A Data Driven Approach to Finding an Edge in the NBA Betting Markets

Reed Peterson

Introduction

Task

- Build a model and develop a strategy to make profitable bets on outcomes of NBA games
- This requires a model that predicts some outcome of an NBA game with higher accuracy than the Vegas betting lines
- Vegas charges a premium for bets (the "vig") which provides them with an opportunity for risk-free profit if they achieve equal sums of money wagered on each side of a line
- Opportunity for exploitation - the line may be statistically biased as a result
- We examine pregame point spreads, pregame point totals, and halftime point spreads

The Kelly Criterion

- Armed with a successful model that presents us with an edge over sportsbooks, it is imperative to employ a near-optimal strategy for how much to wager on each bet
- For this we turn to the Kelly Criterion → this gives a formula for the optimal allocation of capital for a bet given a win-probability and the odds of the bet

$$f^* = p - \frac{q}{b}$$

- Where p is our win probability, q = 1 - p, and b is the odds on the bet
- For sports betting, estimating p with a high degree of certainty is difficult, and *overbetting* is far worse than *underbetting* using a Kelly approach
- Thus, we use *half Kelly betting*, allocating half of what is indicated by the Kelly Criterion

Linear Regression for Totals

Regressing on Box Score Statistics

- Our baseline model will be a multiple linear regression model regressing the total points on box score statistics for each team
- Each team's statistics are aggregated and exponentially smoothed to capture the notion that recent games are more relevant
- R-squared: 0.2202, Mean Absolute Error (MAE): 14.3608 - not sufficient to beat the sportsbooks

Adding Betting Lines to the Regression

- To improve the model we include the opening total line from the sportsbook in the regression
- This should allow us to identify which lines are inconsistent with the box score statistics, indicating an exploitable/inefficient line
- R-squared: 0.2626, MAE: 13.95 - not sufficient to beat the sportsbooks

An Effective Betting Strategy: Margin Betting

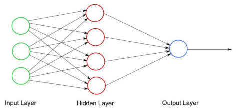
- Margin betting - define some lower bound ℓ and only bet when our model's prediction differs from the betting line by at least ℓ
- After testing different values for ℓ through cross validation we found that ℓ = 2 is a good value for the linear regression model
- This strategy was tested by simulating bets using 5-fold cross validation. The results of this test simulation are shown:

Win Rate	Bankroll Growth
0.5599	74.89%
0.5147	-33.00%
0.5514	68.16%
0.5445	28.83%
0.5155	-30.43%

A Neural Network Approach

Network Architecture

- Neural Networks are more flexible models than can hopefully capture more complex, nonlinear relationships between point totals and box score statistics/betting lines
- We build a neural network for regression, using the same explanatory variables as the linear regression model as inputs



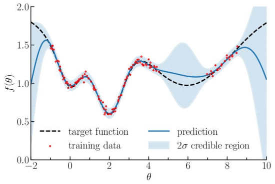
- Our network: 20+ inputs, 2 hidden layers (64, 32 units respectively), ReLu activation, MSE loss

Results

- MAE converges to 13.5-13.9, similar to the linear regression model
- Surprisingly, simulated betting strategies using our "margin betting" approach perform slightly worse using the neural network compared to our linear regression model

Gaussian Process Regression

- Gaussian Processes (GP) allow us to obtain an estimate of uncertainty as well as a prediction
- We implement KNN GP Regression for improved data efficiency
- Cross validation yielded k = 640
- MAE: 13.98, again similar to previous models
- Erratic results for simulated betting - fluctuates between very high and very low win rates
- Hyperparameters are extremely influential in GP models, particularly for the kernel - suboptimal hyperparameter selection could be creating bias in the model and contributing to erratic behavior
- The problem of hyperparameter selection should be examined closely in further research



Brownian Motion for Halftime Point Spreads

Model Description

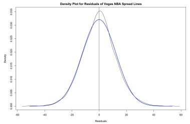
- Previous work (Polson and Stern, 2015) models the point spread of NBA games as Brownian motion and develops the notion of implied volatility of a game
- The probability of team A winning is then a function of its estimated advantage over team B (given by pregame betting markets) as well as the implied volatility of the game
- This can be extended to model the probability of team A *beating the spread*
- In particular, we optimize this model to estimate the probability of team A beating the *halftime spread*, given the scores of the first two quarters
- Treated as a classification task - cross entropy loss function, optimization through backpropagation/SGD

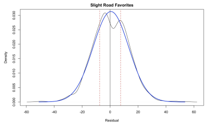
Results

- General win rate of 50-53%, margin betting win rate of 52-56%
- Major drawback - lack of first half box score statistics - this could significantly improve our model

Point Spread Extremes Betting

- Betting on *alternate lines*: original spread ±7.5, but with significantly improved odds
- We can use a normal distribution to approximate the distribution of the true margin of victory around the pregame spread
- Below we show the overall density of these spread deviations, as well as a specific case for when the road team is a slight favorite in the pregame betting lines





- Using these normal approximations we can estimate the probability that the spread of a game will land above or below that indicated by the sports book, and therefore we can identify profitable bets

Conclusion

- From our analysis we found that the NBA betting lines set by Vegas oddsmakers are generally very efficient and difficult to beat
- While no guaranteed winning formula was found, a number of models/strategies yielded promising results, and with further research, optimization, and improved data it seems feasible that a winning system could be produced

Citations

- Stern, H.S. and Polson, N.G. 2015. "*The implied volatility of a sports game.*" Journal of Quantitative Analysis in Sports, Volume 11, Issue 3, Pages 145-153
- Rosenfeld, J.W. 2012. "*An In-Game Win Probability Model for the NBA.*" Thesis, Harvard University
- Kelly, J.L.1956. "*A New Interpretation of Information Rule.*" Bell System Technical Journal. 35.