DECISION SCIENCES INSTITUTE

Is the Football Power Index a Good Bet?

ABSTRACT

Sports analytics have become ubiquitous in print, online and on television. Major media organizations with large, well-funded analytics departments have developed and promoted a type of predictive model known as a Football Power Index. In this paper we examine 3 models, the ESPN FPI, and two versions of the FiveThirtyEight.com Elo model. We discuss how they work, how accurate they are, and how they fare relative to sport betting markets. We also investigate what they tell us about semi-strong efficiency in sports betting markets.

KEYWORDS: Market Efficiency, Sports Analytics, Sports Betting

INTRODUCTION

Sports analytics have become ubiquitous in print, online and on television. Media outlets such as ESPN and Five Thirty Eight have a large analytics staff devoted to developing, refining, and communicating analytics-based predictions. ESPN now regularly provides a win probability for every game in major sports on their chyron generally labelled *according to ESPN Analytics*. The popular website FiveThirtyEight.com, well known for predictive modelling in politics as well as sports, publishes several models that provide forecasted win probabilities in major sports for every game.

Both the ESPN model and the FiveThirtyEight models are variations of a Power Index model. While each model has unique features, they are very similar in structure. Each team has a power ranking that indicates its overall capability. The probability that a team defeats another team is based on the relative values of their power rankings. Based on the outcome of a game power ranking points are taken from the loser and given to the winner, thus over the course of the season power ranking points migrate from poor teams to good teams.

Given the significant level of resources these organizations devote to their modeling it is reasonable to ask how well these model predictions perform. In this paper we will restrict ourselves to the National Football League (NFL). We will review the ESPN Football Power Index (ESPN FPI) and the two FiveThirtyEight Elo based models, the Elo model and the QB Elo model. We will discuss the concepts behind the models and review, to the extent possible, how they work. We will then evaluate the accuracy of these models in a probabilistic sense, comparing them to actual results and each other. Finally, we will compare these models to the probabilistic forecast provided by betting odds. We will test the model's ability to generate positive betting returns by following their recommendations, and in the process test if these models threaten the semi-strong version of the efficient market hypothesis as applied to the betting markets. We will then discuss what, if anything, they tell us about semi-strong efficiency overall.

LITERATURE REVIEW

With the increased application of analytics to various fields of study there is a reasonably large body of research on the use of models to predict sports outcomes. Horvat and Job (2020) provide a detailed review of the application of machine learning to sports prediction. Their analysis covers over 100 papers that cover multiple sports from Cricket to American Football. Some papers address models across sports (Bunker & Thabtah, 2019) while other papers address individual sports such as basketball (Huang & Lin, 2020; Osken & Onay, 2022; Zhao, Du, & Tan, 2023), hockey (Pischedda, 2014), baseball (Huang & Li, 2021; Soto-Valero, 2016), soccer (Dandıl & Ergez, 2018; Hassan, Akl, Hassan, & Sunderland, 2020; Mattera, 2021), collegiate wrestling (Bigsby & Ohlmann, 2017), golf (Chae, Park, & So, 2021), rugby (M. Bennett, Bezodis, Shearer, & Kilduff, 2020) cricket (Balasundaram, Dhandapani, & Ashokkumar, 2022; Patil, Duraphe, Motarwar, Suganya, & Mariappan, 2023; Shakil, Abdullah, Momen, & Mohammed, 2020) and tennis (Wilkens, 2021).

Some papers explicitly consider betting in certain sports such as college football (R. Bennett, 2020; Coleman, 2015) Predictive analytics are very popular in the NFL, with outsiders and insiders. The NFL has itself partnered with organizations such as Amazon Web Services (AWS) to exploit machine learning technology (Learning, 2021). Predictive models for in game win probabilitiy assessment and player evaluation have been published in the lierature and made available as libraries for the popular statistical package R (Yurko, Ventura, & Horowitz, 2019).

<u>Capital Market Efficiency</u>

Our analysis includes an assessment of the efficiency of the sports betting markets. Market Efficiency is a concept first developed in the economics and finance literature as defined by the efficient market hypothesis. (EMH). The efficient-market hypothesis (EMH) is a hypothesis in financial economics that states that asset prices *fully reflect* all available information. A direct implication is that it is impossible to "beat the market" consistently on a risk-adjusted basis since market prices should only react to new information (Wikipedia, 2022).

A key review of the theoretical and empirical literature on the Empirical Market Hypothesis is provided in Fama (1970). Fama analyzes efficient markets relative to three board information sets. Weak form efficiency is based on the use of historical prices, semi-strong efficiency on publicly available information, and strong efficiency is based on all information public and private. In this framework the possibility of trading systems based on the relevant information set generating excess returns, that is returns in excess of equilibrium expected profits, are ruled out. Simply put, excess profits are not possible from a trading system in a market that is efficient. The empirical analysis in this paper provides reasonably strong support for the weak, and semi-strong forms of market efficiency. The analysis, however, identified exceptions to the strong form of market efficiency whereby market makers and corporate insiders can exploit their monopolistic access to information in order to earn returns in excess of the expected risk-adjusted rate.

<u>Sports Betting Market Efficiency</u>

A large body of research published beginning in the 1980s and 90s examined efficiency in sports betting markets. Much of this early research focused on pari-mutual and fixed odds systems in horse racing. A comprehensive review of this literature is provided in (Kuypers, 2000). Kuypers reviews five papers that examine pari-mutual systems, six papers on odds-based systems, and four papers on spread based systems. Seven of the reviewed papers assess weak form efficiency, while six examine semi-strong efficiency and two assess strong-form efficiency.

Kuypers uses the following definitions of efficiency in the sports betting context:

- Weak form: no abnormal returns, either to the bookmaker or the bettor, can be achieved solely from price information. An abnormal return is defined as a return different from the bookmaker's expected take.
- Semi-strong: no abnormal returns can be achieved from odds or any publicly available information.
- Strong: no abnormal returns can be achieved by any group in society incorporating odds publicly available and privately available information.

A reasonably large body of literature examines various aspects of betting efficiency across different sports and different betting markets. The majority of these papers focus on weak form efficiency by testing startaegies based on polint spreads. Gray and Gray (1997) examine efficiency in NFL point-spread markets using a probit model approach to see if the spread is an unbiased predictor. They test various betting strategies including betting on teams where their model forecasts a greater than 52.4% win probability; the implied win probability in most point-spread bets. Efficiency tests for NFL betting have been the subject of several economics thesis projects (Anderson, 2019; Hetherington, 2006; Williams, 2021) Hetherington (2006) looks at NFL prediction markets with a behavioral finance perspective. Stetzka and Winter (2021) investigate the rationality of gambling overall. Anderson (2019) uses OLS regression to evaluate various strategies for betting against the spread in NFL games. (Williams, 2021) uses an econometric approach to analyze rationality and efficiency. The analysis finds evidence of bias but cannot show any market inefficiency.

Robbins (2023) examines weak form efficiency across moneyline bets on multiple North American professional and collegiate sports. He finds technical violations of market efficiency including longshot biases and differential returns on various odds ranges, but nothing that leads to a long term consistent positive return. Arscott (2023) examines efficiency in college football betting.

Most of these analyses are focused on weak-form efficiency. In sports betting weak form efficiency implies that decision rules based simply on the stated odds will not earn a long-term profit, and all odds ranges will earn a similar negative return. Semi-strong efficiency implies that profitable strategies cannot be developed based on publicly available information. One implication of semi-strong efficiency in sports betting is that an analytical model that uses publicly available information will not identify bets that generate a positive long-term return.

BETTING ODDS AND PROBABILITIES

The menu of bets that can be made on sports is very large. Bets can be placed on almost anything related to a game, a team, or even individual performances of players. With minor variations from sport to sport, the main betting options have three different components: totals, spreads and moneyline.

- Totals: the total, or over/under, is a bet on the total points scored in the game. Bettors can bet the total points will be over, or under the stated line.
- Spreads: a bet on a team to win by a certain margin. The underdog is bet with plus points, the favored with negative points.
- Moneyline: a straight bet on what team will win the game. Moneyline bets are made with differential payouts such that a bet on a favorite will risk more than can be won, while a bet on an underdog will return more than the amount risked.

Note that both totals and spread bets are quoted along with moneyline odds so that the payout to a winner is less than the amount risked. Odds are stated in different equivalent formats in different locations and different settings. In the United States odds are most often quoted in American Odds format. Our analysis will focus on moneyline bets for the NFL.

In the American format the odds can be expressed as either a positive number or a negative number. A positive number shows the profit a successful wager will return on a $100 bet. For example, a bettor who wagers $100 at +110 odds and wins, will earn a profit of $110, plus the original wager of $100 for a total payout of $210. Positive odds typically imply the team is an underdog. Conversely, negative odds show how much a bettor must risk to earn a $100 profit. For example, if a bet is made for $120 at -120 odds, the successful bettor will receive a profit of $100, plus the original wager of $120 for a total payout of $220. The favorite team is given negative odds, but in some evenly matched games both teams may have negative odds. More formally the Payout *P* to a wager of stake *S*, at odds *M* are given by equation (1).

$$P = \begin{cases} S \times \dfrac{M}{100} + S & \text{for M} > 0 \\ \dfrac{S}{-M/100} + S & \text{for M} \leq 0 \end{cases} \tag{1}$$

Odds of +100- and -100 are equivalent. In practice *M* is always quoted as a number with an absolute value greater than or equal to 100. So, while odds of -125 and +80 would both return a profit of $80on a $100 bet, the odds are always quoted as -125.

Moneyline odds carry an implied probability of success. The implied probability is the probability at which a bettor is indifferent to taking either side of the bet. The probability calculation in the American odds format again depends on whether the odds are positive or negative. So, for a bet with odds *M*, the implied probability *p* is given by

$$p = \begin{cases} \dfrac{100}{M+100} & \text{for M} > 0 \\ \dfrac{-M}{-M+100} & \text{for M} \leq 0 \end{cases} \tag{2}$$

While equation (2) gives the odds on one side of a bet, the bookmaker quotes odds in pairs. For example, a bookmaker might quote odds of -120 for a favorite and +110 for the underdog. Converting each of these to implied probabilities gives probabilities of 54.5% and 47.6%. These odds are not fair, in the sense that they add up to more than 100%. The excess probability, in this example 2.1%, is the book margin (*k*), sometimes referred to as the *vig* or the *juice*. The book margin exists so that the bookmaker is guaranteed a profit as long as bets are made in the appropriate proportion. Book margins in the range of 3%-5% are common.

In order to convert the bookmaker's odds into meaningful probability estimates the odds must be converted to consistent probabilities. Draws are rare in football. So, if the contest ends in a draw all win-lose bets are effectively cancelled, and bettors are returned their original stake. The most common way to convert the implied probabilities is a simple normalization process. So, for a contest with implied probabilities of p_1 and p_2, the normalized probability that team 1 will win the game and bets will pay is

$$p_{1_n} = \frac{p_1}{p_1 + p_2} \tag{3}$$

The Sportsbook's Margin

Because the implied odds are unfair, they add up to more than one, the sportsbook has a built-in advantage. The excess probability gives the sportsbook a built-in margin, appropriately allocated bets on either side will guarantee the book a profit. The sportsbook's profit margin is proportional to the book sum, the excess implied probability in the stated odds. If we have a two-way bet with implied odds p_1 and p_2, then the book sum *k*, is given by

$$k = p_1 + p_2 - 1 \tag{4}$$

The bookmaker's margin (*m*), also known as the *hold*, is the sportsbook's average profit and can be shown to be

$$m = \frac{k}{k+1} \tag{5}$$

The Bettor's Decision

The bettor is presented with a series of potential bets and must decide which bets to accept. Assume for simplicity and without loss of generality that the odds for a particular bet are presented at positive American Odds *M* and the bettor will make a bet of one dollar.

The expected profit from the bettor's perspective is

$$E[P] = p_b \frac{M}{100} - (1 - p_b) \tag{6}$$

Where p_b is the bettor's subjective probability assessment that his bet will win. The bettor's expected profit is positive if

$$p_b > \frac{100}{M+100} \tag{7}$$

The right-hand side of equation (7) is the implied probability of the odds as expressed in equation (2). The decision of the bettor is then quite simple, accept bets where you assess the probability of victory to be greater than the probability implied by the odds. It is important to note that does not mean the bettor is simply choosing sides on every bet. Because the bookmaker is offering unfair odds there will be a number of bets where neither side is assessed as a positive EV bet. This is especially true if the bettor's odds assessments and those of the market, as expressed by the bookmaker are similar to each other.

FOOTBALL POWER INDEX

The ESPN predictive model is known as the ESPN Football Power Index (ESPN FPI). The ESPN FPI is a rating system developed by ESPN that measures team strength and uses it to forecast game and season results (ESPN). The FPI gives each team a rating number that indicates their strength. The relative strengths of two teams competing against each other are then used to determine the win probability for each team. Based on the results of each game team strength ratings are updated.

The models used by FiveThirtyEight.com are methodologically very similar to ESPN's model. But while ESPN treats the details of their model as proprietary, FiveThirtyEight is generally fairly transparent about how their models work (Silver, 2022). FiveThirtyEight introduced their model, which they called NFL Elo Ratings in 2014 (Silver, 2014) and modified the approach somewhat in 2015 (Paine, 2015). Starting in 2019 FiveThirtyEight introduced a second model similar to the original along with a quarterback adjustment, the QB Elo model. FiveThirtyEight describes their models at a reasonable level of detail and provides access to the data used to in the model on Github (FiveThirtyEight.com). The FiveThirtyEight models are variations on the Elo model, originally developed by Arpad Elo to rank chess players (Wikipdeia). The application of Elo models to sports is described in Aldous (2017).

FiveThirtyEight Elo Model

The basic Elo model was developed by Arpad Elo as a tool for ranking chess players, but has since been adapted to sports, including the NFL. The Elo model assigns each team a numerical ranking. In the NFL the Elo ranking is scaled so that the average team has a ranking of 1505 points. For the 2022 season ranking ranged from a low of 1291 to a high of 1719.

The forecast for each game is expressed as a win probability. The base probability that team *A*, with Elo score R_A will defeat team *B*, with Elo score R_B is given as

$$\Pr(A) = \frac{1}{10^{\frac{R_B - R_A}{400}} + 1} \tag{8}$$

The pre-game Elo point differential is adjusted for several factors including a home field advantage of 4 points per thousand miles travelled buy the visitor, and a 25-point rest adjustment for teams coming off a bye week. In the QB Elo model an additional adjustment is made based on the quarterback.

After the game Elo points are adjusted based on the outcome, in a zero-sum exchange where the winner takes points from the loser. The base level of points transferred is derived based on how likely the team was to win based on the pregame Elo ratings. If E_A is the probability team *A* would win, S_A is 1 if team *A* won, 0 otherwise and *K* is a scaling constant, Then the new Elo rating for team *A* is

$$R_A^{'} = R_A + K(S_A - E_A) \tag{9}$$

K is a factor that determines how quickly ratings react to new results. The larger the *K* value the more points shift. The *K* value for football is approximately 20. Sports with more games, where each individual game has less impact, like baseball or basketball, will have smaller *K* values.

Elo scores are also impacted by the margin of victory (winner's points-loser's points). The effect is implemented as a multiplier. The larger the margin, the larger the multiplier. But the number of points has a declining marginal benefit.

$$MovMultiplier = \ln(WinnerPointDiff + 1) \times \frac{2.2}{WinnerELODiff \times 0.001 + 2.2} \tag{10}$$

Overall, the shift in Elo points per game is relatively small. During the 2022 NFL season the median point shift was 17.13 points.

The ESPN FPI Model

ESPN is significantly less transparent about their model than FiveThirtyEight, providing only general descriptions of the model and the model results. They recently published an overview of the model along with pre-season ratings for the 2023 season (Walder, 2023). Like the FiveThirtyEight model the FPI gives each team a Power Ranking Score. Win probabilities are calculated based on the difference in ranking score. Team rankings are updated during the season based on opponent score, game location, and point differential. The FPI model differs from FiveThirtyEight in that it accounts for the separate impact of offense, defense, and special teams based on the Expected Points Added (EPA) metric. The ESPN model also differs in how it determines pre-season rankings, with the ESPN model utilizing betting odds as an initial factor.

MODEL ACCURACY

The most basic assessment of a predictive model such as the FPI models is how often they are right, i.e., what proportion of the teams the model favors to win actually end up winning. A standard baseline for comparison is the coin-flip which would over the long term be accurate 50% of the time. However, winning percentages in most major sports, including football, are biased in favor of the home team. The home team in the NFL wins about 55% of the time, so simply betting on the home team will result in roughly a 55% win rate. But recall that since bets are not even money, and the home field advantage is built into the odds, winning 55% will not necessarily provide a profit. We will return to this issue in the next section, but for now we compare the winning percentage of the three models with the home field winning percentage. We also consider a second benchmark, the winning percentage of the team favored by the betting odds. Those complete results are presented in Table 1.

Table 1-Winning Percentages

Accuracy Picking Game Winner						
Season	Games	Home	ELO	QBELO	FPI	Odds
2014	267	57.5%	69.2%	-	-	66.2%
2015	267	53.2%	65.9%	-	62.5%	63.3%
2016	267	58.5%	64.5%	-	62.3%	65.3%
2017	267	56.9%	65.2%	-	67.2%	68.9%
2018	267	59.6%	62.6%	-	64.2%	64.5%
2019	267	51.9%	63.9%	64.3%	62.4%	64.3%
2020	269	49.6%	65.3%	68.3%	66.7%	67.2%
2021	284	50.9%	59.0%	60.4%	60.1%	62.2%
2022	271	55.4%	62.5%	62.1%	59.9%	64.7%
Total	2,426	54.8%	64.2%	63.7%	63.1%	65.1%

The home team won a majority of games in every year except 2020, which was the Covid year where attendance restrictions muted the home field advantage. Each of the three models consistently predicted winners with a higher accuracy rate than the naïve home field rule and often by reasonably large margins. The models were, however, generally less accurate than the betting odds. The Elo model outperformed the betting odds in its first two years, but not since. Similarly, the QB Elo model did as well or better than the betting odds in its first two years, but not since. In the aggregate none of the models has been as accurate as the betting odds.

Accuracy is an interesting, but insufficient metric for evaluating the betting utility of the models. It depends how the model does at various odds levels. To begin assessing this we present calibration graphs. To create the calibration graph we divide the predictions into quintiles, twenty evenly sized groups based on the win probability implied by the model. For each quintile we calculate the average implied win probability and the actual win probability. The graph shows these odds in a scatter plot format. If the odds are well calibrated, teams with a 30%-win probability will win 30% of the time and the points on the calibration chart will align on the diagonal.

Figure 1 is the calibration plot for the base Elo model. The model appears to be reasonably well calibrated as none of the points diverge from the diagonal by much. However, the model does appear to exhibit some bias. Of the 10 points with forecasted probabilities less than 50%, 8 are above the line. Conversely points with win probabilities above 50% are generally below the line. The implication is that the Elo forecast is biased. It tends to undervalue underdogs, that is give them a higher lower probability than they achieve and overvalue favorites. Figure 2 shows the corresponding graph for the QB Elo model. The points here diverge from the line more than the Elo model, and the same general bias is evident; underdogs mostly overperform and favorites mostly underperform.   Finally, Figure 3 has the calibration chart for the ESPN FPI model. It exhibits a similar pattern to the Elo models.

Figure 1- Elo Win Probability Calibration

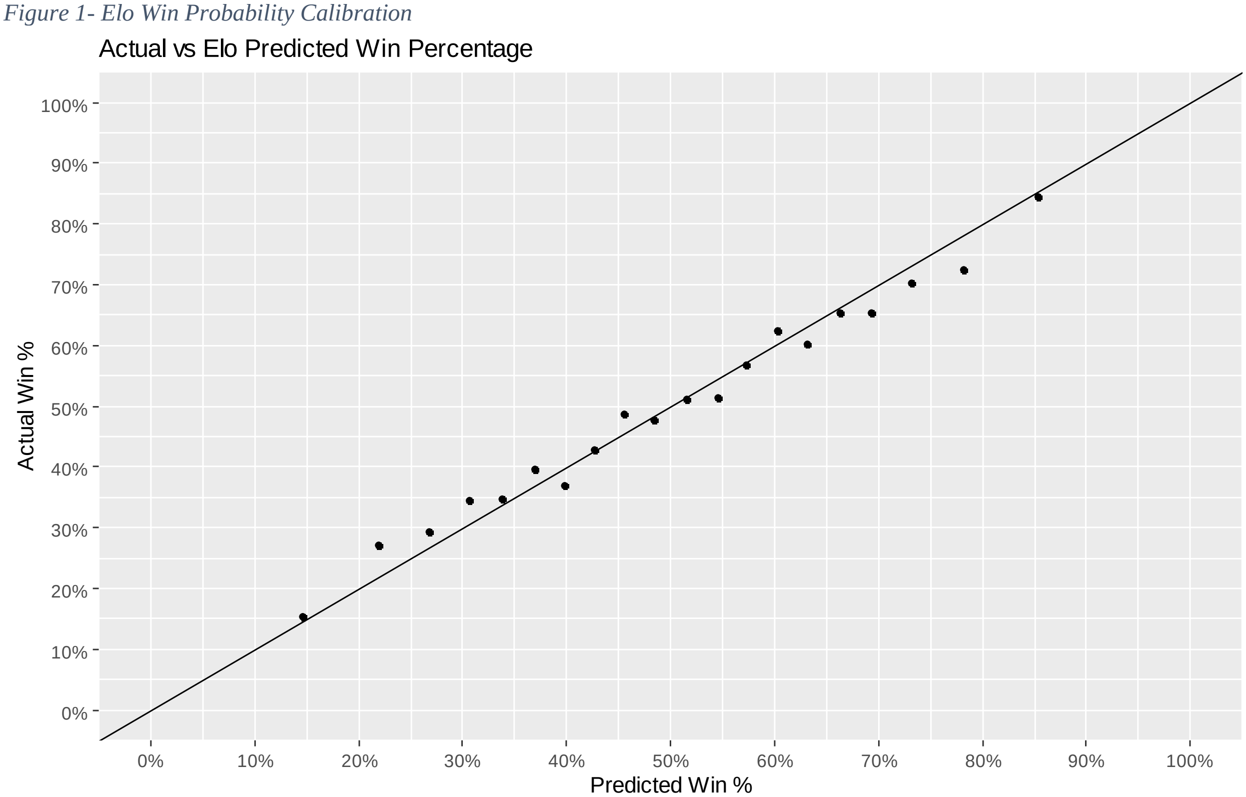


Figure 2-QB Elo Win Probability Calibration

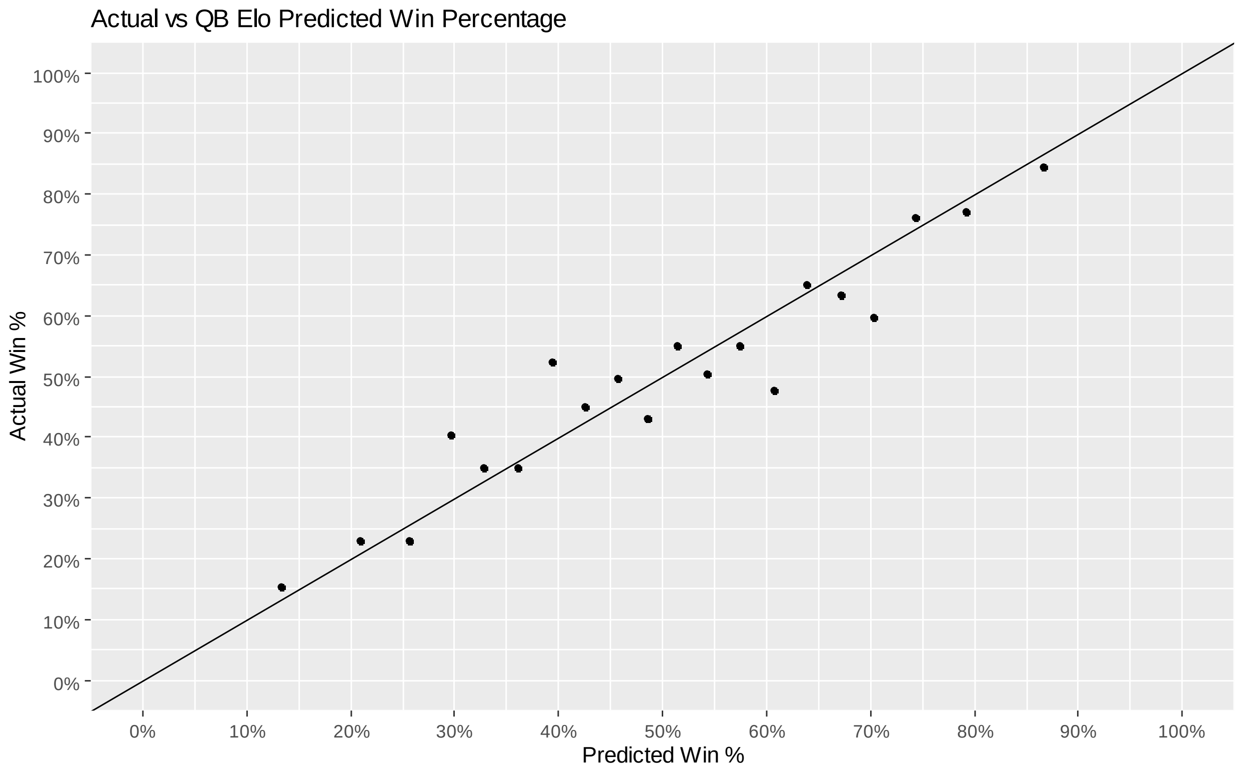
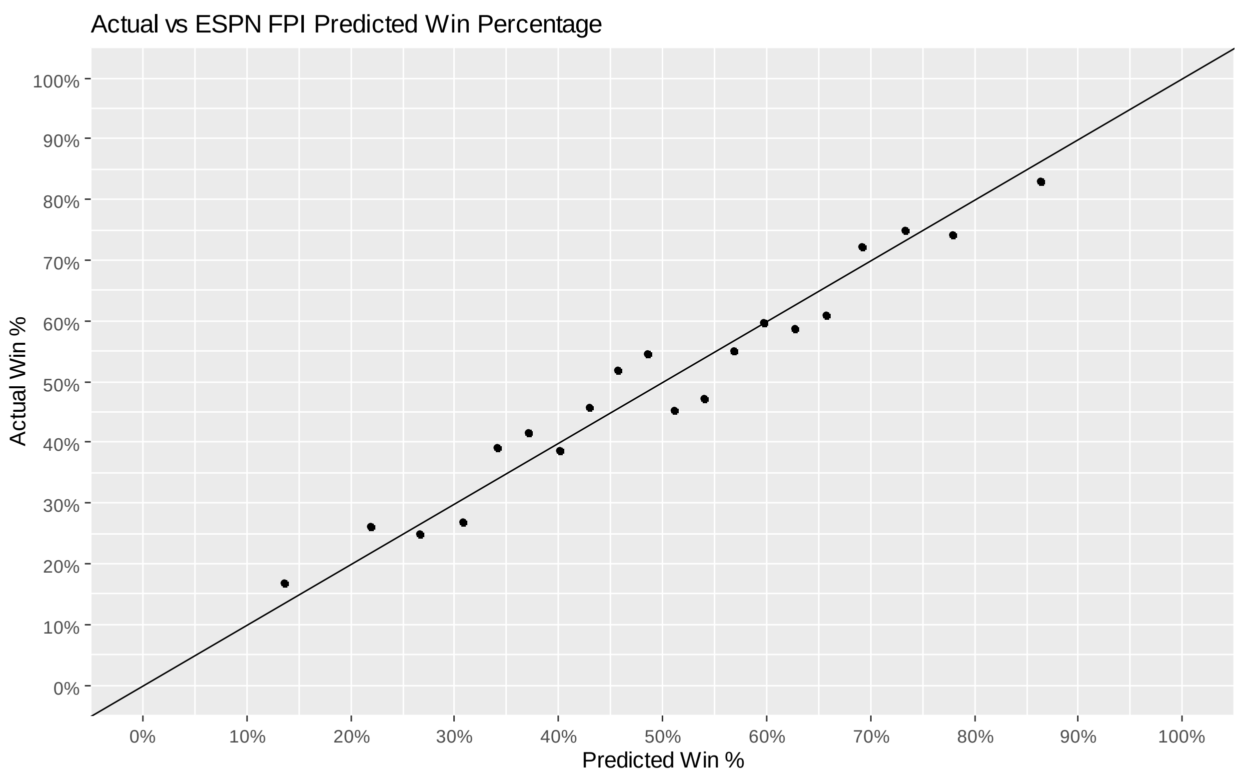


Figure 3-ESPN FPI Win Probability Calibration



The graphical information presented in these figures is summarized numerically in Table 2. From here we can verify that of the ten underdog bins the actual win rate was higher in eight bins for the Elo model, 7 for the AB Elo model, and seven for the ESPN FPI model. This confirms the observation that these models are biased in the sense that they undervalue the under dogs.

To get a sense of the overall accuracy of the probabilistic predictions we calculate the Brier score and show the results by season in Table 3. The Brier score is a metric commonly used to evaluate the accuracy of classification models. For a unidimensional model such as a win-loss model the Brier score is equivalent to the mean squared error of the prediction. The Brier score (B) is given by equation (11), where *f* is the forecasted probability of victory and *o* is an indicator variable that is 1 if the team wins and 0 otherwise, and *N* is the number of predictions made, which is two times the number of games.

$$B = \frac{1}{N}\sum_{i=1}^{n}(f_t - o_t)^2 \qquad (11)$$

A naïve predictor that gives each team a 50% chance of winning will have a Brier score of .25. A perfect Brier score of 0 is possible, but only if the model assigns 100% probability to every prediction and is always correct. So, practically we would hope to see Brier scores below .25. Each of the models outperform the naïve forecast and register Brier scores in the .207 to .235 range. The best performance was the Elo model in its first year, but that performance tended to decline over time rising from .207 in 2014 to .235 in 2022. In the most recent 4 years when all models have been making projections, the ESPN model had the best score in two years (2019 and 2021) and the worst score in 2022. The enhanced Elo model, the QB Elo, outperformed the base Elo model in all 4 years.

Table 2- Win Probabilities

Forecasted and Actual Win Rates by Quintile

For Games with Predictions

	538 Elo			538 QB Elo			ESPN FPI		
Bin	ELO_N	ELO_PWP	ELO_AWP	QBELO_N	QBELO_PWP	QBELO_AWP	ESPN_N	ESPN_PWP	ESPN_AWP
1	243	14.7%	15.2%	110	13.4%	15.5%	215	13.7%	16.7%
2	243	21.9%	27.2%	109	20.9%	22.9%	215	22.0%	26.0%
3	242	26.8%	29.3%	109	25.6%	22.9%	222	26.6%	24.8%
4	243	30.7%	34.6%	109	29.7%	40.4%	213	30.8%	26.8%
5	242	33.7%	34.7%	109	32.8%	34.9%	215	34.0%	39.1%
6	243	36.9%	39.5%	109	36.1%	34.9%	212	37.1%	41.5%
7	242	39.8%	36.8%	109	39.3%	52.3%	218	40.1%	38.5%
8	243	42.7%	42.8%	109	42.5%	45.0%	212	42.9%	45.8%
9	242	45.5%	48.8%	109	45.7%	49.5%	214	45.7%	51.9%
10	243	48.4%	47.7%	109	48.6%	43.1%	218	48.5%	54.6%
11	243	51.6%	51.0%	109	51.4%	55.0%	217	51.2%	45.2%
12	242	54.5%	51.2%	109	54.3%	50.5%	212	54.0%	47.2%
13	243	57.3%	56.8%	109	57.5%	55.0%	218	56.8%	55.0%
14	242	60.2%	62.4%	109	60.7%	47.7%	216	59.7%	59.7%
15	243	63.1%	60.1%	109	63.9%	65.1%	211	62.7%	58.8%
16	242	66.3%	65.3%	109	67.2%	63.3%	213	65.8%	61.0%
17	243	69.3%	65.4%	109	70.3%	59.6%	216	69.1%	72.2%
18	242	73.2%	70.2%	109	74.4%	76.1%	214	73.3%	74.8%
19	243	78.1%	72.4%	109	79.1%	77.1%	217	77.9%	74.2%
20	243	85.3%	84.4%	110	86.6%	84.5%	212	86.3%	83.0%

Table 3- Brier Scores

Season by Season Brier Scores
For Games with Predictions

Season	ELOBrier	QBELOBrier	ESPNBrier
2014	0.207	-	-
2015	0.225	-	0.227
2016	0.219	-	0.221
2017	0.217	-	0.215
2018	0.223	-	0.210
2019	0.225	0.220	0.218
2020	0.219	0.209	0.210
2021	0.235	0.232	0.229
2022	0.226	0.221	0.227

MODEL PREDICTIONS VS BETTING ODDS

The 3 forecasting models outperform the naïve 50-50 model as well as the slightly less naïve model of picking the home team, but how do they do as compared to the betting markets. To evaluate this, we implement a simple betting rule, bet all teams where the model indicates a win probability in excess of the implied probability of the betting odds. This is the definition of a positive expected value bet we outlined in equation (7).

It is important to note that this decision rule will not imply a bet on every game. Because the bookmaker is offering unfair odds, the model must forecast a win probability a few points higher than the betting odds. Stated another way, if the betting odds and the forecasting model assess the same win probability our decision rule would not bet either side.

As a baseline we also test the naïve approach of betting on the home team when odds are available. The results of this analysis are presented in Table 4. This analysis assumes that for each game where the decision rule indicates a bet that $100 is wagered. For each category the table provides the aggregate dollar impact with the average profit per bet. Since each bet is $100, the average represents the percentage return to bets.

Table 4- Results from Rule Based Betting

Profit from Picking Based on Model Forecasts												
Season	Bets	TotalProfit	HomeAvgProfit	ELOBets	TotalELOProfit	ELOPickAvgProfit	QBELOBets	TotalQBELOProfit	QBELOPickAvgProfit	ESPNBets	ESPNTotalProfit	ESPNPickAvgProfit
2014	263	−782	−2.97	217	1,144	5.27	-	-	-	-	-	-
2015	263	−1,915	−7.28	222	1,687	7.60	-	-	-	191	3,384	17.72
2016	265	−679	−2.56	229	−95	−0.42	-	-	-	217	−1,086	−5.01
2017	266	−1,157	−4.35	224	−1,548	−6.91	-	-	-	196	−1,035	−5.28
2018	266	59	0.22	227	−284	−1.25	-	-	-	195	1,477	7.57
2019	264	−3,316	−12.56	217	−3,036	−13.99	215	−1,906	−8.87	197	784	3.98
2020	268	−3,639	−13.58	227	−2,602	−11.46	208	−237	−1.14	207	−5,072	−24.50
2021	280	−1,460	−5.21	245	−858	−3.50	227	−1,917	−8.45	230	11	0.05
2022	267	−354	−1.33	232	−1,977	−8.52	213	−1,728	−8.11	227	−2,668	−11.75
Total	2,402	−13,244	−5.51	2040	−7,571	−3.71	863	−5,788	−6.71	1,660	−4,205	−2.53

The table shows that betting on the home team is generally a poor bet. It has a negative return in all years except for 2018 where it has a very small positive return of 0.22%. Recall that the home team wins a majority of games in most years, but that fact is crucially reflected in the betting odds. The overall loss rate of 5.5% is roughly in line with the expected profit of the bookmaker based on the hold they apply to their odds.

The Elo model makes predictions that differ enough from the betting odds to wager on a large number of games, roughly 85% of the games. The Elo model outperforms the naïve model but over the course of the data set earns a negative return rate of –3.7%. Interestingly the Elo model showed positive results in its first two years and then turned increasingly negative. The QB Elo model which debuted later is somewhat more selective in the number of games where its prediction warrants a bet. Its performance against the odds was better than Elo in three out of four years, but over its 4-year history it earns a negative -6.7% return. It significantly underperformed the naïve rule in the last two seasons.

The ESPN FPI model has the most volatile returns of the three models. Its reported results in 2014, the first year that predictions are available from the ESPN website are extremely positive at +17.7%. But in 2020 it recorded a -24.5% return, far worse than the home field naïve rule. Overall, the ESPN model has the best returns of the three models, but still comes in at -2.5% over the 8-year history.

CONCLUSION

Both ESPN and FiveThirtyEight have significant resources dedicated to analytics across multiple sports, and in the case of FiveThirtyEight across multiple fields. ESPN regularly and actively promotes their individual game predictions on their network. Their models are reasonably accurate, picking the winner as the favorite more than 60% of the time. The models are reasonably well calibrated though it appears that each model is biased in that they tend to underrate underdogs.

While these models are reasonably accurate, none of them can outperform the betting odds on any consistent basis. They all earn negative returns over the long term only moderately better than the naïve rule of betting the home team. These models are reasonably sophisticated, with access to large data sets, both broad in scope and long in history. Betting market odds are in contrast largely the result of the un-coordinated action of many bettors. While sports books may also use analytics to set initial odds, they adjust based on market forces. The betting odds represent a market-based consensus, and the odds makers have the undeniable and very significant advantage of setting unfair odds and implementing a hold. That being said, even ignoring the hold and utilizing the normalized prediction the odds are more accurate than the models.

So, does this analysis tell us anything about market efficiency in the semi-strong format. Remember that if betting markets are semi-strong efficient, no analytical model will generate positive returns. Our analysis shows that at least so far, none of the three models we analyzed generate positive returns. That does not prove that the betting markets are efficient, but it fails to show that they are not. Had we been able to find a consistent positive return from any of these models we could declare the market inefficient. But we have not done so.

It could well be the case that some highly skilled modeler somewhere has developed a model that earns long term positive returns but has kept the model details private for selfish reasons. The real implication of market efficiency is that publicly available information is quickly incorporated into prices by market mechanisms. The implications of this theory are that no matter how good a model such as the ESPN Football Power Index, or any variation of the Elo model becomes, those predictions would be quickly incorporated into the odds pricing out any long-term profit. Since the models we have discussed are so public efficient market theory implies their predictions are incorporated into the betting off their predictions will not earn a positive return. While it is highly speculative, the fact that model predictions, in particular the Elo and ESPN models, had their best results in the first year supports the notion that once these models came into the public domain their predictions were priced in as efficient market theory suggests.

REFERENCES

Aldous, D. (2017). Elo Ratings and the Sports Model: A Neglected Topic in Applied Probability? *Statistical Science, 32*(4), 616-629. Retrieved from http://www.jstor.org/stable/26408861

Anderson, S. (2019). *NFL Betting Market Efficiency: Finding a Profitable Betting Strategy.* Skidmore College, Retrieved from https://creativematter.skidmore.edu/econ_studt_schol/121/

Arscott, R. (2023). Market Efficiency and Censoring Bias in College Football Gambling. *Journal of Sports Economics, 24*(5), 664-689. doi:10.1177/15270025221148991

Balasundaram, A., Dhandapani, K., & Ashokkumar, S. (2022). *Predicting different facets in the game of cricket using machine learning* (Vol. 2418).

Bennett, M., Bezodis, N., Shearer, D., & Kilduff, L. (2020). Predicting performance at the group-phase and knockout-phase of the 2015 Rugby World Cup. *European Journal of Sport Science, 21*, 1-18. doi:10.1080/17461391.2020.1743764

Bennett, R. (2020). Next game reaction to mispriced betting lines in college football. *Applied Economics Letters, 28*, 1-4. doi:10.1080/13504851.2020.1795063

Bigsby, K. G., & Ohlmann, J. W. (2017). Ranking and prediction of collegiate wrestling. *Journal of Sports Analytics,, Forthcoming*.

Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied Computing and Informatics, 15*(1), 27-33. doi:https://doi.org/10.1016/j.aci.2017.09.005

Chae, J., Park, J., & So, W.-Y. (2021). Victory Prediction of Ladies Professional Golf Association Players: Influential Factors and Comparison of Prediction Models. *Journal of Human Kinetics, 77*, 245-259. doi:10.2478/hukin-2021-0023

Coleman, B. (2015). Team Travel Effects and the College Football Betting Market. *Journal of Sports Economics*. doi:10.1177/1527002515574514

Dandıl, E., & Ergez, C. (2018). *Prediction of Soccer Results Using Artificial Neural Networks with Web-Based Automation*.

ESPN. Breaking Down the Football Power Index. Retrieved from https://www.espn.com/video/clip/_/id/20457171

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25*(2), 383-417. doi:10.2307/2325486

FiveThirtyEight.com. NFL Elo. Retrieved from https://github.com/fivethirtyeight/data/tree/master/nfl-elo

Gray, P. K., & Gray, S. F. (1997). Testing Market Efficiency: Evidence From The NFL Sports Betting Market. *The Journal of Finance, 52*(4), 1725-1737. doi:https://doi.org/10.1111/j.1540-6261.1997.tb01129.x

Hassan, A., Akl, A.-R., Hassan, I., & Sunderland, C. (2020). Predicting Wins, Losses and Attributes' Sensitivities in the Soccer World Cup 2018 Using Neural Network Analysis. *Sensors, 20*, 3213. doi:10.3390/s20113213

Hetherington, R. A. (2006). *Hetherington, R. Alexander, Betting Against Efficiency: Behavioral Finance in an Nfl Gambling Exchange (March 26, 2006). Available at SSRN: https://ssrn.com/abstract=881514 or http://dx.doi.org/10.2139/ssrn.881514*. SSRN.

Horvat, T., & Job, J. (2020). The use of machine learning in sport outcome prediction: A review. *WIREs Data Mining and Knowledge Discovery, 10*(5), e1380. doi:https://doi.org/10.1002/widm.1380

Huang, M.-L., & Li, Y.-Z. (2021). Use of Machine Learning and Deep Learning to Predict the Outcomes of Major League Baseball Matches. *Applied Sciences, 11*, 4499. doi:10.3390/app11104499

Huang, M.-L., & Lin, Y.-J. (2020). Regression Tree Model for Predicting Game Scores for the Golden State Warriors in the National Basketball Association. *Symmetry, 12*, 835. doi:10.3390/sym12050835

Kuypers, T. (2000). Information and Efficiency: An Empirical Study of a Fixed Odds Betting Market. *Applied Economics, 32*, 1353-1363. doi:10.1080/000368400050151449

Learning, A. M. (2021). *The NFL "goes long" on machine learning*. Retrieved from https://pages.awscloud.com/hk-machine-learning-nfl.html

Mattera, R. (2021). Forecasting binary outcomes in soccer. *Annals of Operations Research*, 1-20. doi:10.1007/s10479-021-04224-8

Osken, C., & Onay, C. (2022). Predicting the winning team in basketball: A novel approach. *Heliyon, 8*, e12189. doi:10.1016/j.heliyon.2022.e12189

Paine, N. (2015). NFL Elo Ratings Are Back! Retrieved from https://fivethirtyeight.com/features/nfl-elo-ratings-are-back/

Patil, R., Duraphe, A., Motarwar, P., Suganya, G., & Mariappan, P. (2023). Cluster-Centric Based Hybrid Approach for Cricket Sports Analytics Using Machine Learning. In (pp. 268-283).

Pischedda, G. (2014). Predicting NHL Match Outcomes with ML Models. *International Journal of Computer Applications, 101*, 15-22. doi:10.5120/17714-8249

Robbins, T. R. (2023). Weak Form Efficiency in Sports Betting Markets. *American Journal of Management, 23*(2). doi:https://doi.org/10.33423/ajm.v23i2.6051

Shakil, F., Abdullah, A., Momen, S., & Mohammed, N. (2020). Predicting the Result of a Cricket Match by Applying Data Mining Techniques. In (pp. 758-770).

Silver, N. (2014). Introducing NFL Elo Ratings. Retrieved from https://fivethirtyeight.com/features/introducing-nfl-elo-ratings/

Silver, N. (2022). How Our NFL Predictions Work. Retrieved from https://fivethirtyeight.com/methodology/how-our-nfl-predictions-work/

Soto-Valero, C. (2016). Predicting Win-Loss outcomes in MLB regular season games – A comparative study using data mining methods. *International Journal of Computer Science in Sport, 15*, 91-112. doi:10.1515/ijcss-2016-0007

Stetzka, R. M., & Winter, S. (2021). How rational is gambling? *Journal of Economic Surveys, n/a*(n/a). doi:https://doi.org/10.1111/joes.12473

Walder, S. (2023). NFL Football Power Index: 2023 projections, Super Bowl chances. Retrieved from https://www.espn.com/nfl/story/_/id/37681787/nfl-football-power-index-2023-season-projections-super-bowl-chances-playoff-draft

Wikipdeia. Elo rating system. Retrieved from https://en.wikipedia.org/wiki/Elo_rating_system

Wikipedia. (2022). Efficient-market hypothesis. Retrieved from https://en.wikipedia.org/wiki/Efficient-market_hypothesis

Wilkens, S. (2021). Sports prediction and betting models in the machine learning age: The case of tennis. *Journal of Sports Analytics, 7*, 99-117. doi:10.3233/JSA-200463

Williams, A. (2021). *Rationality and Efficiency in NFL Gambling Markets.* Ohio State University, Retrieved from https://kb.osu.edu/handle/1811/92881

Yurko, R., Ventura, S., & Horowitz, M. (2019). nflWAR: a reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports, 15*(3), 163-183. doi:doi:10.1515/jqas-2018-0010

Zhao, K., Du, C., & Tan, G. (2023). Enhancing Basketball Game Outcome Prediction through Fused Graph Convolutional Networks and Random Forest Algorithm. *Entropy, 25*, 765. doi:10.3390/e25050765