Spring 2014

Quarterback Passer Rating System: Accessible for All Who Care

McKenna Mettling
Regis University

QUARTERBACK PASSER RATING SYSTEM: ACCESSIBLE $\forall$ WHO CARE

A thesis submitted to
Regis College
The Honors Program
in partial fulfillment of the requirements
for Graduation with Honors
by
McKenna Mettling

May 2014

Contents

Chapter 1

Introduction

In a society where everything now appears to be a competition, we are constantly looking for a way to make comparisons on who or what is the best. In order to make these comparisons, statisticians have turned to the use of statistics to solve problems in many different fields. Although statistics is prominently used in the fields of finance, insurance, and business; we have seen a growing interest in understanding statistics in the fields of marketing and sports [2]. With marketing we have seen a growing need to analyze statistics in order to determine what customers are interested in, which would allow companies to determine what items they should sell. Although marketing continues to create new and interesting reasons for studying statistics, we prefer to focus on the developing need for statistics in the sports industry.

Statistics have been used for many years in sports in order to provide players with an idea of how well they did during a game; however, baseball, with the use of sabermetrics, was the first sport to take the statistics and use them in a way to get a better sense of what they needed to do in order to win. Due to the recent release of the book and movie *Moneyball* we have seen how statistics can be used in sporting events to develop the teams by picking players based on their statistics. While baseball uses statistics as a way to develop their teams, most of the other sports industries use statistics as a way to better entertain the fans.

This is especially noticeable in the sport football. There have been many statistics systems developed in order for fans to better understand the game, as well as to allow fans to participate in the game through events like fantasy football. Consequently, the National Football League (NFL) continues to work on developing different models that fans and even the coaches can use to better understand the game and the capability of the players. Due to the growing desire to better determine the ability of the players, we need to look at the current system in place for rating quarterbacks so that we can determine its accessibility as well as test a new system that we hope to develop.

Chapter 2

Current Passer Rating System

Although many people believe that the NFL has only just started using statistics for monitoring players ability, the first quarterback passer rating system was used from 1960 until the current system was officially adopted in 1973 by the NFL. The current system that the NFL continues to use today was developed by a special committee led by Dan Smith of the Pro Football Hall of Fame, Seymour Siwoff of the Elias Sports Bureau, and Don Weiss of the NFL. The purpose of this system was to measure the pass efficiency of the quarterbacks in the NFL and to provide a way to compare how a player's performance varied from one season to the next [11]. Admittedly, the system developed by Smith, Siwoff, and Weiss is not the only system currently in use because there are also systems like the systems developed by National Collegiate Athletics Association (NCAA) and ESPN.

In spite of the fact that both the NFL and the NCAA systems used the same stats to develop the models that they use today, they have produced two strikingly different formulas for calculating the passer rating. The stats that both the NFL and NCAA are using include the passing yards (YDS), the number of completions (CP), the number of passing attempts (ATT), the number of touchdown passes (TD), and the number of interceptions thrown (INT) per season. Although both systems use the same statistics, the systems are quite different in how they implement the statistics.

The current NFL system uses a five step formula for determining players' passer ratings where each step looks at a particular aspect of the quarterback's game. The first piece looks at the player's completions per attempt (CP/ATT), the second piece is determined by the yards per attempt (YDS/ATT), then the next piece is determined by the touchdowns per attempt (TD/ATT), after that the fourth piece is determined by the interceptions per attempt (INT/ATT), and in the final piece we combine all of the previous parts. This system also has the stipulation that each step is truncated between 0 and 2.375, therefore the formula is non-linear even though the equation is. Here is the complete formula for the NFL:

$$a = \frac{CP/ATT - 0.3}{0.2}$$
$$b = \frac{YDS/ATT - 3}{4}$$
$$c = \frac{TD/ATT}{0.05}$$
$$d = \frac{0.095 - INT/ATT}{0.04}$$
$$Rate = \frac{a + b + c + d}{6} * 100$$

In contrast, the system for the NCAA was developed as a one step formula. In this system we first form a linear combination of the YDS, the TDs, the CPs, and the INTs so we can then take the total and divide by the ATT. Here is the simpler formula for the NCAA :

$$Rate = \frac{8.4 * YDS + 330 * TD + 100 * CP - 200 * INT}{ATT}$$

These systems appear to be beneficial since both provide a way to use multiple statistics in order to come up with a passer rating that will better evaluate the player as a whole. We have also been using these systems for many years now, so we know that they provide a quarterback rating that is fairly understandable. Another benefit that the current system appears to have is that it has a positive correlation between the teams that have quarterbacks with high ratings and the teams which have a high win percentage. Nonetheless, both of these systems are not without their flaws.

In order to be more accessible, we need to find a balance between the complexity of the NFL system and the simplicity of the NCAA system in order to find a system that is easier to compute. Also, both formulas have elements that are generally confusing to most. For example, why are the systems bounded and what is the reasoning behind these specific truncations? We hope to determine if we can find a system that is less confusing and more straightforward for the fans and teams to use. Regarding the issue of why the NFL committee chose to have each step produce a value that is bounded between 0 and 2.375, we have found that there is very little research on how the current system for the NFL was developed . It is understandable that they would want to bound the formula so that they could reduce range between quarterbacks and possibly remove any outliers, but what is still hard to understand is why they didn't choose to make the highest rate you can get 100 or 200 rather than the rather arbitrary 158.3, which is the current highest achievable rate.

We also must consider the issue that since we are using a different system for the NFL than the NCAA, we do not have a consistent way of measuring the readiness of a quarterback who is transitioning into the NFL. There still does not appear to be a consistent system for selecting and evaluating a player's ability besides a coach's opinion or a joint decision with scouts [7]. If we can develop a system that can be used universally, then we will have a better way of predicting how a rookie will do in their first year.

The other issue with these systems is that we are not actually finding a rating of the quarterback, but rather we are determining the efficiency of the quarterback as a passer. There are many other elements that shape a quarterback as a whole rather than just a passer including statistics like rushing yards, sacks, and fumbles. The difficulty with sacks and fumbles is that they may be difficult to calculate since there are more players involved besides just the quarterback; yet, we should at least consider including rushing yards in the system since we seem to be changing to a league where the quarterback is willing to rush as much as he throws. We need to be able to evaluate a quarterback by his complete skill set rather than just focusing on how he does as a passer.

As a result of this, there have been some new systems developed with the idea of evaluating the whole quarterback, but the NFL and the NCAA have the only systems that are officially used. One of the models that has been developed is the ESPN's Total Quarterback Rating (Total QBR) which looks at the quarterback as a whole instead of just as a passer. In *The New York Times* we get a sense of some of the issues this new system has brought up: "Some did not like the inclusion of subjective factors (will a wide receiver be blamed more for an incompletion if Aaron Rodgers is throwing rather than, say, Joe Flacco?)" [13]. Another system has been recently developed by Cold Hard Football Facts. Their system also attempts to evaluate the quarterback's performance as a whole, and although their system appears to be slightly less confusing there are still some aspects of their system that could possibly be improved upon since the subjectivity factor is still a concern.

The last system that we looked at was a system developed by Chris White and Scott Berry using tiered polychotomous regression. This system focuses on the current system in order to test how their results match up with the results from the current system. Although this system appears to be comparable to the current system there are some concerns that come up: "The biggest hurdle that a ranking like ours must overcome is the complexity that is involved" [5]. A system like this could possibly work for the team's research department, but yet many fans like to use these systems and if they are too complicated they may not get used outside of a university statistics department.

That being the case, we have found that each system has its strengths and weaknesses, still we were able to identify a few reoccurring issues. One of the issues is finding a way to be accessible and easy to understand for the fans. We also found that there is an issue with making sure that the statisitcs being used are not subjective and can be used by many. Finally, is there a way to test the quarterback's total performance or do we have to focus only on his ability to pass? We hope to find a way to create a system that will solve many of these issues in order to give the most accessible system possible.

Chapter 3

New Quarterback Rating Systems

After looking further into the research done on the current quarterback rating system, we worked on creating a new quarterback rating system that is accessible and easy to understand. Our goal was to develop two, possibly three, formulas. Originally, we had two sources of data that could have been used, which included a data set from pro-football-reference.com and a data set from the Gamebook Committee of the Professional Football Researchers Association. We had initially chosen to use the game-by-game data that was collected by the Gamebook Committee of the Professional Football Researchers Association which consists of all of the game- by-game data from 1960 to the current season; however, we found that the layout for the data was not as friendly for the mathematical software we were going to use because there was no easy way to transfer the data into R. Thus, we have chosen to use the season-by-season data given from pro-football-reference.com because they have set up the data so that it is easy to export into a text file which is the easiest way to work with data in R.

We used the mathematical software R to find a linear regression based on the current formula in order to develop the first formula which should be able to predict the rate given from the current formula. Once the first formula was developed, then we will work on developing a new formula based on the linear regression looking at the formula where win percentage is the dependent variable since there should be a positive correlation between the quarterback passer rating and win percentage. For both formulas we will be using the independent variables: Completions Per Attempt, Yards Per Attempt, Touchdowns Per Attempt, and Interceptions Per Attempt.

The final formula that we wanted try to develop was similar to the second formula such that the linear regression to develop the formula would have used the dependent variable of win percentage, but the new formula would have looked at the four independent variables we have been testing in the linear regression as well as the variables rushing yards per attempt and fumbles per attempt to see if there was any relevance between these two variables and finding a quarterback rating. The issue that came up with this final formula is that the data set we chose to use only had passing data so it does not include rushing yardage or fumble data so we chose not to develop this last formula since we would have needed to develop a way to include the rushing data and fumble data without having to manually enter the data for each player being tested. In the future, if possible, we would like to be able to develop this third formula because with the football industry today, we do have a lot more rushing quarterbacks entering the league so a rating of the quarterback as a passer might not truly show the efficiency of a quarterback as a whole.

After we developed these new formulas based off of each linear regression, we tested the formulas with the statistics from the current season to see what the predicted ratings would be. Once we found all of the predicted ratings for the formulas, we created a table that contains the name of the players, the current ratings and the new ratings, which we sorted from highest to least highest rating. We then used this table to determine how these ratings compared. This hopefully told us if the current formula is slightly better or if one of the formulas developed might be more accessible to use.

Chapter 4

Analysis and Results

4.1 First Linear Regression Results

	Estimate	Std. Error	t-value	Pr(>$\|t\|$)
(Intercept)	37.03683	0.94797	39.07	<2e-16
Cmp.Att	58.05913	1.90724	30.44	<2e-16
Y.A	0.83090	0.07021	11.84	<2e-16
TD.Att	47.75470	2.31660	20.61	<2e-16
Int.Att	-48.46816	2.64224	-18.34	<2e-16

Residual standard error: 14.48 on 1332 degrees of freedom
Multiple R-squared: 0.7981, Adjusted R-squared: 0.7975
F-statistic: 1316 on 4 and 1332 DF, p-value: <2.2e-16

We were able to do a linear regression of the data using the formula:

$$Rate = CompletionPerAttempt(x_1) + YardsPerAttempt(x_2)$$
$$+TouchdownPerAttempt(x_3) + InterceptionPerAttempt(x_4)$$

which allowed us to see if the linear regression reflects the current system since the formula for the current system is based off these variables. After running the linear regression we used the coefficients given from the linear regression to create a formula which was:

$$Predicted.Rate = 58.05913x_1 + 0.83090x_2 + 47.75470x_3 - 48.46816x_4 + 37.03683,$$

to see whether or not the formula we developed would accurately reflect the current formula which when simplified to a linear formula we get:

$$Current.Rate = 83.3x_1 + 4.2x_2 + 333.3x_3 - 416.7x_4 + 20.8.$$

Although the formula we derived from the linear regression appears to reflect the current formula fairly well, it was interesting to find that the formula from the linear regression did not match up perfectly to the current formula. After doing more research into the current formula we discovered the reason for this. When we first began looking at the current system, we had originally assumed that the current system follows a linear formula because we speculated that they used linear regression to derive their formula. However, we discovered that they placed bounds on the first four steps so that each step has a value that falls between 0 and 2.375. Because of these bounds, the current formula has to truncate up to 0 or down to 2.375 if any step falls outside of these bounds. So by truncating these values, the current system is no longer a linear formula, and this is reflected in our linear regression. We believe that they still might have originally used linear regression to find the coefficients and then chose to narrow the range that the rating can fall between.

Instead of just looking at how the regression model fit the current model, we also looked at the t-value and p-value for each variable that we tested in order to see if the variables are necessary in determining the rate. For each variable the linear regression produced a t-value and p-value, or probability, based on the null hypothesis that the derived coefficients for each variable had values of zero. When studying the t-values and p-values, for a strong correlation, we would expect the t-values to be fairly large and the p-values to be very close to zero. Based on our results of the linear regression we see that the t-value are indeed rather large and the p-values appear to be really close to zero. Also the residual standard error shows the standard deviation for how close to the true model the linear regression is and the results suggest that the regression fits fairly well to the true model. One reason why there might be a larger residual standard error might be because the current formula is truncated and the linear regression does not take this into consideration.

4.2 Second Linear Regression Results

	Estimate	Std. Error	t-value	Pr(>$\|t\|$)
(Intercept)	0.092887	0.091030	1.020	0.3077
Cmp.Att	0.475001	0.183145	2.594	0.0096
Y.A	-0.006468	0.006742	-0.959	0.3376
TD.Att	-0.336298	0.222453	-1.512	0.1308
Int.Att	-0.078882	0.253724	-0.311	0.7559

Residual standard error: 1.391 on 1332 degrees of freedom
Multiple R-squared: 0.006813, Adjusted R-squared: 0.003831
F-statistic: 2.284 on 4 and 1332 DF, p-value: 0.05833

For the second formula that we developed we did a linear regression with the formula:

$$Win.Record = CompletionPerAttempt(x_1) + YardsPerAttempt(x_2) \\ +TouchdownPerAttempt(x_3) + InterceptionPerAttempt(x_4),$$

since we assumed that there should be a positive correlation between the win percentage of a team and the quarterbacks passer rating. Our results from the linear regression provided the coefficients that allowed us to create the second formula which came out to be:

$$Predicted.Rate_2 = 0.475001x_1 - 0.006468x_2 - 0.336298x_3 - 0.078882x_4 + 0.092887.$$

We were surprised with the formula we developed because the only positive elements in the formula were the completions per attempts and the intercept. The fact that most of the coefficients are negative might pose a problem since we figured that most of the coefficients would be positive except for the interceptions since they negatively impact a quarterback's game. Also we found this formula to have much smaller coefficients than both of the previous formulas which should provide us with a much smaller rate. The benefit of this smaller rate could be that we could easily turn this rate into a percentage which is an easier value for fans to better understand and it could allow us to predict win percentage.

Like with the first linear regression, we wanted to look at how well the linear regression fit the formula that we tested in order to determine the accuracy of the new formula developed from the coefficients given by the linear regression. When looking at the t-values we found that our t-values were fairly large but quite small compared to the t-values we got in the first linear regression. Also our p-valuse or probabilities are considerably larger than the p-values given in the first linear regression. Due to the larger p-values we cannot assume that the variables we tested have a strong correlation to winning; however, this might be due to possible interactions between the variables. Similar to the first linear regression the residual standard error is within 1332 degrees of freeedom; however, the residual standard error is much smaller in the second linear formula than the residual standard error in the first linear regression.

4.3 Comparison

After we were able to develop the two formulas with the data from the past ten seasons, we were then able to use the formulas with the data from the current season to create a table in order to compare the ratings from the current formula with the ratings from the two new formulas.

When comparing the first formula to the current system we found that the formula produced similar ratings. There were only minor differences between the two formulas where some of the players' ratings would interchange; however, there are very few differences that it would seem there is a plausible chance that the new formula would work as well as the current system. Although the new formula has a similar range as the current system, the new system might be slightly more accessible since it is not bounded like the current system and since the linear regression comes from the current system. When comparing the current system to the second formula, based off of the winning percentage, the ratings between the two formulas were less consistent. Of the two new formulas, the second formula seems to be the most accessible since the second formula provides a "predicted win percentage".

Player	C.Rate	Player	P.Rate	Player	P.Rate.2
Darren McFadden	158.3	Ace Sanders	160.299	Shann Schillinger	0.619
Ace Sanders	158.3	Darren McFadden	156.145	Colt McCoy	0.483
Spencer Lanning	152.1	Spencer Lanning	151.990	Antonio Brown	0.470
Tarvaris Jackson	140.2	Maurice Jones-Drew	149.497	Patrick Peterson	0.457
Maurice Jones-Drew	139.6	Mike James	144.512	Marcel Reece	0.425
Nick Foles	119.2	Mat McBriar	120.022	Mohamed Sanu	0.406
Antonio Brown	118.7	Bilal Powell	120.022	Dominique Davis	0.400
Mike James	118.7	Mohamed Sanu	115.868	Brock Osweiler	0.381
Mat McBriar	118.7	Marcel Reece	113.375	TJ Yates	0.376
Colt McCoy	118.7	Patrick Peterson	109.221	Mat McBriar	0.373
Patrick Peterson	118.7	Antonio Brown	107.559	Bilal Powell	0.373
Bila Powell	118.7	Colt McCoy	105.897	Seneca Wallace	0.368
Marcel Reece	118.7	Tarvaris Jackson	95.009	Tarvaris Jackson	0.357
Mohamed Sanu	118.7	Shann Schillinger	88.448	Matt Ryan	0.353
Peyton Manning	115.1	Peyton Manning	86.829	Chase Daniel	0.352
Josh Cribbs	109.7	Josh Cribbs	86.295	Philip Rivers	0.348
Josh McCown	109.0	Philip Rivers	86.021	Drew Brees	0.346
Philip Rivers	105.5	Nick Foles	85.539	Matt Barkley	0.337
Aaron Rodgers	104.9	Drew Brees	85.408	Christian Ponder	0.337
Drew Brees	104.7	Josh McCown	85.025	Josh McCown	0.335
Russell Wilson	101.2	Aaron Rodgers	84.701	Peyton Manning	0.334

Chapter 5

Conclusion

Although we were able to develop the two new formulas that produce accessible ratings like the current formula, we had hoped to have a much better understanding of how the current system was developed in order to have a better understanding of how the two new formulas compared. Through our research we were able to learn more about the current system, but only to the point of learning about when it was created and by who. We also discovered through the research and through developing the first formula that the current system was bounded, which made it even more challenging to determine how the current system was developed because by bounding the formula we could no longer tell whether they had used linear regression to develop their system. However, since the first formula produced ratings that appeared to be consistent with the current system, we could conclude that the current formula was probably a result of a linear regression.

We believe that the ratings from the two new formulas are consistent enough with the current system that either formula could be used as a reliable system. We concede that there are certain inconsistencies within both of the new formulas, but yet it appears that the two new formulas might be more accessible for people to understand and use. The main issue that we found when testing all of the formulas, but especially the second formula, was that some of the quarterbacks only played in one or two games, which could result in that player receiving a higher rating because their rating considers a lower sample size than those quarterbacks who play in every game of the season. This is evident in the second formula developed where Brock Osweiler, the backup quarterback for the Broncos, has a higher rating than Peyton Manning, the starter for the Broncos. Although the system based on win percentage would be the most accessible for people to use, it still has its issues until there is a way to factor in the amount of games each player plays so that the players who play more will still have a higher rating than those who only play in a few games.

One of the main goals for our research was to determine if there was a way to rate the quarterback as a whole rather than just rating the quarterback as a passer since there are many quarterbacks in the league now that not only pass but also rush. Unfortunately, we were unable to find an adequate data set that not only included the players' passing data but also their rushing data. Another issue that developed during our research was that we had trouble with the fact that adding certain variables were difficult because there was a certain amount of subjectivity that could be included. Therefore, we were unable to create a third formula that used variables such as rushing yards, fumbles, and sacks because of issues with finding adequate data or dealing with the subjectivity of the data. Future work could include determining what other factors besides passing are necessary in determining the rating of a quarterback and creating a new formula that takes these variables into consideration. Another area that could use more research is to determine if there is any information about how the current system was developed in order to better develop more accessible formulas.

Appendix A

Comparison Table

Player	C.Rate	Player	P.Rate	Player	P.Rate.2
Darren McFadden	158.3	Ace Sanders	160.299	Shann Schillinger	0.619
Ace Sanders	158.3	Darren McFadden	156.145	Colt McCoy	0.483
Spencer Lanning	152.1	Spencer Lanning	151.990	Antonio Brown	0.470
Tarvaris Jackson	140.2	Maurice Jones-Drew	149.497	Patrick Peterson	0.457
Maurice Jones-Drew	139.6	Mike James	144.512	Marcel Reece	0.425
Nick Foles	119.2	Mat McBriar	120.022	Mohamed Sanu	0.406
Antonio Brown	118.7	Bilal Powell	120.022	Dominique Davis	0.400
Mike James	118.7	Mohamed Sanu	115.868	Brock Osweiler	0.381
Mat McBriar	118.7	Marcel Reece	113.375	TJ Yates	0.376
Colt McCoy	118.7	Patrick Peterson	109.221	Mat McBriar	0.373
Patrick Peterson	118.7	Antonio Brown	107.559	Bilal Powell	0.373
Bila Powell	118.7	Colt McCoy	105.897	Seneca Wallace	0.368
Marcel Reece	118.7	Tarvaris Jackson	95.009	Tarvaris Jackson	0.357
Mohamed Sanu	118.7	Shann Schillinger	88.448	Matt Ryan	0.353
Peyton Manning	115.1	Peyton Manning	86.829	Chase Daniel	0.352
Josh Cribbs	109.7	Josh Cribbs	86.295	Philip Rivers	0.348
Josh McCown	109.0	Philip Rivers	86.021	Drew Brees	0.346
Philip Rivers	105.5	Nick Foles	85.539	Matt Barkley	0.337
Aaron Rodgers	104.9	Drew Brees	85.408	Christian Ponder	0.337
Drew Brees	104.7	Josh McCown	85.025	Josh McCown	0.335
Russell Wilson	101.2	Aaron Rodgers	84.701	Peyton Manning	0.334
Tony Romo	96.7	Dominique Davis	82.579	Kyle Orton	0.333
Ben Roethlisberger	92.0	Matt Ryan	82.563	Ben Roethlisberger	0.332
Colin Kaepernick	91.6	Russell Wilson	82.490	Aaron Rodgers	0.331
Sam Bradford	90.9	Tony Romo	81.994	Tony Romo	0.329
Matt Ryan	89.6	Brock Osweiler	81.854	Matt Schaub	0.328

Jay Cutler	89.2	Ben Roethlisberger	81.511	Chad Henne	0.328
Alex Smith	89.1	Kyle Orton	81.057	Carson Palmer	0.327
Andy Dalton	88.8	Jay Cutler	80.737	Josh Cribbs	0.327
Cam Newton	88.8	Chase Daniel	80.615	Matt Flynn	0.326
Tom Brady	87.3	Carson Palmer	80.151	Ryan Fitzpatrick	0.325
Andrew Luck	87.0	Andy Dalton	80.102	Jay Cutler	0.324
Jake Locker	86.7	Cam Newton	79.869	Scott Tolzien	0.323
Michael Vick	86.5	Matt Flynn	79.548	Alex Smith	0.322
Matt Flynn	85.7	Sam Bradford	79.400	Cam Newton	0.320
Kyle Orton	85.3	Christian Ponder	79.268	Tom Brady	0.320
Matthew Stafford	84.2	Alex Smith	79.133	Sam Bradford	0.320
Brock Osweiler	84.1	Ryan Fitzpatrick	79.098	Andrew Luck	0.320
Carson Palmer	83.9	Jake Locker	79.014	Ryan Tannehill	0.320
Mike Glennon	83.9	Tom Brady	78.953	Jake Locker	0.319
Brian Hoyer	82.6	Colin Kaepernick	78.827	Robert Griffin III	0.319
Robert Griffin III	82.2	Andrew Luck	78.702	Thaddeus Lewis	0.318
Ryan Fitzpatrick	82.0	Seneca Wallace	78.542	Joe Flacco	0.318
Chase Daniel	81.9	Matt Cassel	78.259	Andy Dalton	0.318
Dominique Davis	81.8	Ryan Tannehill	78.204	Mike Glenno	0.317
Ryan Tannehill	81.7	Robert Griffin III	78.139	Russell Wilson	0.316
Matt Cassel	81.6	Mike Glennon	77.876	EJ Manuel	0.316
Thaddeus Lewis	81.0	Matthew Stafford	77.808	Matt Cassel	0.315
Shann Schillinger	79.2	Brian Hoyer	77.799	Kellen Clemens	0.313
Kellen Clemens	78.8	Thaddeus Lewis	77.535	Brian Hoyer	0.313
Case Keenum	78.2	Chad Henne	77.444	Terrelle Pryor	0.310
Christian Ponder	77.9	Matt Schaub	77.392	Nick Foles	0.309
EJ Manuel	77.7	Kellen Clemens	77.014	Jason Campbell	0.307
Jason Campbell	76.9	Scott Tolzien	77.002	Eli Manning	0.306

Chad Henne	76.5	EJ Manuel	76.798	Matthew Stafford	0.305
Matt McGloin	76.1	Michael Vick	76.550	Colin Kaepernick	0.302
Joe Flacco	73.1	TJ Yates	76.453	Geno Smith	0.300
Matt Schaub	73.0	Joe Flacco	76.325	Curtis Painter	0.297
Brandon Weeden	70.3	Matt Hasselbeck	75.839	Matt McGloin	0.295
Eli Manning	69.4	Jason Campbell	75.755	Matt Hasselbeck	0.293
Terrelle Pryor	69.1	Matt McGloin	75.544	Kirk Cousins	0.293
Scott Tolzien	66.8	Eli Manning	75.357	Matt Simms	0.292
Geno Smith	66.5	Terrelle Pryor	75.088	Case Keenum	0.290
Seneca Wallace	64.4	Case Keenum	74.841	Brandon Weeden	0.287
Matt Simms	63.4	Geno Smith	74.137	Michael Vick	0.283
Matt Hasselbeck	61.1	Matt Barkley	73.695	Blaine Gabbert	0.278
Kirk Cousins	58.4	Brandon Weeden	73.074	Jeff Tuel	0.258
Josh Freeman	52.6	Matt Simms	71.134	Josh Freeman	0.256
Jeff Tuel	45.1	Kirk Cousins	70.990	Mike James	0.218
Matt Barkley	44.6	Blaine Gabbert	66.654	Maurice Jones-Drew	0.179
TJ Yates	42.4	Josh Freeman	65.570	Tyrod Taylor	0.169
Michael Koenen	39.6	Jeff Tuel	65.287	Matt Moore	0.168
Tavon Austin	39.6	Curtis Painter	62.999	Spencer Lanning	0.160
Josh Bush	39.6	Matt Moore	47.545	Darren McFadden	0.128
Larry Fitzgerald	39.6	Tyrod Taylor	39.287	Ace Sanders	0.095
John Hekker	39.6	Michael Koenen	37.036	Michael Koenen	0.092
Jeremy Kerley	39.6	Tavon Austin	37.036	Tavon Austin	0.092
Luke McCown	39.6	Josh Bush	37.036	Josh Bush	0.092
Bobby Rainey	39.6	Larry Fitzgerald	37.036	Larry Fitzgerald	0.092
Denard Robinson	39.6	John Hekker	37.036	John Hekker	0.092
Brad Smith	39.6	Jeremy Kerley	37.036	Jeremy Kerley	0.092
Blaine Gabbert	36.0	Luke McCown	37.036	Luke McCown	0.092
Matt Moore	27.1	Bobby Rainey	37.036	Bobby Rainey	0.092
Curtis Painter	19.0	Denard Robinson	37.036	Denard Robinson	0.092
Tyrod Taylor	0.0	Brad Smith	37.036	Brad Smith	0.092

Appendix B

R code

Here is the R code that I used in order to compute the linear regressions, develop the formulas, and compare results:

```
data <- read.table(”2013season.txt”, header = TRUE)
data1 <- as.data.frame(data)
data1[”Cmp.Att”] <- NA
data1 $Cmp.Att <- data1 $Cmp / data1 $Att
data1[”TD.Att”] <- NA
data1 $TD.Att <- data1 $TD / data1 $Att
data1[”Int.Att”] <- NA
data1 $Int.Att <- data1 $Int / data1 $Att
data1[”Pred.Rate”] <- NA
data1 $Pred.Rate <- 58.05913*data1 $Cmp.Att + 0.83090*data1 $Y.A + 47.75470*data1
$TD.Att - 48.46816*data1 $Int.Att + 37.03683
data1[”Pred.Rate.2”] <- NA
data1 $Pred.Rate.2 <- 0.475001*data1 $Cmp.Att - 0.006468*data1 $Y.A - 0.336298*data1
$TD.Att - 0.078882*data1 $Int.Att + 0.092887
reg <- lm(Rate ~ Cmp.Att + Y.A + TD.Att + Int.Att, data = data1)
reg2 <-lm(W.Rec ~ Cmp.Att + Y.A + TD.Att + Int.Att, data = data1)
data.r <- data1[,c(”FirstName”, ”LastName”, ”Rate”)]
data.reg <- data1[,c(”FirstName”, ”LastName”, ”Pred.Rate”)]
data.reg.2 <- data1[,c(”FirstName”, ”LastName”, ”Pred.Rate.2”)]
results <- data.r[order(-data.r[,”Rate”]),]
results.1 <- data.reg[order(-data.reg[,”Pred.Rate”]),]
results.2 <- data.reg.2[order(-data.reg.2[,”Pred.Rate.2”]),]
my.results <- cbind(results, results.1, results.2)
as.data.frame(my.results)
summary(reg)
```

head(my.results)
head(results.2)
write.table(my.results, ”C:/Users/Terril/Documents/my.results.txt”, sep=”\t”, col.names = TRUE)

Bibliography

[1] Allerhand, Mike. *A Tiny Handbook of R*. Springer-Verlag Berlin Heidelberg, Berlin, 2011.

[2] Ayres, Ian. *Super crunchers: Why thinking-by-numbers is the new way to be smart*. Bantam Books, New York, 2007.

[3] Barra, Allen. ”Passing Grade.” Wall Street Journal - Eastern Edition, 1999.

[4] Barra, Allen and Rob Neyer. ”Max tackles the system for rating NFL passing.” Wall Street Journal - Eastern Edition, 1997.

[5] Berry, Scott and Chris White. ”Tiered polychotomous regression: Ranking NFL quarterbacks.” *The American Statistician*, 2002.

[6] Gardener, Mark. *Beginning R [electronic resource]: the statistical programming language: programming and software development*. John Wiley & Sons, Indianapolis, 2012.

[7] Libkuman,Terry M., Kevin G. Love, and Paul D. Donn. ”An Empirically Based Selection and Evaluation System for Collegiate Football.” Journal of Sport Management, 12.3, 1998.

[8] Matloff, Norman. *Art of R programming*. No Starch Press, San Francisco, 2011.

[9] McHale, Ian and Phil Scarf. ”Ranking Football Players.” Significance, 2.2, 2005.

[10] Mendenhall, William and Terry Sincich. *A Second Course in Statistics: Regression Analysis*. Pearson Education, Inc., Boston, Seventh Edition, 2012.

[11] ”NFL’s Passer Rating.” *History Release*. Football Hall of Fame, Web. 2014. $http://www.profootballhof.com/history/release.aspx?release_id = 1303$.

[12] Pace, Larry. *Beginning R [electronic resource]: An Introduction to Statistical Programming*. Apress, Berkeley, 2012.

[13] Stuart, Chase. ”Inside the Ratings for Quarterbacks.” New York Times, 2011.