting specifically is Putting Average (PA). This is the average number of putts a golfer takes when he reaches the green in regulation. A typical professional Putting Average is around 1.7 putts per green in regulation. We show later that this metric is biased to favor golfers with better approach shots.

1.2 Strokes-to-Go and Shot Value

A large vocabulary has grown on the golf course to attempt to describe the worth of particular shots.¹ Birdies, eagles, and albatrosses are good; shanking, skulling, and chunking are not. There are shots named after conservative republicans and liberal democrats (think about shots erroneously flying right or left, respectively).

¹ Pennington, Bill. "Golf's Colorful Language Goes with Any Green." *New York Times* 11 May 2009: D7.

But despite this extensive and colorful vocabulary, there exists no widespread method of objectively evaluating how good a good shot is, or how bad a bad shot is. Building on the work of Landsberger (1994) and Broadie (2008), we explicitly quantify the value of an individual shot.

We use the term *strokes-to-go* to refer to the expected number of strokes it takes to get from a particular location on a particular hole to reach the hole itself. Here, expectation is taken over the field of PGA tour professionals. This differs slightly from previous definitions in that the baseline is based on the field of PGA TOUR golfers rather than on a scratch golfer (cf. Broadie 2008).

As outlined in Broadie (2008), the value of an individual shot can be calculated by comparing the *strokes-to-go* before the shot with the *strokes-to-go* after the shot. On average, we would expect *strokes-to-go* to decrease by one with each shot taken. Any decrease more than this is a good shot and any decrease less than this a bad shot.

$$\text{shot value} = \text{strokes-to-go}_{\text{Before}} - \text{strokes-to-go}_{\text{After}} - 1$$

As an example, let's look at the par 4 14th hole at Quail Hollow, illustrated in Figure 1.



Figure 1: Example of a hypothetical golfer's shots on the 14th hole at Quail Hollow.

For this hole, the expected *strokes-to-go* from the tee is 4.1 (regardless of the fact that the course designers set par at 4). We would expect the average score on this hole to be 4.1 were it to be played by a field of average PGA golfers. A hypothetical golfer might drive the ball off the tee to a place on the fairway where the *strokes-to-go* is 3.2.² (See Location 2 in the figure.) The *strokes-to-go* before the shot was 4.1 and after the shot 3.2, so that shot was worth $(4.1 - 3.2 - 1 =) -0.1$ strokes. The value is less than zero, implying it was not a good shot. He then hits the ball onto the green about 15 feet from the hole (Location 3 on the figure). From our model, we know the expected *strokes-to-go* for this location on the green is 1.8, so the golfer has just picked up $(3.2 - 1.8 - 1 =) 0.4$ strokes. In an excellent putt, he then manages to sink this 15-footer. This final putt has a shot value of $(1.8 - 0 - 1 =) 0.8$. In all, he's gained $(-0.1 + 0.4 + 0.8 =) 1.1$ strokes on this hole, relative to the expectation of the field of PGA tour golfers. We could have just as easily calculated this by comparing his actual strokes (3) to the average strokes for the hole among PGA TOUR golfers (4.1). But the benefit of the *shot value* metric is that we know exactly where he gained or lost strokes. Our hypothetical golfer had a great putt (picking up 0.8 strokes), a good approach shot (picking up 0.4 strokes), and a worse-than-average drive (losing 0.1 strokes). These values are summarized in Table 1.

Shot Location	Strokes-to-Go	Shot value of the stroke taken from this location
Tee	4.1	-0.1
2	3.2	0.4
3	1.8	0.8
Hole	0	NA

Table 1: Example of shot values of a hypothetical golfer's shots on the 14th hole at Quail Hollow.

The putting model described in this paper predicts *strokes-to-go* for any position on the green, for any player, and for any hole. Subsequently, since we

² Although this paper does not include a model for *strokes-to-go* off the green, we have developed a model to predict *strokes-to-go* from anywhere on a golf course including from bunkers, the rough, and the fairway. All values off the green for this example are hypothetical and for illustrative purposes.

are focusing solely on the green, we refer to this value as *putts-to-go*. Knowing the *shot value* of every putt allows us to compare player putting performance in a fair, unbiased way.

1.3 ShotLink™ Dataset

In 2001, IBM and the PGA TOUR teamed up to introduce an advanced data collection system known as ShotLink. Using about 250 volunteers per tournament, ShotLink captures the ball location for every shot taken on the PGA TOUR. When a ball lands on the golf course, volunteers point a survey laser at it to pinpoint its location and elevation to within 1 cm on the green and within 1 foot on the fairway. Simultaneously, walking scorers use handheld devices to capture non-locational data about each ball, such as quality of lie, player, ground slope, score, etc. For example, on the final round of the 2008 Crowne Plaza Invitational at Colonial Country Club, Phil Mickelson's drive on the 18th hole took him to a location in the rough 140 yards from the hole and ten inches below it. On his second stroke, he hit the ball 137 yards onto the green, landing him 9 feet 3 inches from the hole laterally and 3/16 of an inch above the hole vertically. He went on to make this 9 foot putt for birdie winning the tournament and \$1.098 million dollars.

There are 45 PGA TOUR events per year, each held at a different course. In this paper we focus on 30 courses.³ These include the tournaments with the best qualities of field as measured by the World Golf Rankings of their participants. The dataset on which we base our analysis includes every putt on these thirty courses from 2003 to 2008, approximately 2 million rows of data. This roughly corresponds to [100 players] * [6 years] * [30 courses] * [4 rounds per tournament] * [18 holes per tournament] * [about 1.6 putts per hole]. These 2 million putts were taken by more than 1000 distinct players. In order to validate our model, we hold out 15% of these putts as a test set.

³ Annandale Golf Club, Bay Hill Club, Brown Deer Park Golf Club, Cog Hill Golf and Country Club, Colonial Country Club, Congressional Country Club, Doral Resort and Country Club, East Lake Golf Club, Firestone Country Club (South Course), Harbour Town Golf Links, La Cantera Golf Club, Magnolia Golf Club, Montreux Golf and Country Club, Muirfield Village Golf Club, Plantation Course at Kapalua, Quail Hollow Country Club, Riviera Country Club, Torrey Pines (South Course), TPC Boston, TPC Deere Run, TPC Las Colinas, TPC River Highlands, TPC Sawgrass, TPC Scottsdale, TPC Southwind, TPC Sugarloaf, Waialae Country Club, Warwick Hills Golf and Country Club, Westchester Country Club, and Westin Innisbrook – Copperhead.

1.4 Literature Review

Golfers and statisticians have been trying to model the game of golf for decades. Although analysis has covered modeling everything from total score (see for example, Brown, 2007) to the impact of doubling the hole size (Landsberger 1998, Bansal and Broadie 2008), the green has been the most prolific area of study. Data from tournaments and controlled putting experiments have been collected since the 1960s. Cochran and Stobbs (1968) measured every ball location for every round for 6 holes in the Dunlop Masters Tournament at Royal Birkdale in 1964. They used this data to plot the sigmoidal relationship between distance and probability of holing-out as well as the monotonically increasing relationship between average number of putts versus starting distance. While the scale has changed since then, the shape of the plots has not changed much over the last 45 years. Soley (1977) investigated the relationship between the starting distance of putts that missed and the distance remaining to the pin. He collected data on the 25th, 50th, and 75th percentiles of distances-to-go from missed putts, finding a roughly linear relationship between starting distance and remaining distance quartile for first putts longer than 11 feet.

While these early texts represent important advancements, many of them concentrate on understanding the empirical data. More recent works, on the other hand, have started to build parametric models that not only predict the outcomes, but whose parameters give insight into the game itself. Hoadley (1994) presented a model in which error entered the putting process both in the initial angle of the putter and in the intended distance (both errors were modeled as normal distributions). Tierney and Coop (1998) took the alternate path of modeling the final position of the ball as a bivariate normal distribution centered in back of the hole with asymmetrical “x” and “y” errors (the distance error is more severe than the side-to-side error). Carnahan (2002) carried out a controlled experiment for which “vintage potables ...[were]....offered as prizes” in which he attempted to quantify the effect of slope, distance, and break. Although he mentions logistic regression as a possible next step (which is unpublished as far as we know), his analysis focuses primarily on determining if break and slope significantly affect the probability of holing-out, and reporting the hole-out percentage of the different distances in the experiment.

Pope and Schweitzer (2010) predict the probability of making the putt using a logistic regression model. The explanatory variables include functions of distance as well as dummy variables for eagle, birdie, par, bogey, and double bogey situations. Their findings support the hypothesis that players try harder and take more risks on putts for par as opposed to putts for birdie, even though it is the cumulative score over the entire tournament that determines their positions and earnings.

Landsberger (1994) attached a value to individual strokes for his “Golf Stroke Value System.” On the green, the value of a stroke is the difference in expectation between the starting and ending position of the ball. Broadie developed this idea further in 2008 with his idea of “fractional par,” which is the expected number of strokes remaining for a scratch golfer from anywhere on the course. He then uses fractional par to value individual shots, where the value of a stroke is the beginning fractional par minus the ending fractional par minus 1. Both of these models are based on empirical data.

In terms of looking at professional golf metrics and trying to rank players in an accurate way, less work has been done. Cochran and Stobbs (1968) remark on the biases of the Putting Average metric. Larkey and Smith (1998) discuss some of the shortcomings of other current golf metrics, and the need for improvement. This work is a follow up of Larkey’s (1994) earlier work on ranking professional golfers by skill. He had identified the difficulty in ranking players who do not all play in the same tournaments. Larkey develops a metric based on a player’s earnings relative to purse size as a starting point for measuring skill (to normalize for inflation). Connolly and Rendleman (2008) also emphasize the importance of including player skill in their work on the role of luck and streaky play in golf. They use as data professional golfers’ aggregated golf scores per round of golf.

1.5 Contributions

The primary contribution of our work is the development of a new putting performance metric, *putts gained per round*, that eliminates the most significant biases inherent in existing PGA TOUR putting statistics including Putting Average. We subsequently develop a simple model to predict *strokes-to-go* off the tee and use this to evaluate off-green performance. For both our on-green and off-green models, we develop an iterative algorithm that allows the full 30 tournaments worth of data to be considered in a computationally efficient fashion. In addition, we show that by applying the concepts of *putts gained* more broadly, many interesting questions can be answered about the game of golf. We believe that incorporating *putts gained per round* into the reported PGA TOUR statistics would not only help existing fans better understand the game, but would also generate new interest in the sport.

2 Statistical Modeling

In this section, we describe our statistical modeling process and the resulting models. In the first three subsections, we develop the statistical models that provide the core of our subsequent analysis. First, we model the probability of *holing out*, that is, of making the putt, using a logistic regression. Next, assuming the putt is missed, we model the conditional distribution of *distance-to-go* using a gamma regression. Third, using a simple Markov model, we combine the *holing out* model with the *distance-to-go* model to estimate the distribution of *putts-to-go* for the field of PGA TOUR golfers.

The two regression models, *holing out* and *distance-to-go*, utilize distance and simple transformations of distance (i.e., logarithm and polynomials) as the primary covariates. Thus, we ignore features such as green speed and the contours of the green. In Section 2.4, we argue that although these features are statistically significant, they are practically irrelevant for our purposes because their exclusion does not introduce any systematic biases when measuring and comparing putting performance.

Though we feel it is safe to exclude green-specific features in our model, this does not mean we can ignore green differences entirely. As mentioned previously, different golfers play in different tournaments with different putting greens and corresponding difficulty levels. Thus, in order to effectively compare putting performance either between players or situations, we need to control for this variation. To do so, we introduce adjustments for green difficulty into our two regression models. In order to separate the impacts of green difficulty and quality of field, we need to simultaneously consider adjustments for player skill level. We describe this process and its benefits in Section 2.5.

Lastly, for practical relevance our approach needs to be computationally tractable to allow for inclusion all of the PGA TOUR events. In Section 2.6, we describe an iterative training procedure for estimating the green and player adjustments. This procedure allows us to consider segments of the dataset independently, thus dramatically improving performance. We use this approach to include 30 courses in the results we report throughout the paper.

2.1 Holing Out Model

As reviewed in Section 1.4, the probability of *holing out* has been discussed in the literature. For many questions, an empirical analysis of this probability against distance is sufficient as the data are remarkably stable, up to 60 feet or more. In Figure 2, we plot the empirical probability of *holing out* versus starting distance. This plot is based on 30 golf courses, and starting distances are bucketed into two-

foot intervals. The starting distance point used for each bucket in the plot is the empirical mean of the starting distances within that bucket.

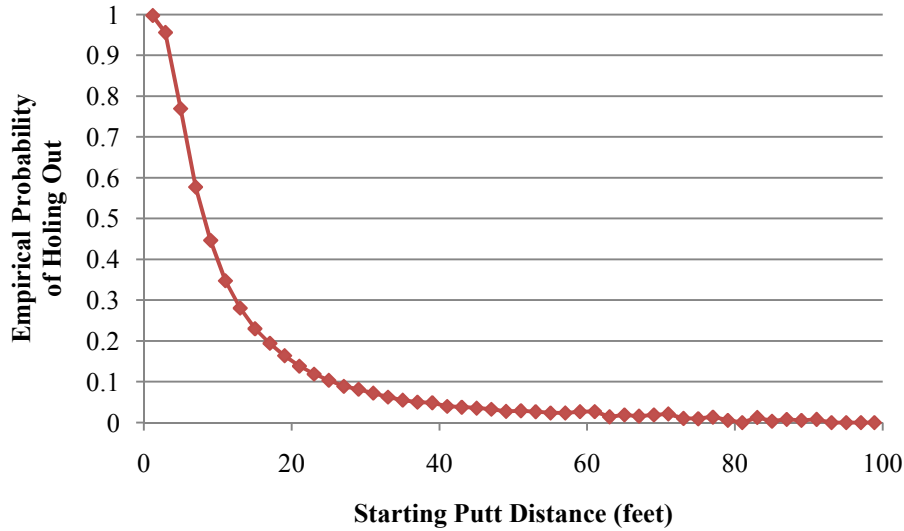


Figure 2: Empirical probabilities of holing out plotted against putt distance.

We desire a parametric model of holing out for two reasons: to smooth out the curve in areas in which the data are sparse, as with long distance putts, and, most important, to analyze additional impacts on holing out probabilities like green difficulty and quality of field. With these goals in mind, a binomial logistic regression model is a natural choice. For our first model, we include only an intercept term and five, distance-based terms corresponding to the logarithm of distance and a fourth-order polynomial of distance. Thus, we write the probability of one-putting as:

$$P[Y=1|d]=\left[1+e^{-(\beta_0+\beta_1d+\beta_2d^2+\beta_3d^3+\beta_4d^4+\beta_5\log d)}\right]^{-1} \quad (1)$$

where Y is the number of *putts-to-go* and d is the distance to the pin. In Section 2.5, we include additional dummy variables to incorporate the effects of green difficulty and quality of field. The decision to include the fourth-order polynomial and natural logarithm of distance was determined through 10-fold cross validation (Hastie, Tibsharani, and Friedman 2001).

The parameters are fit in order to maximize the likelihood of observing the training data, which is equivalent to minimizing the deviance (McCullagh and Nelder 1989). The fitted values for these parameters, along with the corresponding Wald test z-statistics and p-values, are listed in Table 2. Each of the parameters is significantly different from 0 at an extremely small significance level. This model leads to a reduction in deviance from 2,199,471 for the null model to 988,732 for the model with 5 parameters, which is also extremely significant. In Figure 3, using these fitted parameter values, we plot the model against the empirical results averaged over two foot-buckets. Based on a visual inspection, it is clear that the model fits the empirical data well.

	Estimate	Standard Error	z-statistic	p-value
(Intercept)	7.31E+00	2.48E-02	294.71	< 2E-16
Logarithm distance	-5.58E+00	4.59E-02	-121.61	< 2E-16
First-order distance	6.76E-01	1.28E-02	52.99	< 2E-16
Second-order distance	-1.97E-02	5.21E-04	-37.80	< 2E-16
Third-order distance	2.93E-04	1.02E-05	28.66	< 2E-16
Fourth-order distance	-1.62E-06	7.13E-08	-22.71	< 2E-16

Table 2: Fitted parameters of holing out logistic regression model.

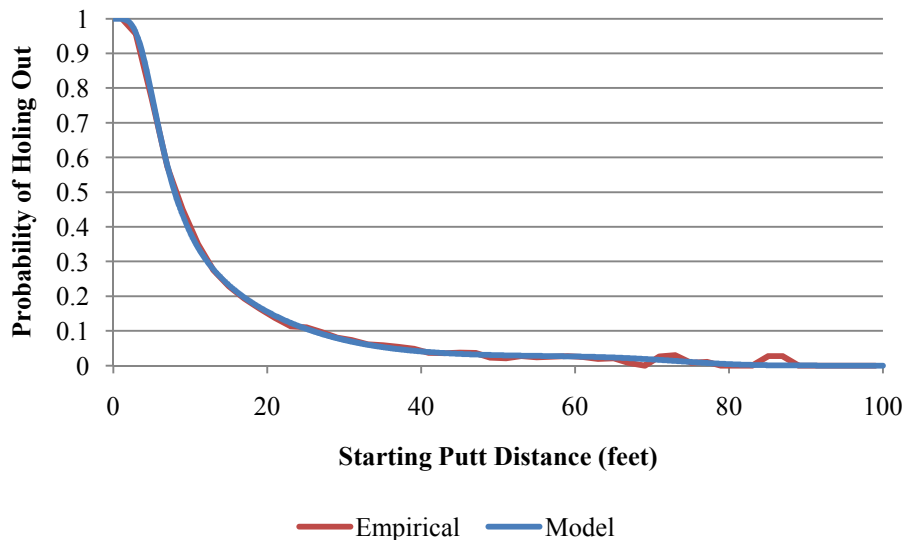


Figure 3: Empirical probabilities of holing out compared to fitted logistic regression model.

2.2 Distance-to-Go Model

In this section, we are concerned with where the ball comes to rest if the putt misses. That is, if a professional golfer misses a putt from distance d , what is the distribution of *distance-to-go*, the remaining distance to the pin? In Figure 4, we note a strong, roughly linear relationship between putt distance and the mean of *distance-to-go*. Soley (1977) describes a similar trend for each of the quartiles of *distance-to-go*. Note that beyond 80 feet this relationship starts to break down due to the small number of matching observations.

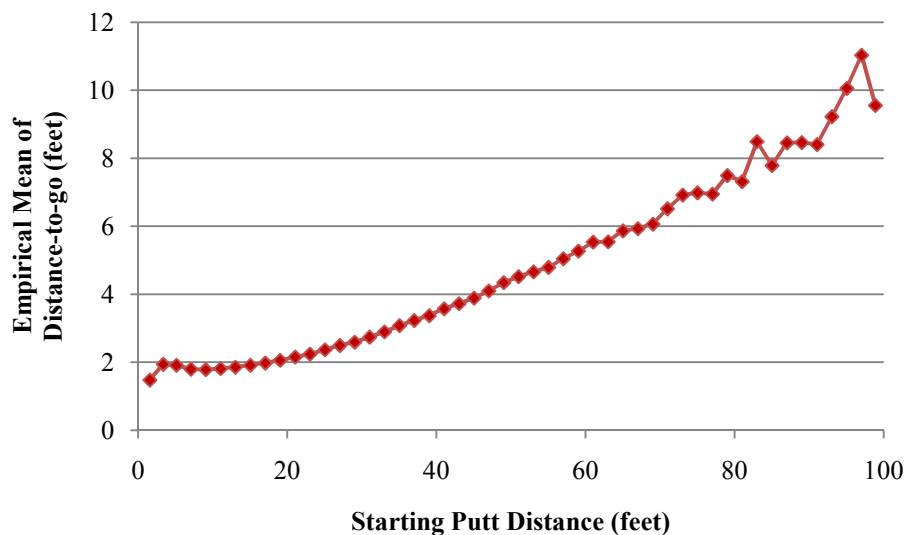


Figure 4: Empirical mean of *distance-to-go* conditional on missing, plotted against initial putt distance.

Regarding the data, there are two points worth noting. First, the empirical standard deviation of *distance-to-go* is increasing relative to putt distance, as displayed in Figure 5, with a roughly constant coefficient of variation.⁴ Second, *distance-to-go* is strictly positive, with a positively skewed distribution and a strictly positive mode that is typically between 2 and 3 feet from the hole. For example, in Figure 6 we chart the empirical distribution of *distance-to-go* for 10-foot, 20-foot, 30-foot, and 40-foot putts. With this in mind, we immediately rule out an Ordinary Least-Squares (OLS) regression since the assumptions of homoscedasticity and normally-distributed errors are violated.

⁴ The coefficient of variation equals the standard deviation divided by the mean.

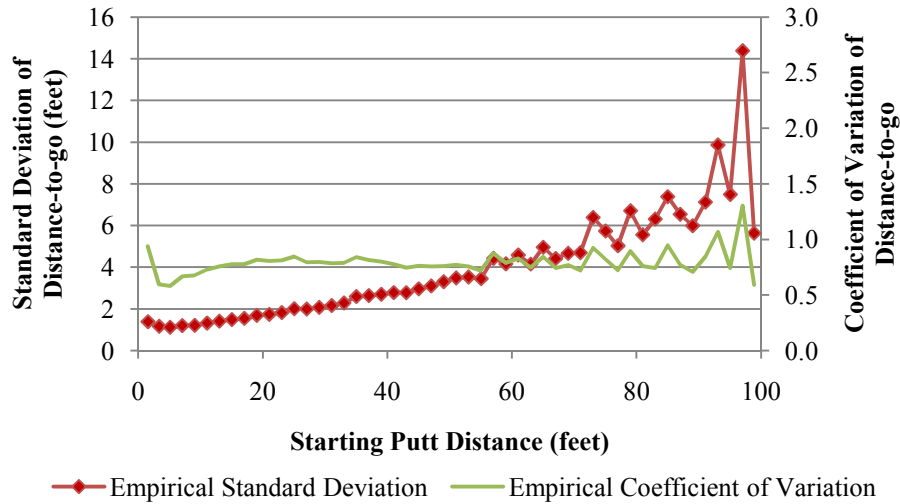


Figure 5: Empirical standard deviation and coefficient of variation of *distance-to-go* conditional on missing plotted against initial putt distance.

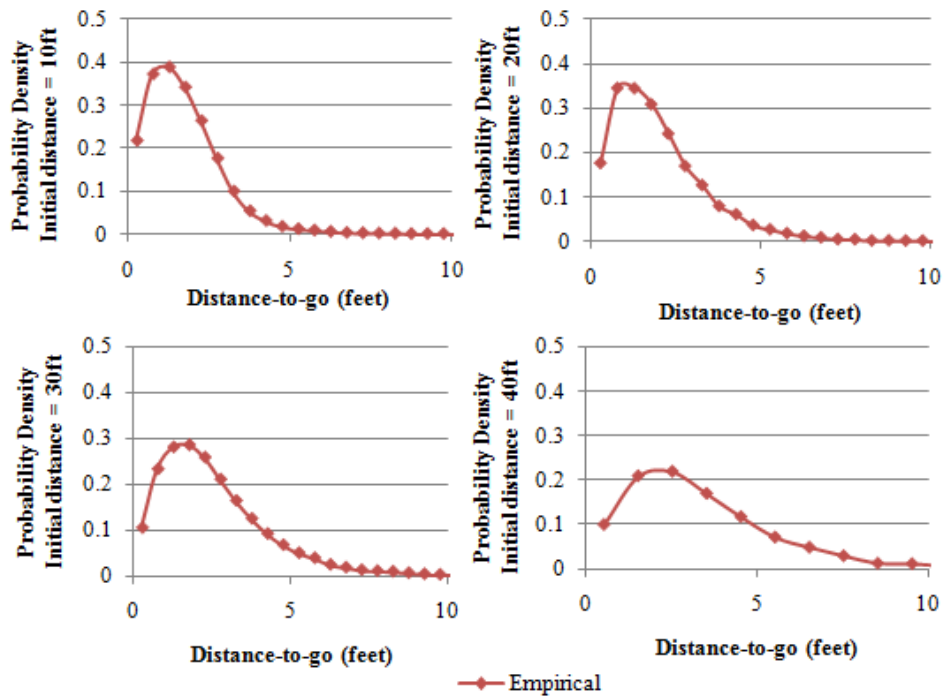


Figure 6: Empirical distributions of *distance-to-go* conditional on missing for initial putt distances in the one-foot buckets centered at 10 feet, 20 feet, 30 feet, and 40 feet.

A gamma regression model, one of the standard, generalized linear models, is an appropriate choice for data with a constant coefficient of variation and error terms that are approximately gamma distributed (McCullagh and Nelder 1989). The gamma distribution, like our data, is strictly positive, positively skewed, and peaked away from 0, at least for appropriate values of the shape parameter. For a generalized linear model, such as a gamma regression, the link function determines the functional relationship between the linear-in-parameter regression function and the response variable. For instance, if a generalized model has a log link function, the linear-in-parameter regression function predicts the natural logarithm of the mean of the response variable. In our testing, an identity or log link function fits the data equally well. We choose to report on the results utilizing a log link for two reasons: the log link maps the strictly positive *distance-to-go* to the real domain of the linear-in-parameter regression function; and the log link enables a nice interpretation of the results, which we discuss further in Section 2.5. As with our holing out model, we consider only an intercept and distance-related terms in our set of independent variables. In this case, we utilize a second-order polynomial of distance instead of a fourth-order polynomial. Additionally, a gamma regression model requires us to fit the shape, k , of the gamma distribution. Thus, for the *distance-to-go* model, the fitted parameters include the coefficients of the logarithm of distance, the second-order polynomial of distance, and the shape, k , of the gamma distribution. We write the density function of *distance-to-go*, $f(z|d;\beta,k)$, as follows:

$$\mu_d = e^{\{\beta_0 + \beta_1 d + \beta_2 d^2 + \beta_3 \log d\}} \quad (2)$$

$$f(z|d;\beta,k) = kz^{k-1} \frac{e^{-zk/\mu_d}}{\Gamma(k)\mu_d} \quad (3)$$

In Equations 2 and 3 above, μ_d represents the conditional mean of *distance-to-go* from distance d , $f(z|d;\beta,k)$ is the equation for the gamma density with shape k and scale μ_d/k , and $\Gamma(\cdot)$ is the gamma function. As with the *holing out* model, the decision to include the second-order polynomial and natural logarithm of distance was determined through 10-fold cross validation.

As with the logistic regression, we estimate these parameters to minimize the model deviance. The estimates for these parameters, along with the robust standard errors and t -values, are listed in Table 3.⁵ Using these parameter

⁵ Although we recognize that using t -tests to determine significance of independent variables in a model with non-normal errors is only an approximate approach, we are nonetheless comfortable

estimates, we plot the mean and standard deviations from the fitted model against the empirical results in Figure 7 and Figure 8. In Figure 9, we similarly plot the conditional distribution of *distance-to-go* based on our model against the empirical distributions for a sample of initial putt distances. As with the holding-out model, we see a close fit between our *distance-to-go* model and the empirical data.

	Estimate	Standard Error	t-value	p-value
(Intercept)	9.5E-01	9.6E-03	99	< 2E-16
Logarithm distance	-3.5E-01	6.6E-03	-53	< 2E-16
First-order distance	4.6E-02	6.3E-04	73	< 2E-16
Second-order distance	-1.6E-04	5.8E-06	-28	< 2E-16
Shape (k)	2.132	0.004		

Table 3: Fitted parameters of *distance-to-go* gamma regression model.

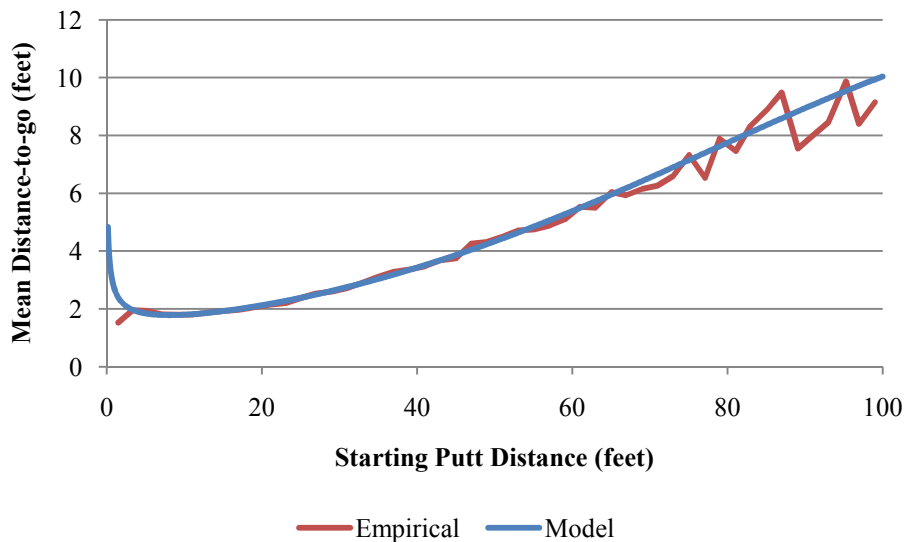


Figure 7: Empirical mean of *distance-to-go* conditional on missing compared to fitted gamma regression model.

concluding that the coefficients are significant because their standard errors are more than an order of magnitude smaller than their coefficient estimates.

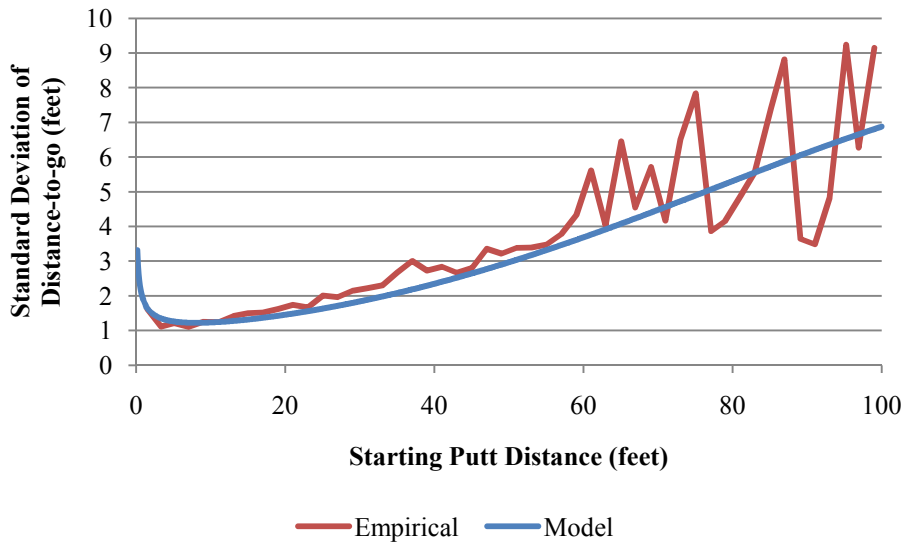


Figure 8: Empirical standard deviation of *distance-to-go* conditional on missing compared to fitted gamma regression model.

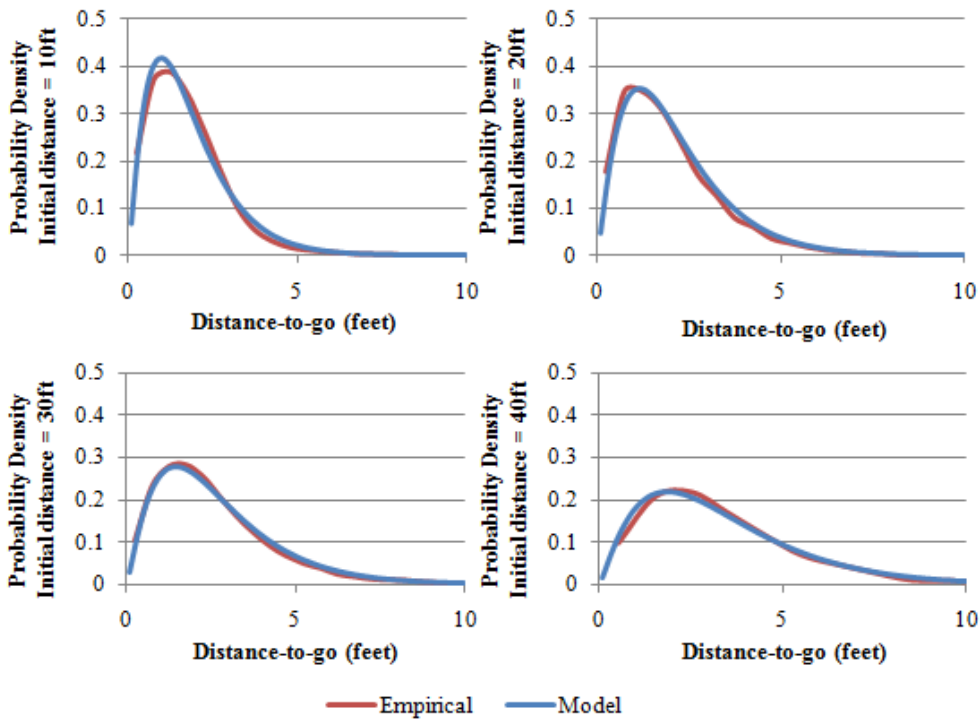


Figure 9: Empirical distributions of *distance-to-go* conditional on missing for initial putt distances in the one-foot buckets centered at 10 feet, 20 feet, 30 feet, and 40 feet compared to fitted gamma regression model.

2.3 Putts-to-Go Model

In this section, we combine our holing-out logistic regression model with our *distance-to-go* gamma regression model to produce a Markov model of *putts-to-go*. Although we require only the expected *putts-to-go* for each distance in order to calculate the performance metrics, we desire the distribution of *putts-to-go* in order to calculate confidence intervals around these results. With this in mind, in Figure 10 and Figure 11 we plot both the empirical mean and standard deviation of *putts-to-go* for each initial putt distance. It is interesting to note that between 0 feet and 10 feet, fewer than 1% of holes played are 3-putted by a professional golfer. Thus, we can roughly consider the *putts-to-go* as a Bernoulli trial in this range. This explains the peak in standard deviation around 7 feet, at which the likelihood of one-putting or two-putting is each approximately 50%. Between 10 feet and 30 feet, the likelihood that a professional golfer will 2-putt a hole increases rapidly, while the chance of 3-putting remains relatively low. This explains the corresponding dip in standard deviation. The standard deviation begins to rise again beyond 30 feet as the proportion of 3-putts increases at a steady rate. This non-monotonic behavior of the standard deviation is due to the discrete nature of the data.

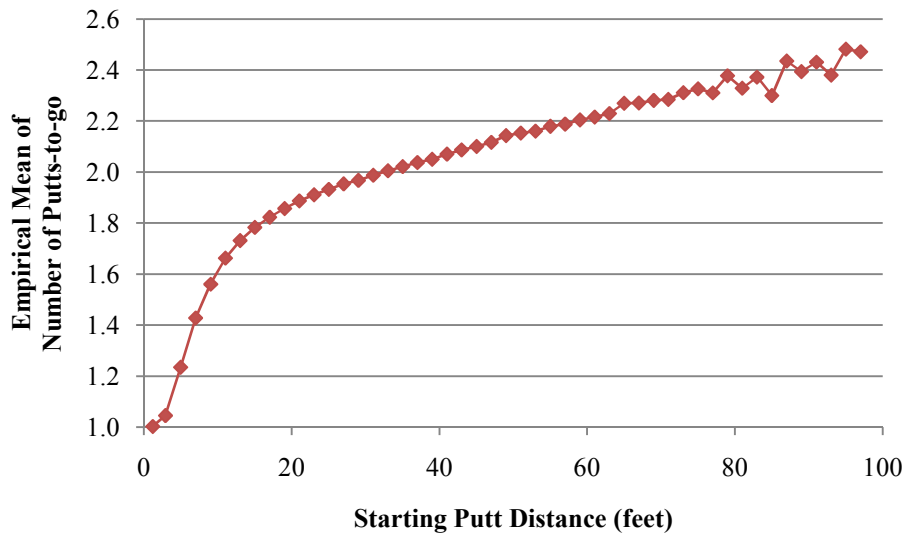


Figure 10: Empirical mean of putts-to-go plotted against initial putt distance.

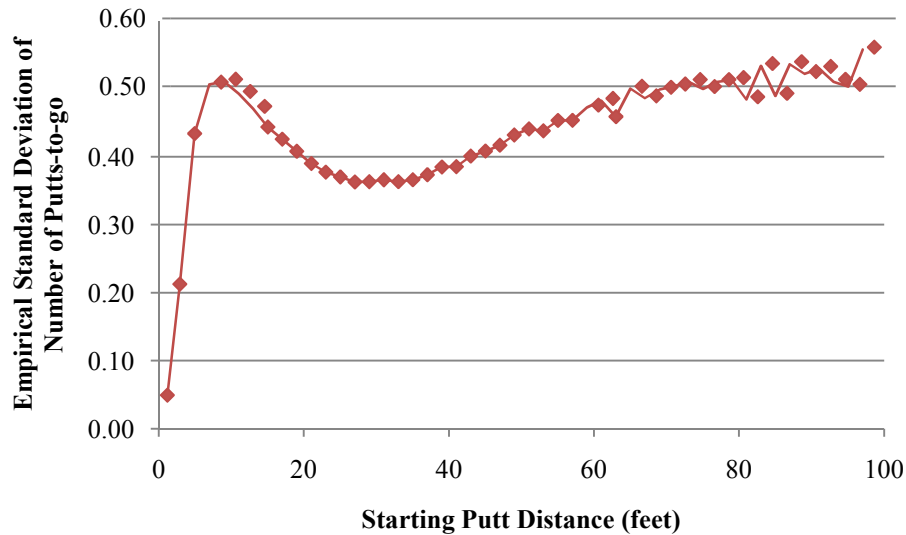


Figure 11: Empirical standard deviation of putts-to-go plotted against initial putt distance.

Although a multinomial logistic regression or Poisson regression would model the discrete data directly, we instead choose to take an approach that more closely represents the physical act of putting. With the holing out and *distance-to-go* models at our disposal, we describe the process of putting through a simple Markov chain. The state space is the positive real line representing the distance to the hole combined with an absorbing state, H , representing the hole. From any state $d \in \mathbf{R}^+$, the probability of transitioning to state H is equal to the holing out probability, and the probability density associated with transitioning to state $z \in \mathbf{R}^+$ is equal to the probability of missing the first putt times the density represented by the *distance-to-go* model. In terms of the Markov model, the probability of reaching the hole in n putts from distance d corresponds to the probability of first reaching the absorbing state, H , in exactly n transitions from state d . More formally, we have:

$$P[Y = n | d] = \begin{cases} P[Y = 1 | d] & n = 1 \\ P[Y > 1 | d] \int_0^\infty P[Y = 1 | z] f(z | d) dz & n = 2 \\ P[Y > 1 | d] \int_0^\infty P[Y = 2 | z] f(z | d) dz & n = 3 \\ \dots & \dots \end{cases} \quad (4)$$

where, as above, Y is the number of putts, d is the starting putt distance, z is the conditional *distance-to-go*, and $f(z|d)$ is the gamma density function.

This model requires two key assumptions. First, a putt that starts on the green ends either in the hole or on the green. For professional golfers, this is a very reasonable assumption, as only 0.036% of putts in the ShotLink data set ended up off the green. Second, we need to satisfy the Markov property, which in our context states that the probability of reaching the hole in n -putts from a given distance is independent of the previous putts the golfer has taken on the hole. For individual golfers, we would expect the model performance to degrade somewhat as the number of putts taken previously is correlated with the skill level of the golfer. This concern is addressed to an extent by the adjustments we make to our model in Section 2.5. Additionally, the Markov property suggests that from the same distance a second putt should be just as difficult as a first putt. In fact, as we show in Section 3.3, the second putt is consistently easier than the first putt. We choose not to model this directly in order to maintain the simplicity of the model as the basis for an easily interpretable putting performance metric.

Using Equations 1, 3, and 4, we can easily calculate the mean and standard deviation of *putts-to-go*. These results are plotted against the empirical values in Figure 12 and Figure 13. Additionally, in Figure 14, we plot the probabilities of reaching the hole from distance d within n putts against the empirical proportions. Our model suggests that even at 100 feet, the probability of requiring 4 or more putts to reach the hole is less than 3.5%.

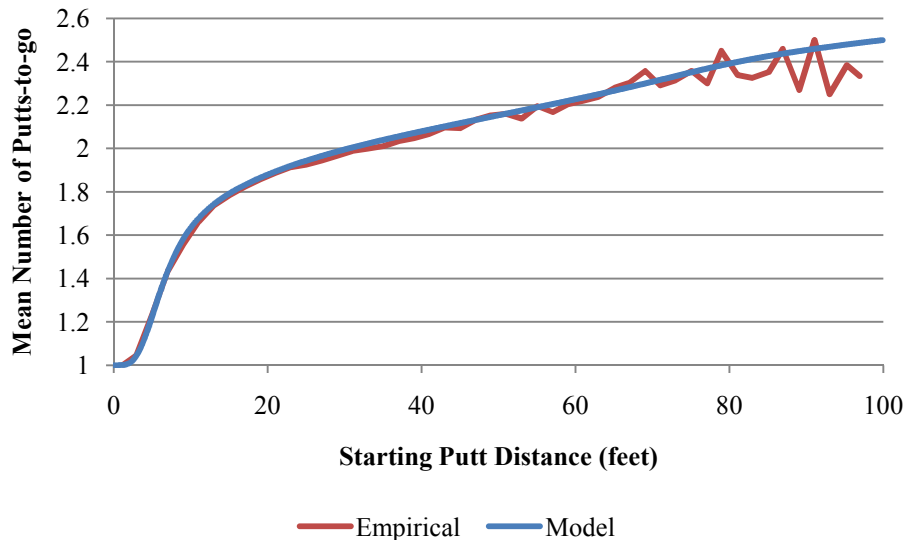


Figure 12: Empirical mean of putts-to-go compared to fitted Markov model.

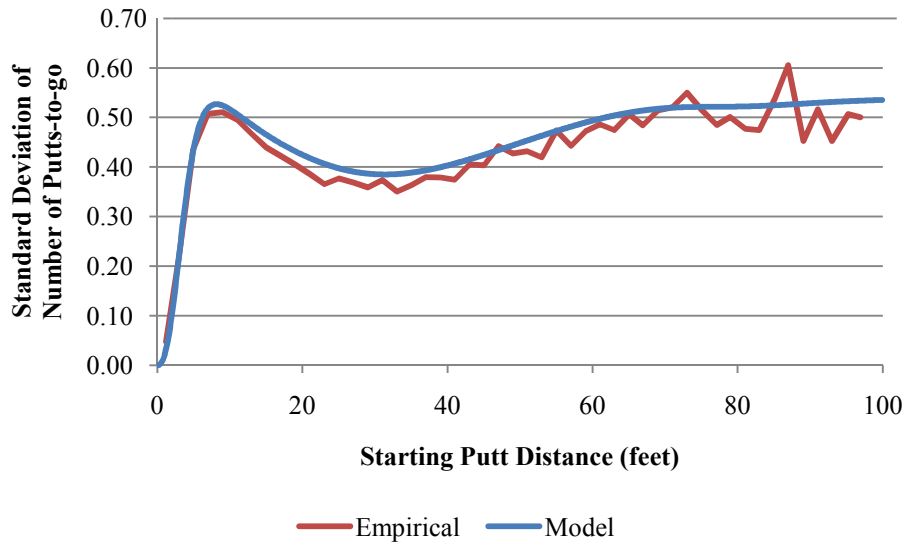


Figure 13: Empirical standard deviation of putts-to-go compared to fitted Markov model.

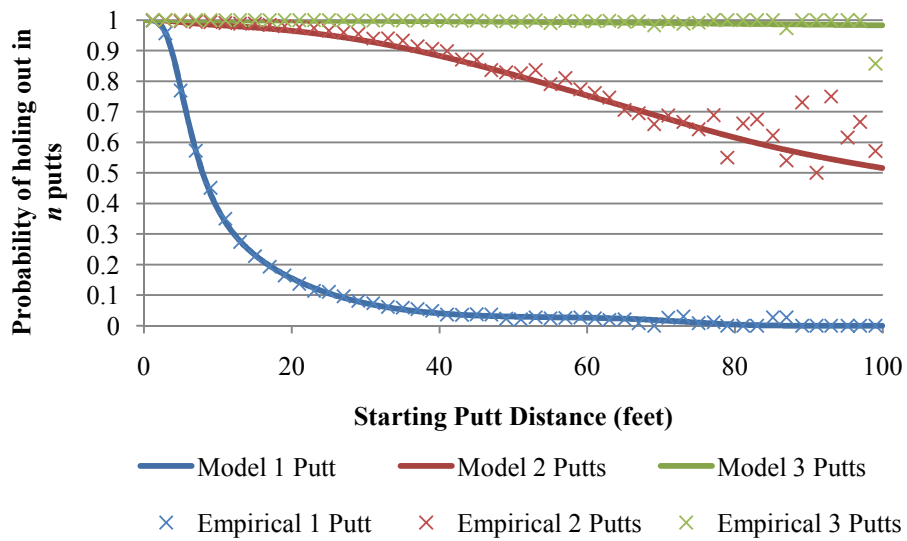


Figure 14: Empirical "Within n putts" probabilities compared to estimates based on fitted Markov model.

2.4 Considering Additional Features

In our efforts to build a performance metric for putting, we have made decisions about which features to include in our model and which to exclude. These decisions are driven by the goal of creating a simple model, but also of including enough features to allow for the unbiased evaluation of putting performance.

Features that affect putting (and therefore can introduce bias into any putting metric) can be thought of as residing in two groups. The first group contains features that are relatively consistent for a given hole. These include grass type, grass length, hole difficulty and terrain, and to some extent weather. Because every golfer does not play every hole (in fact, golfers get to choose in which tournaments they play), it is necessary to correct for this. A golfer who always plays on courses with fast, tougher greens should not be penalized over the golfer who chooses to play on slower, easier greens. In Section 2.5, we describe how we correct for individual hole difficulty to address this first source of bias.

The second group of potentially bias-inducing features differs from putt to putt, even within a single green. The most obvious is distance. In fact, our motivation for creating a better putting metric is driven in part by the observation that some golfers consistently putt from shorter distances than others, as we demonstrate in Section 3.1. This makes the popular Putting Average metric ineffective for comparing putting performance.

Another potentially bias-inducing feature in this category is slope. It is conventional wisdom among the golfing community (and supported by our analysis) that uphill putts are easier to make than flat putts, and that flat putts are easier to make than downhill putts. If some golfers have the skill to consistently place the ball downhill of the pin more often than their peers, then their putting metrics may look inflated because, for a given distance, they are actually often facing easier (uphill) putts. However, if all golfers have about the same distribution of uphill versus downhill putts, we can safely exclude this feature and still argue that our putting metric is fair and unbiased. In fact, we show that it is the case that there is an insignificant difference in the distribution of uphill versus downhill putts for all golfers.

To test the hypothesis that players' slope distributions are similar, we look at a sub-sample of ten courses from our overall sample of 30. For these courses, we bucket the putts into one-foot increments by distance. For each bucket, we run an OLS regression on the corresponding subset of data with slope as the dependent variable and player as the independent variable. Slope is defined in the conventional sense as elevation of the pin above the ball divided by the distance from the ball to the pin. (Note that slope is negative if the ball is above the pin.) The F-statistic from this regression, then, provides us with the means to test the null hypothesis that "All players' slope coefficients are equal." For distance

buckets farther than 3 feet from the hole, regressions on all distance subsets, save one, fail the F-test.⁶ Although the 8- to 9-foot bucket indicated a statistically significant difference, we find it unlikely that player performance differs at this distance while being less significant at every other distance from 3 to 24 feet. Thus, we conclude there is not sufficient evidence to believe that slopes vary significantly from player to player beyond 3 feet.⁷

For observations 3 feet and closer, the regressions show that some players' coefficients are statistically significant. However, practically, these players' coefficients are irrelevant because (i) the magnitude of the effect is so small, (ii) golfers almost always make these short putts, regardless of exogenous factors such as slope, and (iii) only a small number of golfers have statistically significant slope coefficients.

For the subset of golfers whose slope coefficients are statistically significant (about 20 out of 600 golfers), their ability to consistently place the ball downhill from the pin improves their probability of holing out at 3 feet from about 95.4% to 95.8%. Under the assumption that putters sink their balls in either 1 or 2 putts from this distance, this implies that their expected *putts-to-go* improves from 1.046 to 1.042. This equates to a potential bias of 0.04 putts per round (assuming that about a third of all putts are taken from within three feet, and that there are about 1.6 putts per green). As we will see later in section 3.1, the order of magnitude of the standard deviation of *putts gained per round* is 0.1, over twice the size of this potential bias.

In addition to linear slope, second order contours such as hills and dips also have some effect on a golfer's ability to *hole out*. However, these complicated contours are beyond the scope of this analysis. This resulting paucity of features has the benefit of making the model easier to execute and understand, without sacrificing our goal of an unbiased putting metric.

2.5 Green Difficulty and Quality of Field

As discussed previously, variation in green difficulty levels is an important factor to consider when comparing player performance. A 10-foot putt on a slower,

⁶ The resulting p-values were 0.00, 0.00, 0.00, 0.25, 0.07, 0.06, 0.17, 0.13, 0.00, 0.52, 1.00, 0.31, 0.64, 0.74, 0.33, 0.49, 0.13, 0.07, 0.21, 0.15, 0.20, 0.29, 0.69, and 0.20 for the 0- to 1-foot, 1- to 2-foot, ..., 23- to 24-foot buckets, respectively.

⁷ The slope data is not exactly normally distributed within each distance bucket. Therefore, we also performed the non-parametric Kruskal-Wallis rank sum test on the slope data. The results of this test were similar to those for the F-test. With the rank-sum test, for most distance buckets beyond four feet the null hypothesis that "All players have the same slope" failed at the 5% level, except for the 8- to 9-foot bucket and 10- to 11-foot bucket.

level green is considerably easier than a 10-foot putt on a fast, highly-contoured green. Rather than consider the specific attributes of the green in our model, we adjust the individual parameters of our two statistical models on a hole-by-hole basis. An advantage of this approach is that it does not require a description of the green's physical layout. However, it creates a new challenge to be addressed: to separate green difficulty from quality of field. As we have mentioned before, each tournament is played by a different set of PGA TOUR professionals. Thus, if we only adjust the model parameters for each hole, it is unclear whether the adjustments capture the difficulty of the green or the putting ability of the golfers who played the corresponding tournament. To develop a true reading of the hole-specific parameter adjustments, we must simultaneously fit both hole-specific and player-specific parameter adjustments. We do this by adding hole-specific and player-specific dummy variables to the model. Equation 5 below lists the hole and player-adjusted logistic regression function with a logit link for the probability of making a putt, $P(Y = 1)$, by player p on hole h from distance d , where n_p and n_h represent the number of putts that match the corresponding player or hole in the training data. Equation 6 below lists the hole- and player-adjusted gamma regression function with a log link for mean distance-to-go, μ_d , conditioned on a miss by player p on hole h from distance d , where n_p^{γ} and n_h^{γ} represent the number of missed putts that match the corresponding player or hole in the training data.⁸ The indicator variables $I_{\{condition\}}$ equal 1 if the condition is true and 0 otherwise. Thus, a player receives an adjustment to a model parameter only if that player has a large number of observations. The more observations a player has, the more parameters will be adjusted. For a player with very few observations, the parameters will be identical to the general model parameters. Note that Equations 5 and 6 are similar to Equations 1 and 2, with additional terms added corresponding to player skill and hole difficulty. Each of these equations ignores the impact of pin placement on the hole under the assumption that variations in difficulty due to pin placement are averaged out over the course of multiple rounds' and years' worth of play.

$$\log\left(\frac{P(Y = 1)}{P(Y > 1)}\right) = \beta_0 + \beta_1 d + \beta_2 d^2 + \beta_3 d^3 + \beta_4 d^4 + \beta_5 \log d + I_{\{n_p \geq 2000\}} \beta_{0p} + I_{\{n_p \geq 8000\}} \beta_{1p} \log d + I_{\{n_h \geq 750\}} \beta_{0h} \quad (5)$$

⁸In adding player and hole specific adjustment terms, we assume that player performance and hole difficulty are relatively stable for the 6-year period spanned by our data.

$$\log(\mu_d) = \beta_0 + \beta_1 d + \beta_2 d^2 + \beta_3 \log d + I_{\{n_p' \geq 50\}} \beta_{0p} + I_{\{n_p' \geq 1000\}} \beta_{1p} d + I_{\{n_h' \geq 300\}} \beta_{0h} + I_{\{n_h' \geq 1000\}} \beta_{1h} d \quad (6)$$

We determine the appropriate parameters to adjust for each hole and player by using our 15% holdout sample to protect against over-fitting the data. More specifically, the problem that we face is determining for each hole or player which parameters from the base model described in Sections 2.1 and 2.2 should be adjusted. If we adjust all parameters, including the logarithm term and full polynomial, we are likely to over-fit the data, as this would be equivalent to training separate models on each subset of the data. If we adjust too few, we are leaving biases in the model that should be addressed. Thus, our desire is to find threshold observation counts above which the data supports the corresponding adjustment. For example, for player-specific adjustments to the logistic regression model, we adjust the intercept term for players who have taken more than 2,000 putts and the logarithm of distance term for players who have taken more than 8,000 putts. Note that traditional statistical tests, such as a Chi-squared test, can be used to determine whether individual parameter estimates are significant. However, these tests are sensitive to outliers. A player with 1 observation who makes an extraordinary 60-foot putt would receive his own adjustment parameter, whereas a player who has thousands of observations and is slightly (but surely) better than his peers could receive no adjustment.

Although there are other equally valid approaches, we choose to consider this problem in a staged fashion to limit the computational burden. That is, we determine the observation count threshold for adjusting the intercept parameter before determining the threshold for the first-order distance parameter, and we calculate the thresholds for the hole-specific adjustments prior to the player-specific adjustments. Thus, the first adjustment we consider is hole-specific adjustment to the intercept parameter for both statistical models. Because our two statistical models are trained independently, there is no need to require any ordering between them, and, because each adjustment introduces more complexity, we require the thresholds to be monotonically increasing. Thus, the threshold for adjusting the intercept parameter will be lower than the threshold for adjusting the first-order distance parameter.

As mentioned above, in order to determine the appropriate threshold for each parameter we utilize the out-of-sample deviance on the withheld 15% test sample. Ideally, we would calculate the out-of-sample deviance for separate models trained using each possible threshold value, but this is not feasible for computational reasons because each model takes between hours and days to train

on the full data set, and the possible threshold values range from 50 to 10,000. To address this issue, we complete the following steps.

1. We train the model *without* the parameter adjustment and calculate the out-of-sample deviance for each hole or player in the data set, sorted by the number of matching observations, descending.
2. We train the model *with* the parameter adjustment, and a threshold value equal to the threshold of the previous parameter plus 50 (or 50 in the case of the first parameter). As in the previous step, we then calculate the out-of-sample deviance for each hole or player in the data set and sort the results according to the number of matching observations descending.
3. We approximate the total out-of-sample deviance for each possible threshold value by using the out-of-sample deviance from the second model for each hole or player above the threshold and the out-of-sample deviance from the first model for each hole or player below the threshold.

With this procedure, calculating the approximate out-of-sample deviance for each possible threshold value requires the model to be trained only twice. Plotting this approximate out-of-sample deviance against each possible threshold value provides a roughly convex curve for all the parameters we evaluated. For each parameter, we choose a threshold value for our final model that approximately minimizes the out-of-sample deviance in this plot. There is still a significant amount of noise in the results. Thus, we do not focus on choosing the true minimum, but rather on choosing a subjectively reasonable threshold that is close to this minimum.

A sample of these plots is provided in Figure 15 and Figure 16. The first figure compares the estimated deviance plots that correspond to the hole first-order distance and player intercept parameter adjustments for the gamma regression. In the hole first-order distance plot, the deviance decreases slightly before flattening out and then starting to increase at around 1,200 observations. In this case, we choose to set the observation count threshold to 1,000. For the player intercept plot, the deviance steadily increases as the threshold value increases, thus we keep the observation count threshold at 50. The second figure compares the estimated deviance plots that correspond to the hole log-distance and player log-distance and parameter adjustments for the logistic regression. In the hole log-distance plot, the deviance steadily decreases as the threshold value increases. Thus, we do not adjust the corresponding parameter. For the player log-distance plot, the deviance flattens out around 8,000 observations before

beginning to increase again. In this case, we choose to set the observation count threshold to 8,000.

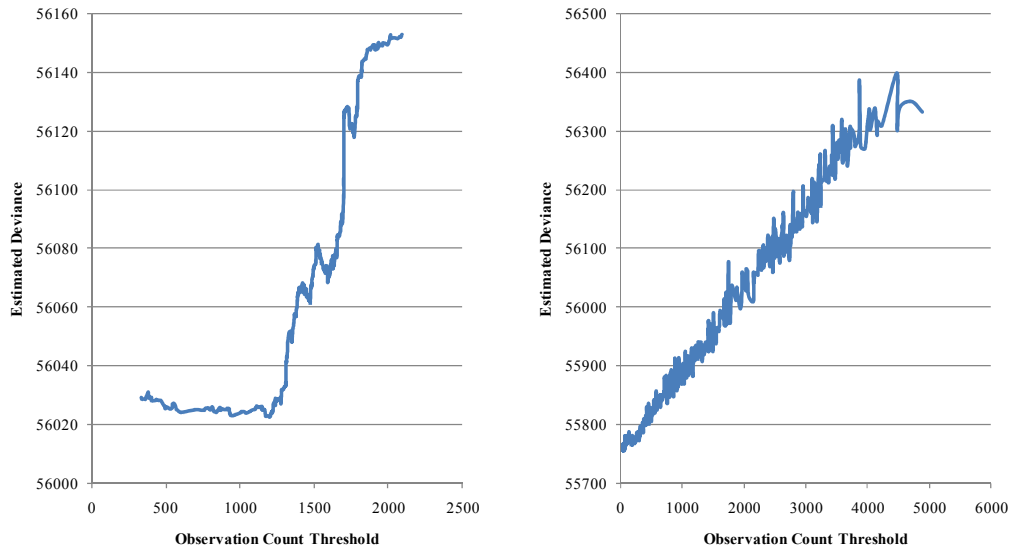


Figure 15: Estimated out-of-sample deviance plots for the hole first-order distance adjustment (left) and player intercept adjustment (right) for the gamma regression model.

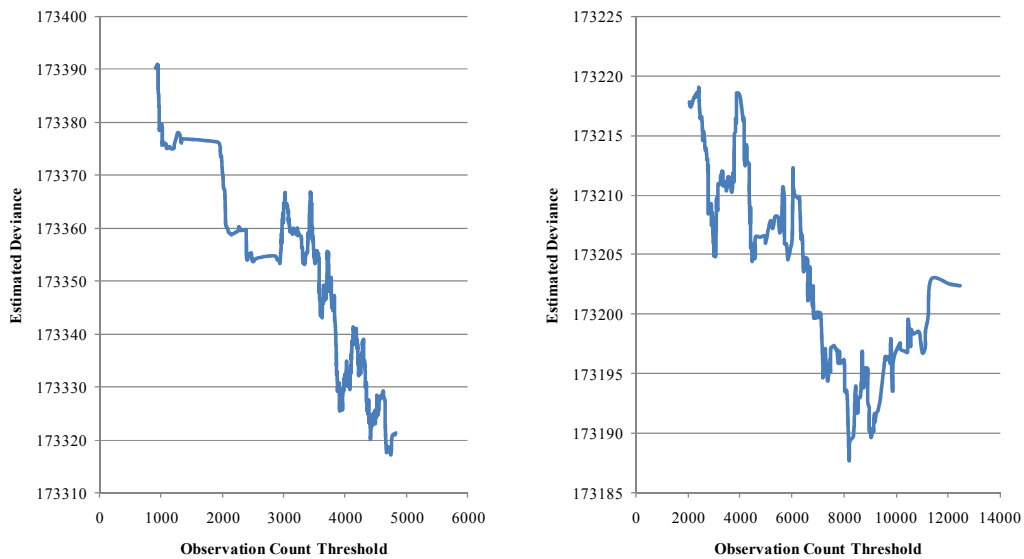


Figure 16: Estimated out-of-sample deviance plots for the hole log-distance adjustment (left) and player log-distance adjustment (right) for the logistic regression model.

In Table 4 and Table 5, we provide the thresholds for adjusting each of the listed parameters along with the number of matching observations. Parameter adjustments that are not supported by the data are not listed in these tables. In the logistic and gamma regression models, we adjust the hole intercept term for all 540 holes, for which the threshold listed corresponds to the minimum number of observations (rounded down to the nearest 50). Note that for the gamma regression model we adjust the first-order distance parameter for the player, whereas for the logistic regression model we adjust the parameter for the log of distance term. For the player-specific adjustments, we tested both the log-distance parameter and first-order distance parameter for each model.

Logistic Regression	Observation Count Threshold	Number of Parameter Adjustments	Number of Matching Samples
Hole Intercept	750	540	1,644,998
Player Intercept	2,000	272	1,410,646
Player Logarithm	8,000	28	248,738

Table 4: Hole and player-specific parameter adjustment thresholds for holding out logistic regression model.

Gamma Regression	Observation Count Threshold	Number of Parameter Adjustments	Number of Matching Samples
Hole Intercept	300	540	640,525
Hole First-Order	1,000	439	573,130
Player Intercept	50	566	629,519
Player First-Order	1,000	237	517,088

Table 5: Hole and player-specific parameter adjustment thresholds for distance-to-go gamma regression model.

At the beginning of this section, we mentioned that the purpose of adding the player-specific adjustments is to tease out the quality of field effects from the hole-specific adjustments. In order to complete this task, we need to normalize the player-specific parameter adjustments so that 0 represents an average quality of field across all of our tournaments. We do this by forcing the sum of the player-specific parameter adjustments weighted by the corresponding observation

counts to equal 0. For instance, for the player-specific intercept adjustment of either model, we first calculate the weighted average of the estimated parameters: $w_{\hat{\beta}_0} = \sum_p n_p \hat{\beta}_{0p} / N$, where n_i is the number of observations for player p and N is the total number of observations. Next, we subtract this value from each of the player-specific intercept terms, $\hat{\beta}_{0p} = \hat{\beta}_{0p} - w_{\hat{\beta}_0}$, and add it to the base intercept term, $\hat{\beta}_0 = \hat{\beta}_0 + w_{\hat{\beta}_0}$. Note that this normalization does not change the result for either the logistic or gamma regression functions, because we have added and subtracted the same value. Thus, the estimated likelihood of our original model is maintained. We follow the same approach for the player-specific adjustments to the first-order and logarithm parameters. For ease of interpretation, we also normalize the hole-specific adjustments so that 0 represents a green of average difficulty. All parameter estimates reported subsequently are based on this normalization scheme.

One nice feature of the gamma regression adjustments with the log link function is that they are more easily interpretable. For instance, in the case of the player-specific parameter adjustments, we can rewrite the mean distance-to-go from Equation 2 as follows: $\mu_d = e^{(\beta_{0p} + \beta_{1p}d)} e^{\{\beta_0 + \beta_1d + \beta_2d^2 + \beta_3 \log d\}}$, where the first term represents a player-specific scaling factor relative to the overall mean *distance-to-go*. For example, Tiger Woods has an estimated intercept adjustment of $\beta_{0p} = 0.3878$ and first-order distance adjustment of $\beta_{1p} = -0.0044$. This means that when Tiger misses a putt from 20 feet, we would expect him to end up 5% closer to the hole than the average across the field (i.e. $1 - e^{(0.3878 - 20 \cdot 0.0044)} = 0.05$).

In Table 6 and Table 7, we provide sample parameter adjustments for two players, John Huston and Brent Geiberger, and two holes, the first hole at Sawgrass and the ninth hole at Bay Hill.⁹ In order to visualize how these adjustments affect our models, in Figure 17 and Figure 18 we plot the adjusted curves against the normalized field averages. Notice that Brent Geiberger is a better putter overall, as his logistic regression curve lies below the mean and his gamma regression curve lies above it, leading to a smaller than average *putts-to-go*.

⁹ Note that the coefficients in these tables are based on the field-adjusted model, which was normalized as described earlier in this section.

Logistic Regression	(Intercept)	Logarithm distance	First-order distance	Second-order distance	Third-order distance	Fourth-order distance
Baseline for normalized model	7.32	-5.573	0.67	-0.019	0.0003	-1.6E-06
<i>(Adjustments to Baseline)</i>						
John Huston	-0.11	0.005				
Brent Geiberger	0.11	0.005				
TPC Sawgrass Hole 1	-0.10					
Bay Hill Hole 9	0.19					

Table 6: Coefficient adjustments for the logistic regression holing out model.

Gamma Regression	(Intercept)	Logarithm distance	First-order distance	Second-order distance
Baseline for normalized model	0.93	-0.342	0.04548	-0.0002
<i>(Adjustments to Baseline)</i>				
John Huston	0.10		-0.00149	
Brent Geiberger	-0.11		0.00085	
TPC Sawgrass Hole 1	0.11		-0.00005	
Bay Hill Hole 9	-0.15		-0.00005	

Table 7: Coefficient adjustments for the gamma regression distance-to-go model.

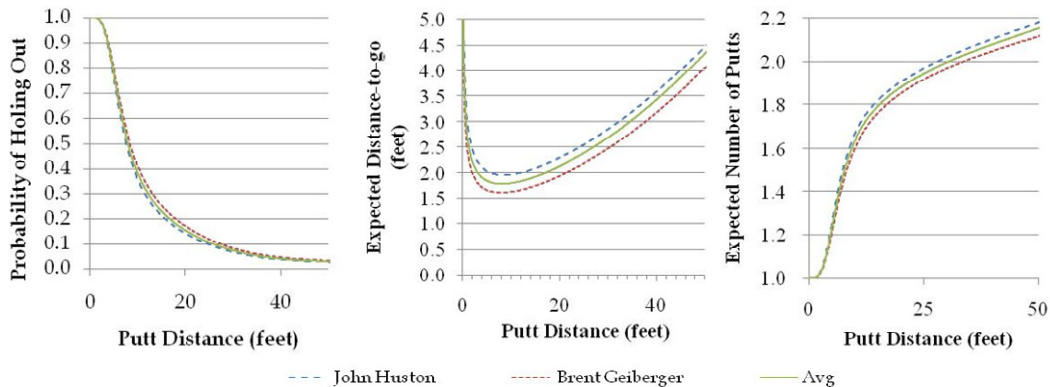


Figure 17: Comparison of two player adjustment effects for the fitted logistic regression (left), gamma regression (center), and Markov (right) models. These adjustments assume each player is facing an average hole.

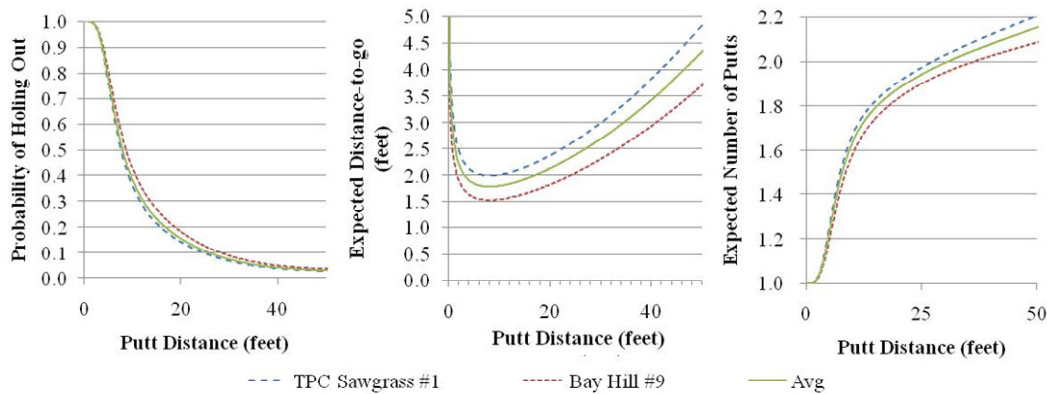


Figure 18: Comparison of two hole adjustment effects for the fitted logistic regression (left), gamma regression (center), and Markov (right) models. These adjustments assume an average golfer is facing each hole.

2.6 An Iterative Training Approach

After adding the hole-specific and player-specific dummy terms and interaction terms described in the previous section, training the model becomes a significant computational challenge due to the number of data columns for each observation. All of our models are trained and evaluated in the statistical computing language R.¹⁰ In R, training the final model directly on 30 courses exceeds the 8 GB of RAM in our computational environment because each observation requires 2,632 columns of data due to the adjustment terms. Ideally, we would like to be able to train the model for all tournaments available within the ShotLink data set. To address this, we have utilized an iterative approach to training our model.

Algorithms that fit generalized linear models typically optimize all coefficients simultaneously in an iterative fashion until the relative improvement in the model deviance is below some threshold (e.g., 0.000001%). Instead of optimizing all 2,632 coefficients simultaneously, we train the model on a small subset of coefficients at a time. Thus, in a given overall iteration, our algorithm performs multiple sub-iterations optimizing over disjoint subsets of the 2,632 coefficients. In each of these sub-iterations, we use the current estimates for the coefficients outside of the subset to calculate an offset that is subtracted from each observation’s response variable. Then we fit the generalized linear model using the adjusted response variable and the subset of coefficients of interest. The algorithm then moves on to the next subset of coefficients, and the next, until

¹⁰ R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

every coefficient has been updated. This completes a single overall iteration. As with the traditional approach, our algorithm iterates until the relative improvement in model deviance is below some threshold (0.000001%), at which point it terminates.

Using this approach, we are able to train the model on 30 courses' worth of data in less than two days per model. The performance improvement is due to the fact that, by optimizing across a subset of the coefficients at a time, we need to consider only the subset of the data that corresponds to those coefficients. Note that this procedure does not affect the optimality guarantees, or lack thereof, for the original problem because the procedure is equivalent to a nonlinear, directional, local search. That is, if the original likelihood function is concave in the region of interest, either procedure will converge to the maximum likelihood. The other interesting feature of this approach is that approximate convergence occurs extremely quickly. Table 8 and Table 9 list the resulting model deviance for the first five iterations of this procedure as well as the final model deviance for the logistic and gamma regression respectively. We chose to report on five iterations, because five training iterations can be completed in under an hour for each model. The improvement in deviance for the logistic regression from the 5th to 27th iteration is just 0.002%, whereas the improvement in deviance for the gamma regression from the 5th to 178th iteration is just 0.014%. These results suggest that in a practical setting, the model could be incrementally updated as new data are available by applying just a few iterations of this procedure.

Logit Iteration	Cumulative Duration	Logit Deviance
0		988,732
1	0:09:58	985,577
2	0:18:46	985,557
3	0:27:51	985,546
4	0:36:43	985,537
5	0:45:31	985,530
27	4:08:45	985,511

Table 8: Convergence data for iteratively training the logistic regression model.

Gamma Iteration	Cumulative Duration	Gamma Deviance
0		323,376
1	0:10:30	314,471
2	0:20:52	314,433
3	0:31:13	314,427
4	0:41:33	314,423
5	0:51:46	314,419
178	30:13:49	314,376

Table 9: Convergence data for iteratively training the gamma regression model.

3 Results

In Section 3.1, we use the field-adjusted putts-to-go model described above to evaluate putting performance. We use the concept of *shot value* as described in Section 1.2 to evaluate the stroke gain associated with each putt taken. We aggregate these values for each player to calculate and compare each player's *putts gained per putt* and *per round*. In Section 3.2, we use *putts gained per round* to determine each player's off-green performance. To do this, we calculate the field-adjusted scoring average for each hole using an OLS regression model. This allows us to determine the total strokes gained, which can be described as the sum of the putts gained and the off-green strokes gained.

In Section 3.3, we use the player-adjusted *putts-to-go* model to evaluate situational putting performance. First, we evaluate the difference in putting performance on putts for birdie relative to putts for par. We show that although there is a statistically significant difference between the two, the magnitude of this difference in terms of a player's overall score is much smaller than has been previously suggested. Lastly, we evaluate putting performance on the final round of play in the aggregate and for Tiger Woods to test for the impact of psychological pressures.

3.1 Putts Gained

In Section 1.2, we defined the concept of *shot value*, the strokes gained or lost on a particular shot (i.e., $shot\ value = strokes\ to\ go_{Before} - strokes\ to\ go_{After} - 1$). Using the field normalized, hole-specific expected *putts-to-go* function as a baseline, *shot value* (or *putt value*) allows us to calculate the quality of each individual putt of each player. Averaging this *shot value* across all putts of a particular golfer, we develop the metric *putts gained per putt*:

$$putts-gained-per-putt_p = \frac{1}{TotalPutts_p} \sum_{i=1}^{TotalPutts_p} putt-value_{ip},$$

where p represents the player in question and $putt\ value_{ip}$ is the *shot value* of player p 's i^{th} putt. This *putts gained per putt* metric is unbiased and corrected for hole difficulty.

The metric *putts gained per round* is similar to the above, but instead of being normalized by the total number of putts taken the metric is normalized by the number of rounds played:

$$putts-gained-per-round_p = \frac{1}{TotalRounds_p} \sum_{i=1}^{TotalPutts_p} putt-value_{ip}$$

where, again, p denotes the specific player. Note that this is equivalent to multiplying *putts gained per putt* by the average number of putts the player takes in each round. Although there is no guarantee that ordering players by these two different metrics will provide identical rankings (because players’ putts per round averages may differ), in practice, the rankings are very close. In fact, the top 19 putters ranked by each method are identical, and the maximum difference in rank between the two methods is 5 golfers, or 1.5 percentiles. We use *putts gained per round* to evaluate players going forward over *putts gained per putt* because the former is more intuitive and its impact on a tournament’s outcome is clearer.

As desired, *putts gained per round* is a less biased way of judging putting skill than Putting Average (PA). Recall from the introduction that Putting Average is the average number of putts a golfer takes on greens reached in regulation (i.e., reached in no more than $(par - 2)$ strokes). This metric is also sometimes known as Putts Per Green in Regulation (PPGIR), and it is the most popular metric by which golfers’ putting skills are judged. However, Putting Average can be biased by the quality of golfers’ approach shots. Golfers whose first putts are closer to the hole will have lower Putting Averages, even if they are not better putters. Alternatively, great putters may have terrible Putting Averages because they start farther from the hole, on average, but actually make amazing putts from these longer distances. The metric *putts gained per round* corrects for this bias by explicitly correcting for the distance of each putt. In addition, *putts gained per round* corrects for green difficulty, so that sinking a 10-foot putt on a difficult green is valued more than doing the same on an easy green.

The advantage of *putts gained per round* over Putting Average can be demonstrated through a simple example. Consider two golfers, Ernie Els and Stephen Leaney. Ernie Els is ranked 15th in Putting Average (5th percentile), but is ranked 283rd in *putts gained per round* (88th percentile). On the other hand, Stephen Leaney is ranked 191st in Putting Average (59th percentile), but 28th in *putts gained per round* (9th percentile). Table 10 shows the putting statistics and rankings of each player.

Player	Number of Rounds Played	Putting Average Percentile	Putts Gained per Round Percentile	Putting Average	Putts Gained per Round
Stephen Leaney	264	59%	9%	1.79	0.26
Ernie Els	212	5%	88%	1.75	-0.63

Table 10: Comparison of Ernie Els’ and Stephen Leaney’s performance using two putting performance metrics.

As discussed earlier, an hypothesis for the discrepancy between these two golfers' rankings between the Putting Average and *putts gained per round* metrics is that Ernie Els starts closer to the hole, on average, meaning that his putts are easier. This would inflate his Putting Average rank. By this same hypothesis, we would expect Stephen Leaney's average first putt distance to be farther than the average golfer's. In fact, this is the case. Figure 19 shows the cumulative distribution function for first putt distance for Ernie Els, Stephen Leaney, and the average for the field of golfers. These distributions and averages consider only those first putt distances for which the golfer reached the green in regulation, since Putting Average considers only these putts. Ernie Els' first putt on average is 20.0 feet, almost two feet closer to the hole than the average first putt distance across the field, which is 21.8 feet. On the other hand, Stephen Leaney's average first putt distance is 23.0 feet, more than a foot farther from the hole than the average.

Although at first glance, Ernie Els may appear to be a great putter by the Putting Average metric, the data seem to indicate that his true skill is on the approach. Actually, he is a worse-than-average putter with respect to *putts gained per round*. After correcting for distance and green difficulty, we find that 88% of his peers putt better than he does.

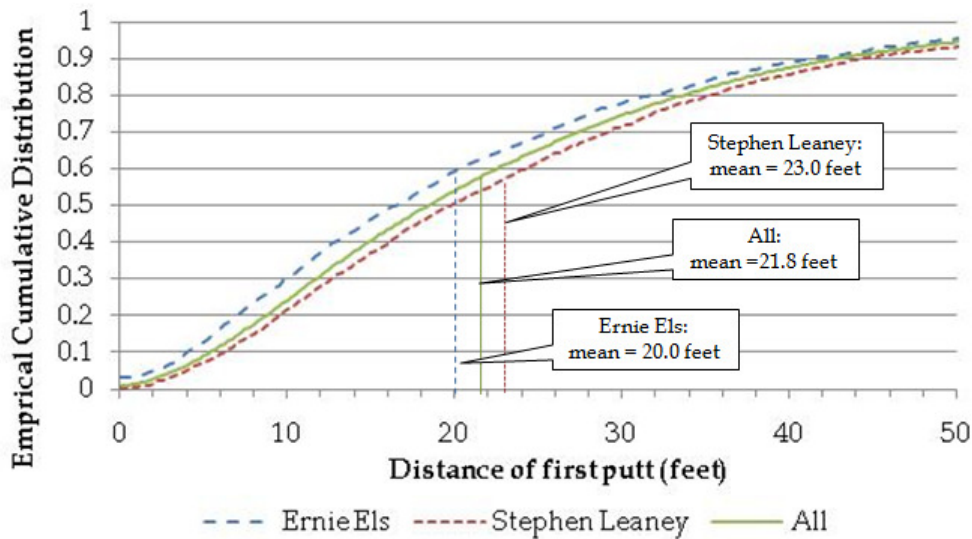


Figure 19: Empirical cumulative distribution of Ernie Els' and Stephen Leaney's first putt distances for greens reached in regulation.

An additional benefit of *putts gained per round* is that it utilizes all available putting data. Putting Average includes only those putts for greens reached in regulation, excluding putts made by golfers who took at least an extra stroke to get to the green. Because it excludes putts on greens not reached in regulation, approximately 30% of all putts taken, and the corresponding information these putts provide, Putting Average has a higher intrinsic error or variability.

Table 11 and Table 12 show the top twenty putters ranked by each metric. The first table ranks the top 20 golfers by Putting Average. The second table ranks the top twenty players with respect to *putts gained per round*. Notice the large discrepancy between the rankings. Aside from Tiger Woods, who is first in both rankings, no other golfer has the same ranking using either metric, and many golfers' ranks differ by a large amount.

Player	Number of Rounds Played	Putting Average Rank	Putts Gained per Round Rank	Putting Average	Putts Gained per Round	Putts Gained per Round Standard Deviation
Tiger Woods	230	1	1	1.71	0.69	0.12
Greg Chalmers	136	2	31	1.73	0.25	0.15
Aaron Baddeley	303	3	5	1.74	0.53	0.10
Fredrik Jacobson	248	4	3	1.74	0.56	0.11
Nathan Green	197	5	4	1.74	0.55	0.12
Scott Verplank	316	6	76	1.74	0.06	0.10
Tim Clark	296	7	54	1.74	0.15	0.10
Vijay Singh	387	8	218	1.75	-0.36	0.09
Aaron Barber	79	9	300	1.75	-0.83	0.20
Vaughn Taylor	301	10	57	1.75	0.14	0.10
Padraig Harrington	134	11	45	1.75	0.17	0.15
Stewart Cink	375	12	7	1.75	0.49	0.09
Bob Heintz	111	13	60	1.75	0.13	0.17
David Toms	268	14	50	1.75	0.16	0.11
Ernie Els	212	15	283	1.75	-0.63	0.12
Brian Gay	388	16	23	1.75	0.28	0.09
Ben Crane	273	17	9	1.75	0.44	0.11
Brenden Pappas	206	18	103	1.75	0.00	0.12
Darren Clarke	107	19	8	1.75	0.45	0.17
Jim Furyk	310	20	101	1.75	0.00	0.10

Table 11: Top 20 putters ordered by the PGA TOUR's Putting Average statistic for golfers with at least 50 rounds of play.

Player	Number of Rounds Played	Putting Average Rank	Putts Gained per Round Rank	Putting Average	Putts Gained per Round	Putts Gained per Round Standard Deviation
Tiger Woods	230	1	1	1.71	0.69	0.12
David Frost	113	60	2	1.77	0.67	0.16
Fredrik Jacobson	248	4	3	1.74	0.56	0.11
Nathan Green	197	5	4	1.74	0.55	0.12
Aaron Baddeley	303	3	5	1.74	0.53	0.10
Jesper Parnevik	315	47	6	1.76	0.50	0.10
Stewart Cink	375	12	7	1.75	0.49	0.09
Darren Clarke	107	19	8	1.75	0.45	0.17
Ben Crane	273	17	9	1.75	0.44	0.11
Willie Wood	72	92	10	1.77	0.42	0.20
Brad Faxon	246	30	11	1.76	0.41	0.11
Dean Wilson	356	58	12	1.77	0.39	0.09
Jay Delsing	132	52	13	1.77	0.38	0.15
Jay Haas	149	22	14	1.75	0.38	0.15
Steve Stricker	237	26	15	1.75	0.37	0.11
Scott McCarron	197	62	16	1.77	0.37	0.13
Zach Johnson	313	41	17	1.76	0.37	0.10
Daniel Chopra	312	25	18	1.75	0.36	0.10
Stephen Ames	304	40	19	1.76	0.36	0.10
Bob Tway	309	54	20	1.77	0.31	0.10

Table 12: Top 20 putters ordered by *putts gained per round* for golfers with at least 50 rounds of play.

Using the *putts gained per round*, we are also able to determine the statistical significance between any two golfers’ putting skill. For instance, Tiger Woods’ performance is two standard deviations above every putter not in the top 10, and 9 standard deviations above Vijay Singh’s *putts gained per round*.

We assert that *putts gained per round* is a more accurate metric than Putting Average. Due to its lack of bias, intuitive interpretation, and ease of calculation, we suggest that putting skill should be evaluated with respect to *putts gained per round* rather than Putting Average.

We can also use *putts gained per round* to evaluate putting performance on a tournament-by-tournament basis. For example, Table 13 lists the top 10 tournament performances with respect to *putts gained per round*. As above, we restrict our analysis to only golfers with at least 50 rounds of play, and we include only golfers who played all rounds of the tournament (i.e., made the cut).

Rank	Player	Course Name	Year	Putts Gained per Round
1	Hal Sutton	Harbour Town GL	2003	3.97
2	Ben Crane	TPC Sugarloaf	2003	3.39
3	Aaron Baddeley	Bay Hill Club	2005	3.27
4	David Peoples	TPC Southwind	2003	3.14
5	Marc Tumesa	Annandale GC	2008	3.09
6	Jay Haas	TPC Sawgrass	2003	3.09
7	Fredrik Jacobson	Muirfield Village GC	2007	3.05
8	Scott McCarron	TPC Scottsdale	2003	2.99
9	Dean Wilson	Quail Hollow CC	2007	2.99
10	Geoff Ogilvy	TPC Sawgrass	2003	2.98

Table 13: Top 10 tournament level performances with respect to *putts gained per round*.

In 2003, Hal Sutton finished third at the MCI Heritage played at the Harbour Town Golf Links, but his putting performance over the four rounds of play was extraordinary. In the first round, Sutton gained just over 5.1 putts on the field. In the second round, he gained 1.6 putts on the field, but followed that up by picking up 3.4 putts in round 3. In the last round, Sutton gained almost 5.9 putts on the field, exceeding his already spectacular mark in round 1. As a point of reference, in our 2003–2008 data set, fewer than 1 out of every 400 rounds played resulted in a golfer gaining more than 5.0 putts in a round. Hal Sutton accomplished this feat twice in a single tournament, with Fredrik Jacobson being the only other golfer to do so (in the seventh-ranked tournament level putting performance). Over the course of the tournament, Sutton gained at least half a putt on 22 different holes, including making one putt from 53 feet (gaining 1.22 strokes on the field) and two putts from just over 24 feet (gaining 0.94 and 0.92 strokes). As these results suggest, analyzing putts gained can provide an interesting backdrop for putting performance on a tournament-by-tournament, or even round-by-round basis.

3.2 Off-Green Performance

Once we determine each player’s total strokes gained, we can use our model for putts gained to calculate how many strokes are gained off the green (i.e., total strokes gained minus putts gained). The challenge is that for consistency, the total strokes gained should be calculated relative to the same average field of PGA TOUR professionals as putts gained. Strictly averaging the number of

strokes taken for each hole is insufficient because it would incorporate biases based on each tournament's quality of field.

To address this issue, we perform an OLS regression to predict the number of strokes taken for each player's hole of play. In this model, the independent variables we include are an intercept term, the distance from the tee to the pin, a dummy variable for each player, and a dummy variable for the interaction of each hole with the day of play. We utilize the interaction of the hole with the day of play in order to pick up day-specific effects such as weather. We include the distance from tee to pin to provide stability for tournaments with fewer daily observations. The full linear regression function is described in Equation 7:

$$E[Y] = \beta_0 + \beta_1 d + \beta_{0p} + \beta_{0hd} \quad (7)$$

where Y represents the strokes-to-go from the tee for player p on hole and day combination hd . After normalizing the model using the same technique as described in Section 2.5, we estimate the expected scoring average for each hole and day of play based on an average quality of field. By using the same adjustment and normalization approach, we ensure that this corresponds to the average quality of field utilized by the two putting regression models. Lastly, we calculate the strokes gained for each hole of play by subtracting the expected scoring average from the player's hole score. As with putts gained, we choose to report off-green strokes gained using a per round average because the top 20 golfers remain the same regardless of whether we normalize by the number of strokes taken or the number of rounds. As with putts gained, we recommend *off-green strokes gained per round* due to its value as a more easily interpretable metric.

In Table 14, we provide the breakdown of *putts gained per round* and *off-green strokes gained per round* for the top 20 golfers as measured by total strokes gained per round. The most interesting thing to note regarding these results is that there is significant variation in how the best golfers pick up strokes on the field. For instance, it is interesting to compare Tiger Woods and Vijay Singh, two elite PGA TOUR golfers, over the 6-year period. Tiger excels due to his all-around performance, whereas Vijay is the top player off the green, but his putting performance is well below average. Another observation is that all the top twenty golfers are better than average off-green performers, but roughly a third are worse than average putters. This suggests that bad putting can be salvaged by great performances off the green, but not the other way around.

Player	Number of Rounds Played	Strokes Gained per Round Rank	Strokes Gained per Round	Putts Gained per Round	Off-Green Strokes Gained per Round
Tiger Woods	230	1	3.22	0.69	2.53
Vijay Singh	387	2	2.29	-0.36	2.65
Jim Furyk	310	3	2.03	0.00	2.03
Phil Mickelson	243	4	1.94	0.19	1.74
Ernie Els	212	5	1.85	-0.63	2.48
Adam Scott	220	6	1.77	0.08	1.69
Sergio Garcia	259	7	1.52	-0.67	2.20
David Toms	268	8	1.43	0.16	1.27
Retief Goosen	214	9	1.40	-0.44	1.84
Stewart Cink	375	10	1.39	0.49	0.89
Padraig Harrington	134	11	1.37	0.17	1.19
Kenny Perry	358	12	1.36	-0.28	1.65
Darren Clarke	107	13	1.35	0.45	0.90
Robert Allenby	303	14	1.32	-0.34	1.66
Scott Verplank	316	15	1.31	0.06	1.24
Jay Haas	149	16	1.27	0.38	0.89
Luke Donald	255	17	1.23	-0.06	1.29
Stephen Ames	304	18	1.22	0.36	0.86
Anthony Kim	117	19	1.22	0.01	1.21
Zach Johnson	313	20	1.14	0.37	0.78

Table 14: Top 20 golfers as measured by total strokes gained per round for golfers with at least 50 rounds of play.

Extending the previous section’s discussion of Hal Sutton’s putting performance in the 2003 MCI Heritage, we can now look at his off-green performance in the same tournament. Over the course of the four rounds of the tournament, Sutton lost 6.1 strokes to the field off the green compared to the 15.9 strokes he gained on the green. Sutton finished the tournament just 1 stroke behind the leaders Woody Austin and Davis Love III, with Love III winning the subsequent playoff. Assuming a field-average performance of 0 strokes gained off the green, Hal Sutton would have won the tournament by 5 strokes. The differences between Sutton’s 6.1 strokes lost off the green and the 6.2 and 5.4 off-green strokes gained by Austin and Love III, respectively, cancel out most of Sutton’s putting gains. Even the best putting performance can still result in a tournament loss.

3.3 Situational Putting Performance

An additional benefit of our *putts-to-go* model is that it allows us to evaluate situational putting performance by comparing the putts gained in each situation relative to player-specific baselines. That is, by building our Markov model using player-specific transition probabilities based on the models with player adjustment terms, we are able to determine player-specific *putts-to-go* expectations for each putt taken. For instance, from nine feet on the tenth hole of Quail Hollow, Tiger Wood’s expected *putts-to-go* is 1.54 and Vijay Singh’s expected *putts-to-go* 1.59. This baseline explicitly controls for the difficulty of the green *and* the quality of the putter, unlike the expected *putts-to-go* for the field described in Section 3.1. Next, by aggregating the putts gained relative to these player-specific baselines, we can calculate player-specific *putts gained per putt* for each situation.

We illustrate this approach by analyzing two situational putting breakdowns: putting performance on putts for birdie as compared to putts for par, and putting performance on the fourth round of play. The benefit of our approach is that once player-specific putts gained have been calculated for each putt taken, new situational analyses require only aggregating these values. Thus, our approach makes it easy to test new hypotheses without having to train additional models. Additionally, this approach allows us to calculate the impact of situational effects on the golfer’s total score based on the difference between the *putts gained per putt*, using the standard deviations to determine the significance.

Pope and Schweitzer (2010) have demonstrated that there is a significant difference in putting performance between par and birdie putts. The conclusion of the paper is that the difference in behavior between birdie and par putts costs golfers on the order of 1 putt per tournament. In Table 15, we estimate the *putts gained per putt* on putts for birdie as well as putts for par, resulting in a difference of 0.014 *putts gained per putt*. On average, golfers make 11.3 putts for birdie per round. Thus, this difference corresponds to a loss of 0.64 strokes per tournament, which is similar to the magnitude estimated by Pope and Schweitzer.

	Putt Count	Putts Gained per Putt	Putts Gained per Putt Deviation
For Birdie	756,643	-0.00741	0.00023
For Par	914,652	0.00673	0.00014
Difference		0.01414	0.00027

Table 15: Comparison of *putts gained per putt* on par and birdie putts.

An alternative hypothesis also investigated is that this difference is instead due to the difference between first putts and following putts. This line of reasoning suggests that first putts are more difficult because the golfer has no information about the break of the green and less control over the secondary characteristics of the putt. For instance, if a 10-foot first putt is the result of a 100-yard approach shot, the golfer has less control over the exact location on the green. A 10-foot second putt, on the other hand, is likely to be from a more desirable location, for instance, downhill of the hole, because it is the result of a more controllable shot (i.e., the first putt). To test this hypothesis, we perform a similar analysis as above, this time splitting the data into first putts and subsequent putts. The results are listed in Table 16.

	Putt Count	Putts Gained per Putt	Putts Gained per Putt Deviation
First Putt	1,182,012	-0.00561	0.00018
Follow-on Putts	752,185	0.01216	0.00010
Difference		0.01777	0.00021

Table 16: Comparison of *putts gained per putt* on first putt and following putts.

Although there is a significant difference in *putts gained per putt* for both cases (Table 15 and Table 16), the larger difference in the second case suggests that this may be the more significant effect. To tease out these two correlated effects, we perform an OLS regression to predict putts gained with an intercept term as well as indicators for first putts and birdie putts as independent variables, restricting the data to only putts for par or birdie. Table 17 summarizes the fitted parameters of this regression. Based on these parameters, the estimated difference between birdie and par putts is 0.0054 putts gained per putt, or approximately 0.2425 strokes per tournament. This supports the conclusion that professional golfers behave differently with respect to putting for birdie or par, but suggests that the magnitude of the effect is much smaller than previously estimated.

In addition, the above results violate the Markov assumption underlying Equation 4, which we used to generate the *putts-to-go* baselines. To address this issue, we could extend the Markov chain state space to include the putt number

(e.g., first, second, third) and/or the putt score (e.g., putt for a birdie, par, bogey). We leave this to future research as a potential improvement over the current model.

	Estimate	Standard Error	t-value	p-value
(Intercept)	0.0130	0.0005	28.7350	< 2E-16
First Putt Indicator	-0.0154	0.0007	-22.3420	< 2E-16
Birdie Putt Indicator	-0.0054	0.0007	-8.1850	2.71E-16

Table 17: Fitted parameter of OLS regression to analyze impact of first putts and birdie putts.

Next, we consider putting performance on the fourth round of play: Sunday, the last day of a tournament. In particular, we are interested in determining whether putting performance suffers due to psychological pressure. To that end, in Table 18 we first compare player-specific putts gained per putt in the third round versus the fourth round. The results indicate that the difference is insignificant as compared to the deviations of either the third or fourth round estimates. Still, the lack of a significant difference does not exclude the possibility of psychological factors because many players in the fourth round are out of contention by the time the round begins.

	Putt Count	Putts Gained per Putt	Putts Gained per Putt Deviation
3rd Round	359,079	0.00237	0.00027
4th Round	353,979	0.00246	0.00027
Difference		0.00009	0.00038

Table 18: Comparison of putts gained per putt on putts in the 3rd round and 4th round of play.

Thus, we next restrict our focus to the players that are in contention. We say golfers are in contention if they end the third round of play within six strokes of the leader. This corresponds to approximately 10 golfers per tournament. Using this definition, in Table 19 we divide fourth round putting performance by comparing players in contention against those who are out of contention. Under this segmentation of the data, we find that players who are in contention tend to mildly underperform those who are out of contention, but the magnitude of this effect is smaller than the deviation of the in-contention putts gained per putt

estimate. Note that even if this difference was statistically more significant, it would correspond to a loss of just 0.07 strokes per tournament. Thus, it seems that, at least in aggregate, professional golfers are relatively immune to the pressures associated with the fourth round of play.

	Putt Count	Putts Gained per Putt	Putts Gained per Putt Deviation
In Contention	46,918	0.00194	0.00073
Out of Contention	307,061	0.00254	0.00029
Difference		0.00060	0.00079

Table 19: Comparison of putts gained per putt during the 4th round when the player is in or out of contention.

Along a similar vein, we can use this approach to evaluate the fourth round play of individual golfers. For instance, a common perception is that one of Tiger Wood’s strengths is his ability to close out the final round. At least with regards to putting, this is easy to test using our model. In Table 20, we break down Wood’s *putts gained per putt* by round of play. These results suggest that Wood’s best round of putting is in fact the second round, and that his fourth round performance is very close to his player-specific expectation, namely, 0.69 strokes better than the field.

	Putt Count	Putts Gained per Putt	Putts Gained per Putt Deviation
1st Round	1,614	0.00036	0.00386
2nd Round	1,589	0.00847	0.00395
3rd Round	1,654	-0.00293	0.00375
4th Round	1,671	-0.00022	0.00380

Table 20: Comparison of Tiger Wood's putts gained per putt for each round of play.

Note that the purpose of these analyses is not to make unassailable claims, but rather to illustrate the simplicity of performing these tests using the model we have developed. As should be apparent, there are many other questions of a

similar nature that would benefit from the application of our approach. At a minimum, this approach provides a method for quickly evaluating which questions would benefit from further investigation.

4 Conclusion

In this paper, we analyze putting performance on the PGA TOUR. Our analysis builds off of two generalized linear models that describe the process of putting: a logistic regression to predict the probability of making the putt, and a gamma regression to predict the distribution of distance-to-go when the putt misses. Each of these models fits the empirical data exceedingly well. Additionally, we develop a Markov chain modeling approach that utilizes these fitted regressions to derive an estimated distribution of *putts-to-go*. The success of this model is made possible by the ability of our gamma regression to accurately predict not only the mean, but also the distribution, of *distance-to-go*.

We adjust our models to control for green difficulty and player skill level simultaneously. A unified approach is necessary in order to separate green difficulty from each tournament's quality of field. Considering all of these adjustments at once vastly increases the computational burden associated with training the model. Thus, we develop an algorithm that allows these models to be trained with reduced memory requirements and improved performance.

These efforts allow us to develop an enhanced putting performance metric, *putts gained per round*, and demonstrate its benefits over Putting Average, the most commonly reported putting statistic. Due to its lack of bias, ease of interpretation, and computational tractability we believe that the PGA TOUR should add *putts gained per round* to its set of reported statistics. Once the expected *putts-to-go* plot has been derived (perhaps using complex mathematical machinery), calculating *putts gained per round* requires nothing more than the plot for each hole and the distance of each putt taken. We hope that improved performance metrics such as this will help the PGA TOUR develop a diehard statistical following much like other popular sports. Combined with *off-green strokes gained per round*, our analysis of putting performance suggests that there are multiple ways to succeed in professional golf. Unfortunately for the competition, Tiger Woods excels at all of them, making him very difficult to catch.

5 References

- Bansal, M. & M. Broadie. (2008). "A Simulation Model To Analyze the Impact of Hole Size on Putting in Golf." In S.J. Mason, R.R. Hill, L. Moench, and O. Rose (eds.), *Proceedings of the 2008 Winter Simulation Conference*. Winter Simulation Conference.
- Broadie, M. (2008). "Assessing Golfer Performance Using Golfmetrics." In D. Crews and R. Lutz (eds.), *Science and Golf V: Proceedings of the 2008 World Scientific Congress of Golf*, 253-262. Energy in Motion, Inc., Mesa, Arizona.
- Brown, J. (2007). "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars." Working paper.
- Carnahan, J.V. (2002). "Experimental Study of Effects of Distance, Slope, and Break on Putting Performance for Active Golfers." In Eric Thain (ed.), *Science and Golf IV: Proceedings of the World Scientific Congress on Golf*, 113-126. Routledge, London.
- Cornish, G. & R. Whitten. (1993). *The Architects of Golf*. Harper Collins, New York.
- Cochran, A. & J. Stobbs (1968). *The Search for the Perfect Swing*. J.B. Lippincott Company, Philadelphia.
- Connolly, R.A. & R.J. Rendleman Jr. (2008). "Skill, Luck, and Streaky Play on the PGA TOUR." *Journal of the American Statistical Association*, March 2008, Vol. 103 No 481, 74-88.
- Hastie, T., R. Tibsharani, & J. Friedman. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.
- Hoadley, B. (1994). "How to Improve Your Putting Score without Improving." In A.J. Cochran and M.R. Farrally (eds.) *Science and Golf II: Proceedings of the World Scientific Congress on Golf*, 186-192. E&FN Spon, London.
- Landsberger, L.M. (1994). "A Unified Golf Stroke Value scale for Quantitative Stroke-by-stroke Assessment." In A.J. Cochran and M.R. Farrally (eds.) *Science and Golf II: Proceedings of the World Scientific Congress on Golf*, 216-221. E&FN Spon, London.

- Landsberger, L.M. (1998). "Hole Size, Luck, and the Cruelty of Putting: A Thought Experiment on the Impact of Quantization in Golf." In A.J. Cochran and M.R. Farrally (eds.), *Science and Golf III: Proceedings of the World Scientific Congress of Golf*, 363-370. Human Kinetics, UK.
- Larkey, P.D. (1994). "Comparing Players in Professional Golf." In A.J. Cochran and M.R. Farrally (eds.) *Science and Golf II: Proceedings of the World Scientific Congress on Golf*, 193-198. E&FN Spon, London.
- Larkey, P.D. & A. Smith. (1998). "All Around Improvements." In A.J. Cochran and M.R. Farrally (eds.) *Science and Golf III: Proceedings of the World Scientific Congress of Golf*, 377-384. Human Kinetics, UK.
- McCullagh, P. & J.A. Nelder. (1989). *Generalized Linear Models (Second Edition)*. Chapman & Hall, New York.
- Pelz, D. & J. Frank. (2000). *Dave Pelz's Putting Bible*. Doubleday, New York.
- Pelz, D. (1994). "A Study of Golfers' Abilities to Read Greens." In A.J. Cochran and M.R. Farrally (eds.) *Science and Golf II: Proceedings of the World Scientific Congress on Golf*, 186-192. E&FN Spon, London.
- Pennington, B. "Golf's Colorful Language Goes with Any Green." New York Times 11 May 2009: D7.
- Pope, D. & M. Schweitzer (2010). "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes." *American Economic Review*, forthcoming.
- R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Soley, C. (1977). *How Well Should you Putt? A Search for a Putting Standard*. Soley Golf Bureau, San Jose.
- Tierney, D.E. & R.H. Coop. (1998). "A Bivariate Probability Model for Putting Efficiency." In A.J. Cochran and M.R. Farrally (eds.), *Science and Golf III: Proceedings of the World Scientific Congress of Golf*, 385-394. Human Kinetics, UK.