



# A Summer Project in NFL Analytics

Adam Lynch

May 13, 2025 - May 13, 2025

# Acknowledgements

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Point Rating Systems . . . . .	3
1.2	Prediction Systems . . . . .	4
1.3	Probabilistic Systems . . . . .	4
<b>2</b>	<b>Simple Rating System</b>	<b>6</b>
2.1	The Simple System . . . . .	6
2.2	Rating Breakdown . . . . .	6
2.3	A Probabilistic Framework . . . . .	7
2.4	Performance . . . . .	7
<b>3</b>	<b>Linear Predictive Point Rating System</b>	<b>8</b>
<b>4</b>	<b>Generalized Non-Linear Predictive System</b>	<b>9</b>

# Chapter 1

## Introduction

### 1.1 Point Rating Systems

One class of models we explore in this report are what we call *point rating systems*. This is a term we’ve made-up to describe the type of models that rates teams based on a point differential scale. A key property of these models is the ability to predict game spreads by subtracting team ratings (potentially with an adjustment for home field advantage). For example, if team A has a rating of 10 and team B has a rating of  $-5$ , the system’s predicted spread for a game between the two teams would be  $10 - (-5) = 15$ .

One of the most well-known point rating systems is ESPN’s Football Power Index (FPI) [1], which measures team strength as “how many points above or below average a team is”. An interesting feature of the FPI is the rating break down into three components: offense, defense, and special teams. For example, to close out the end of the 2024 NFL season, ESPN had the Baltimore Ravens atop the FPI with a rating of 8.1 where 6.8 of that came from offensive, 1.5 from defense, and  $-0.1$  from special teams (rounding seems to have caused some error here, but with more precision the three components should always sum to the overall rating). It’s worth noting that the FPI components can also be thought of as point rating systems themselves. For example, if Baltimore were to faceoff against the New York Giants, whose offensive FPI is  $-6.5$ , we can think of  $8.1 - (-6.5) = 14.6$  as *the part of the total spread attributable to the discrepancy in both teams’ offenses*. In this case, these components can be useful in helping us understand that the Ravens-Giants spread is predicted to be large not just because the Ravens put up a lot of points, but also because the Giants struggle to score.

Of course, point rating systems have their limitations. Fundamentally, they assume there is a *scale of betterness*, meaning if A is ranked higher than B, and B is higher than C, then A *must* be higher than C. Although, this might not always be the case in the NFL. Team C might be more equipped to expose A’s weakness, and we have the parity of A beats B who beats C who then beats A. Maybe even in a hypothetical world where each team plays each other hundreds of times we still see this parity in the long-run.

### Evaluation

We can evaluate the accuracy of point rating systems a few different ways. The most obvious being to compare the predicted spreads against the actual spreads – analyzing the root-mean-square deviation (RMSD) for example. We can also analyze accuracy or win percentage, the rate is which the system correctly predicts game winners. A more interesting evaluation methodology – particularly for the sinners among us – would be to craft betting strategies against the line and analyze performance over time. Throughout this report, we will typically present the following five evaluation metrics for any point rating system:

#### 1. Spread Error (SpE)

RMSD of the predicted spread against the actual spread:

$$SpE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - A_i)^2} \quad (1.1)$$

Here  $P_i$  is the predicted spread and  $A_i$  is the actual spread.

#### 2. Spread Bias (SpB):

Bias of the predicted spread against the actual spread, in units of points:

$$SpB = \frac{1}{N} \sum_{i=1}^N (P_i - A_i) \quad (1.2)$$

Ideally, this should be close to 0.

### 3. Win Percentage (W%):

The percentage of games the system correctly predicts the winner.

$$W\% = \frac{1}{N} \sum_{i=1}^N W_i \quad (1.3)$$

Here  $W_i$  is an indicator that takes 1 if the system predicts correctly and 0 otherwise.

### 4. Unweighted Betting Strategy (UBS):

The per-game return on a betting strategy in which we take a  $-115$  over/under bet on the line. For example, if we predict team A by 10.5 and the line has A by 12.5, we bet the under on the line.

$$UBS = -115 + 200 \times \frac{1}{N} \sum_{i=1}^N O_i \quad (1.4)$$

Here  $O_i$  is the outcome of the bet (1 for win, 0 for loss).

## 1.2 Prediction Systems

## 1.3 Probabilistic Systems

Both point rating systems and prediction systems can be formulated probabilistically. That is, for a particular game, rather than predicting a single number for the point spread, we predict a distribution of possible spreads. This can be useful for understanding the uncertainty of our models and even determining the confidence in our predictions. We present a few additional evaluation metrics for probabilistic systems:

### 1. Weighted Win Percentage (WW%):

The win percentage of the system, weighted by the system's prediction confidence. The prediction confidence is simply the probability the system assigns to its prediction (the probability that it thinks it will be right).

$$WW\% = \frac{\sum_{i=1}^N W_i \times Pr_i}{\sum_{i=1}^N Pr_i} \quad (1.5)$$

Here  $W_i$  is an indicator that takes 1 the system predicts correctly and 0 otherwise while  $Pr_i$  is the prediction confidence.

### 2. Weighted Betting Strategy (WBS):

The per-game return on a betting strategy in which we take a  $-115$  over/under bet on the line, weighted by the system's prediction confidence. The prediction confidence for a betting strategy is the probability the system assigns to its *prediction on the line*, rather than its prediction on the game.

$$WBS = -115 + 200 \times \frac{\sum_{i=1}^N O_i \times Pr_i}{\sum_{i=1}^N Pr_i} \quad (1.6)$$

### 3. Probabilistic Calibration (PC):

Arguably the most important evaluation metric for probabilistic systems is calibration. Calibration is a measure of how well the system's predicted probabilities match the frequency of actual outcomes. For example, a system could on average be predicting the spread quite well, but if it is consistently predicting non-zero probabilities for the spread to be  $+100$ , then we would say the system is poorly calibrated since a spread of  $+100$  is nearly impossible. Additionally, if a system assigns a 10% probability for its prediction to be wrong by 12 or more points, we should expect to see this amount of error in approximately 10% of real games (if the system is well calibrated).

We define probabilistic calibration as the root-mean-square deviation (RMSD) of a P-P plot from  $y = x$ . A P-P plot is a plot of the predicted probabilities against the actual probabilities where a perfectly calibrated model should look like a straight line  $y = x$ . The RMSD of a P-P plot from  $y = x$  will tell us how far a system is from perfect calibration ( $PC = 0$ ). The calculation of the actual probabilities involves ordering all the predicted probabilities and then calculating the actual probabilities as the cumulative distribution function (CDF) of the ordered predicted probabilities. For example, if the predicted probability for a given game is 0.95, and 75% of the games have a predicted probability less than 0.95, then the actual probability for that game is 0.75.

$$PC = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - \hat{p}_i)^2} \quad (1.7)$$

Here  $p_i$  and  $\hat{p}_i$  is the actual and predicted probabilities, respectively.

## Chapter 2

# Simple Rating System

### 2.1 The Simple System

The Simple Rating System (SRS) is a straightforward method used to evaluate teams based on their average point differential adjusted for the strength of their opponents. Though its exact origins are unclear in my research, Doug Drinen of Sports Reference discussed the SRS and its calculation in a series of blog posts in 2006 [2, 3]. Drinen provides an excellent overview of the SRS which we recommend reading. We present Drinen's *transitivity* formulation of the SRS here, which will be useful as a base to build upon later.

The argument of transitivity is that if team A is better than team B, and team B is better than team C, then team A is better than team C. In fact, in a perfect world, team A beats team B by  $X$  points, team B beats team C by  $Y$  points, and team A beats team C by  $X + Y$  points. Following this argument, if we define the SRS as *a team's point differential in a game against an average opponent*, we can set up the following system of equations:

$$\begin{aligned}(P_{1,A} - HFA) - P_{1,B} &= SRS_A - SRS_B \\ (P_{2,B} - HFA) - P_{2,C} &= SRS_B - SRS_C \\ (P_{3,C} - HFA) - P_{3,A} &= SRS_C - SRS_A\end{aligned}\tag{2.1}$$

Here we use the notation  $P_{i,j}$  to represent the points scored by team  $j$  in game  $i$ . In addition,  $HFA$  represents the home field advantage, which we subtract from the home team's points scored in each game (We calculate  $HFA$  as the average home team points minus the average away team points for all completed games in the season. In the case of a game played at a neutral site like The Super Bowl, we would simply leave out the  $HFA$  term). Unfortunately, this system might not be solvable – especially if two teams play each other more than once. The more appropriate system instead includes an error term  $\epsilon$ , where the SRS ratings are such that the sum of squares of the error terms is minimized (i.e. minimize  $\epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2$ )<sup>1</sup>.

$$\begin{aligned}(P_{1,A} - HFA) - P_{1,B} &= SRS_A - SRS_B + \epsilon_1 \\ (P_{2,B} - HFA) - P_{2,C} &= SRS_B - SRS_C + \epsilon_2 \\ (P_{3,C} - HFA) - P_{3,A} &= SRS_C - SRS_A + \epsilon_3\end{aligned}\tag{2.2}$$

### 2.2 Rating Breakdown

We can break down a team's SRS into three components: offensive ( $O$ ), defensive ( $D$ ), and special teams ( $ST$ ).

$$SRS = SRS_O + SRS_D + SRS_{ST}\tag{2.3}$$

To do this, we must first break down the points scored by a team into offensive, defensive, and special teams points. The distribution of team points across the three components should be relatively straightforward. Offensive points include all points scored by the offense excluding extra points and field goals. Defensive points include all interception/fumble touchdowns and safeties. Special teams points include all extra points, field goals, and kick/punt return touchdowns. Once we have broken down the score components, determining the SRS components follows a similar approach to **Equation 2.2**, except for each game we have three equations instead of one.

$$(P_{A,O} - HFA_O) - P_{B,D} = SRS_{A,O} - SRS_{B,D} + \epsilon_1\tag{2.4}$$

$$P_{B,O} - (P_{A,D} - HFA_D) = SRS_{B,O} - SRS_{A,D} + \epsilon_2\tag{2.5}$$

$$(P_{A,ST} - HFA_{ST}) - P_{B,ST} = SRS_{A,ST} - SRS_{B,ST} + \epsilon_3\tag{2.6}$$

These equations simply formulate that teams A's offense will face off against team B's defense, teams A's defense will face off against team B's offense, and both teams special teams will face off. Keep in mind the equations are written for

---

<sup>1</sup>This is the same as using least squares on the overdetermined/unsolvable system in **Equation 2.1**.

a game where team A is home and team B is away. The final week 18 SRS ratings and breakdowns for the 2024 NFL season are shown in **Table 2.1**.

Table 2.1: 2024 NFL Final Week 18 SRS Ratings

Team	MoV	SoS	SRS	OFF	DEF	ST	Team	MoV	SoS	SRS	OFF	DEF	ST
DET	13.1	0.7	13.8	9.8	2.1	1.9	HOU	0.0	-0.7	-0.7	-2.4	0.5	1.2
BAL	9.2	0.6	9.9	7.8	0.6	1.5	SF	-2.8	1.6	-1.2	0.9	-2.2	0.2
GB	7.2	0.9	8.1	4.3	2.0	1.8	ATL	-2.0	-0.2	-2.2	-1.1	-1.2	0.2
BUF	9.2	-1.1	8.1	7.8	0.0	0.3	CHI	-3.5	1.2	-2.3	-2.4	2.8	-2.7
PHI	9.4	-1.7	7.7	3.1	2.6	1.9	MIA	-1.1	-1.9	-3.0	-3.4	1.5	-1.1
TB	6.9	-0.4	6.4	5.3	0.7	0.4	IND	-2.9	-0.7	-3.7	-0.3	-2.5	-0.9
DEN	6.7	-0.3	6.4	0.4	4.3	1.7	NO	-3.5	-0.6	-4.1	-3.1	0.2	-1.2
MIN	5.9	0.3	6.2	1.3	2.5	2.4	NYJ	-3.9	-0.5	-4.3	-0.4	-0.6	-3.3
LAC	5.9	-0.6	5.3	0.0	3.0	2.2	DAL	-6.9	0.6	-6.3	-5.2	-3.4	2.3
KC	3.5	0.7	4.2	0.3	1.6	2.3	LV	-7.4	1.0	-6.4	-4.6	-0.7	-1.1
WAS	5.5	-1.8	3.7	4.7	-1.9	1.0	JAX	-6.8	-0.8	-7.5	-2.5	-2.8	-2.2
ARI	1.2	0.9	2.1	0.5	1.9	-0.3	NYG	-8.4	0.3	-8.0	-6.9	0.2	-1.4
PIT	1.9	0.1	2.1	-3.1	1.5	3.7	NE	-7.5	-0.6	-8.1	-4.3	-1.1	-2.7
CIN	2.2	-0.8	1.4	3.6	-3.8	1.6	TEN	-8.8	0.4	-8.4	-4.2	-1.9	-2.3
SEA	0.4	0.8	1.3	-0.7	2.9	-0.9	CLE	-10.4	1.2	-9.2	-5.1	-2.2	-1.9
LA	-1.1	1.1	-0.1	-0.3	0.7	-0.4	CAR	-11.4	0.4	-11.0	-0.1	-7.2	-3.7

#### Glossary

MoV	–	Margin of Victory (Points Scored - Points Allowed) / Games Played
SoS	–	Strength of Schedule ( $SRS - MoV$ )
SRS	–	Simple Rating System
OFF	–	Offensive SRS
DEF	–	Defensive SRS
ST	–	Special Teams SRS

## 2.3 A Probabilistic Framework

## 2.4 Performance



## Chapter 3

# Linear Predictive Point Rating System

## Chapter 4

# Generalized Non-Linear Predictive System

# Bibliography

- [1] “NFL Football Power Index (FPI)”, *ESPN*, <https://www.espn.com/nfl/fpi>
- [2] Drinen D. (2006), “A very simple ranking system”, *Pro-Football-Reference*, <https://web.archive.org/web/20161031224357/http://www.pro-football-reference.com/blog/index4837.html>
- [3] Drinen D. (2006), “Another ranking system”, *Pro-Football-Reference*, <https://web.archive.org/web/20161102124021/http://www.pro-football-reference.com/blog/indexba52.html?p=39>