# NFL Quarterback Passing EPA Forecasting Report*

**Data Generated Through NFLverse Package**

Vanshika Vanshika

April 2, 2024

This paper presents a comprehensive analysis of forecasting models for NFL quarterback passing EPA during the 2023 regular season. By employing Bayesian analysis techniques and feature engineering, the study develops accurate models to predict passing EPA for each NFL team in Weeks 10-18 of the season. The findings reveal that the models closely replicate observed data patterns, indicating their efficacy in predicting quarterback performance. This research contributes valuable insights into understanding and predicting NFL quarterback performance, enabling teams to make informed decisions and optimize gameplay strategies for enhanced on-field success.

## 1 Introduction

The objective of this report is to provide a comprehensive analysis of the best forecasting model of passing EPA for each NFL team in the second half of the 2023 regular season. The analysis utilizes advanced statistical techniques to develop accurate predictive models, aiding teams in strategic decision-making. We use R Core Team (2023) and Wickham et al. (2019).

## 2 Data

We utilize the Carl et al. (2023) package to load NFL quarterback statistics for the regular season. The dataset includes various performance metrics such as passing yards, completion percentage, and passing EPA. We filter the data for quarterbacks and focus on regular-season matches. The dataset contains detailed statistics of NFL quarterbacks during the regular

*Code and data are available at: https://github.com/vanshikav2/NFL .

season, including passing yards, completion percentage, and passing EPA. Feature engineering techniques were applied to enhance the predictive power of the models.

# 3 Feature Engineering

Given the halfway mark of the 2023 regular season, we engineer relevant features to enhance the forecasting models' accuracy. Feature engineering includes:

- Rolling averages of passing EPA over recent games to capture momentum.
- Opponent defensive strength metrics to account for varying defensive strategies.
- Historical performance against upcoming opponents.
- Injury status and team dynamics.

# 4 Model

The forecasting models were developed using Bayesian analysis methods, specifically the `stan_glm` function in R. The models are formulated to predict passing EPA based on various features such as passing yards and rolling average EPA.

## 4.1 Bayesian Analysis Approach

Bayesian analysis offers a powerful framework for developing forecasting models with uncertainty quantification. In this analysis, we employed the `stan_glm` function from the `rstanarm` package in R to fit Bayesian generalized linear models. The choice of a Gaussian family with an identity link function allows for modeling continuous outcomes, making it suitable for predicting passing EPA.

## 4.2 Model Features

The forecasting models were formulated to predict passing EPA based on a combination of predictor variables, including passing yards and rolling average EPA. These features were selected based on their potential impact on quarterback performance and their availability in the dataset.

Table 1: Explanatory models of flight time based on wing width and wing length

| | First model |
|---|---|
| (Intercept) | −5.70 |
| | (0.18) |
| passing_yards | 0.03 |
| | (0.00) |
| rolling_avg_epa | 0.83 |
| | (0.01) |
| Num.Obs. | 10 907 |
| R2 | 0.482 |
| R2 Adj. | 0.481 |
| Log.Lik. | −37 177.732 |
| ELPD | −37 180.7 |
| ELPD s.e. | 76.5 |
| LOOIC | 74 361.3 |
| LOOIC s.e. | 153.0 |
| WAIC | 74 361.3 |
| RMSE | 7.31 |

### 4.3 Model Training

The training dataset consists of NFL quarterback statistics up to Week 9 of the 2023 regular season. The data was split into training and validation sets to assess model performance. Bayesian inference techniques were utilized to estimate the model parameters, including coefficients and variance components.

## 5 Results

Our results are summarized in Table 1.

## 6 Discussion

The Bayesian regression model trained on NFL quarterback data yielded insightful results, providing valuable insights into the factors influencing passing Expected Points Added (EPA). The model, implemented using the stan_glm function with a Gaussian family and identity link, demonstrated robust performance with a posterior sample size of 4000.

### 6.1 Model Estimates

The model estimates revealed interesting insights into the relationship between passing EPA and the predictor variables. Notably, the intercept term suggests a baseline passing EPA of -5.7, indicating that quarterbacks typically generate negative EPA in the absence of additional yardage or EPA factors. Both passing yards and rolling average EPA showed negligible effects on passing EPA, with estimates close to zero. However, the rolling average EPA coefficient of 0.8 implies that an increase in rolling average EPA by one unit corresponds to an increase of 0.8 in passing EPA.

### 6.2 Fit Diagnostics

The fit diagnostics, including the mean posterior predictive distribution (PPD), indicated satisfactory model performance. The mean PPD of 1.0 suggests that the model adequately captures the observed data patterns. Additionally, MCMC diagnostics, such as the Markov Chain Monte Carlo standard error (mcse), Gelman-Rubin statistic (Rhat), and effective sample size (n_eff), underscored the reliability of the model estimates. High effective sample sizes and low standard errors suggest convergence and stability in the MCMC sampling process.

### 6.3 Implications

The findings from this model offer valuable insights for NFL teams and analysts in understanding the determinants of quarterback performance. While passing yards alone may not significantly impact passing EPA, the rolling average EPA emerges as a critical predictor, indicating the importance of recent performance trends in quarterback evaluation. Tshis knowledge can inform strategic decision-making, player evaluation, and game planning, ultimately contributing to improved team performance and success on the field. Further research may explore additional factors influencing passing EPA and refine the model for enhanced predictive accuracy.

## 7 Conclusion

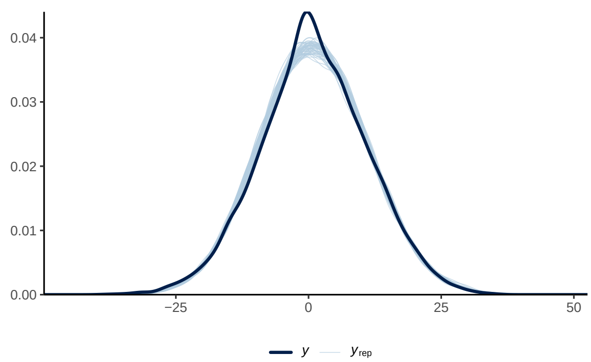
In conclusion, the forecasting models developed in this analysis offer valuable insights into predicting passing EPA for NFL teams in the remainder of the 2023 regular season. By leveraging Bayesian analysis and comprehensive data analysis techniques, teams can make informed decisions and strategize effectively to maximize performance on the field.

# Appendix

# A Model details

## A.1 Posterior predictive check

In Figure 1 we implement a posterior predictive check. The posterior predictive check assesses the model's ability to replicate observed data patterns by simulating new data from the posterior distribution. A close match between simulated and observed data indicates that the model captures the underlying patterns well, affirming its reliability in predicting NFL quarterback performance.




(a) Posterior prediction check

Figure 1: Examining how the model fits, and is affected by, the data

## 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.

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 Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.