

Forseeing the rest of the 2023 NFL season*

Predicting each team's EPA

Janssen Myer Rambaud

April 4, 2024

1 Introduction

This paper looks to predict how the rest of the 2023 NFL regular season will play out after the halfway mark. Using the data collected from Week 1 up to Week 9, we aim to predict how the passing EPA will look for each team until the regular season ends. Passing EPA, according to nflverse is defined as the “total expected points added on pass attempts and sacks” (Carl et al. 2023).

2 Data

This data was first collected through Carl et al. (2023), where it was selected to retrieve recorded games by a QB in the 2023 season up to week 9. Using R Core Team (2023), the package was downloaded, cleaned, and fitted into a model with the help of Wickham et al. (2023), Firke (2023), Wickham et al. (2019), Arel-Bundock (2022) and Kuhn and Wickham (2020). Taking a look at the provided statistics from Carl et al. (2023), the best analytics to predict passing EPA appeared to be: completions, attempts, passing_yards, passing_yards_after_catch, passing_tds (passing touchdowns), and sacks, according to their definition of EPA.

Completions are successful passes (caught) that did not hit the ground/go out of bounds. Pass attempts are just any attempt at passing the ball. Passing yards are the distance (in yards) the ball has covered from where it was thrown to where it was caught. Passing yards after catch are technically rushing yards, but as it is influenced by the throw of the QB, it can remain here. Passing touchdowns are when the QB passes the ball to a receiver at the endzone, scoring a point. Sacks are when the QB are taken down by a defender prior to making a complete pass.

*Code and data are available at: <https://github.com/foreverwoods/nfl-prediction>.

```
# A tibble: 6 x 9
  recent_team passing_tds week completions attempts passing_yards
  <chr>          <dbl> <dbl>          <dbl>      <dbl>          <dbl>
1 NYJ              0     1              0         1              0
2 LA                0     1             24        38             334
3 LA                1     2             34        55             307
4 LA                1     3             18        33             269
5 LA                1     4             27        40             319
6 LA                2     5             21        37             222
# i 3 more variables: passing_yards_after_catch <dbl>, sacks <dbl>,
#   passing_epa <dbl>
```

3 Model

The goal of our modelling strategy is to use every available metric that involves the QB's passing, so in this case, passing yards after catch is included, as well as sacks. This is enough to fit and hopefully predict the future EPA for weeks 10-18.

3.1 Model set-up

Define y_i as the passing EPA recorded for a QB for a game. Then the rest of the factors are the important QB passing-influenced metrics as described in the data section.

$$\begin{aligned}
y_i | \mu_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \beta_0 + \beta_1 \times \text{completions}_i + \beta_2 \times \text{attempts}_i + \beta_3 \times \text{passing-tds}_i \\
&\quad + \beta_4 \times \text{passing-yards}_i + \beta_5 \times \text{passing-yards-after-catch}_i + \beta_6 \times \text{sacks}_i
\end{aligned}$$

4 Results

Our results are summarized in Table 1.

Table 1: Predicted Passing EPA

	(1)
(Intercept)	−1.080 (0.715)
completions	0.697 (0.115)
attempts	−0.982 (0.072)
passing_yards	0.102 (0.008)
passing_yards_after_catch	−0.023 (0.011)
passing_tds	2.140 (0.344)
sacks	−2.151 (0.160)
Num.Obs.	318
R2	0.776
R2 Adj.	0.772
AIC	1926.2
BIC	1956.3
RMSE	4.88

References

- Arel-Bundock, Vincent. 2022. “modelssummary: Data and Model Summaries in R.” *Journal of Statistical Software* 103 (1): 1–23. <https://doi.org/10.18637/jss.v103.i01>.
- Carl, Sebastian, Ben Baldwin, Lee Sharpe, Tan Ho, and John Edwards. 2023. *Nflverse: Easily Install and Load the 'Nflverse'*. <https://nflverse.nflverse.com/>.
- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://github.com/sfirke/janitor>.
- Kuhn, Max, and Hadley Wickham. 2020. *Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles*. <https://www.tidymodels.org>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.