<div align="center">**The Predictability of NBA and MLB Games**</div>

Introduction

Following professional sports is a popular hobby in the world. Not only is it fun for fans to cheer on their favorite teams, but many also find pleasure betting on the outcome of games. This paper investigates how well bookmakers predict the results of MLB and NBA games. This project was inspired by the legalization of sports betting in the United States by the Supreme Court in 2018. It is expected that sports betting will become increasingly popular in the US as more and more states legalize sports betting, particularly on the MLB and NBA, two of the four major American leagues. In order for bookmakers to make a profit, their odds must be predictive of the outcome of the games. Otherwise, bettors would be able to exploit odds that are "poorly calculated". For example, for a game between Team A and Team B, if a sportsbook lists team A to be the favorite where in reality, Team A is the underdog, more informed bettors can bet on Team B and have a very good chance of receiving a large profit. The questions this paper will examine are: How well do the odds provided by various bookkeepers predict the outcomes of MLB and NBA games? Is MLB more or less predictable than the NBA? Are there temporal trends in the predictability of MLB and NBA games?

Data Scraping

Data was scraped from https://www.sportsbookreview.com/ using the Selenium package in Python. For the MLB, the closing moneyline odds[1] for every game in the 2019 season (including preseason and postseason games) was collected across 26 different sportsbooks. For the NBA, the closing moneyline odds for every game from October, 2019 (i.e. the beginning of the 2019-2020 season) to March 5th, 2020 was collected across the same 26 sportsbooks. Separate Python files were written to scrape the MLB and NBA (MLB_Moneylines.PY and NBA_Moneylines.PY respectively). The MLB scraper writes data into two files: a games file (sbr_mlb_games.xlsx) and an odds file (sbr_mlb_odds.xlsx). The games file includes details about every game (i.e. the home team ID, the away team ID, year, month, day, gametime, home score, and away score).[2] A unique game ID was generated for each game based off of the home team ID, away team ID, year, month, day, and game's starting hour. The sbr_mlb_odds file is a dataset of all moneyline odds for the MLB. It has four columns: the game ID, the book ID, the outcome (i.e. the team ID of which the odd corresponds to), and the closing decimal odds the book offered for the team in question for the given game.[3] The NBA scraper also writes data into a games file (sbr_nba_games.xlsx) and an odds file (sbr_nba_odds.xlsx). The only difference between the NBA games file and the MLB games file is that the NBA games file does not record the game time because in the MLB, two teams can play each other on the same day, but this is not possible in the NBA. Thus, for NBA games, a unique game ID was generated for each game based only on the home team ID, away team ID, year, month, and day. The NBA odds files has the same structure as the MLB odds file.

Predictability of NBA and MLB Games

Figure 1 shows the percent of NBA and MLB games correctly predicted for ten of the sportsbooks.[4] For each game, a sportsbook classifies one of the teams as the favorite and the other as the underdog. The favorite has a

---

[1] The moneyline market is a market to bet on the winner of the game. Closing odds are the odds immediately before betting on a particular game is locked (e.g. right before the game begins).

[2] See the team_ids.xlsx file for a list of all team IDs and their corresponding team names (for the MLB and NBA).

[3] See the book_ids.xlsx file for a list of all book IDs and their corresponding bookmaker names

[4] Only ten of the sportsbooks were chosen in order to not overcrowd the figure. Choosing any of the sportsbooks would yield similar results.
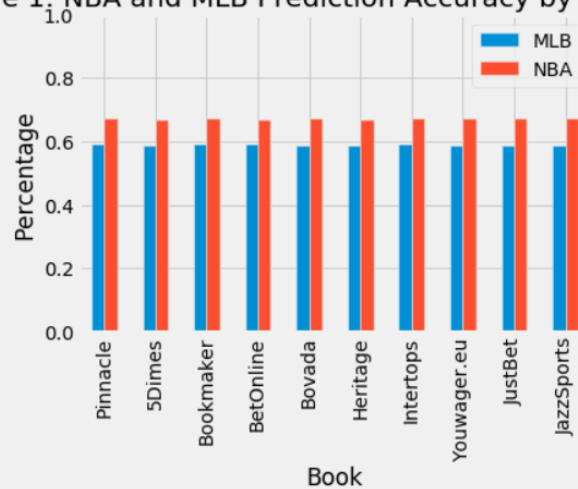
greater implied probability[5] of winning than the underdog. A correct prediction is when the favorite wins the game. The results of Figure 1 shows that all ten sportsbooks correctly predicted the winner of NBA games about 66% of

the time, but only about 58% of the time for MLB games. Thus, for both sports, sportsbooks predict the outcome of the games considerably better than randomly choosing.

The results indicate that NBA games are more predictable than MLB games. This makes sense as luck generally plays a greater role in the NBA than the MLB. In the NBA, a player can get lucky and make a tough three-pointer. However, the impact of that made shot on the game result is very small as each team generally scores over 100 points per game. In the MLB, a player can get lucky and hit a home run.

Figure 1: NBA and MLB Prediction Accuracy by Sportsbook

That home run has a large impact on the game result as generally, teams score less than 10 runs per game. In fact, it is not completely unusual for an MLB game to end 1-0.

The results of Figure 1 also indicates that each of the sportsbooks perform similarly to each other. This suggests that the closing moneyline odds of the sportsbooks for each game are nearly identical to each other. This is not surprising because if the closing odds differed significantly from each other, there would be numerous arbitrage opportunities for bettors, which would cause many of the sportsbooks to lose money.[6] One extension for future analysis would be to also scrape the opening odds for all of the games for each sportsbook. It might be possible for the opening odds to differ wildly by sportsbook, and each sportsbook would need to adjust their odds in order to compete with other sportsbooks, ultimately converging to similar closing odds.

Predictability of NBA and MLB Games Using Linear Regressions

Book Name	NBA R-Squared	MLB R-Squared
Pinnacle	0.161587	0.048774
5Dimes	0.158915	0.047969
Bookmaker	0.160273	0.048054
BetOnline	0.158822	0.048663
Bovada	0.173355	0.047293
Heritage	0.160676	0.047795
Intertops	0.168262	0.049700
Youwager.eu	0.164340	0.047780
JustBet	0.167524	0.047940
JazzSports	0.161620	0.045841

[5] When cleaning the data, all decimal odds were converted to implied probability (Implied Probability $= \frac{1}{\text{Decimal Odds}}$).

[6] In sports betting, arbitrage is when a bettor can bet on all outcomes and earn a riskless profit. For example, consider a game between Team A and Team B. Sportsbook 1 has +200 for Team A winning (bet $100 to profit $200) and -250 for Team B winning (bet $250 to profit $100). Sportsbook 2 has -250 for Team A winning and +200 for Team B winning. A bettor can exploit the two sportsbooks by betting $100 on Team A for Sportsbook 1 and $100 on Team B for Sportsbook 2. Regardless as to who wins, the bettor will profit $100 (one of the bets will result in a loss of $100 and the other bet will result in a $200 profit).

Another method to measure how well sportsbooks predict the winner of NBA and MLB games is with linear regressions. For ten of the sportsbooks, the following linear regression was run separately for NBA and MLB:

$$Win_i = a_0 + a_1 * Odds_i + U_i$$

The *Win* variable is 1 if the team won the game; 0 otherwise. The *Odds* variable is the implied probability the sportsbook believes the team has of winning the game (i.e. is continuous on a scale between 0 and 1). For each of the 20 regressions, the R-Squared was calculated, which are all reported in the table above. For reference, an R-Squared of 0 indicates that the sportsbook's ability to predict the winner of a game is as good as flipping a coin. From the results, the R-Squared for each of the sportsbooks for predicting NBA games was around 0.16, and they were around 0.048 for predicting MLB games. This provides further evidence that the sportsbooks predict the winners of NBA games better than MLB games.

Temporal Trends in the Accuracy of Bookmakers for Predicting NBA and MLB Games

Figure 2: Accuracy of Odds for Different Sportsbooks for NBA Games by Month
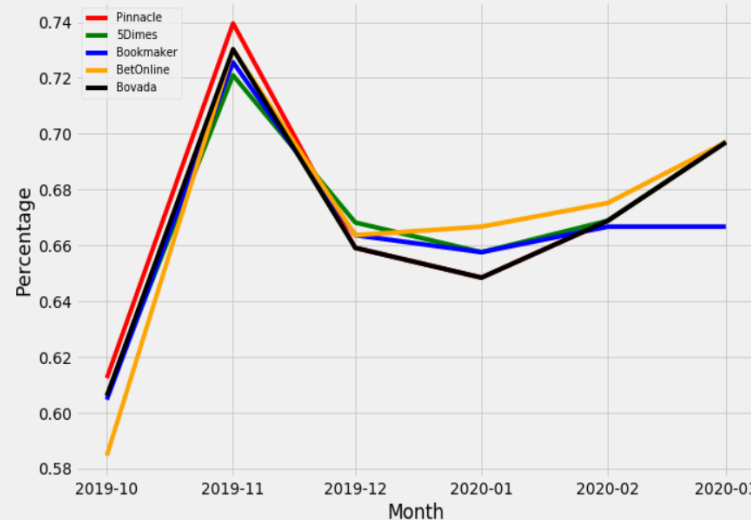

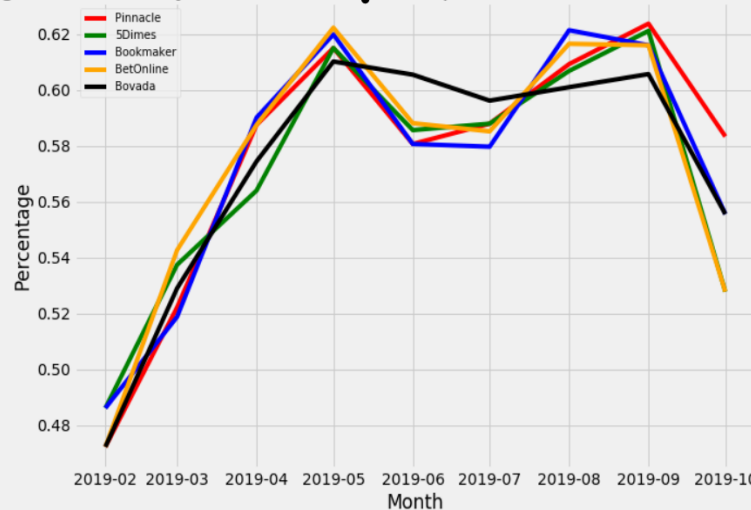Figure 3: Accuracy of Odds for Different Sportsbooks for MLB Games by Month

Figure 2 shows that the temporal trends for the five sportsbooks in predicting results of NBA games (2019-2020 season) were very similar to each other.[7] The trends observed showed a clear low in October, a peak in November, a slight drop-off from November to December, followed by a slight upwards progression between January and March, where the season currently sits.[8] Predicting NBA games in October was the hardest as the vast majority of games were pre-season games. Players who play in pre-season games are often unpredictable and can differ wildly than in regular season games as each team tries to experiment with different lineups. Then, in November, predicting NBA games becomes much easier as teams revert back to their regular season lineups. Predicting NBA games in December was more difficult than in November. This can be due to the fact that some teams may begin performing unexpectedly well or poorly as the season goes on. For example, coaches may have devised strategies that are surprisingly effective (or ineffective) which go unaccounted by bookmakers. NBA games become slightly easier to predict from January onwards, likely because bookmakers began accounting for any "surprise performances" by teams earlier in the season.

Figure 3, similar to Figure 2, shows that the temporal trends for the five sportsbooks in predicting results of MLB games (2019 season) were very similar to each other. All the bookmakers did a very poor job at predicting games in February (i.e. preseason games). In fact, the accuracy of all of the bookmakers were below 50% for that month, meaning that they would have performed better by flipping a coin. Thus, it may be a profitable strategy to always bet on the underdog for preseason games in MLB, and it would be interesting to see temporal data from other MLB seasons to see if the same results hold. From March to September, the accuracy of each of the sportsbooks hovered between 58% and 62%. However, the accuracy of each of the five sportsbooks dropped substantially from September to October. In fact, the MLB postseason began on October 1st, so upsets were more likely to happen in postseason games than in regular season games. In the postseason, teams are playing with much higher stakes (i.e. they can be quickly eliminated by losing consecutive games in a series). It might be the case that some teams tend to perform well under pressure and others do not, and this is a factor that bookmakers have trouble accounting for when determining the favorites of the games.

Distribution of the Work

All three members of the group made significant contributions to the project. Jonathan did the data scraping and analyzed the MLB dataset. Chloe analyzed the NBA dataset and helped out with the data visualizations. Charlie assisted in the analysis of the MLB and NBA datasets and wrote much of the write up.

---

[7] In order to not overcrowd Figures 3 and 4, only five sportsbooks were chosen for analysis. Choosing any of the sportsbooks would yield similar temporal trends for the MLB and NBA. In both figures, accuracy is defined as the percentage of games where the projected favorite wins.

[8] Note that March games only reflects those from 3/1/2020 to 3/5/2020.