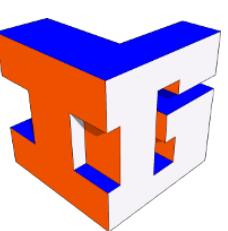


PREDICTABILITY IN COLLEGE SPORTS:

COMPARING THE ACCURACIES OF PREDICTION MODELS FOR COLLEGE FOOTBALL AND BASKETBALL GAMES



IML SCHOLARS: YIHAN GAO, SAM LAM, DAVID XIA

MENTORS: PROF. A.J. HILDEBRAND, BINGYAN LIU

ILLINOIS MATHEMATICS LAB, UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

INTRODUCTION AND MOTIVATION

College football and basketball are a huge deal. The recent mens' and womens' NCAA basketball championship games both drew well over 10 million viewers. Broadcasting rights deals for college sports approach \$1 billion, and billions of dollars are bet each year on college football and basketball. In our project, we focus on questions such as the following:

- How effective is the betting market in predicting the outcomes of games compared to more naïve prediction methods?
- How does the predictability of college football compare to that of college basketball?
- How much of an advantage is it to be the home team?
- How do we measure and compare the accuracies of probabilistic predictions?

DATA SOURCES

- **College Football Data (CFBD)**, collegefootballdata.com: Contains game results, Elo ratings, betting market data and poll rankings for college football games going back to 1980.
- **College Poll Archive**, collegepollarchive.com: Contains every AP Top 25 Poll for college football and basketball going back to 1949.
- **Sportsbook Reviews Online**, sportsbookreviewsonline.com: Contains college football and basketball game result data, as well as betting data, going back to 2007.
- **Sports Reference**, sports-reference.com: Contains game result data for college football and basketball, going back to 1950.

PREDICTION METHODS

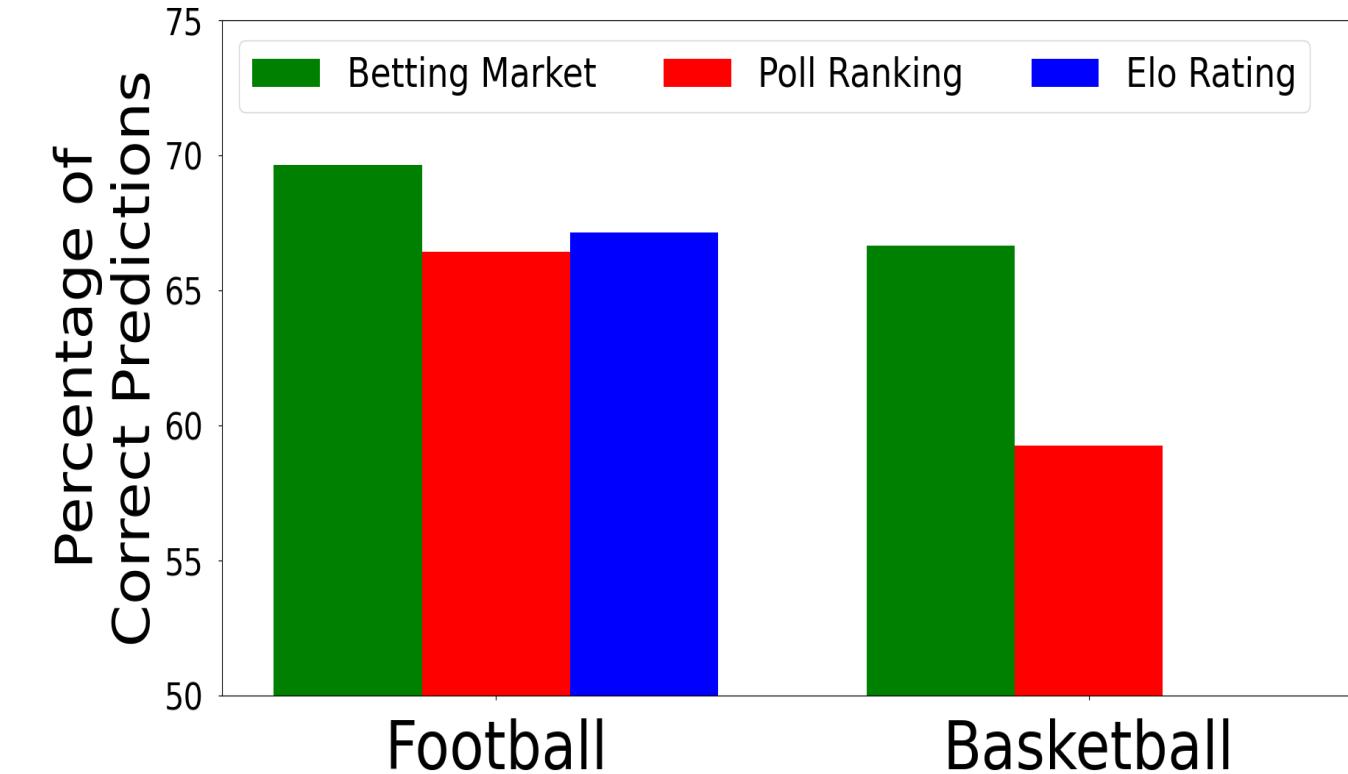
We analyzed 3 ways to predict the outcome of a game:

- **Betting Market [1,4]**: The team favored to win by the betting market is predicted to win. We use moneyline scores as a proxy for betting market predictions. Moneyline scores come in pairs such as (+110, -135) that correlate to the amount of money won or lost by betting on each of the teams. The team with the negative moneyline score is the betting market favorite.
- **Poll Based Predictions [1,2,5]**: The AP Top 25 Poll is a weekly poll ranking college sports teams, which aggregates the individual rankings by approximately 60 sports journalists. The higher ranked team is predicted to win.
- **Elo Ratings [1,3]**: A team's Elo rating is a point system updated after every game depending on the strength of its opponent (1500 is an average team, 2000 represents a very good team). It was originally developed to rate chess players. The team with the higher Elo rating is predicted to win.

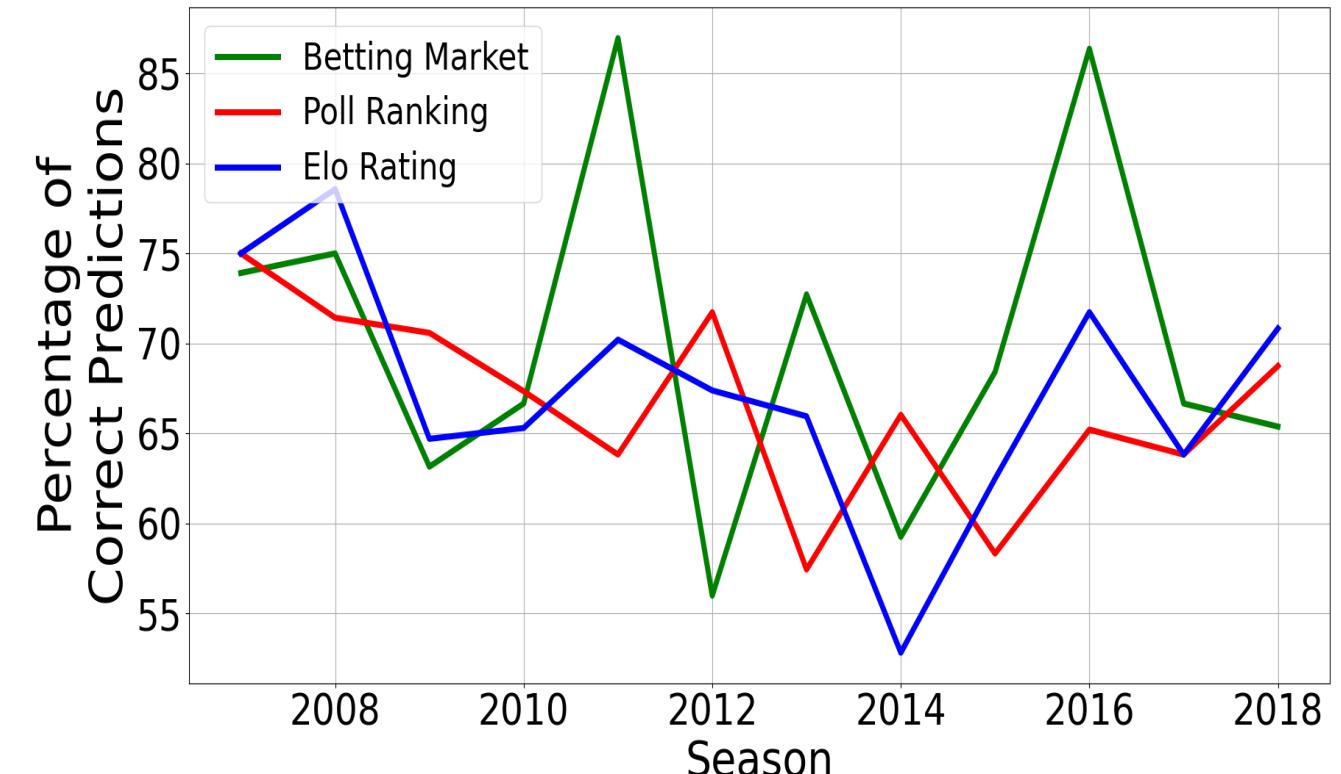
Elo ratings were obtained from collegefootballdata.com (college football only).

PREDICTION ACCURACIES IN COLLEGE FOOTBALL AND COLLEGE BASKETBALL

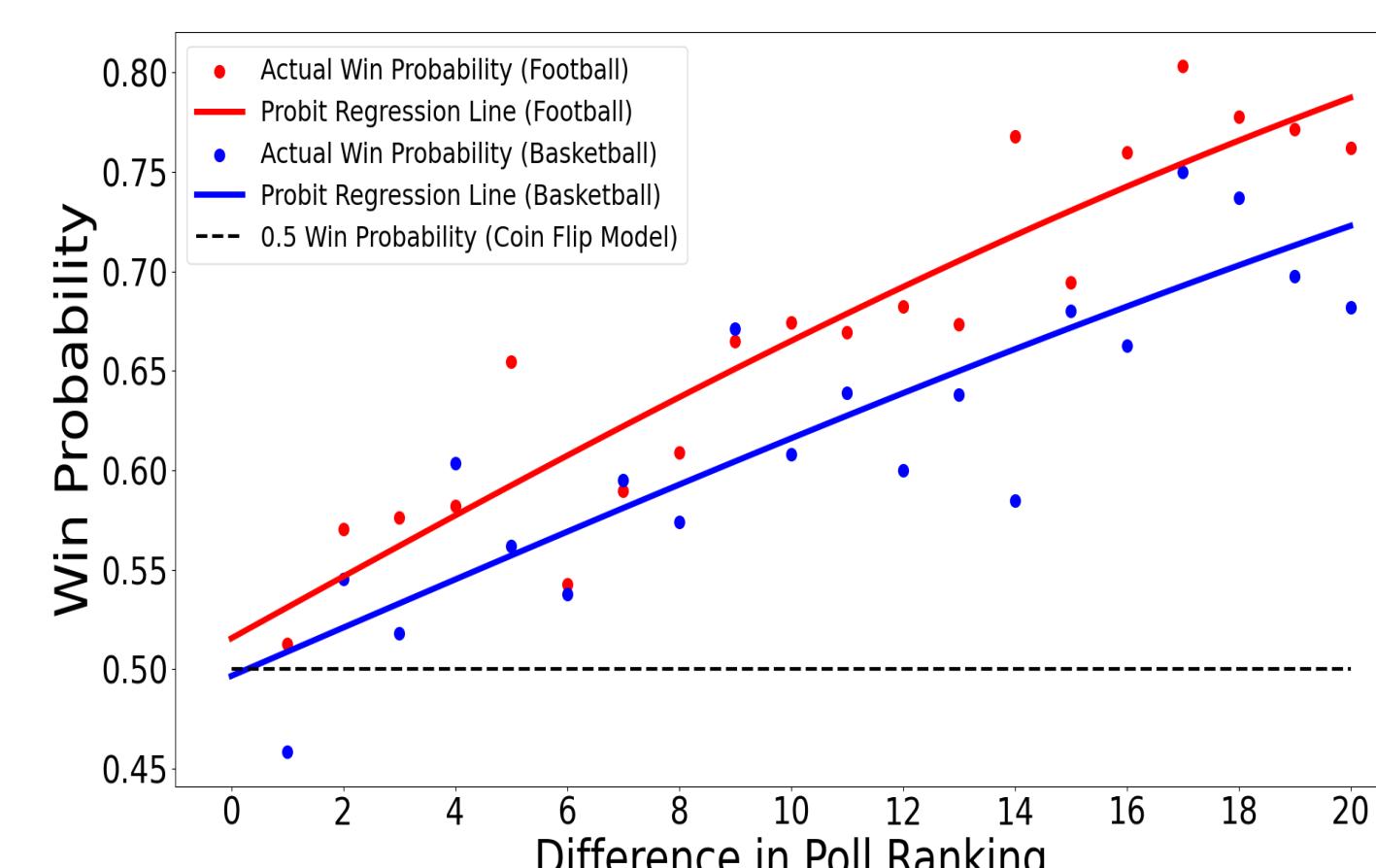
Accuracies by Method and Sport



Yearly Variations in Accuracies (College Football)



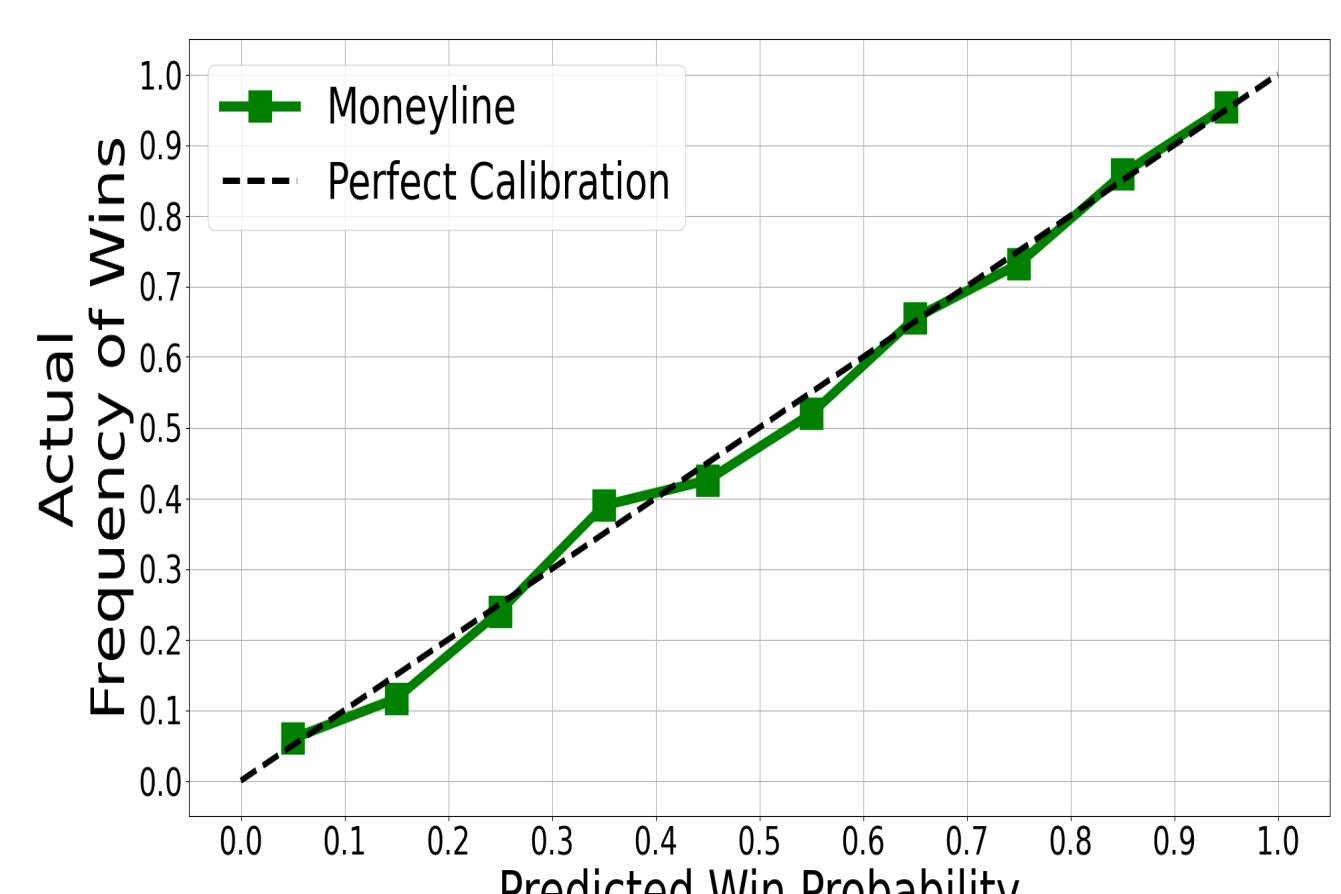
Prediction Accuracy vs Rank Difference



Proportion of games won by higher ranked team in AP Top 25 matchups in college football (1936 to present) and college basketball (1950 to present).

- There is significant variation in the win probabilities by rank difference. This is likely due to the small sample sizes of games (for large rank differences, the sample sizes were as small as 21 games).
- The regression lines show that the accuracy of the predictions grows approximately linearly as a function of the rank difference, and is higher for college football than college basketball.

Accuracy of Probabilistic Predictions



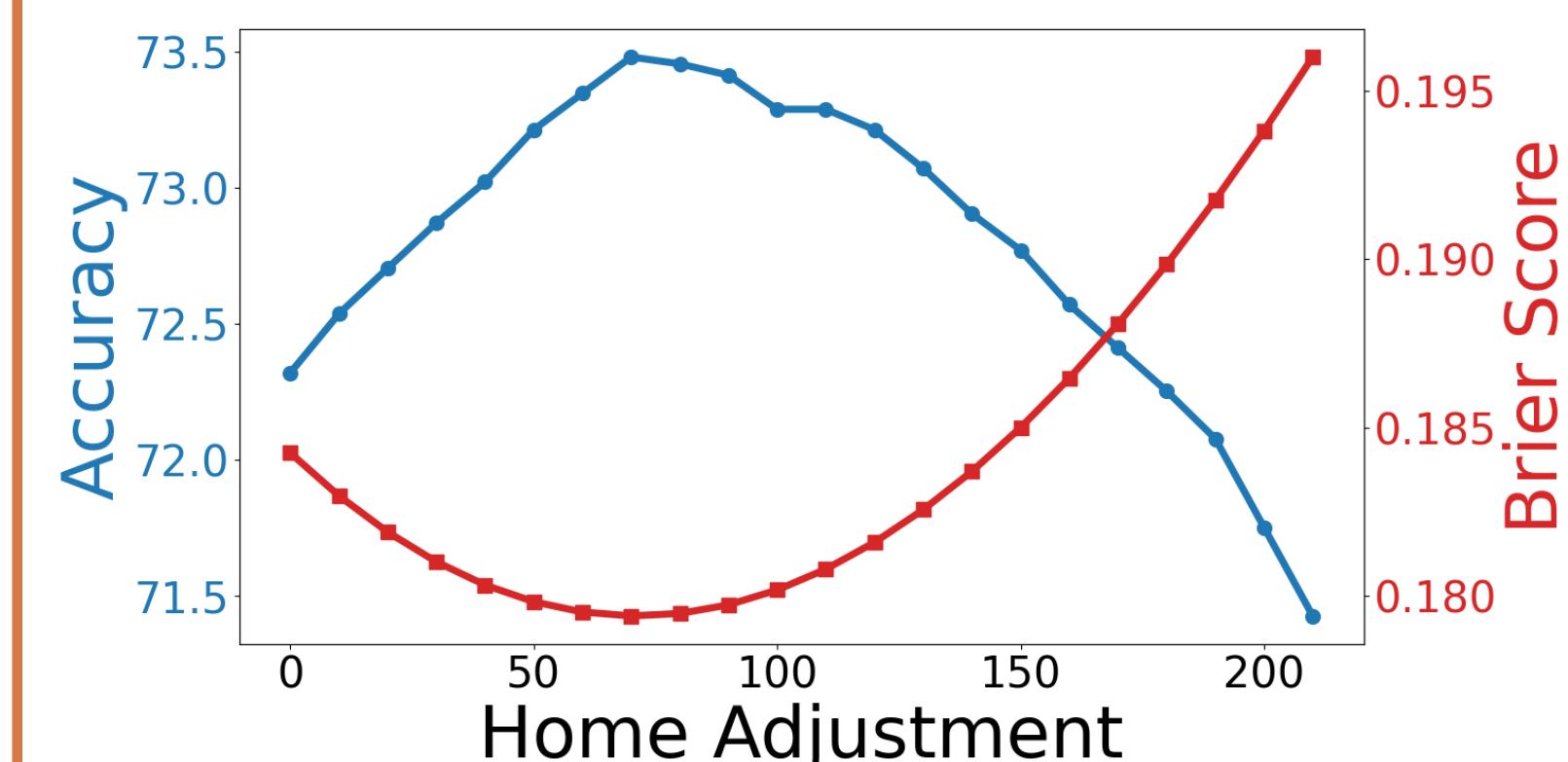
Actual proportion of games won vs implied win probabilities based on Elo ratings and moneyline scores.

- Elo ratings and moneyline scores are converted to win probabilities, which are compared with actual win frequencies using bins of width 0.1.
- The dotted line represents a perfect calibration of the probabilistic predictions in which the actual proportion of games won is equal to the predicted win probability. The graph shows that both Elo and moneyline probabilistic predictions are very close to perfect calibration.

INCORPORATING HOME TEAM ADVANTAGE

Of the three prediction methods we analyzed, only the betting market accounts for a possible home team advantage, which may explain in part the higher accuracy of predictions based on the betting market.

It's natural to ask if other methods like Elo rating can be improved by giving the home team an appropriate Elo rating boost. The graph below shows the improvement in prediction accuracy and Brier score (which measures the accuracy of probabilistic predictions – the lower the score the better it is) as a function of the Elo rating boost. We found the optimal home team boost is 73 Elo rating points.



EVALUATING PROBABILISTIC PREDICTIONS

We use two approaches to measure and compare the accuracies of probabilistic predictions.

- **Calibration Plot**: Plots the actual proportion of games won against the predicted win probability. (See graph on left for an example).
- **Brier Score**: Numerical quantity representing the average deviation between predicted and actual probability. The lower the Brier score, the more accurate the probabilistic prediction.

REFERENCES

1. College Football Data, <https://www.cfbdb.com>
2. College Poll Archive, <https://www.collegepollarchive.com>
3. Langville, A. N., Meyer, C. D. (2012). "Who's # 1? The science of rating and ranking", Princeton University Press.
4. Sportsbook Reviews Online, <https://www.sportsbookreviewsonline.com>
5. Sports Reference, <https://www.sports-reference.com>
6. Stern, H. S. (1997). "How accurately can sports outcomes be predicted?". *Chance* 10, 19-23.