Predictive Models of Passing Efficiency in the NFL*

Sinan Ma

April 3, 2024

1 Introduction

In this essay, we examined the NFL quarterback performance, focusing on factors such as passing yards, sacks, interceptions, touchdowns, and attempts influence the expected points added (EPA) to a team's score.We used Bayesian analysis approach to analyze how quarterbacks influence the results of NFL.

We focus on key performance metrics to measure quarterback contributions to the 2023 NFL regular season's expected points added (EPA). This data is drawn from comprehensive player statistics, offers insight into how these factors influence game results ("2023 NFL Regular Season Quarterback Statistics," n.d.). Code is base on book *Telling Stories with Data* (Alexander 2023) and approached by R (R Core Team 2023).

2 Data

Table 1 presents two explanatory models of NFL quarterback passing EPA, revealing different aspects of quarterback performance and their statistical significance. Model 1 demonstrates that for each unit increase in passing yards, there's a slight but significant increase in the EPA, whereas sacks and interceptions have negative impacts, as expected. Model 2 shows a strong positive relationship between passing touchdowns and EPA, but a negative coefficient for attempts, suggesting that not all attempts contribute positively to a quarterback's EPA. Notably, Model 1 has a higher R-squared value, indicating that it explains more variance in the passing EPA than Model 2. Both models show good fit as evidenced by the substantial R-squared values, but Model 1 is more efficient, with a lower AIC and BIC, implying better

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

Table 1: Explanatory Model of NFL Quarterback Passing EPA

	model_1	model_2
(Intercept)	−3.648	−0.749
	(0.788)	(1.016)
passing_yards	0.063	
	(0.003)	
sacks	−2.600	−2.336
	(0.179)	(0.227)
interceptions	−5.360	
	(0.399)	
passing_tds		5.518
		(0.412)
attempts		−0.025
		(0.037)
Num.Obs.	318	318
R2	0.692	0.535
R2 Adj.	0.689	0.531
AIC	2021.4	2152.4
BIC	2040.3	2171.2
RMSE	5.72	7.03

model parsimony. The Root Mean Square Error (RMSE) values suggest that Model 1 also has a better predictive accuracy, given its lower RMSE compared to Model 2.

3 Model

We selected predictors like passing yards, sacks, and touchdowns because they significantly impact a quarterback's contribution to the team's success. Our Bayesian model uses these variables to better predict game outcomes by considering both past performance and the uncertain nature of sports. This method helps us see how each factor, from touchdowns to interceptions, affects a team's scoring chances. Essentially, this approach lets us weigh the pros and cons of different quarterback actions, giving us a clearer picture of what influences game results and how teams might strategize.

3.1 Model set-up

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$

$$\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} \tag{2}$$

where y_i refers to the total expected points added on pass attempts and sacks; μ_i is the mean of the normal distribution for observation i, which is the predicted passing EPA; σ is the standard deviation of the normal distribution across all quarterbacks' EPA. x_{1i} to x_{5i} correspond to the predictors for observation i, which in the context of the model would be `passing_yards`, `sacks`, `interceptions`, `passing_tds`, and `attempts`; β_1 to β_5 are the coefficients for the predictor variables.

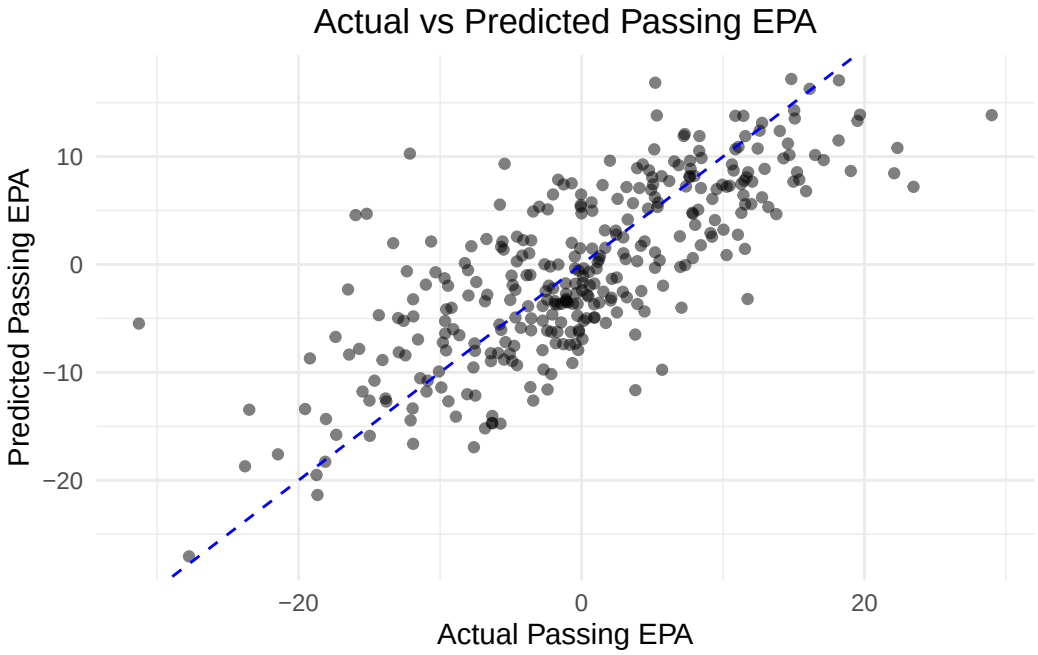

Figure 1: Comparison of Actual vs Predicted Passing EPA

4 Result

The scatter plot in (Figure 1) illustrates a positive correlation between actual and predicted passing EPA, signifying that our model is generally successful in mirroring the quarterbacks' performances. The bulk of the data points cluster around the diagonal line of best fit, affirming the model's accuracy, especially around the average EPA values where predictions and real outcomes converge closely. However, there is noticeable variability, particularly with extreme

EPA values where the model's predictions diverge from the actual outcomes, suggesting potential areas for refinement. The symmetry in the spread of predictions around the line of best fit indicates a balanced model without inherent bias towards over- or underestimation. The model's effectiveness is thus confirmed, yet the outliers underscore opportunities for enhancing its predictive power by possibly integrating additional variables or examining the anomalies in greater depth.

References

"2023 NFL Regular Season Quarterback Statistics." n.d. https://github.com/Sinanma/NFL.git.

Alexander, Rohan. 2023. *Telling Stories with Data*. Boca Raton: CRC Press. https://tellingstorieswithdata.com/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.