Quarterback Evaluation in the National Football League

by

Matthew Reyers

B.Sc., Simon Fraser University, 2018

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2020

Approval

Name: **Matthew Reyers**

Degree: **Master of Science (Statistics)**

Title: **Quarterback Evaluation in the National Football League**

Examining Committee:

Chair: Gary Parker
Professor

Tim Swartz
Senior Supervisor
Professor

Derek Bingham
Supervisor
Professor

Harsha Perera
Internal Examiner
Lecturer

Date Defended: **August 20, 2020**

Abstract

This project evaluates quarterback performance in the National Football League. With the availability of player tracking data, there exists the capability to assess various options that are available to quarterbacks and the expected points resulting from each option. The quarterback's execution is then measured against the optimal available option. Since decision making does not rely on the quality of teammates, a quarterback metric is introduced that provides a novel perspective on an understudied aspect of quarterback assessment.

Keywords: Sports Analytics, Expected Points, Machine Learning, Model Validation, Player Tracking Data

Table of Contents

Dedication

In loving memory of my grandfather, Clarke Akeroyd

Acknowledgements

This project would not have been possible without the support and mentorship of my supervisor Dr. Tim Swartz. His effort and patience have helped propel me through my graduate studies and my project. Despite an ever increasing workload, Dr. Swartz has been a stalwart in the creation of this work. I am beyond grateful.

Further, an enormous thank you to Dr. Michael Lopez of Skidmore College and the NFL for his role in developing my fascination with Sports Analytics and in the facilitation of this project. All tracking data made available for this project comes by way of him and the NFL Big Data Bowl, the event that inspired this project.

Thank you to the wonderful instructors I have had the chance to work with closely during my graduate studies including Dr. Derek Bingham, Dr. Richard Lockhart, Dr. Jinko Graham, Dr. Rachel Altman, Dr. Tom Loughin, and Ian Bercovitz. I also want to thank Charlene Bradbury, Sadika Jungic, Jay Young, and Kelly Jay for their undeniable ability to keep me organized and informed. From outside the department, a thank you to Dr. Tamon Stephen for his guidance and support.

To my colleagues, thank you for making my time at SFU memorable. The lunch meetings, the late night discussions, and the ever expanding list of projects has only enhanced the experience. I am appreciative for all the time we had together.

Finally, the biggest thank you of all to my family for their love, support, and laughter. No matter the place in sports analytics, no team is stronger than family!

Chapter 1

Introduction

The National Football League (NFL) is the top revenue league in world sport (Raul 2016) with an average team revenue of $453,000,000 in the 2017 season (Gough 2018). Despite the big money nature of the NFL, football analytics trails some of the other “big” professional sports including basketball (the National Basketball Association), soccer (major European leagues) and baseball (Major League Baseball). For a survey of some of the work that has been done in sports analytics, see Albert, Glickman, Swartz and Koning (2017).

The analytics landscape in the NFL is beginning to change as Next-Gen-Stats’ player tracking data was made available to all 32 NFL teams in 2019. Player tracking data is detailed spatio-temporal data where the locations of each player on the field are recorded 10 times per second. This type of data leads to analytics opportunities that were previously unthinkable in the era of boxscore data. Subsets of the data have been released by the NFL in a yearly competition known as the Big Data Bowl (https://operations.nfl.com/the-game/big-data-bowl/) which is an analytics event held in conjunction with the NFL Scouting Combine. The availability of the player tracking data has led to a flurry of recent NFL analytics research and includes Burke (2019), Chu et al. (2020), Deshpande and Evans (2020), Yam and Lopez (2020) and Yurko et al. (2020).

A traditional NFL statistic that is widely reported and is endorsed by the NFL is the quarterback passer rating. In the 2019 season, Patrick Mahomes of the Super Bowl Champion Kansas City Chiefs was one of the top quarterbacks in the NFL with a rating of 105.3 (see `www.nfl.com/stats/categorystats?statisticCategory=PASSING`). The quarterback passer rating (Zilavy 2018) is a complex formula for which there is a minimum rating of 0 and a maximum rating of 158.3. The components of the formula involve aspects of passing performance (e.g. completions, passing yards, interceptions, etc.) that all make “football sense”. Yet, these components are combined in an ad-hoc manner to yield a rating that does not admit straightforward interpretation.

It is a curious fact that amongst the top 12 quarterbacks in 2019 (according to passer rating), all 12 of these quarterbacks played for the 12 NFL playoff teams. An immediate reaction to this observation is that a team must have an outstanding quarterback in order to make the playoffs.

Another explanation is that a quarterback's rating is highly dependent on the quality of his team. This is the motivation for our research; we attempt to introduce a quarterback metric which is less dependent on the performance of one's teammates. It is possible that there are some excellent quarterbacks who play on weaker teams and their worth is not fully appreciated.

To investigate quarterback performance, we use the player tracking data previously mentioned. On each passing down, we identify options that are available to the quarterback. This is a novel investigation as it requires the enumeration of options (both passing and running) prior to the quarterback's actual decision, and even options that may have eventuated after his decision. Then, each of these options are assessed an EPV (expected point value) depending on the state of the game. Further, probabilities of the successful execution of these options are estimated. These components then allow us to formulate a metric which compares actual outcomes versus optimal options. An advantage of this approach is that the quarterback metric is less dependent on one's teammates - some quarterbacks will have better options than other quarterbacks, and the metric is formulated such that a quarterback's performance is only compared against his available options. Another advantage is that the metric takes running into account unlike the traditional quarterback passing rating. Clearly, the running abilities of quarterbacks such as Russell Wilson, Lamar Jackson, Deshaun Watson and Michael Vick have been a great benefit to their respective NFL teams.

In terms of completing a catch, there have been various investigations (Yurko et al. 2020, Deshpande and Evans 2020) and vendors (Next Gen Stats Team 2018) that provide probabilities of pass completion. These approaches typically look at the circumstances at the time of the catch (e.g. the location of defenders, where the ball is thrown, etc.). We emphasize an important novelty of our approach where completion probability is assessed *at the time that the ball is thrown* - this involves more uncertainty as it is unclear how the play will develop downfield.

The only closely related approach to our work is that by Burke (2019). Burke (2019) uses neural networks to predict the targeted receiver. The covariates chosen are different from ours and consideration is only given to passing options at the instant that the pass is made. In assessing quarterbacks, Burke (2019) uses expected yards; we believe that EPV is a more relevant measure in football since it incorporates context. For example, gaining 7 yards on first down and 10 yards is a much better outcome than gaining 7 yards on third down and 10 yards. Unlike Burke (2019), we also introduce some stochasticity in our approach; this facilitates error assessment. Importantly, our approach also takes into account non-designed quarterback runs and interceptions; intercepted passes form a critical component of the outcome of matches.

In Section 2, we begin with a broad overview of our approach. The basic idea is that we identify options that are available to a quarterback, we assign value to these options, and we then compare the actual results versus the optimal options. This results in a quarterback rating that is less dependent on teammates. In Section 3, details of the procedure are provided. We describe the player tracking data, propose covariates and then use machine learning algorithms to determine the probabilities associated with the options available to the quarterback. The probabilities are validated against fresh data. In Section 4, the methods are applied and ratings are obtained for NFL quarterbacks. The ratings generally agree with popular opinion although they reveal some surprises; there are some quarterbacks held in high esteem who are not rated so highly, and vice-versa. We consider a discussion of results in Section 5, comparing our evaluation criterion with a process-based alternative. We conclude with a brief summary of outcomes and future considerations in Section 6.

Further, the work presented in this MSc project is an expansion of Reyers and Swartz (2020).

Chapter 2

Overview of the Approach

Consider a particular quarterback and all of his passing and running options on a play that was not a designed run. For the ith play, the quarterback executes a decision at time t_i. For the time interval $t \in (0, t_i + \epsilon]$, we consider all $j = 1, \ldots, n_i$ options that were available to the quarterback. No doubt, inferences become more difficult for larger values of ϵ since players alter their patterns once $t > t_i$. For example, players slow down once a pass is initiated and they realize that they will not be active in the play. In Section 4, we set a small window $\epsilon = 0.5$ seconds.

We denote p_{ij} as the probability that the jth option on the ith play is a success where all running plays are successes and passing plays are only successes if they result in a completion. The quantity p_{ij} is an unknown parameter which we estimate by $\hat{p}_{ij}$ using machine learning methods. We let G_{ij} denote the corresponding expected points gained from the successful execution of option j on play i. Expected point values are obtained from Yurko, Ventura and Horowitz (2019) and take into account both field position and game situation. For example, suppose that your team is faced with first down and 10 yards at your own 20 yard line early in the first quarter. The EPV is 0.40, indicating that on average a team will gain 0.4 points on the set of possessions following this state. Your team then completes a 6 yard pass and is faced with second down and 4 yards at your own 26 yard line. The EPV of the updated state is 0.69, and therefore the expected points gained from the completed pass is $G = 0.69 - 0.40 = 0.29$. Therefore, the EPV gained from the optimal decision by the quarterback on play i is given by

$$Y_i = \max\left[\hat{p}_{i1}\mathrm{G}_{i1},\ \hat{p}_{i2}\mathrm{G}_{i2},\ \ldots, \hat{p}_{in_i}\mathrm{G}_{in_i}\right] \ . \tag{2.1}$$

Now, corresponding to play $i = 1, \ldots, N$, we can calculate the actual expected points gained A_i. This is obtained by taking the difference between the EPV value before and after the play. We therefore propose the quarterback metric

$$Q = \left(\frac{\sum_{i=1}^{N} A_i}{\sum_{i=1}^{N} Y_i}\right) 100\% \ . \tag{2.2}$$

We note that Q is nicely calibrated. Recall that the sum $\sum_i Y_i$ denotes the maximal EPV gained by an average quarterback (in terms of execution) who is always making the best decisions. Therefore, a score of Q represents the execution percentage relative to this hypothetical quarterback. As a measure of quarterback performance, the metric Q combines both the fundamental elements of decision making and execution. And again, we emphasize that a feature of the metric (2.2) is that the basis of comparison involves the options that are available to the quarterback. Different quarterbacks have different options, and we would expect quarterbacks on better teams to have better options.

Chapter 3

Details of the Approach

3.1 Data

The data used in this investigation were provided by Next Gen Stats. Released in 2019, the data cover the first six weeks of the 2017 NFL season. This subset of the season includes five or six games per team, dependent on whether teams had been assigned a bye week. This leads to a total of 91 games for which there are 6960 passing plays. These plays were augmented with 252 non-designed quarterback runs and 452 sacks. After removing problematic tracking data, the cleaned dataset consists of 6727 plays of interest.

At a more granular level, each play contains measurements on 23 unique actors on the field: 11 offensive players, 11 defensive players and the football. Measurements for each actor were recorded 10 times per second. The data were collected by Next Gen Stats and its partner organizations Zebra Technologies and Wilson Sporting Goods (see `https://operations.nfl.com/thegame/technology/nfl-next-gen-stats/`). Each player measurement includes detailed information about movement including velocity, direction, distance travelled since the last frame, acceleration, and position. Similar measurements are available for the football in the same format.

The primary motivation of this project is the assessment of quarterbacks. However, there is a contextual aspect to the evaluation where it is well-known that teams have dramatically different styles depending on the circumstances of the game. For example, in "garbage time", a team will stop throwing the ball when they lead by an insurmountable margin. To address these less competitive situations, we use the win probability calculation in nflscrapR (Horowitz, Yurko and Ventura 2020), and we omit plays where the win probability falls outside the range $(0.1, 0.9)$. This further reduces the number of pass attempts and non-designed runs in our dataset to 5276.

3.2 Covariates

A core problem in the development of our methods involves the estimation of the success probability p_{ij} corresponding to the jth option on the ith play. Our modeling will attempt to capture the covariates that influence the completion of a pass attempt. Since the eventual goal concerns quarterback evaluation involving decision making, completion probability is assessed at the time the ball is released rather than when it arrives. Therefore, some variables that are relevant at the time when the ball arrives (e.g. receiver separation from defenders) will be estimated at the time of release.

Previous work (Next Gen Stats Team 2018) has explored the modeling of completion probability. Their work highlights the relationship between factors such as pass air distance, air yards, receiver separation, pass rush separation, and the speed of the quarterback at release. There are other covariates included in their modeling but these have not been publicly disclosed. Unfortunately, many of the modeling details remain proprietary and cannot be reviewed.

Deshpande and Evans (2020) also model completion probability. They leverage a collection of factors including receiver separation from the nearest defender and from the ball, receiver movement vectors, and cumulative distance covered by the receiver during the game. These covariates are essentially doubled up, being measured both at the time of release and at the estimated time of pass arrival. Their model which is based on Bayesian additive regression trees generates 90% prediction accuracy.

We use similar covariates to the aforementioned work with a few additions. However, we emphasize that we only make use of covariates that were measurable at the time of the throw since we wish to focus on quarterback decision making. We now introduce the covariates such that for every play i and option j, there is a specified potential receiver. For the time being, we omit running options.

3.2.1 Football covariates

The two football covariates that we consider are similar to those used in previous completion probability models. The first football covariate is air distance. Given the intended receiver, we calculate the Euclidean distance that the football needs to travel. This is a simple measurement between final ball location and quarterback location at the time of the pass. Intuitively, longer passes have lower probabilities of completion. The final ball location is estimated using the receiver velocity and the ball velocity. We use a fixed value of 20 yards per second for all ball velocity calculations.

The second covariate is yards downfield. This is similar to air distance but only considers yardline distance. The covariate adds football context as the number of yards gained is relevant to scoring. Also, the probability of a pass completion may depend on the angle that the ball is thrown. For example, an angled pass 10 yards to the side of the quarterback typically has a higher pass completion probability than a 10 yard pass directly downfield.

3.2.2 Receiver covariates

Generally, the more open the receiver, the higher the completion probability. We attempt to characterize openness with three covariates. The first two are similar to those in other completion probability models whereas the remaining covariate is novel.

The first covariate is receiver separation from the nearest defender. This is obtained by calculating the minimum Euclidean distance between the receiver and all players on defence at the time that the pass is initiated.

A second covariate is the sideline separation distance at the time of release. A pass is complete only if the receiver establishes control of the ball inbounds and the sideline is used to mark the edge of the inbounds surface. If there is little space along the sideline, this reduces the completion probability.

Although receiver separation provides information on openness, we also introduce a field ownership metric which utilizes the positions and velocities of receivers and defenders. The resultant covariate extends the notion of receiver separation beyond the consideration of a single defender. The field ownership metric is adapted using ideas from Fernandez and Bornn (2018) which were developed for soccer. We begin by estimating the probability densities of the location of players at the time of ball arrival. The densities are based on kinesiological ideas such as the recognition that it is more difficult for players to change directions at higher speeds. A team's ownership at a given location is then the sum of the individual densities for that team's players at that location. Influence at a given location is then calibrated on the interval $[0, 1]$ where a value of 0.5 is interpreted as equal location ownership by both teams. An owned cell by the offensive team is one for which influence > 0.5. The influence measure is then used to generate the covariate capturing the total influence of cells owned by the offense within five yards of the estimated ball arrival location.

3.2.3 Quarterback covariates

The success of a passing play depends on more than just the receiver and his ability to get open. In addition, there is a reliance on the offensive line to provide ample time for the quarterback while also minimizing required quarterback movement. We aim to capture these notions via the four following quarterback covariates which are similar to existing covariates in the literature. Calculation of the covariates is done on a frame by frame basis to assess hypothetical passes.

We define the covariate rush separation as the Euclidean distance between the quarterback and the nearest defensive opponent. This accounts solely for physical closeness and does not consider the estimated time it takes the defender to reach the quarterback.

We also measure the time to throw covariate which is the time from the snap to the current observed frame. Generally, a quarterback is under more duress as time progresses.

The quarterback speed covariate is estimated from his change in position between the current frame and the frame observed 0.5 seconds prior. It is generally more difficult for a quarterback to complete a pass when he is moving faster.

Finally, the distance from the pocket covariate uses the positioning of the quarterback relative to a 7 yard by 7 yard square bordering the line of scrimmage. The covariate is set to 0 if the quarterback is within the pocket; otherwise it is the minimum distance for the quarterback to re-enter the pocket. The intuition is that it is easier to make a pass from the pocket.

3.3 Modeling and Estimation

3.3.1 Estimation of the p_{ij}'s

The estimation of the completion probabilities p_{ij} requires a statistical learning approach that is flexible (e.g. non-linear) to accommodate the non-linearity and multicollinearity of the covariates. We utilize a Stacking algorithm built on an ensemble of base learners including random forests, gradient boosting, general linear models, logistic regression, neural networks and naive Bayes. At the super learner level we incorporate a gradient boosting model. This treats the cross-validated predictions generated by the base learners as covariates (van der Laan, Polley and Hubbard 2007). Although there are other choices at the super learner level, we found that gradient boosting offers the best predictive performance for our problem (Naimi and Balzer 2017). Note that the prediction exercise is more challenging in our context where covariates were obtained at the time of the throw rather than at the time of arrival of the pass.

3.3.2 Estimation of yards gained after the catch

To model the yards gained after the catch, we restrict the dataset to the 3933 instances where the pass was completed. In addition to original covariates used in the completion probability model (Section 3.3.1), we introduce two new covariates that describe the presence of tacklers "downfield" where downfield encompasses all defenders beyond the receiver. The first covariate estimates the distance of the nearest downfield defender to the intended receiver at the time of ball arrival. This is based on the velocities of the two players and the average speed of a pass. The second covariate is the estimated number of defenders downfield at the time of ball arrival. With more separation from the nearest downfield defender and fewer tacklers downfield, there is an expectation of a greater number of yards after the catch.

We use the same class of base learners as in the completion probability model (Section 3.3.1) with slight modifications for a regression task rather than a classification task. We use a non-negative generalized linear model (GLM) as the super learner which combines the base learners. The root mean squared error corresponding to the fitted yards after the catch compared to the actual yards after the catch is 2.96 yards.

3.3.3 Estimation of yards gained from non-designed runs

Non-designed quarterback runs make up a small proportion of our observed plays (only 252 plays). Therefore, building a training and testing set to assess model fit would likely lead to overfitting. Instead, we treat the yards gained from non-designed quarterback runs as similar to yards gained after the catch, and we derive our estimates from the respective model. The root mean squared error corresponding to these plays is 3.99 yards.

3.3.4 Handling interceptions

Modeling thus far has considered a pass outcome as binary - either a completion or an incompletion. This was formulated with interceptions treated as incomplete passes. Although this is sensible from the perspective of estimating completion probability, it is inadequate to equate incompletions with interceptions in terms of EPV. Generally, an interception is far more damaging to the offensive team than an incompletion.

The introduction of interceptions complicates the simple formulation (2.1) involving the optimal expected points gained on the ith play. Denote $\hat{q}_{ij}$ as the estimated probability of an interception corresponding to passing option j on play i. Then equation (2.1) is modified by replacing the jth term $\hat{p}_{ij}G_{ij}$ in (2.1) by

$$\hat{p}_{ij}G_{ij}^{(\text{comp})} + \hat{q}_{ij}G_{ij}^{(\text{int})}$$

where $G_{ij}^{(\text{comp})}$ is new notation for the expected points gained from a completion and $G_{ij}^{(\text{int})}$ is the expected points gained from an interception. Note that the expected points corresponding to an incompletion is constant across all options on a given play and as such is absent from this equation.

With our restricted dataset involving only 158 interceptions, it is challenging to estimate the probabilities q_{ij} of an interception with a comprehensive categorical model that includes completions, incompletions and interceptions. For this reason, we analyze interceptions separately using the same approach and covariates as in the completion probability model of Section 3.3.1.

Due to the lack of data, it is also difficult to reliably estimate yards gained after an interception. Therefore, we assign no yards gained following an interception. Although this is an unrealistic assumption, we note that interceptions are rare events where the probabilities q_{ij} are small and do not affect Y_i in (2.1) greatly. With more data, yards gained after an interception could be better estimated with a larger dataset using the ideas from Section 3.3.2.

The same principles can be applied for the analysis of quarterback fumbles for non-designed runs in Section 3.3.3. Fumbles on non-designed quarterback runs are even more rare in our dataset with only a single occurrence resulting from 252 runs. For the time being, we omit the consideration of this possibility.

3.4 Validation

For the completion probability model (Section 3.3.1), we randomly split the data into a training set (85%) and a validation set (15%) where base learners and weights were determined using 10-fold cross-validation on the training data. Recall that a gradient boosting super learner was utilized. Model performance was then tested on the held-out validation data which generated an accuracy rate of 75.4% using $p = 0.5$ as the cutoff for classification.

For the yards after the catch model (Section 3.3.2), we again randomly split the data into a training set (85%) and a validation set (15%) where base learners and weights were determined using 10-fold cross-validation on the training data. Recall that a non-negative GLM super learner was utilized. Model performance was then tested on the held-out validation data which generated a root mean squared error of 3.64 yards.

Chapter 4

Results

4.1 Using Evaluation Criterion Q

Using the proposed models, we predict the completion probability and the yards gained after the catch for each option on all passing plays. Then using the EPV tables, this permits the calculation of the quarterback execution metric Q given by (2.2).

To provide some additional insight, we calculate Q under two conditions to highlight the impact of mobile quarterbacks through non-designed quarterback runs:

- Q_1: non-designed runs removed from the dataset
- Q_2: all potential passing plays (i.e. pass plays and non-designed runs)

In Table 4.1, we report the statistics Q_1 and Q_2 for the 29 quarterbacks who had at least 100 potential passing plays and a valid NFL Passer Rating [1] in the first six weeks of the 2017 NFL season. The statistic Q_1 corresponds to pure passing whereas the statistic Q_2 incorporates both passing and running. One of our first observations from Table 4.1 is that there is some disagreement between Q_1 and the NFL Passer Rating. If we look at the six teams who had quarterbacks with passer ratings exceeding 100, we observe that these teams had fast starts in 2017. Specifically, after the first six weeks of the season, Kansas City was 5-0, Philadelphia was 5-1, New England was 4-2, New Orleans was 3-2 and the LA Rams were 4-2. This is again suggestive that the NFL Passer Rating is partially a function of team success rather than pure quarterback performance. On the other hand, our statistic Q_1 incorporates performance with decision making. We see that the top quarterback according to pure passing is Dak Prescott with $Q_1 = 44.5$ and at the bottom of the list is DeShone Kizer with $Q_1 = 24.5$. With Dak Prescott, the interpretation of the statistic Q_1 is that over the first six weeks of the 2017 NFL season, in pure passing plays, his EPV contribution was 44.5% of the hypothetical quarterback

[1] Only Brian Hoyer of the otherwise valid quarterbacks falls below the threshold for attempts per game set by Pro Football Reference

who made optimal decisions on every play. We also observe that Q_1 does not correlate strongly with the NFL Passer Rating ($r = 0.51$).

When we look at the overall quarterback rating Q_2 in Table 4.1 which includes non-designed runs, we observe that Russell Wilson has the greatest increase in Q_2 over Q_1. This corresponds to the widespread opinion that Russell Wilson has great value as a scrambling quarterback. It is probably surprising to many football fans to see that Eli Manning's Q_2 statistic also suggests that he makes valuable runs. We bear in mind that we have a limited dataset, and in the first six weeks of the 2017 season, we have only three recorded Eli Manning runs. Generally, the differences between Q_1 and Q_2 are not great, and this demonstrates that passing (decision making and execution) remains the fundamental contribution of quarterbacks.

The calculations obtained in Table 4.1 were based on the allowance of an additional $\epsilon = 0.5$ seconds from the time of the release of the pass or until the quarterback had passed the line of scrimmage in a non-designed run (see Section 2). Although this is a feature of the methods, we need to be sensitive to the reality that prediction of potential player actions beyond $\epsilon = 0$ seconds becomes increasingly difficult for larger ϵ. We therefore repeated the analyses in Table 4.1 using $\epsilon = 0$ seconds and found that the sample correlation using the two evaluations was $r = 0.99$ for Q_1 and $r = 0.99$ for Q_2.

4.2 Using Modified Evaluation Criterion

In the calculation of our original evaluation criterion Q, we allow for the estimate to be built on the observed results in relation to an average quarterback who makes optimal decisions. This comparison establishes a consistent connection with other outcome-based metrics currently in use by the NFL and its affiliates such as Passer Rating.

It is possible to modify the evaluation criterion Q to move away from the current outcome-based metrics and establish a process-based alternative. Define A_i^* to be the estimated value associated with the decision made on play i at the time of the throw, where the estimated value is calculated as in section 3.3.4. Note that A_i^* is independent of the play result: it has the same value whether the pass was caught, dropped, fumbled, or intercepted. We then create a similarly modified evaluation criterion Q^*, which we define as

$$Q^* = \left(\frac{\sum_{i=1}^{N} A_i^*}{\sum_{i=1}^{N} Y_i} \right) 100\%$$

Naturally, we calculate Q_1^* and Q_2^* in this process-based framework. The results are presented in Table 4.2. We retain $\epsilon = 0.5$ for comparison purposes with our original criterion.

One situation in which we most closely observe the differences between Q and Q^* is in the inclusion of running plays. We have oft referred to Russell Wilson, quarterback for the Seattle Seahawks in 2017, as an example of a player who adds value through non-designed runs. We note this behaviour in his evaluation of $Q_2 > Q_1$. Further his difference in value between Q_2

and Q_1 of 2.1 marks the largest single improvement throughout our collection of quarterbacks. There were 12 quarterbacks that observed a larger Q_2 than Q_1.

If we instead consider this from the perspective of Q_1^* and Q_2^*, we find a different landscape in which only two quarterbacks improved their rating when non-designed runs were included. Neither of these quarterbacks is Russell Wilson. Rather, Russell Wilson observes a discrepancy of -9.4 going from Q_1^* to Q_2^*. We note that much like many other quarterbacks in our sample, very few plays for Russell Wilson have rushing options that offer an expected value similar to that of at least one of the passing options on a given play. The infrequent occurrences in which rushing offers similar values is on short yardage passing plays or goal-line situations. These situations offer an artificial cap on the expected value of the optimal decision for the play, thereby allowing a rushing decision to look better than in other plays. This is not to say that choosing to rush on a play is a bad decision, rather that the play is frequently estimated to be suboptimal relative to passing the ball in terms of expected value added. We reconcile this notion by reminding the reader that our original metric observes quarterbacks adding value through running, highlighting the importance of execution on running decisions.

QB	Team	# Plays	Q_1	Q_2	Passer Rating
D Prescott	Dallas	140	44.5	43.9	86.6
K Cousins	Washington	154	42.8	43.4	93.8
J Winston	Tampa Bay	118	40.4	40.4	92.2
A Smith	Kansas City	196	39.9	39.3	104.7
M Ryan	Atlanta	164	39.5	39.7	91.4
D Carr	Oakland	117	39.4	39.3	86.4
C Wentz	Philadelphia	205	38.6	38.1	101.9
T Brady	New England	180	38.5	38.0	102.8
J McCown	NY Jets	179	37.8	38.0	94.5
P Rivers	San Diego	214	37.3	37.3	96.0
A Dalton	Cincinnati	135	37.1	36.2	86.6
D Brees	New Orleans	114	36.4	36.4	103.9
B Roethlisberger	Pittsburgh	193	35.7	35.1	93.4
C Keenum	Minnesota	136	35.0	36.0	98.3
E Manning	NY Giants	201	34.5	35.3	80.4
C Newton	Carolina	184	34.0	34.0	80.7
J Goff	LA Rams	170	33.4	33.3	100.5
T Siemian	Denver	162	32.0	31.5	73.3
A Rodgers	Green Bay	164	31.7	31.5	97.2
M Mariota	Tennessee	128	31.6	31.0	79.3
M Stafford	Detroit	182	31.2	31.3	99.3
J Brissett	Indianapolis	166	30.9	31.8	81.7
R Wilson	Seattle	179	28.9	31.0	95.4
T Taylor	Buffalo	163	28.8	28.3	89.2
C Palmer	Arizona	190	28.6	28.5	84.5
B Bortles	Jacksonville	129	28.1	28.8	84.7
J Flacco	Baltimore	146	26.0	26.0	80.4
J Cutler	Miami	126	25.6	26.4	80.8
D Kizer	Cleveland	128	24.5	24.8	60.5

Table 4.1: NFL Passer Ratings and rankings based on the Q metrics for the first six weeks of the 2017 NFL season.

QB	Team	# Plays	Q_1^*	Q_2^*	Passer Rating
K Cousins	Washington	154	56.2	46.9	93.8
P Rivers	San Diego	214	55.9	52.1	96.0
C Newton	Carolina	184	55.2	47.0	80.7
C Keenum	Minnesota	136	55.0	51.1	98.3
M Mariota	Tennessee	128	54.9	46.5	79.3
R Wilson	Seattle	179	54.7	45.3	95.4
D Prescott	Dallas	140	54.6	48.6	86.6
M Ryan	Atlanta	164	54.4	47.9	91.4
J Winston	Tampa Bay	118	54.1	48.2	92.2
J Brissett	Indianapolis	166	53.8	49.5	81.7
D Brees	New Orleans	114	53.6	47.9	103.9
E Manning	NY Giants	201	53.0	44.2	80.4
J Goff	LA Rams	170	52.7	45.5	100.5
T Brady	New England	180	52.6	51.7	102.8
J Flacco	Baltimore	146	52.5	46.6	80.4
B Roethlisberger	Pittsburgh	193	51.9	52.6	93.4
A Rodgers	Green Bay	164	51.8	46.9	97.2
J Cutler	Miami	126	51.6	42.2	80.8
J McCown	NY Jets	179	51.5	47.3	94.5
B Bortles	Jacksonville	129	51.5	44.4	84.7
T Taylor	Buffalo	163	51.1	46.6	89.2
A Dalton	Cincinnati	135	50.9	47.4	86.6
C Wentz	Philadelphia	205	50.6	46.2	101.9
M Stafford	Detroit	182	50.1	45.8	99.3
T Siemian	Denver	162	48.8	45.2	73.3
D Carr	Oakland	117	48.1	51.3	86.4
A Smith	Kansas City	196	47.9	41.7	104.7
D Kizer	Cleveland	128	47.8	45.1	60.5
C Palmer	Arizona	190	47.7	45.7	84.5

Table 4.2: NFL Passer Ratings and rankings based on the Q^* metrics for the first six weeks of the 2017 NFL season.

Chapter 5

Discussion

Adjusting from Q to Q^* requires a re-framing of our original question. We switch from evaluating the execution of a quarterback's decisions against an average quarterback who is making optimal decisions to evaluating the decisions made from among a collection of possible decisions. Both approaches have merit with the former expressing results more closely coinciding with observable play and the latter expressing results more closely controlled for disparities of talent at other team positions.

To further explore the differences between our execution based and our purely decision based metrics, we consider the following figures. Figure 5.1 demonstrates the differences between Q_1 and Q_1^*. Figure 5.2 demonstrates the differences between Q_2 and Q_2^*.

The largest discrepancies in each of these figures exists for Derek Carr and Alex Smith. Both of these quarterbacks rank in the top 10 with respect to our original metric while ranking in the bottom 5 for our modified evaluation criterion. Their deviations are not trivial to map back to a singular root cause. Instead, these deviations may be functions of a player's decision making, a coach's limits placed on the player, or the game's situation in which the pass existed. Although we filter out extreme win probability situations, there are still many situations that remain where targeting the estimated maximum value target is not necessarily optimal from a coaching perspective.

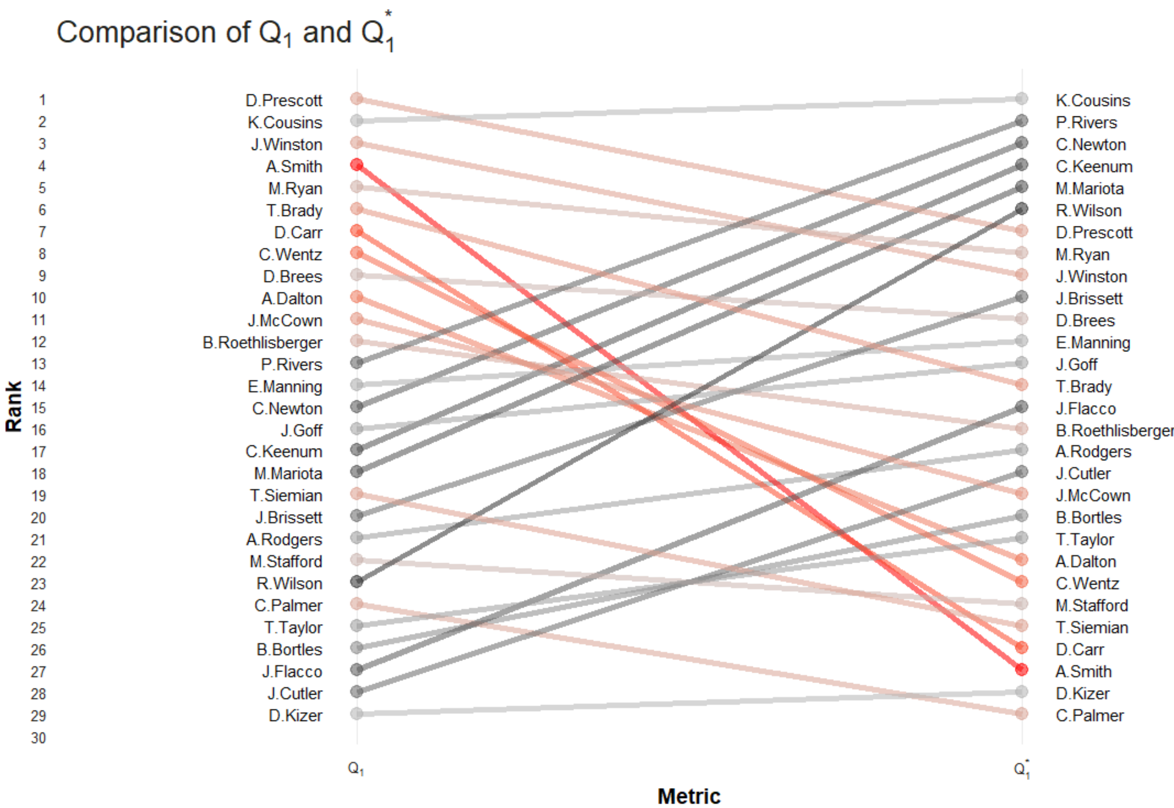

Figure 5.1: Player ranks by Q_1 and Q_1^* values. Colour intensity is proportional to rank increase (black) or decrease (red) going from Q_1 to Q_1^*

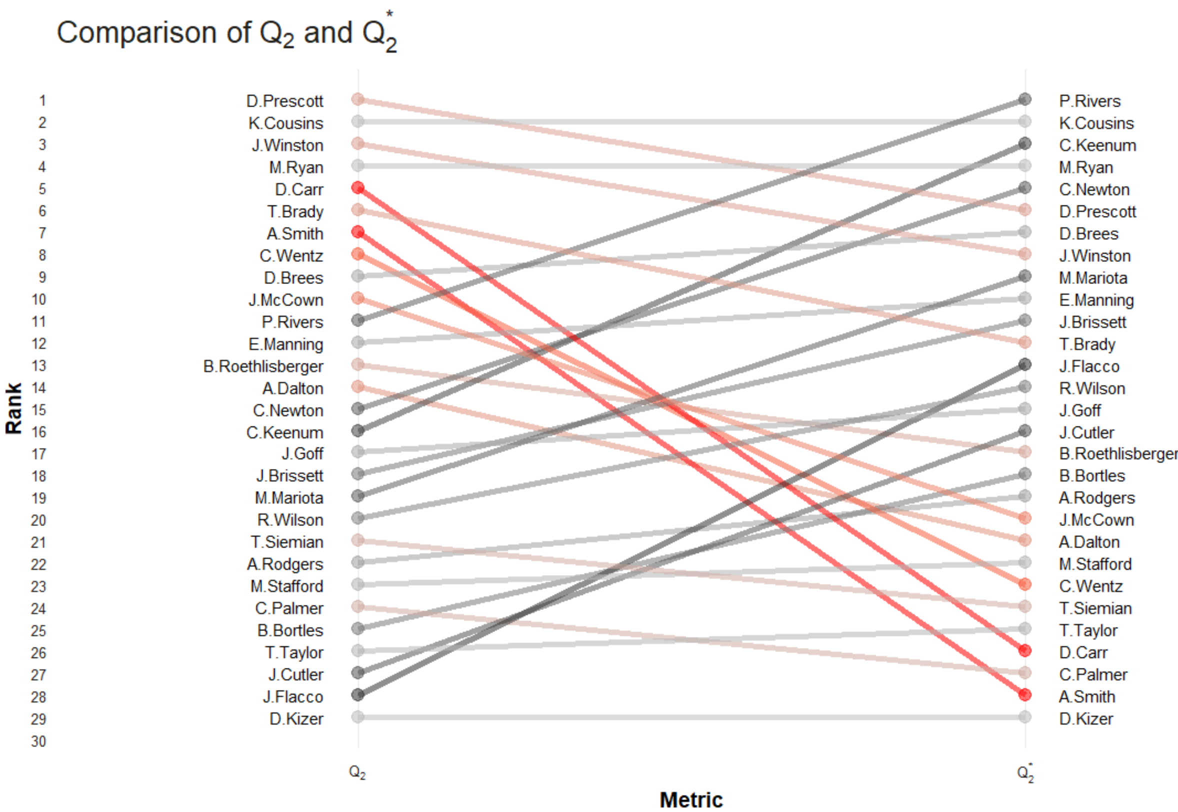

Figure 5.2: Player ranks by Q_2 and Q_2^* values. Colour intensity is proportional to rank increase (black) or decrease (red) going from Q_2 to Q_2^*

Chapter 6

Conclusion

In the NFL, the quarterback is generally regarded as the most important player on a team. The quarterback touches the ball on nearly every offensive possession and his decision making is critical to team success. Yet, the way that quarterbacks are evaluated in the media is not nuanced. Generally, their assessment is determined by basic match statistics.

This paper attempts to use the rich potential of spatio-temporal data to evaluate quarterbacks at a deeper level. The player tracking data used in this analysis considers the locations and velocities of all players on the field in increments of 0.1 seconds. With this wealth of information, we develop interpretable statistics that are based on what a quarterback actually did compared to what they might have done. The statistics use machine learning techniques for the primary purpose of predicting what might have happened had the quarterback chosen a different option. We are not suggesting that our statistics ought to become the standard for quarterback evaluation. Rather, we suggest that they provide a nuanced view involving decision making where quarterbacks on weaker teams are provided a more balanced appraisal.

Although we believe that Tables 4.1 and 4.2 are interesting, we recognize that these tables are based on only six weeks of available data during the 2017 regular season of the NFL. The main purpose of the paper is to explore the possibilities involving quarterback evaluation. Accordingly, there are both limitations and potential future research directions associated with our work.

One limitation that we do not know how to resolve is that quarterbacks are sometimes limited in their freedom to make decisions. Therefore, it is not genuine that all options evaluated by our statistic Q in (2.2) are realistic options. It may be the case that coaches provide experienced quarterbacks more leeway in decision making than inexperienced quarterbacks. Therefore, it might be argued that the statistics developed in this paper are also a function of coaching. Another limitation of the methods is that we have not provided standard errors associated with the statistics. With larger datasets, this may be remedied by some sort of bootstrapping procedure.

For future research, we see various potential enhancements and extensions. First, a greater exploration of ϵ outlined in Section 2 could be investigated. Recall that ϵ is the amount of time that we consider after a pass attempt to assess alternative quarterback options. Another avenue for future work is the consideration of player specific traits. Currently, for example, the catch probability model is based on the concept of an average receiver. A quarterback's decision making may change depending on the quality of a potential receiver. Additionally the data available for this project pre-dates some of the NFL's most prolific running quarterbacks such as Lamar Jackson, Deshaun Watson, and Josh Allen. Given the quality of running quarterbacks now in the league we may be able to achieve better estimates of running ability and, subsequently, different results in comparing Q_1 and Q_2. Finally, individual quarterback performance in the plays that fall beyond our win probability interval of $(0.1, 0.9)$ are not identical. Some teams and quarterbacks have a reputation for excellence in desperate comeback situations. We have ignored these possibilities within this work due to data size restrictions. In a larger sample space, exploring alternatives to this exclusion may offer a more robust understanding of quarterback decision making and execution.

References

Albert, J.A., Glickman, M.E., Swartz, T.B. and Koning, R.H., Editors (2017). *Handbook of Statistical Methods and Analyses in Sports*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, Boca Raton.

Burke, B. (2019). "DeepQB: Deep learning with player tracking to quantify quarterback decision-making and performance", *MIT Sloan Analytics Conference*, retrieved May 18/20 at `www.sloansportsconference.com/wp-content/uploads/2019/02/DeepQB.pdf`

Chu, D., Reyers, M., Thomson, J. and Wu, L.Y. (2020). "Route identification in the National Football League", *Journal of Quantitative Analysis in Sports*, 16, 121-132.

Deshpande, S. and Evans, K. (2020). "Expected hypothetical completion probability", *Journal of Quantitative Analysis in Sports*, To appear.

Fernandez, J. and Bornn, L. (2018). Wide open spaces: A statistical technique for measuring space creation in professional soccer. In *12th Sloan Sports Analytics Conference*, retrieved May 14, 2020 at `www.sloansportsconference.com/wp-content/uploads/2018/03/1003.pdf`

Gough, C. (2018). "National Football League (NFL) - Statistics and facts", *Statista*, retrieved March 18/20 at `https://www.statista.com/topics/963/national-football-league/`

Horowitz, M., Yurko, R. and Ventura, S.M. (2020). "nflscrapR: Compiling the NFL play-by-play API for easy use in R", R package version 1.8.3, `https://github.com/maksimhorowitz/nflscrapR`.

Naimi, A. and Balzer, L. (2017). "Stacked generalization: An introduction to super learning", retrieved June 26/20 at `http://www.jstatsoft.org/v61/io8/`.

Next Gen Stats Team (2018). "Next Gen Stats introduction to completion probability", *NGS Photo Essays*, retrieved May 1/20 at `www.nfl.com/news/story/0ap3000000964655/article/nextgen-stats-introduction-to-completion-probability`

Raul (editor), (2016). "Which professional sports leagues make the most money", *howmuch.net*, retrieved March 18/20 at `https://howmuch.net/articles/sports-leagues-by-revenue`

Reyers, M.J. and Swartz, T.B. "Quarterback Evaluation in the National Football League". *Submitted.*

van der Laan, M.J., Polley, E.C. and Hubbard, A.E. (2007). "Super Learner", *UC Berkeley Division of Biostatistics Working Paper Series*, Working Paper 222, retrieved June 26/20 at `https://biostats.bepress.com/ucbbiostat/paper222`.

Yam, D.R. and Lopez, M.J. (2020). "What was lost? A causal estimate of fourth down behavior in the National Football League", *Journal of Sports Analytics*, 5, 153-167.

Yurko, R., Matano, F., Richardson, L.F., Granered, N., Pospisil, T., Pelechrinis, K. and Ventura, S. (2020). "Going deep: models for continuous-time within-play valuation of game outcomes in American football with tracking data", *Journal of Quantitative Analysis in Sports*, 16, 163-182.

Yurko, R., Ventura, S. and Horowitz, M. (2019). "nflWAR: a reproducible method for offensive player evaluation in football", *Journal of Quantitative Analysis in Sports*, 15, 163-183.

Zilavy, G. (2018). "How to calculate NFL passer rating using a formula in Excel or Google Sheets", *Medium Data Science*, retrieved March 18/20 at `https://medium.com/@gzil/how-to-calculate-nfl-passer-rating-using-a-formula-in-excel-or-google-sheets-54eb07246d1e`