concordance=TRUE

RUSH INDEPENDENT PASSING PLAYER EFFICIENCY NUMBER (RIPPEN)

By Gregory J. Matthews

Skidmore College

AND

By Russell Cain

Loyola University Chicago

AND

By Donald Stolz

Loyola University Chicago

RIPPEN, Rush Independent Passing Player Efficiency Number, is a new measurement of passer performance. In a simulated world, how would a passer perform starting from their twenty yard line and only performing pass plays? The aspects of each play are simulated using a Bayesian model. This allows rookies and backups with minimal data to be fairly evaluated. Drives would end in a touchdown, field goal or turnover. A player's RIPPEN is the average number of points they would be expected to score per game. Our metric improves on existing passer rating systems because it is updated to current NFL data, does not weight situational factors, and it is able to be more intuitively understood.

1. Introduction to Current Passer Rating Systems. Quantifying the performance of a quarterback is one of the most persistent challenges in sports analytics. While numerous metrics exist, most rely on arbitrary weighting systems or proprietary formulas that lack transparency. In this section, we review the three most prominent passer rating systems—NFL Passer Rating, NCAA Passing Efficiency, and Total QBR—and discuss their limitations.

1.1. *NFL Passer Rating.* The National Football League (NFL) passer rating formula has been the official standard since the 1973 season. Developed by a committee headed by Don Smith and Seymour Siwoff, the metric was designed to compare passers against a fixed performance standard based on data from the 1960-1970 era.

The formula relies on four components: completion percentage, yards per attempt, touchdown percentage, and interception percentage. Each component is calculated independently, bounded by a floor of zero and a cap of 2.375, and then combined using the following formula:

All of this leads to the NFL formula for passer rating (http://www.nfl.com/help/quarterbackratingformula). Using the notation from van Dohlen (2011):

$$QBR = \left(\frac{\frac{C}{A} - 0.3}{0.2} + \frac{\frac{Y}{A} - 3}{4} + \frac{\frac{T}{A}}{0.05} + \frac{0.095 - \frac{I}{A}}{0.04} \right) \left(\frac{100}{6} \right)$$

where C = Number of Completions
Y = Number of Yards
A = Number of Attempts
T = Number of Touchdowns
I = Number of Interceptions

This calculation results in a rating between 0 and a maximum of 158.3.

1.2. *NCAA Passer Rating.* College football utilizes a similar but distinct formula known as NCAA Passing Efficiency. Like the NFL metric, it uses completions, yards, touchdowns, and interceptions. However, the NCAA formula does not cap individual components, allowing for a much wider variance in scores. The formula is calculated as:

$$\frac{8.4Y + 330T + 100C - 200I}{A}$$

where C = Number of Completions
Y = Number of Yards
A = Number of Attempts
T = Number of Touchdowns
I = Number of Interceptions

Because there are no artificial upper limits on the inputs, the theoretical range of the NCAA rating is significantly wider than the NFL's, ranging from -731.6 to 1,261.6.

1.3. *Total QBR.* In 2011, ESPN introduced the Total Quarterback Rating (Total QBR) to address the limitations of traditional box-score metrics. Unlike the NFL and NCAA ratings, Total QBR attempts to incorporate contextual elements by utilizing Expected Points Added (EPA). This approach evaluates the difference in expected points before and after a play, rewarding quarterbacks for plays that increase the team's probability of scoring. Total QBR also accounts for rushing contributions, sacks, fumbles, and penalties, and includes adjustments for opponent strength and "garbage time" scenarios.

1.4. *DVOA and DYAR TODO:.* Football Outsiders introduced two prominent advanced metrics: Defense-adjusted Value Over Average (DVOA) and Defense-adjusted

Yards Above Replacement (DYAR). DVOA measures a quarterback's per-play efficiency by comparing each down and distance situation to a league baseline, subsequently adjusting for the strength of the opposing defense. While DVOA acts as a rate statistic expressed as a percentage above or below the league average, DYAR converts this efficiency into a cumulative volume metric, quantifying the total yards a quarterback contributes over a replacement-level alternative.

1.5. *Criticism of Existing Passer Rating Systems.* Despite their widespread use, existing metrics suffer from significant methodological flaws:

- Arbitrary Weighting and Caps: The NFL Passer Rating utilizes weights derived from 1970s defensive standards, which are increasingly irrelevant in the modern offensive era. Furthermore, the imposition of a 2.375 cap on component statistics is arbitrary; a quarterback who completes 77.5
- Lack of Interpretability: Both the NFL (0–158.3) and NCAA scales are difficult to interpret intuitively. A difference between a rating of 85 and 95 does not translate to a tangible unit of football value (e.g., points or yards).
- Lack of Transparency: While Total QBR and DVOA attempt to add context, they are largely "black box" metrics. The exact coefficients, replacement-level baselines, and opponent adjustment formulas for DVOA are proprietary. This lack of transparency makes it impossible for independent researchers to verify, reproduce, or fully audit the results.
- Contextual Blindness: Traditional passer ratings treat all yards equally. A 10-yard completion on 3rd-and-15 is treated identically to a 10-yard completion on 3rd-and-3, despite the vastly different utility of those outcomes.
- Complexity: Unlike simple arithmetic formulas, DVOA requires complex play-by-play data parsing and subjective baseline determinations, making it inaccessible for general calculation without specialized databases.

1.6. *RIPPEN: A New Measure of Efficiency.* To address these limitations, we propose a new metric: the Rush Independent Passing Player Efficiency Number (RIPPEN). Unlike traditional ratings that rely on linear weights, RIPPEN utilizes a Bayesian model to simulate the distribution of play outcomes for a specific quarterback.

RIPPEN answers a fundamental question: In a simulated world where a team starts on their own 20-yard line and only performs pass plays, how many points would the quarterback be expected to score?

By simulating drive outcomes (Touchdown, Field Goal, or Turnover) based on a quarterback's specific posterior distributions for completion, yardage, and interception rates, we derive a metric that is both statistically robust and intuitively interpretable. RIPPEN represents the average number of points a player contributes per ten possessions. This metric improves upon existing systems because it is updated dynamically to current NFL data, relies entirely on open-source data and code, and provides a clear, unit-based measure of performance (points per game) rather than an arbitrary

index.

2. Methods.   The idea of this method is to model the distribution of play outcomes from a certain quarterback. Once we have a model for the ditribution of play outcomes, drives can be simulated based on this distribution. One could then look at drive outcomes and rate quarterbacks based on a summary of these simulated outcomes. We refer to this measure as Rush Independent Passing Player Efficiency Number, or RIPPEN, which is formally defined below.

Drives are simulated using the following rules:

- Drive starts on 20 yard line at 1st and 10.
- Simulate if the pass was an interception or not.
    - Yes: Drive ends. Return 0.
    - No: Simulate pass.
- Simulate if the pass was complete.
    - Yes: Simulate yards.
    - No: 0 yards. Update down.
- Simulate Yards
    - Update down and distance to go.
- This gets repeated until we reach an ending state.
    - Interception: return 0.
    - Reach 4th down: return 3 with probability a function of the yardage left, 0 otherwise.
    - Touchdown: return 7.

Let the random variables D_j be the value of simulated drive j. D_j takes on values of either 0, 3, or 7, and $\bar{D} = \frac{\sum_j^{n_s} D_j}{n_s}$ where n_s is the number of simulated drived. We then define our measure as follows:

$$RIPPEN = 10\bar{D}$$

.

This can then be interpreted as the average number of points a team would score in ten possessions if their play outcomes were entirely based upon the performance of the quarterback's passing plays.

3. Model.

3.1. *Completions.* For a given quarterback over a given period of time, let n_{comp} and n_{att} be the number of completed passes and the number of attempted passes, respectively. We then have the following Bayesian model for completion percentage:

$$n_{comp} \sim Binomial(n_{att}, p_{comp})$$

with the following prior

$$p_{comp} \sim \beta(\alpha_c, \beta_c)$$

.

This yields the following posterior distribution for p_{comp}:

$$p_{comp}|n_{comp}, n_{att} \sim Beta(\alpha_c + n_{comp}, \beta_c + n_{att} - n_{comp})$$

.

3.2. *Interceptions.* Similarly, for a given quarterback over a given period of time, let n_{int} and n_{inc} be the number of intercepted passes and the number of incomplete passes (interceptions are considered incomplete passes), respectively. We then have the following Bayesian model for interception percentage given an incomplete pass:

$$n_{int} \sim Binomial(n_{inc}, p_{int})$$

with the following prior

$$p_{int} \sim Beta(\alpha_i, \beta_i)$$

. This yields the following posterior distribution for p_{int}:

$$p_{int}|n_{int}, n_{inc} \sim Beta(\alpha_i + n_{int}, \beta_i + n_{inc} - n_i)$$

. It is important to note that p_{int} is estimating the probability of an interception given that the pass was incomplete, not the interception rate across all attempted passes.

In all simulations, we chose to use non-informative priors and set all hyperparameters of these models equal to 1. Specifically,

$$\alpha_c = \beta_c = \alpha_i = \beta_i = 1$$

.

(Another possible idea here is to use empircal Bayes where these priors are based on league average completion and interception rates.)

3.3. *Model for yardage given completion.* Let y_i be the yards gained on the i-th completed pass, $y_i^{\star} = log(y_i + 1)$, and TD_i is an indicator equal to 1 if the i-th completion is a touchdown and 0 otherwise. Since we are trying to model yards given a completed pass, we consider touchdowns to be censoring events. That is, if a player

throws a 10 yard touchdown pass, we know that the play we at least ten yards. This leads to the following likelihood function for modeling yards:

$$L(\mu, \sigma^2|\mathbf{y}^\star) \propto \prod_{i=1}^{n} f(y_i^\star|\mu, \sigma^2)^{1-TD_i} S(y_i^\star|\mu, \sigma^2)^{TD_i}$$

where f is the pdf of a normal distribution and S is the survival function of a normal distribution both with parameters μ and σ^2.

For priors on μ amd σ^2, we use:

$$\mu \sim Normal(0, 10^6)$$

$$\sigma^2 \sim Uniform(0, 100)$$

.

3.4. *Data - Open Source.*

4. Data.   We are using play-by-play data from nflscrapr [MORE DETAILS]

In this pursuit of an understandable and intuitive passer ranking system, it makes sense to use the simplest statistics which surround a quarterback's time on the field. Further, as this strives to remain an open source project, the variables pulled in must remain easily accessible and, likewise, public. For this reason, the data pulled in for each quarterback when all was said and done were completions, yards, interceptions and touchdowns, for each time they were snapped the ball and opted to throw.

A pleasant duality of this data decision lies within how closely it mirrors the NFL's passer rating formula described above. In so much as this newly improved metric looks to build upon and redefine the NFL's method, it is not an attempt at reinventing the wheel.

4.0.1. *nflscrapr.*   The data used and simulated upon within RIPPEN is scraped from and publically available in another open-source R package, *nflscrapr*. This project pulls, parses, and groups data from the NFL API for easy use. Although many of the added capabilities were not used for this paper, the building block data for our simulations was. Before diving into talk of simulations, the variables gathered should be ironed out and explained. Below is an example of 4 successive rows in our table, from a game between the Steelers and the Colts, one which had "Big Ben" throwing quite well until an interception gave Collins a chance to toss around the old pigskin.

This table houses the name of the quartback, binary variables for whether the pass was complete or imcomplete, intercepted or not, fumbled or not, and an integer value of yards obtained on the play. An interesting subdivision which *nflscrapr* has to offer is the breakdown of TotalYards into "air yards" and "yards after reception". Although future versions of RIPPEN might factor in these variables separately, it was deemed

wisest to aggregate them for TotalYards as a good quarterback can be recognized by his ability to pick a receiver in the most advantageous receiving position.

```
## Warning in file(file, "rt"): cannot open file '/Users/gregorymatthews/Dropbox/RIPPENgi
No such file or directory
## Error in file(file, "rt"): cannot open the connection
## Error: object 'nfl' not found
## Error: object 'nfl' not found
## Error: object 'tab' not found
```

Additionally, solely tracking air yards would hurt the rating of quarterbacks who are effective in deploying a short pitch play to a receiver now open to run 20 yards. Altogether, as RIPPEN looks to capture the effect the throwing quarterback had on the team's state at that moment in the game, these are the variables chosen, simple as they may be.

4.1. *How we use our data.* Taking this data, broken down by player over seasons we looked to implement a sampling notion, allowing us to build upon our finite examples and imagine a world in which each team put their quarterbacks on the field to throw their hearts out. The only downside to generating this data is that we lose context. Therefore, we needed to structure a proxy measure of whether or not a scoring drive (now just a series of yards gained or imcomplete/intercepted passes) led to a touchdown or a favorable position on the field for a fieldgoal. As such, we need to try to crunch and fit a simulated array of data into a football framework; imposing conditions on successive substrings which, on the field, translate to a continued drive.

Looking at our aforementioned data example, let's try to picture what the game looked like assuming there were no rushing plays betwixt our rows. The game starts, more or less on the Steelers' twenty yard line with Roethlisberger's first play resulting in a five yard gain. Therefore, we are at "second and five", with the drive continuing on. Were they not to make ten yards within this "down series", successive first, second, third, and fourth downs, then the drive is over, either resulting in a punt or a field goal attempt. As Ben's next throw is a moderate bomb of thirty yards he does not need to worry about a third down just yet, as the Steelers are back to first down in a brand new down series for the same drive. Great! So now the Steelers are over half-field and looking to put some points on the board, until Ben goes and throws an interception. Regardless of where the current drive and down series state, an interception is an automatic end to the drive, resulting in a returned value of zero points for the quarterback and the team. Hopefully, even if football is a foreign sport to you, this colorful description helped you identify a few criterion a passed array of play results must meet for a drive to stay alive and to identify the drive's down series at any step along the way.

4.1.1. *Simulation! Bayseian?.* With this ability to map decontextualized data into a football framework, RIPPEN is capable of utilizing simulated data; deepening the pool of observations upon which to gauge a quarterback's efficacy. To generate these sampled observations, **PLEASE SAVE ME**.

4.1.2. *Markov Chain Notion:.* Formalizing our mapping from raw data into a drive and down series framework requires us to apply some notion of ordering and series-dependency into our array. A prominent way of reworking this into a generalized probability, is through Markov Chains. This matrix allows you to map out the probabilities of moving from one state to another, probabilities which will sum to one as everyone leaving a state must be on their way to another one. After you finish simmering over that metaphysical tidbit, you may recognize that the states are fully contained to any generic down series, with the states randing from first to fourth down. As a drive is mortal and can end with either a failed fouth-down attempt or an interception, we also need to include an absorbing state for a dead drive. For clean rendering, any non-zero or one values are encapsulated in variables which will be described beneath:

-Markov-	Down 1	Down 2	Down 3	Down 4	Over
Down 1	a	b	0	0	c
Down 2	d	0	e	0	f
Down 3	g	0	0	h	i
Down 4	0	0	0	1	j
Over	0	0	0	0	1

$$a = Pr(y_{d,1} > 10)$$
$$b = 1 - a$$
$$c = Pr(y_{d,2} > 10 - y_{d,1})$$
$$d = 1 - c$$
$$e = Pr(y_{d,3} > 10 - y_{d,2} - y_{d,1})$$
$$f = 1 - e$$
$$g =$$
$$h = \textbf{Will fill these out in a minute}$$
$$i =$$
$$j =$$

Might try to stagger these out by row, making them two or three wide.
To make sense of the variables above, we need to iron out some notation. As commonly used, $Pr()$ stands for the probability that the given value will occur. This makes sense as Markov Chains are our way of generalizing probabilities within any given state. More **arcane(@luc.edu)** is our notation for the drive state and down: $Pr(y_{d,n} > w)$.

This notation captures the probability that the number of yards yielded from down series d and down n will be greater than w yards. As this $y_{d,n}$ notation pervades, we should pin down constraints on d and n. As our simulations start the drive on the offensive team's twenty yard line and it takes at least ten yards for a new down series to begin within a drive, d is always an integer ranging from 1 to 8, while the nested n is always contained to the set {1,2,3,4}. We also need to account for the ever-present possibility of an interception, an event contained within the I indicator function.

4.1.3. *Variable description! (More i's than Mississippi).* Dedicated to the variables noted in Markov Chain – make sense of each one and explain significance.

1. G: The result of the drive/simulation. Either 7 for TD, 3 for FG or 0 for interception or missed FG.

2. I(...): Indicator function: ...

3. $C_{d,i}$ ...

4. $I_{d,i}$: E[I(D = 4)] = P(D = 4)

5. $C_{d,i}$: $t_1' \cdot M = t_2' = [a \ b \ 0 \ 0]$

6. .... : $t_2' \cdot M = t_3' = [a^2 + bc \ ab \ bd \ 0]$

7.
$$Pr(G_j = 3) = Pr(FG \cap (\sum_{i=1}^{n-1} I(D_n = 4) = 0) \cap (\sum_{i=1}^{n-1} y_i < 80 | Q = \sum_{i=1}^{n-1} y_i)) \cdot P(Q = q)$$

   ... Pr(FG ∩ Q = q)

8.
$$Pr(G = 7) = \sum_{n=1}^{\infty} Pr(\sum_{i=1}^{n} y_i > 80 | \sum_{i=1}^{n} I(D_i = 4) = 0) \cdot P(...)$$

4.2. *How we visualize, parse our analyses?.* Idk, Look at other sections of this paper and prep for that. Suppose we could at least speak to breaking it down by season, game, player and whatnot.

4.3. *Theoretical Results.* Do we have any? I guess, in theory, we do?
****** **end of Rusty's current contributions** *****

	Comp %.	Yds/Att	TD/Att	Int/Att	QBR	rippen
T. Brady	65.8%	7.6	0.0509	0.0193	97.7	2.688
R. Wilson	65.6%	8.1	**0.0820**	0.0164	110.9	2.583
J. Allen	52.8%	6.5	0.0313	0.037	67.9	1.445
J.Driskel	59.7%	5.7	0.034	**0.0114**	82.2	1.406

```
   ## Warning in readChar(con, 5L, useBytes = TRUE): cannot open compressed file
 '/home/don/Dropbox/RIPPENgit/RIPPEN_2018_season_df.RData', probable reason 'No such
                              file or directory'
    ## Error in readChar(con, 5L, useBytes = TRUE): cannot open the connection
                 ## Error: object 'rippen2018_season' not found
                 ## Error: object 'rippen2018_season' not found
                 ## Error: object 'rippen2018_season' not found
```

FIG 1. *Comparing passer rating and RIPPEN for the 2018 season*

```
## Warning in readChar(con, 5L, useBytes = TRUE): cannot open compressed
file '/home/don/Dropbox/RIPPENgit/RIPPEN_2018_season_df.RData', probable
reason 'No such file or directory'
## Error in readChar(con, 5L, useBytes = TRUE): cannot open the connection
## Error: object 'rippen2018_season' not found
## Error: object 'rippen2018_season' not found
## Error: object 'rippen2018_season' not found
## Error: object 'rippen2018_season' not found
## Error: object 'rippen2018_season' not found
## Error: object 'tab' not found
## Error: object 'tab' not found
```

```
## Error in readChar(con, 5L, useBytes = TRUE): cannot open the connection
## Error: object 'results' not found
## Error in dist$X7: object of type 'closure' is not subsettable
## Error in dist$X0: object of type 'closure' is not subsettable
## Error in dist$X3: object of type 'closure' is not subsettable
## Error in dist$X7: object of type 'closure' is not subsettable
## Error in dist$name <- row.names(dist): object of type 'closure' is not
subsettable
## Error in 'ggplot()':
## ! 'data' cannot be a function.
## i Have you misspelled the 'data' argument in 'ggplot()'?
## Error: object 'gg' not found
```

5. Results.

5.1. *Correlation between RIPPEN and winning.* Compare RIPPEN and winning to QBR and winning.

5.2. *Preliminary Results & Notes.*

5.3. *Bayesian Posterior Distributions Stuff.* What do the posterior parameters look like?

5.4. *Rodgers vs Tebow Example. .*

5.5. *Distribtuion of RIPPEN.*

5.6. *Best Games/Seasons.*

6. Conclusion and Future Work. RIPPEN is good. We will do more eventually.

Adding a defensive adjustment.

Do we even want to add these things? How do we deal with pass interference? Defensive Holding? Sacks? Add another layer. Fumbles? Could treat similar to interceptions? Should interceptions ever result in negative numbers? How do we assign the negative numbers for interceptions?

```
## Warning in readChar(con, 5L, useBytes = TRUE): cannot open compressed
file '/home/don/Dropbox/RIPPENgit/RIPPEN_2018_season_df.RData', probable
reason 'No such file or directory'
## Error in readChar(con, 5L, useBytes = TRUE): cannot open the connection
## Error: object 'rippen2018_season' not found
## Error: object 'rippen2018_season' not found
## Error: object 'tab' not found
## Error: object 'tab' not found
```

7. Appendix.

References.

van Dohlen, P. (2011). Tweaking the NFL's Quarterback Passer Rating for Better Results. *Journal of Quantitiative Analysis in Sports* **7** 22.

Tim Tebow example of why QBR is bad:

Read more about this (pareto-frontier). Might be interesting
- Would we add something like this to our results

DYAR and DVOA:

nih: charles poliquin

JQAS

NFL Passer rating:

Don Steinberg: How I Learned to Stop Worrying and Love the Bomb

Total QBR calculation:

http://www.math.montana.edu/graduate/writing-projects/2017/Gomez17.pdf
A Statistical Analysis of NFL Quarterback Rating Variables
Derek Stimel, Journal of Quantitative Analysis in Sports

The Quarterback Prediction Problem: Forecasting the Performance of College Quarterbacks Selected in the NFL Draft
Julian Wolfson et al., Journal of Quantitative Analysis in Sports

Analyzing dependence matrices to investigate relationships between national football league combine event performances
Brook T. Russell et al., Journal of Quantitative Analysis in Sports

Isolating the Effect of Individual Linemen on the Passing Game in the National Football League
Benjamin C Alamar et al., Journal of Quantitative Analysis in Sports

Quantifying NFL Coaching: A Proof of New Growth Theory
Kevin P. Braig, Journal of Quantitative Analysis in Sports

CITE Passer Rating
CITE QBR

Don Steinberg: How I Learned to Stop Worrying and Love the Bomb

Quarterback Rating:

NFL Passer rating:

College Passer efficiency:

Defending Passer rating: Kerry Byrne

PRO FOOTBALL; The N.F.L.'s Passer Rating, Arcane and Misunderstood

7.0.1. *Criticism of QBR.* Arbitrary scale (0 to 158.3??) Hard to interpret (What does 121.6 mean?) QBR overly credits QBs for scoring TDs – discuss whether or not this is entirely wrong. Something to be said for "getting er done", but they weight this a bit too much for a metric which assesses QB efficacy.

> [Don] Smith thought it would be more meaningful if an excellent score came to around 100, just like in school. "I think our attitude was that 100 was an A," he recalls. "And anything above 100, that was an A-plus." So, in a move that made sense at the time and has had everyone else confused for three decades, he multiplied the raw total by 100 and divided by 6, turning a statistically average performance – 1s across the board – into 66.7. It also made the maximum rating a ridiculous 158.2. (http://www.donsteinberg.com/qbrating.htm)

E-mail: gmatthews1@luc.edu

E-mail: rcain@luc.edu

E-mail: dstolz@luc.edu