

Finding Edge in Daily Fantasy Basketball

Brandon Shimiaie
brandonshimiaie@gmail.com

2020-07-20

1 Introduction

The objective of this study is to develop a strategy that generates positive ROI by entering National Basketball Association (NBA) Daily Fantasy Sports (DFS) contests. My attempt to complete this objective is motivated by the ostensible alpha in DFS competitions, caused by "retail players" who possess limited domain knowledge, employ minimal strategy, and comprise approximately 80% of all DFS players [1]. Advancing prior work that aimed to generate a profit in DFS, namely Haugh and Singal (2018) and Hunter, Vielma and Zaman (2019), I describe both an approximation of the optimal solution to the DFS problem, and the concrete pipeline that backtests a time-optimized greedy strategy.

1.1 Overview of Daily Fantasy Sports

On May 14, 2018, the United States Supreme Court paved the way for the expansion of legalized sports gambling with its decision in *Murphy v. Nat'l Collegiate Athletic Ass'n*, 138 S. Ct. 1461 (2018). The court held that the Professional and Amateur Sports Protection Act ("PASPA") — the federal law that for over twenty-five years prohibited states from passing any new laws authorizing gambling on professional or amateur sporting events — was an unconstitutional violation of states' rights [2]. Currently, some form of DFS is legal and operational in 43 states and the District of Columbia. As a result, the amount of public involvement in DFS has steadily increased. In a December presentation to investors, DraftKings reported \$213 million in revenue in 2019 with 60% of the market share, indicating the DFS industry brought in more than \$350 million in revenue in 2019 [3]. Moreover, the fantasy sports market size is expected to reach more than \$1.5 billion by 2024 [5]. As the interest in DFS has grown, so has the competitiveness of the market. Many websites have been created with the purpose of providing DFS analysis to the public, such as Fantasy Cruncher, Fantasy Labs, and Rotogrinders. These sites provide up-to-date player news, player performance projections, lineup optimizers, and more. Despite the public's access to advanced analysis and strategy, "sharks" who use sophisticated and proprietary techniques continue to make a consistent profit. The ruling of DFS as a skill-based game is strongly evidenced by the fact that the top 1% take 90% of the profits and pay 40% of the entry fees [4]. Currently, the two dominant DFS platforms are DraftKings and FanDuel.

On its surface, Daily Fantasy Sports is a simple game. The goal is to enter a lineup that consists of a subset of an available player pool into a contest. Each player in the pool is assigned a position and a salary based on recent and prospective performance. A lineup is subject to constraints such as the number of players required in each position and the sum of the salaries of the players. The limit of each constraint depends on the platform hosting the contest. For example, in Classic DraftKings NBA contests, a valid lineup must not exceed the salary cap of \$50,000, cannot contain duplicate players, and must contain exactly 1 PG, 1 SG, 1 SF, 1 PF, 1 C, 1 G (PG or SG), 1 F (SF or PF), and

1 Util (PG, SG, SF, PF, or C). The score of a player is represented by fantasy points, which is a function of that player’s boxscore statistics that is calculated differently on each platform. The score of a lineup is the sum of the scores of its players. Regular fantasy sports leagues typically last an entire season, and are played by groups of friends or colleagues. Daily fantasy sports contests, however, take place over a single round of games and are played online against hundreds to hundreds of thousands of unknown users. DraftKings and Fanduel offer a plethora of contests every day varying in size, rules, entry fees, and payouts.

There are several different contests available on each platform, although most of them can be broadly categorized into two specific types: cash games and tournaments. Each contest type has its own payoff structure. Cash games distribute the entry fees (less the vig) equally to the lineups that score in the top 50% of the field. They can consist of as little as 2 entries (known as head-to-head contests), to as big as tens of thousands of entries. In tournaments, 20% of lineups typically win money, with the payouts exponentially increasing with respect to the rank of the lineup. The strategy required for tournaments is very different from the strategy required for cash games. To win first, second, or third place in a large field of competitors requires a different drafting strategy than trying to land in the top 50%.

Despite its apparent simplicity, picking winning lineups is a challenging problem. The performance of an athlete has high variance on a game-to-game basis and the salary of each player in a pool usually reflects how well that player is expected to perform, making it hard to find value. Therefore, a bigger advantage can be gained by strategically drafting lineups than by attempting to perfectly predict performance.

1.2 Previous Work

As of the time of this writing, two notable studies on the problem of constructing portfolios exist in the DFS literature. The work of Hunter, Vielma and Zaman (2019) applies the framework of the picking winners problem, introduced by Hunter, Saini, and Zaman (2018), to DFS. In particular, the study considers the problem of selecting a portfolio of entries of fixed cardinality for contests with top-heavy payoff structures, i.e. most of the winnings go to the top-ranked entries [7]. They focus on the optimization problem of maximizing the probability that at least one entry in a tournament wins the contest. As a result of the computational complexity of finding an optimal solution to this problem, they show that this probability can be approximated using only pairwise marginal probabilities of the entries winning when the probability of each entry winning is low and the entries do not have strong correlations with each other. The study then extracts a set of heuristic principles from the approximation to form an integer programming formulation to construct the entries. Namely, the entries should have a high expected value, high variance, and a low correlation with

each other. The study conducted by Haugh and Singal (2018) seeks to maximize the expected reward subject to portfolio feasibility constraints in both cash and tournament contests, while explicitly accounting for opponent behavior. They leverage the mean-variance literature on outperforming stochastic benchmarks to reduce each DFS problem to solving a series of binary quadratic equations [8].

This study will develop the state of the DFS literature by addressing the shortcomings of each study summarized above. The objective the work of Hunter, Vielma and Zaman (2019) is to construct a set of lineups that maximizes the probability that at least one lineup wins a contest. The study attempts to achieve this objective by using integer programming to create entries that maximize expected points subject to the variance of each entry i being larger than the constant ϵ_i and the covariance between each entry pair (i, j) being less than the constant γ_{ij} . Three major problems exist in this approach. First, As noted in Haugh and Singal (2018), this work fails to account for opponent behavior. It does not consider that a lineup's absolute score is irrelevant and it is rather its score relative to all opponent lineups that is important. Contestants' scores are all based on the performance of the same pool of players, and the performance of any given player affects a contestant's score only relative to other contestants who did not draft the player in question. Therefore, the performance of low-ownership players on one's team has vastly more impact on one's overall position in the standings than that of one's high-ownership players [9]. Second, the study estimates the value of a lineup as the probability that it wins a contest. However, the true value of a lineup is its expected PNL from entering a contest. As mentioned in the previous section, 20% of lineups typically win money in tournaments and many payouts below 1st place are non-negligible. The entire payout structure of a tournament should be taken into account when assigning value to a lineup. Lastly, Hunter, Vielma and Zaman do not prove that the heuristic they use to construct lineups maximizes the probability of winning a contest. Each principle used in this study is widely known and employed, which is a large disadvantage when trying to win a large contest. Furthermore, each ϵ_i and γ_{ij} is hand chosen rather than programmatically determined. The work of Haugh and Singal (2018) addresses the shortcomings of Hunter, Vielma and Zaman (2019) by both modeling opponent behavior and maximizing the expected PNL of each entry, rather than its probability of winning a contest. However, the focus of this study is to construct a single entry that maximizes expected profit in a contest. Haugh and Singal briefly discuss a multi-entry extension of the single-entry solution, but it is a naive algorithm that ensures that each portfolio i can not have more than γ players in common with each of the previous $i - 1$ portfolios. Moreover, the optimization algorithm for the tournament problem with a single entry does not consider a diverse enough set of entries when determining the one with maximum value. In contrast, this study considers a large and diverse set of entries, selecting the subset with the highest expected PNL from entering a contest.

1.3 My Contributions

I begin by representing each DFS problem as an optimization problem whose objective is to select a set of portfolios that maximizes expected returns while subject to a system of constraints imposed by the party hosting the competition. I employ a novel approach to solving this optimization problem by modeling opponent lineups, constructing a pool of diverse potential lineups with high expected point totals, and simulating the performance of each potential lineup against the set of opponent lineups I constructed. Moreover, I describe my proprietary methodology for modeling opponent player ownership and expected player distributions.

A major limitation of previous studies that explore the problem of constructing DFS portfolios is an insufficient sample size of contests entered with lineups derived from a proposed strategy. To address this limitation and more thoroughly assess the effectiveness of my approach to entering contests, I developed a pipeline to backtest its historical performance. The pipeline begins by collecting the data necessary to construct a set of portfolios and enter past competitions. Engineering the optimal set of portfolios to enter into a given contest requires knowledge of the distribution of possible performance outcomes and the opponent ownership projection for each player active in the contest. Player performance in DFS is represented by fantasy points, which is a linear combination of various player boxscore statistics. This study employs a novel approach to determining the mean of the distribution of possible fantasy point outcomes: individually predict each statistic then derive expected fantasy points. Each player statistic prediction is generated by first creating predictive features using recent player performances and extraneous variables such as opponent defense and Vegas odds, followed by training a model using past data, and finally plugging the current features into the trained model. The variance of a player's potential fantasy point distribution is determined by the model's squared residual of predicted fantasy points for that player in that season. Opponent player ownership projections are modeled on a slate level, using features that capture the past ownership of a player in similar contests within the same season and the expectation of the performance of a player relative to the field. The backtest iterates through one season at a time, using all previous seasons to train the models used to determine fantasy point distributions and opponent ownership projections. Then for each contest within a season, each active player's fantasy point distribution and opponent ownership projection is modeled, the optimal set of portfolios is generated, and the PNL produced by each portfolio is logged. Finally, I analyze the net ROI of the strategy proposed by this study.

I evaluate my methodology on Classic DraftKings NBA contests starting from the beginning of the 2017-2018 NBA season. I enter the maximum amount of lineups allowed into each contest, with the exception of those entries that violate self-imposed budget constraints. I then visualize the cumulative PNL of my strategy. I finally identify the subset of contests, considering the number of

games in the slate, the entry fee, etc., that is best suited for my strategy.

2 Strategy

2.1 Problem Formulation

Let the pool of n players available in a slate be represented by the set

$$\{P_i\}_{i=1}^n \quad (1)$$

where

$$P_i \sim N(\mu_i, \sigma_i^2) \quad (2)$$

A single entry in a contest consisting of a subset of K players in this pool whose salaries total less than the budget B can be represented as

$$x \in \{0, 1\}^n \quad (3)$$

such that

$$\sum_{i=1}^n x_i = K \quad (4)$$

$$\sum_{i=1}^n s_i x_i \leq B \quad (5)$$

Additional constraints are also imposed by the contest organizers, such as the number of players allowed of each position. I use X to denote the set of binary vectors x that satisfy each of these constraints.

I represent the expected PNL of an entry X_i entered into a contest c as EV_i^c . If a contest allows m total entries per person, the goal is to select a set of entries $S \subset X$ of size m that maximizes the total expected PNL. Finally, the DFS problem is formally described by the following optimization problem

$$\max_{S \subset X, |S|=m} \sum_{i \in S} EV_i^c \quad (6)$$

2.2 Cash Game Strategy

2.3 Tournament Strategy

References

1. Pramuk, J. 2015. <https://www.cnbc.com/2015/10/06/former-pro-poker-player-makes-a-living-on-fantasy-sports.html>.
2. Quinn Emanuel Urquhart & Sullivan, LLP. 2018. <https://www.jdsupra.com/legalnews/august-2018-the-implications-of-the-19446/>
3. Nover, Scott. 2020. <https://www.vox.com/2020/1/29/21112491/daily-fantasy-sports-betting-dfs-merch-analysis-weatherman>
4. Harwell, Drew. 2015. <http://www.dailyherald.com/article/20151012/business/151019683/>
5. Reporterlink. 2019. <https://markets.businessinsider.com/news/stocks/the-fantasy-sports-market-size-is-expected-to-reach-more-than-1-5-billion-by-2024-1028445951>
6. Hunter, Saini, and Zaman. 2018. <https://arxiv.org/pdf/1706.04229.pdf>
7. Hunter, Vielma and Zaman. 2019. <http://www.mit.edu/~jvielma/publications/Picking-Winners.pdf>
8. Haugh and Singal. 2018. <http://www.sloansportsconference.com/wp-content/uploads/2018/02/1001.pdf>
9. Weldon, Alex. 2015. <https://www.parttimepoker.com/dkleak-scandal-why-are-ownership-percentages-so-important>