2023 Mid-Season NFL Quarterback Performance Projection*

Alexander Sun

April 3, 2024

Introduction

This report presents an analysis of quarterbacks' performance in the National Football League (NFL) with a focus on Expected Points Added (EPA) from passing plays. EPA is a metric that quantifies the contribution of each play to the potential points scored by a team. The objective of this study is to employ statistical modeling to forecast quarterbacks' passing EPA for the remaining games of the 2023 regular season.

To accomplish this, data sourced from the NFL's official statistics via the `nflverse` package has been utilized. The dataset comprises regular season performance metrics for quarterbacks and has been filtered to remove incomplete records. The primary variable of interest for predicting future EPA is the average passing EPA achieved by quarterbacks up to Week 9 of the current season. This measure serves as a predictor in a linear regression model, selected for its simplicity and interpretability.

Data

For the analysis, a comprehensive dataset was compiled using the `nflverse` package, which aggregates official NFL player statistics(Carl et al. 2023). The dataset was constrained to include only quarterbacks (QB) from the regular season, with further refinement to consider performance data up to and including Week 9 of the 2023 season. The initial dataset included

*https://github.com/alexandersunliang/nflverse

various performance indicators, from which **passing_epa** was extracted for use as the response variable in the model.

The primary feature selected for prediction was the average **passing_epa** per QB, computed from data up to Week 9 of the season. This feature was chosen for its potential to encapsulate a QB's performance trajectory and its relevance to the outcome variable. The decision to use this feature aligns with the objective of the analysis to provide a clear measure of a quarterback's contribution to the team's scoring potential.

Following feature selection, the dataset was split into training and validation subsets. The training subset was employed to develop the regression model, while the validation subset was reserved to assess the model's predictive performance. Care was taken to partition the data in a manner that avoids temporal leakage, ensuring that the model's training phase is not influenced by future data points.

The regression analysis was conducted to determine the predictive power of the average passing EPA on the quarterbacks' future performance. The model employed was a simple linear regression, utilizing the average passing EPA as the sole predictor. The results are depicted below:

The results indicated a statistically significant relationship between the average passing EPA and the outcome variable, with a p-value less than 0.05, suggesting that the average passing EPA is a reliable predictor of future passing EPA performance. The regression coefficient for the average passing EPA was positive, confirming the expected directional relationship where a higher average passing EPA is associated with a greater expected EPA in subsequent games. The intercept term was not statistically significant, which is coherent with the nature of the EPA metric that centers around zero, indicating no particular bias when the average passing EPA is zero. The model's R-squared value was approximately 0.27, implying that around 27% of the variability in the passing EPA can be explained by the average passing EPA up to Week 9. While this indicates a modest predictive capability, it is noteworthy given the complexity and variability inherent in NFL game outcomes. The adjusted R-squared value, which accounts for the number of predictors in the model, affirmed the explanatory power of the model.

Discussion and Recommendation

The significant relationship between average passing EPA and future performance provides a basis for using past quarterback performance as an indicator of future success. However, the modest R-squared value suggests that while average passing EPA is informative, there are other factors at play that the model does not capture. This is expected in the context of NFL games, where outcomes are influenced by a myriad of factors, including but not limited to, opposing defenses, game plans, and player conditions.

```
Residuals:
     Min       1Q   Median       3Q      Max
-23.3058  -6.3077   0.0536   5.7225  19.3618

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.04325    0.56468  -0.077    0.939
avg_passing_epa  1.01548    0.10473   9.696   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.989 on 254 degrees of freedom
Multiple R-squared:  0.2701,    Adjusted R-squared:  0.2673
F-statistic: 94.01 on 1 and 254 DF,  p-value: < 2.2e-16

[1] "MAE: 6.45499857908627"

[1] "RMSE: 8.45410666619718"
```

Figure 1: Model Results

The insignificance of the intercept term aligns with the expectation that when a quarterback's average passing EPA is zero, the baseline level of expected points added in future games is also zero. Based on the model results, we have the following suggestions:

1. **Quarterback Selection and Rotation Strategy:** The significant predictive relationship between a quarterback's average passing EPA and their future performance suggests that this metric can be an essential factor in determining starting quarterbacks. Management should consider a quarterback's current average passing EPA trend when making rotation decisions, particularly when choosing between quarterbacks with similar skill sets.
2. **Contract Negotiations and Player Development:** The findings can provide a quantitative basis for contract negotiations. The model can identify quarterbacks who consistently contribute to the team's scoring chances, thus warranting investment in their development.
3. **Game Planning:** Performance trends can inform offensive game planning. Coaches may design plays that leverage the strengths of quarterbacks with higher average passing EPA, potentially increasing the team's overall scoring efficiency.
4. **Scouting and Recruitment:** Scouts can use the average passing EPA as a benchmark when evaluating potential new quarterback recruits, both from college drafts and free agency. Prioritizing recruits with a strong history of high passing EPA could bring valuable assets to teams.

We suggest that this analysis be given to every front office in the NFL, to provide another advanced statistic for teams to base lineup decisions around. Further testing with this analysis by teams can be employed and empirical evidence can be collected. Furthermore, we contend that a team dedicated to monitoring advanced statistics similar to the NBA should be developed for the National Football League.

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.