Amateur Golf Handicapping: A Bayesian Approach

Zachary Dawson
Computer Science
Georgia Institute of Technology
Atlanta, United States
zdawson7@gatech.edu

I. Introduction

Golf is one of the only sports where players of differing abilities can compete with each other in a way that is still enjoyable for all participants. This is because you are ultimately competing against the course you are playing, and you can compare scores to your opponents after a round. Other sports require players to compete directly against each other, making a competition between unequal opponents either boring or frustrating for both players. Golf has had variants of handicapping systems almost as long as it has been a game [1]. The golf handicap system that is around today has been largely unchanged for more than 40 years [2].

In this paper we will outline the current handicapping system and its shortcomings. We will outline a proposed Bayesian solution to address the shortcomings of the current system. We will then explore the performance of these Bayesian models and outline how it compares to the current handicapping system.

This paper utilizes a dataset of scores derived from my home golf club. We will be leaving out any personally identifying information for privacy reasons. There are two courses at my home club: 'Lakeland' and 'Highland'. Highland is a very difficult course, and Lakeland is only of moderate difficulty. We will be using the scores of the 190 golfers at the club that had at least 10 full round played at the club in the previous year, with a total of 2688 rounds. We will be grouping players into three bins based on their USGA Handicap Index: low-handicappers, mid-handicappers, and high-handicappers each having a handicap of less than 10, between 10 and 20, and greater than 20 respectively.

A. USGA Handicapping

The USGA is one of the two main systems for handicapping golf along with the R&A. The USGA attempts to make golf competitions fair by creating a number known as a players Course Handicap. This Course Handicap is derived from the course rating, course slope, and the players Handicap Index. The course rating outlines how many stroked a 'scratch' golfer would take to complete the course. The course slope outlines how the difficulty of the course scales as a players handicap increases. A higher slope means a player with a high handicap will struggle more than on a course with the same course rating but a lower slope. A player's Handicap Index is derived by taking the average of the 8 best differentials in their last 20 rounds. A differential is calculated by subtracting the players score from the course rating and adjusting for the course slope [3]. The main benefit of this system is that it is relatively easy to understand and can be intuited by someone with no higher-math background.

There are many shortcomings to this system however, mainly stemming from the distributions of scores for players of differing abilities. Good players tend to have scores with very low variance. Bad players tend to have much more variance in their scores. Because of this, it is very uncommon for a good player to have an unusually low differential, but a bad player is much more likely to. If you are familiar with golf, you will know that someone who routinely shoots around 77 is not going to magically shoot a 67. A golfer who shoots around 97 can have rounds where they get lucky and shoot 87, but they are also liable to shoot more rounds above 107. A golfer who routinely shoots 77 will very rarely shoot above 87. These quirks in scoring distributions lead to inequity when it comes to competitions with handicaps. Good golfers tend to have an advantage in one-on-one matches because their scoring distribution variance is very low; they will beat a poor golfer every time their opponents have a bad round, which happens often. In large field tournaments, high-handicap golfers win events above the frequency that they should. The high variance in their scores increases the probability of a very low score differential given enough players. Ideally, each competition, regardless of the size, should aim for each player having a 1/N probability of winning (N is the number of players) [4].

Figure 1 shows the distribution of winners by handicap in simulated tournaments based on handicap on the Highland course. In all simulated tournaments we use actual scores from the dataset and apply the handicap that the player would have received on the day of that competition. In small fields, high-handicappers under-perform relative to their expected win rate. As the field size increases, high-handicappers perform much better than their expected win rate and low-handicappers under-perform. This effect is especially concerning because large field tournaments are normally those that people care the most about, and oftentimes have cash prizes.

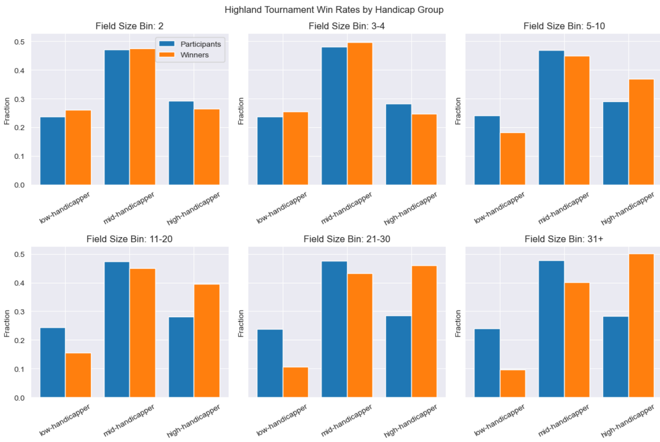

Fig. 1: Each handicap group's performance in simulated tournaments of different field sizes on the Highland Course.

Figure 2 shows a similar trend for the Lakeland Course but exhibits a less severe decrease in low-handicap winners as field size increases. This behavior is expected, as the variability on an easier course is expected to be lower than on a more difficult course, limiting the high-handicapper's advantage in large field tournaments.

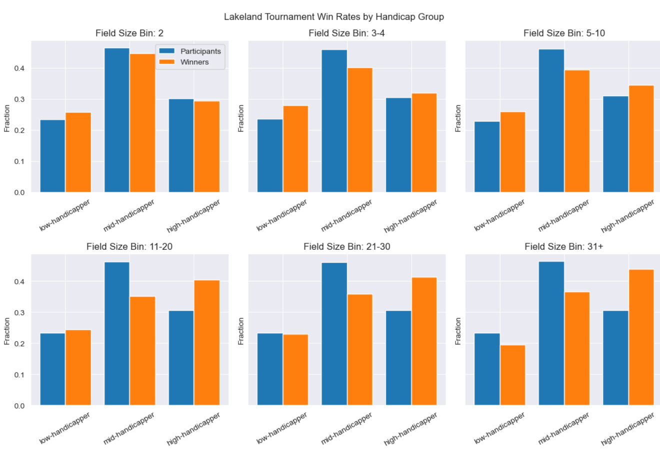

Fig. 2: Each handicap group's performance in simulated tournaments of different field sizes on the Lakeland Course.

B. *Proposed Bayesian Solution*

In order to address the shortcomings of the current handicapping procedure, we propose a Bayesian alternative that will predict a distribution of scores for each player and use those distributions to determine a more equitable handicap for all players.

Our handicapping system involves two steps. The first step is determining a predictive distribution of scores for each player. To do this we build a hierarchical model that uses historical data to predict new scores for each player. We then use the posterior predictive distribution for scores to build a handicap.

We will explore multiple ways to calculate the handicaps from the distribution of predicted scores. The simplest way is to take the score at the 1/N quartile and use that as a player's baseline winning score. You can then subtract each player's actual score from their baseline winning score to determine a differential. The player with the lowest differential wins. This strategy benefits in its simplicity, but does not fully address the concerns related to differing variance.

The 'one-over-n' method only predicts the probability of a player shooting better than 1/N of their predictive scores, not shooting lower than their competitors' score. If the other players have larger variances among their scores, then the 1/N assumption does not predict the probability that a player's score is lower than all of their competitors [4]. In order to address this, we will use a Robbins-Monro update rule [5] to simulate tournaments using the posterior predictive distributions for each player to find the optimal handicap for each tournament. We will use the update rule below:

$$h_i^{(t+1)} \;=\; h_i^{(t)} +\; \gamma_t\Big(w_i^{(t)} - \tfrac{1}{N}\Big),$$

where $w_i^{(t)}$ is the fraction of the M simulated tournaments player i wins using handicaps $h^{(t)}$, and γ_t is a decreasing step size.

After a set number of iterations or once the handicaps have converged, we will use the final handicap values to determine the differentials for the observed scores and determine a winner. This system, while more complex and computationally expensive, should be the most fair way of determining a player's handicap for any competition. We will call this handicapping method 'Simulated' handicaps.

We will be exploring two Bayesian models in our attempts to build a better handicapping system: a Random Walk Model and a Bayesian Linear Model.

The structure of the data is a time series of scores shot by each player with relevant round statistics, like the player's handicap at the time, the course rating, and the course slope.

The AR1 random walk model is outlined below:

$$\begin{aligned}
&\alpha_i \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2), \quad \mu_\alpha \sim \mathcal{N}(85, 7.5), \quad \sigma_\alpha \sim HalfNormal(2.0)\\
&\gamma_c \sim \mathcal{N}(0, \sigma_\gamma^2), \quad \sigma_\gamma^2 \sim HalfNormal(2.0)\\
&\delta_{i,t} = \rho\delta_{i,t-1} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0, \tau^2)\\
&\rho \sim \mathcal{N}(0, 1), \quad \tau^2 \sim HalfNormal(1.5)\\
&\sigma^2 \sim HalfNormal(3.0)\\
&y_{i,t} \sim \mathcal{N}\big(\alpha_i + \gamma_{c(i,t)} + \delta_{i,t}, \quad \sigma^2\big), \quad i = 1, \ldots, N., \quad t = 1, \ldots, T
\end{aligned}$$

where N is the number of players in the training set, and T is the number of rounds for each player.

Each player's score at time t is defined as the player's partial pooling factor (to model player-specific skill), the course effect (to model the differing difficulty of courses), and the random walk component that utilizes previous scores to model streaks while using a correlation component to regress to the mean of previous scores.

The Bayesian linear model is defined below:

$$
\begin{aligned}
&\beta_0 \sim \mathcal{N}(0,\ 5)\\
&\beta_1 \sim \mathcal{N}(0,\ 1)\\
&\sigma_{\text{course}} \sim HalfNormal(1)\\
&\gamma_c \sim \mathcal{N}(0,\ \sigma^2_{\text{course}})\\
&\sigma_{\text{player}} \sim HalfNormal(2)\\
&\alpha_j \sim \mathcal{N}(0,\ \sigma^2_{\text{player}})\\
&\mu_i \ =\ \beta_0\ +\ \alpha_j\ +\ \beta_1 m_i\ +\ \gamma_c
\end{aligned}
$$

$$
\begin{aligned}
&\log\sigma_0 \sim \mathcal{N}\big(\log 3,\ 1^2\big)\\
&\beta_2 \sim \mathcal{N}(0,\ 0.3^2)\\
&\log\sigma_i = \log\sigma_0 + \beta_2\, s_i\\
&\sigma_i = \exp\!\big(\log\sigma_i\big)
\end{aligned}
$$

$$
y_i \ \sim\ \mathcal{N}\big(\mu_i,\ \sigma_i^2\big), \qquad i = 1, \ldots, N.
$$

The linear model defines a player and course specific mean and player specific variance. It uses a rolling average (m_i) and standard deviation (s_i) of the previous rounds up to time T. Both of these inputs were scaled to allow for z-score based parameterization of each of the coefficients in the linear model. This model is simpler than the Random Walk Model, as it does not attempt to model streakiness of golfers. It simply uses the player specific rolling standard deviation as a surrogate for player variability. This simplicity should allow it to train faster than the random walk model and be easier to sample after training.

II. Methods

We will train each model on an identical dataset. We will use a 60-40 train test split with the training set coming from the first 60% of rounds for each player and the testing coming from the remaining 40%. This distinction is important, as we do not want any data leakage from future scores improving the models performance beyond what it could reasonably do on unseen future data. After each model is trained, we will simulate tournaments of size 2, 4, 5, 10, 15, 20, and 30. Each tournament will be simulated with each handicap type (One-Over-N and Simulated). The handicaps will be calculated using posterior predictive distributions for the test dataset, and the observed test scores (which the model has no knowledge of) will be used to determine the winner. We will then compare the winning frequencies of each handicap group in the different tournament sizes. Note, we will use the Handicap Index from the USGA to make the groups to allow for easier comparison to the original data.

III. Results

A. *Random Walk Model*

The training of the Random Walk model took 6 minutes and resulted in no divergences for any of the parameters. Figure 3 shows the PPC plot for the Random Walk model. It shows that the predictive power of the model aligns well with the observed data. The overall playing ability of the players μ_α was 90.5 with a 95% HDI of (87.97, 93.22). The variability among players σ_α was 7.5 strokes with a 95% confidence interval of (6.8, 8.3). This indicated that most scores fall between 71 and 110, which aligns with our expectations. The course effects were non-significant with Highland having a course effect of 0.48 and Lakeland a course effect of -0.36. The confidence intervals for each parameter overlap significantly, (-2.0, 2.9) and (-2.0, 2.9) respectively. The average temporal effect delta was about 1 stroke, significantly less than player-dependent variability. There was also a 0.86 mean (HDI (0.54, 0.99)) for ρ, which modeled the factor of mean regression. It highlights that there is a large dependence of deviation between consecutive rounds.

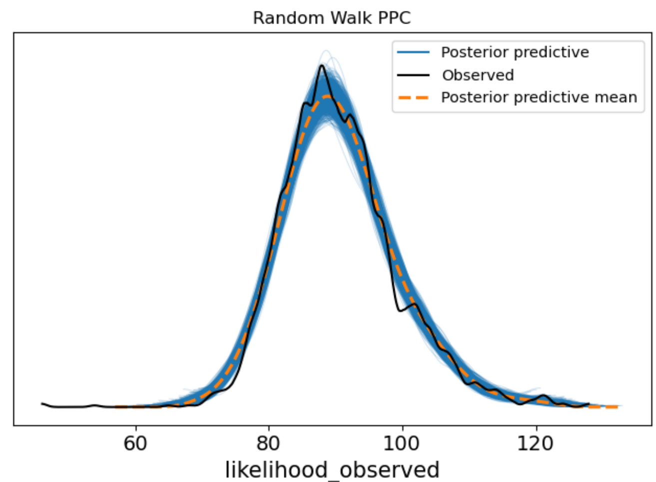

Fig. 3: PPC for Random Walk Model.

As we can see in Figure 7, this predictive power leads to a significant improvement in the fairness of tournament simulations, when compared to the standard USGA handicapping process. The levels are nearly perfect at low field sizes, and while there are still some discrepancies as field size increases to its highest levels, the advantage given to high-handicap players is significantly reduced.

The winning rate for high-handicappers in Lakeland tournament simulations is 0.43 compared to its expected 0.31. On Highland the discrepancy is much smaller, with high-handicappers winning at a rate of 0.2 compared to their expected 0.23. These discrepancies are caused by a non-standard distribution of high-handicap scores on Lakeland. This can be seen in comparing Figure 4 and Figure 5. Lakeland's high-handicap distribution has two modes and more low differentials relative to the total number of available samples. A small number of abnormally low differentials will lead to a group always winning when they are in a large field simulation. With more data, these low differentials would be sampled in the tournament field less often and the results would likely converge to a more fair level. Due to the lack of data, many tournaments of large size have a similar playing pool and the same winner.

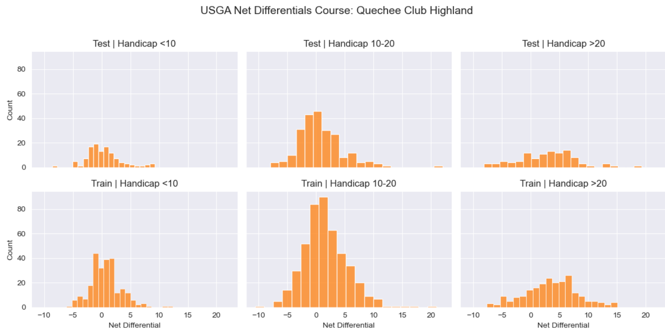

Fig. 4: USGA differential distribution on Highland for each handicap group.

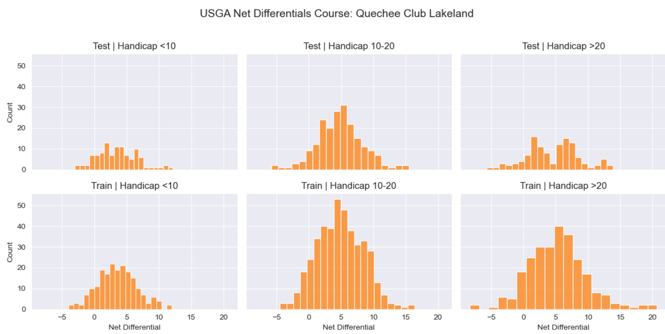

Fig. 5: USGA differential distribution on Lakeland for each handicap group.

B. Bayesian Linear Model

Similarly to the Random Walk model, the Bayesian Linear Model was trained with no divergences in 6 minutes. However, the PPC plot does not fit the observed data as well as the Random Walk Model, which can be seen in Figure 6. It underestimates the performance of mid-handicappers and it overestimates the performance of low-handicappers. This could be an explanation for why low-handicappers under perform even more using this model. Mid-handicappers also over-perform in large field tournament simulations.

The parameters for the model do not align well with our expectations. β_0 had a posterior mean 0f 90.3 and 95% HID of (88.7,92.1). This value for an intercept does not align with our expectation, as you would expect the intercept to be closer to the par of the course rather than a higher than average score. β_1 had a mean of -0.20 and confidence interval of (-0.35, 0.05). This also does not align with expectations, as you would expect a higher moving average to lead to a higher score, not a lower one. β_2 had an insignificant effect with a mean of 0.002 confidence interval of (-0.039, 0.047). The course effects are also insignificant in this model. The inconsistency with reasonable values is a possible explanation for why the model fits the data worse and performs poorly in simulations compared to the Random Walk Model.

As seen in Figure 8, the simulated tournaments for the model perform better than the USGA simulated tournaments, but they perform worse than the Random Walk Model's simulations, especially at large field sizes.

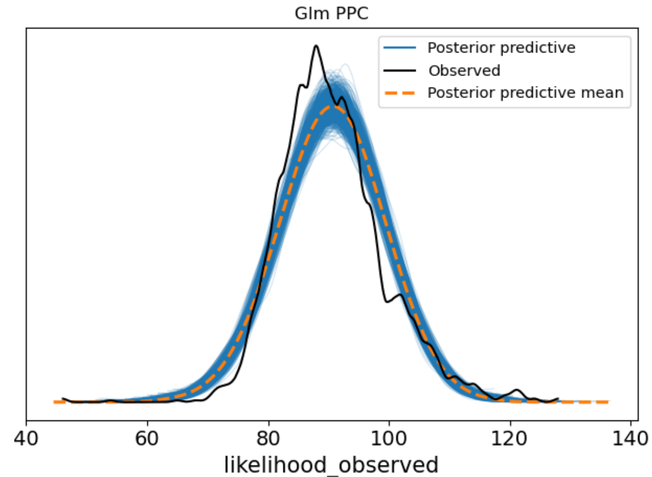

Fig. 6: PPC for Bayesian Linear Model.

IV. Discussion

The PPCs provide insight into the performance of each model. The Random Walk model demonstrates strong alignment between the posterior predictive samples and the observed data. In contrast, the Bayesian Linear Model struggles to capture the sharp peak around scores of 88–92, instead producing a broader and flatter distribution. This suggests that the model underestimates the concentration of scores near the mode and overestimates the frequency of mid-tail scores. The GLM's heavier tails further indicate a misestimation of player variance, likely due to its lack of temporal dependence. These differences contribute to the performance of the Random Walk model in tournament simulations, particularly in capturing low-score outliers more accurately.

The disparity in win rates, especially for large field sizes, can be directly tied to the model's ability to capture variance and tail behavior in score predictions. The Random Walk model's posterior predictive distribution closely tracks the observed scores, capturing both the sharp mode and the heavier tails. This enables more accurate simulation of the left-tail performance, which is critical in large tournaments.

Also, a small distinction emerges between the One-over-N and Simulated handicap approaches in terms of fairness. While the One-over-N method provides a simple adjustment, it does not account for the variance of score distributions between competitors. In contrast, the Simulated handicaps produce more equitable win distributions. These results emphasize the importance of tailoring handicaps to a player's own scoring distribution as well as the distributions of their competitors.

V. Conclusion

The results of this project show that the Random Walk model paired with Simulated handicaps provides the most equitable handicapping system across a wide range of tournament sizes. Its ability to model temporal trends, player-specific variance, and course effects gives it a distinct advantage over the Bayesian Linear Model, which tends to produce flatter,

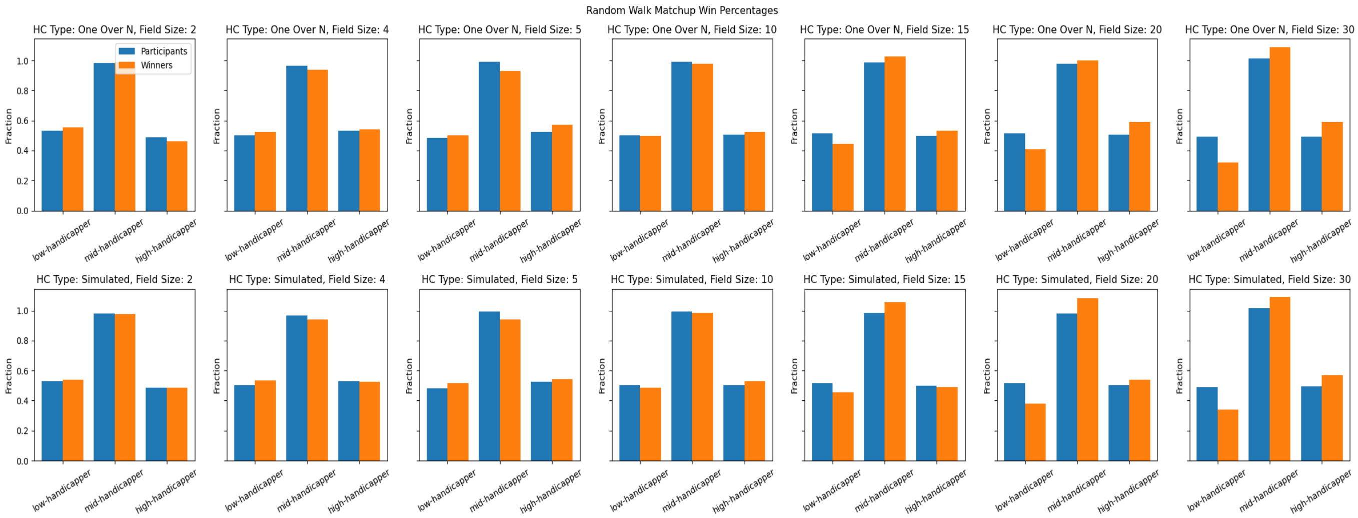

Fig. 7: Each handicap group's performance in simulated tournaments of different field sizes using the Bayesian Random Walk on the Highland Course.

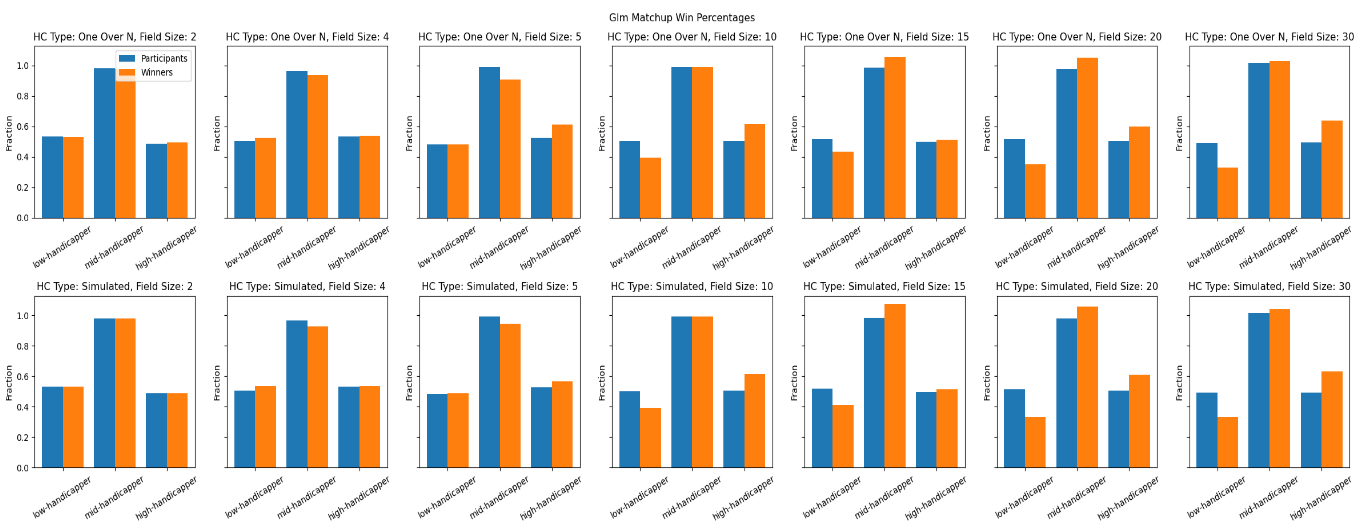

Fig. 8: Each handicap group's performance in simulated tournaments of different field sizes using the Bayesian Random Walk Model on the Highland Course.

less accurate posterior predictive distributions. Similarly, the Simulated handicap approach outperforms the One-Over-N method by directly optimizing each player's win rate toward the ideal 1/N target win rate, especially in larger fields where tail outcomes matter most.

That said, the simplicity of the GLM and the One-over-N method should be considered. Their structure and interpretation are much easier to explain to a general golfing audience, and they are less computationally expensive. These qualities would make them more appealing in a real-world implementation.

A limitation of this analysis is the relatively small dataset available. A handful of unusually low scores by high-handicap players in the Lakeland data skewed results in large field simulations. With more data, these anomalies would likely have a diminished impact, allowing the model's fairness to improve further and stabilize across repeated samples. It would also be beneficial to test the methods using more than two courses and a player set that goes beyond a specific golf club.

Finally, this framework could be extended beyond amateur play. Professional tours now track detailed strokes-gained statistics, which offers a much richer input for modeling golfer performance. With accurate predictive distributions built using this data, a similar system could be used for improving betting odds in sports markets.

VI. References

References

[1] H. Part, "History Of Handicapping, Part I: Roots Of The System," USGA, Oct. 06, 2011. https://www.usga.org/content/usga/home-

page/articles/2011/10/history-of-handicapping-part-i–roots-of-the-system-21474843620.html (accessed Apr. 19, 2025).

[2] H. Part, “History Of Handicapping, Part IV: The Rise Of The Slope System,” USGA, Oct. 25, 2011. https://www.usga.org/content/usga/home-page/articles/2011/10/history-of-handicapping-part-iv-the-rise-of-the-slope-system-21474843751.html (accessed Apr. 19, 2025).

[3] The, “The Handicap Index Calculation,” USGA, Oct. 29, 2019. https://www.usga.org/content/usga/home-page/handicapping/world-handicap-system/topics/handicap-index-calculation.html

[4] T. B. Swartz, “A New Handicapping System for Golf,” Journal of Quantitative Analysis in Sports, vol. 5, no. 2, Jan. 2009, doi: https://doi.org/10.2202/1559-0410.1168.

[5] Wikipedia Contributors, “Stochastic,” Wikipedia, Aug. 16, 2019. https://en.wikipedia.org/wiki/Stochastic