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 Committee Member Professor
	Harsha Perera External 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

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 perfor-

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 $t \in (0, t_i + \delta]$, 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 δ 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 $\delta = 0.5$ 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

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

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

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

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 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.

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.

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 ^

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

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 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

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

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.

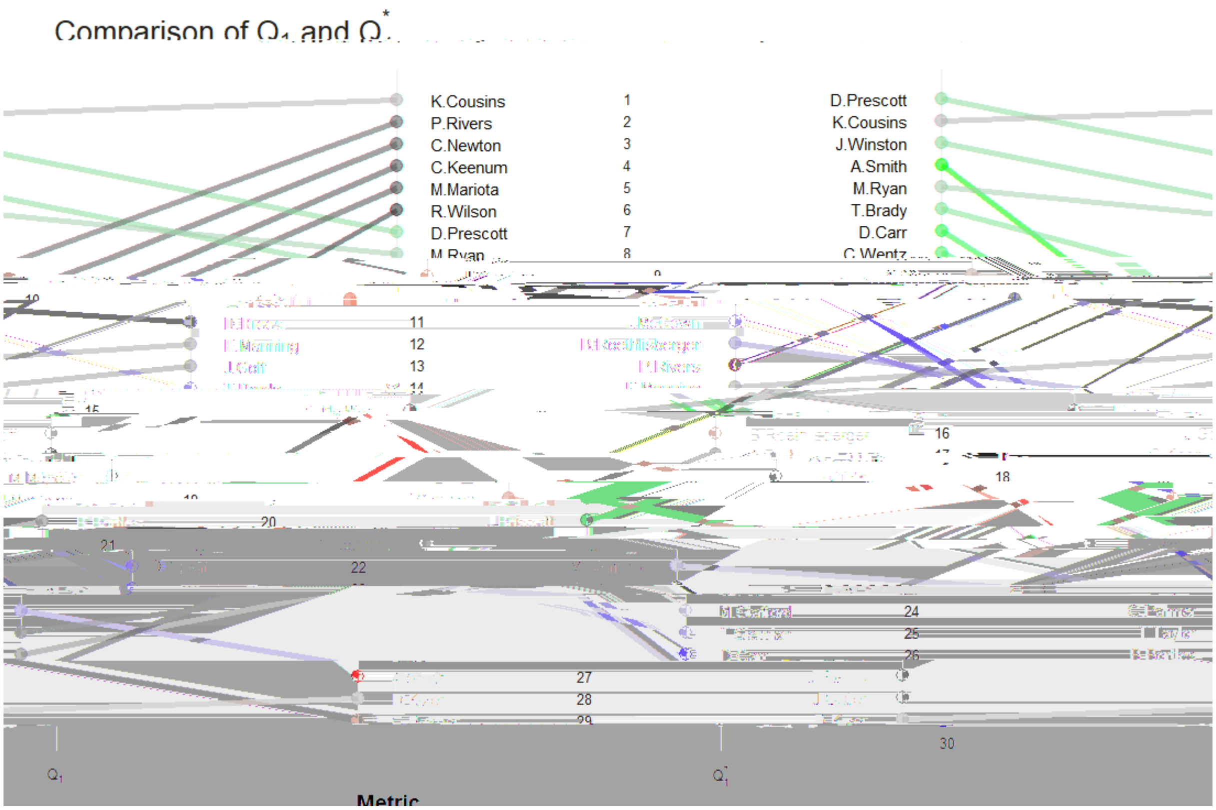

Figure 5.1: Player ranks by Q_1 and Q_1 values. Colour intensity is proportional to rank increase (black) or decrease (red) going from Q_1 to Q_1

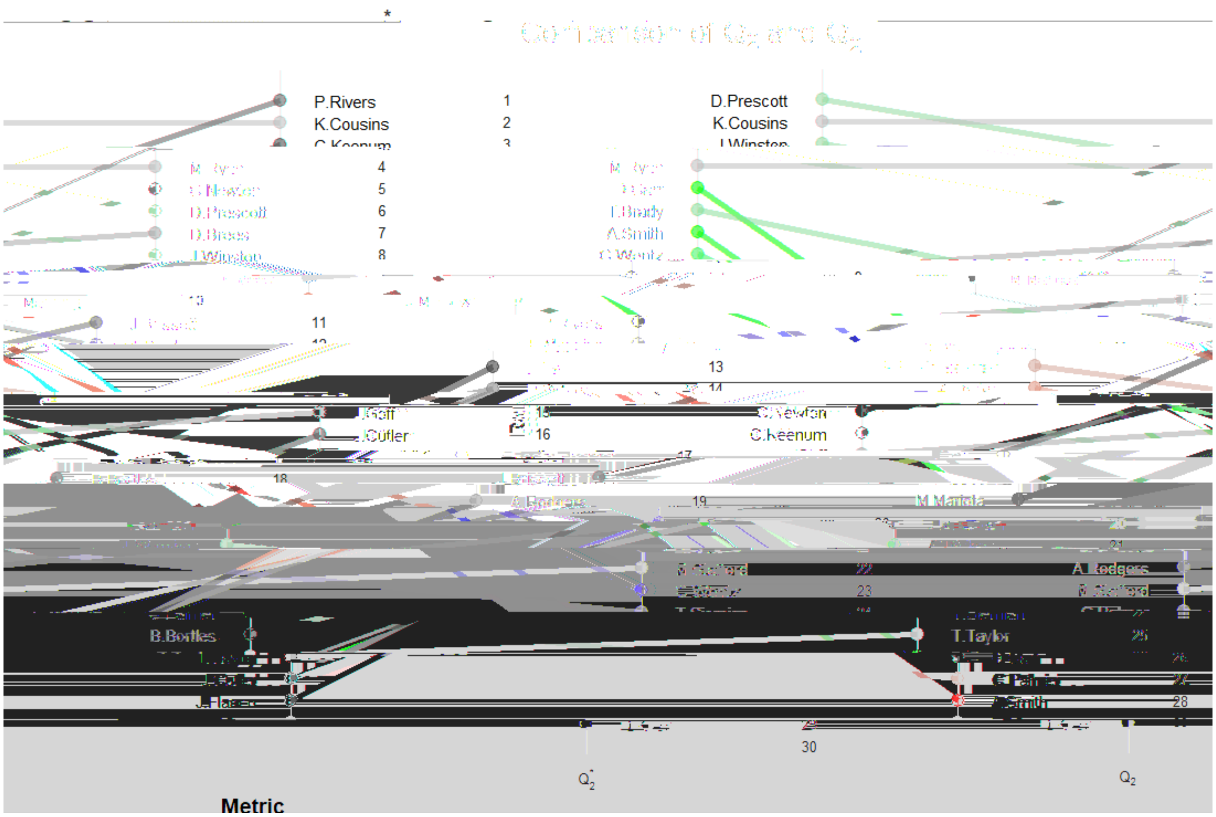

Figure 5.2: Player ranks by Q_2 and Q_2 values. Colour intensity is proportional to rank increase

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 outlined in Section 2 could be investigated. Recall that 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

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

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