# Predicting the 2023 NFL season EPA*

### The Impact of Past Performance on Future EPA

Raghav Bhatia

April 2, 2024

## Table of contents



## 1 Dataset Construction

To build a prediction model which aims to predict team wise EPA for the remainder of the 2023 season, we need to use factors which have some explanatory power. The model we've constructed uses factors such as Air Yards/Attempt, Sack Rate, DAKOTA, and weeks 1-9 EPA in order to predict the EPA for weeks 10-18. The reasons why each factor has explanatory power is the following:

1. Air Yards/Attempt : This factor gives the average depth on each pass. It indicates whether a team prefers plays with a higher or lower depth. Higher depth plays are considered riskier, thereby leading to either more yards gained and a higher positive EPA, or more interceptions and a negative lower EPA. On the other hand, lower depth plays are safer and therefore lead to smaller positive/negative EPA values. This measure can be great in explaining variability in EPA across teams depending on whether they make riskier or safer plays.

*Code and data are available at: https://github.com/1raghav-bhatia/nfl-analytics.git

2. Sack Rate: The sack rate indicates poor QB performance and lower overall EPA due to interceptions. How often a QB gets sacked can be a great predictor of EPA as it leads to a big decrease in its value.

3. DAKOTA: This is a composite metric which has CPOE (Completed passes over expectation) as a major component. CPOE is a count of the number of passes completed adjusted for the difficulty of the pass. A high CPOE indicates that a team made much harder passes than expected, increasing the EPA. DAKOTA takes CPOE along with certain other EPA predicting components into account in order to give a good predictor of EPA.

4. EPA Prior (Weeks 1-9 EPA): This gives the average EPA for the first half of the season. It can be a good predictor of the EPA for the second half as it gives the trajectory a team is on in the current season.

We use the load_player_stats dataset from the nflverse (Carl et al. 2023) library to retrieve data on the above factors. The data is grouped by team and season, and averages are taken for each variable for each team and season. The season data used to train the model was from the 2018-2022 dataset. This subset of the dataset was specifically used because the time period is very recent while at the same time we have 160 observations to use while training, reducing model bias significantly.

We use the first half of the 2023 dataset as inputs into our trained model in order to predict the EPA for the remaining season.

# 2 Model

## 2.1 Model set-up

In this section, we specify the model and use it on the 2023 dataset in order to get predictions for team wise EPA.

### 2.1.1 Model Specifications

The model used in this analysis is a multi-linear regression model. The variables along with their distributions are given as below:

$$\text{future EPA}_i|\mu_i \sim \text{Normal}(\mu_i, \sigma_i) \tag{1}$$

$$\mu_i = \beta_0 + \beta_1 \times \text{air yards per attempt}_i + \beta_2 \times \text{sack rate}_i + \beta_3 \times \text{dakota}_i \tag{2}$$

$$+ \beta_4 \times \text{EPA prior}_i \tag{3}$$

$$\beta_0 \sim \text{Normal}(0, 2.5) \tag{4}$$

$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{5}$$

$$\beta_2 \sim \text{Normal}(0, 2.5) \tag{6}$$

$$\beta_3 \sim \text{Normal}(0, 2.5) \tag{7}$$

$$\beta_4 \sim \text{Normal}(0, 2.5) \tag{8}$$

where:

1. future EPA is the predicted EPA for team i for weeks 10-18
2. $\mu_i$ is the expected future EPA given respondent $i$
3. Normal$(\mu_i, \sigma_i)$ is the Normal Distribution with parameters $\mu_i$ and $\sigma_i$
4. air yards per attempt$_i$ is the $i^{\text{th}}$ teams air yards per attempt.
5. sack rate$_i$ is the $i^{\text{th}}$ teams sack rate.
6. dakota$_i$ is the $i^{\text{th}}$ teams Dakota.
7. EPA prior$_i$ is the $i^{\text{th}}$ teams average EPA from the weeks 1-9.
8. $\beta_0$ is the intercept of the multi-linear regression equation.
9. $\beta_1$ is the coefficient of the air yards per attempt variable.
10. $\beta_2$ is the coefficient of the sack rate variable.
11. $\beta_3$ is the coefficient of the dakota variable.
12. $\beta_4$ is the coefficient of the EPA prior variable.

#### 2.1.1.1 Model Predictions

The above model is regressed on the data for the regular season for each team from 2018-22. From out estimated model, we can now use it alongwith the 2023 dataset with only weeks 1-9 data to predict the EPA for weeks 10-18 by team. The table with the models predictions for the EPA is given below:

Table 1: Team wise predictions for Future EPA

| Team | Air Yards | Sack Rate | DAKOTA | Prior EPA | Predicted EPA |
|---|---|---|---|---|---|
| ARI | 7.18 | 0.06 | 0.06 | -8.23 | 0.90 |
| ATL | 8.02 | 0.09 | 0.06 | -2.86 | -2.37 |
| BAL | 9.41 | 0.07 | 0.13 | 1.80 | 4.05 |
| BUF | 8.78 | 0.03 | 0.15 | 5.78 | 5.09 |
| CAR | 7.61 | 0.11 | 0.04 | -7.48 | -2.82 |

| Team | Air Yards | Sack Rate | DAKOTA | Prior EPA | Predicted EPA |
|---|---|---|---|---|---|
| CHI | 7.47 | 0.11 | 0.07 | -4.21 | -1.33 |
| CIN | 6.06 | 0.09 | 0.11 | -0.14 | 1.25 |
| CLE | 8.31 | 0.08 | 0.02 | -6.87 | -3.96 |
| DAL | 8.50 | 0.07 | 0.16 | 2.41 | 6.71 |
| DEN | 7.23 | 0.11 | 0.10 | -0.92 | 0.25 |
| DET | 6.77 | 0.05 | 0.12 | 3.73 | 3.01 |
| GB | 9.86 | 0.05 | 0.13 | -0.25 | 5.46 |
| HOU | 8.12 | 0.07 | 0.08 | 6.77 | -3.51 |
| IND | 7.03 | 0.06 | 0.07 | -0.52 | -1.17 |
| JAX | 7.59 | 0.06 | 0.12 | 0.18 | 3.10 |
| KC | 6.40 | 0.04 | 0.10 | 5.10 | 0.53 |
| LA | 7.61 | 0.06 | 0.07 | -0.90 | -0.80 |
| LAC | 7.56 | 0.08 | 0.06 | 4.58 | -4.41 |
| LV | 7.89 | 0.11 | 0.07 | -2.36 | -1.73 |
| MIA | 6.57 | 0.04 | 0.15 | 7.32 | 4.44 |
| MIN | 7.56 | 0.09 | 0.11 | 3.13 | 0.47 |
| NE | 7.54 | 0.08 | 0.02 | -4.34 | -5.49 |
| NO | 9.45 | 0.06 | 0.09 | -0.23 | 1.43 |
| NYG | 6.45 | 0.16 | 0.06 | -6.33 | -3.79 |
| NYJ | 7.21 | 0.17 | 0.00 | -9.51 | -8.54 |
| PHI | 8.43 | 0.07 | 0.12 | 4.90 | 1.74 |
| PIT | 7.26 | 0.08 | 0.07 | -4.27 | -0.29 |
| SEA | 7.56 | 0.06 | 0.09 | 0.01 | 0.23 |
| SF | 7.41 | 0.05 | 0.20 | 5.11 | 10.29 |
| TB | 8.32 | 0.06 | 0.11 | 3.17 | 0.96 |
| TEN | 9.38 | 0.16 | 0.07 | -1.64 | -3.54 |
| WAS | 7.57 | 0.10 | 0.09 | -3.05 | 0.06 |

We also construct plot predicted EPA against prior EPA to show the predicted improvement or decline in each teams EPA:

The plot shows that on average most teams will increase their EPA while the relative difference won't be very different from the current EPA. Some teams such as LAC and HOU are predicted to see a big decline in their EPA while teams such as ARI and GB are predicted to see a big improvement.

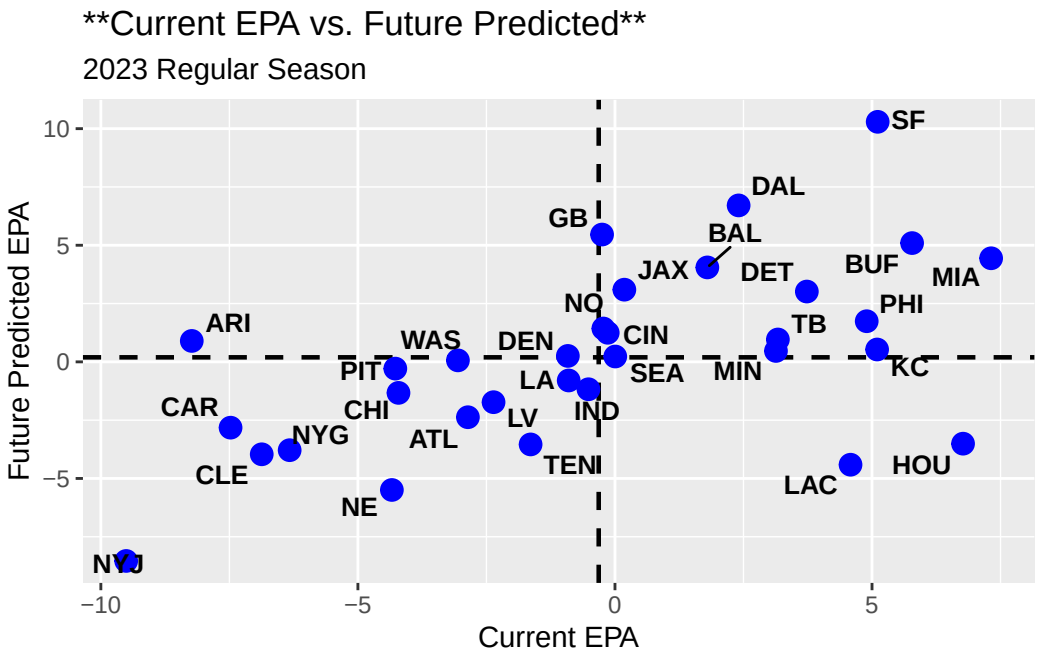


Figure 1: Prior EPA against Predicted EPA

## References

Carl, Sebastian, Ben Baldwin, Lee Sharpe, Tan Ho, and John Edwards. 2023. *Nflverse: Easily Install and Load the 'Nflverse'*. https://CRAN.R-project.org/package=nflverse.