NFL Player Evaluation Using Expected Points Added with nflscrapR

Ron Yurko    Sam Ventura    Max Horowitz

Department of Statistics
Carnegie Mellon University

Great Lakes Analytics in Sports Conference, 2017
Game of Yards?
Tale of Two Runs

Arizona Cardinals line up on 3rd-and-20 on their own 10 yard line:
- David Johnson gains 11 yards, team punts next play

Pittsburgh Steelers line up on 3rd-and-1 on their own 10 yard line:
Le'Veon Bell gains 4 yards, converts first down

Johnson gained nearly 3 times more yards than Bell...

NOT ALL YARDS ARE CREATED EQUAL

Ron Yurko (@Stat_Ron)
NFL Player Evaluation
Great Lakes Analytics, 2017
Tale of Two Runs

Arizona Cardinals line up on 3rd-and-20 on their own 10 yard line:
- David Johnson gains 11 yards, team punts next play

Pittsburgh Steelers line up on 3rd-and-1 on their own 10 yard line:
- Le’Veon Bell gains 4 yards, converts first down
Tale of Two Runs

Arizona Cardinals line up on 3rd-and-20 on their own 10 yard line:
- David Johnson gains 11 yards, team punts next play

Pittsburgh Steelers line up on 3rd-and-1 on their own 10 yard line:
- Le’Veon Bell gains 4 yards, converts first down

Johnson gained nearly 3 times more yards than Bell...
Tale of Two Runs

Arizona Cardinals line up on 3rd-and-20 on their own 10 yard line:
- David Johnson gains 11 yards, team punts next play

Pittsburgh Steelers line up on 3rd-and-1 on their own 10 yard line:
- Le’Veon Bell gains 4 yards, converts first down

Johnson gained nearly 3 times more yards than Bell...

NOT ALL YARDS ARE CREATED EQUAL
How do we properly evaluate a play?

3rd-and-20, CLE 10, (10:48) (Shotgun) D.Johnson up the middle to CLE 21 for 11 yards (R.McLeod).
How do we properly evaluate a play?
3rd-and-20, CLE 10, (10:48) (Shotgun) D.Johnson up the middle to CLE 21 for 11 yards (R.McLeod).

Consider a play-by-play dataset:

- **Down**: 4 downs to advance the ball 10 (or more) yards
- **Yards to go**: distance in yards to convert first down
- **Yard line**: distance in yards away from opponent’s endzone (100 to 0) - the field position
- **Time remaining**: seconds remaining in half, each half is 1800 seconds long (overtime is 900)
How do we properly evaluate a play?

3rd-and-20, CLE 10, (10:48) (Shotgun) D.Johnson up the middle to CLE 21 for 11 yards (R.McLeod).

Consider a **play-by-play** dataset:

- **Down**: 4 downs to advance the ball 10 (or more) yards
- **Yards to go**: distance in yards to convert first down
- **Yard line**: distance in yards away from opponent’s endzone (100 to 0) - the field position
- **Time remaining**: seconds remaining in half, each half is 1800 seconds long (overtime is 900)

Value of a play depends on the situation

- Former BYU and NFL quarterback Virgil Carter
- Took 2852 1st-and-10 plays, turned the field into 10 yard buckets, **averaged the value of the next scoring event**
We Know This Already

*Operations Research on Football (Carter and Machol, 1971)*
- Former BYU and NFL quarterback Virgil Carter
- Took 2852 1st-and-10 plays, turned the field into 10 yard buckets, **averaged the value of the next scoring event**

*The Hidden Game of Football (Carroll et al., 1988)*
- Classic work and first deep dive into football statistics
- Play’s success is a function of down and yards to go
- Linear **expected points** model from -2 on team’s goal line to +6 on opponent’s, every 25 yards leads to 2 more expected points
We Know This Already

*Operations Research on Football (Carter and Machol, 1971)*

- Former BYU and NFL quarterback Virgil Carter
- Took 2852 1st-and-10 plays, turned the field into 10 yard buckets, **averaged the value of the next scoring event**

*The Hidden Game of Football (Carroll et al., 1988)*

- Classic work and first deep dive into football statistics
- Play’s success is a function of down and yards to go
- Linear **expected points** model from -2 on team’s goal line to +6 on opponent’s, every 25 yards leads to 2 more expected points

Recent developments by Aaron Schatz at Football Outsiders, Brian Burke at ESPN, and Keith Goldner at numberFire
Recent work in football analytics is not easily reproducible:
- Reliance on proprietary and costly data sources
- Data quality relies on potentially biased human judgement
Reproducible with nflscrapR

Recent work in football analytics is not easily reproducible:
- Reliance on proprietary and costly data sources
- Data quality relies on potentially biased human judgement

nflscrapR:
- R package created by Maksim Horowitz to enable easy data access and promote reproducible NFL research
- Collects play-by-play data from NFL.com and formats into R data frames
- Data is available for all games starting in 2009

Available on Github, install with:
devtools::install_github(repo=maksimhorowitz/nflscrapR)
How to Model Expected Points

Developed a novel multinomial logistic regression model built on nflscrapR data from 2009-2016 seasons, using the nnet package.
Expected Points Model

Model is generating probabilities, agnostic of value associated with each next score type

Next Score: \( Y \in \{\text{Touchdown (7), Field Goal (3), Safety (2), No Score (0), -Safety (-2), -Field Goal (-3), -Touchdown (-7)}\} \)

Situation: \( X \in \{\text{down, yards to go, field position, time remaining}\} \)

Outcome probabilities: \( P(Y = y | X) \)

**Expected Points (EP)** = \( E(Y|X) = \sum_y P(Y = y | X) \times y \)
Expected Points Added (EPA) estimates a play’s value based on the change in situation, providing a point value

\[ EPA_{\text{play}_i} = EP_{\text{play}_{i+1}} - EP_{\text{play}_i} \]
Expected Points Added (EPA) estimates a play’s value based on the change in situation, providing a point value

\[ EPA_{\text{play}_i} = EP_{\text{play}_{i+1}} - EP_{\text{play}_i} \]

David Johnson’s 11 yards leading to 4th down?

Le’Veon Bell’s 4 yards converting 1st down?

His 4 yards are almost 100 times more valuable!
Expected Points Added (EPA) estimates a play’s value based on the change in situation, providing a point value

\[ EPA_{\text{play}_i} = EP_{\text{play}_i+1} - EP_{\text{play}_i} \]

David Johnson’s 11 yards leading to 4th down? 0.0096 EPA

Le’Veon Bell’s 4 yards converting 1st down? 0.9225 EPA

His 4 yards are almost 100 times more valuable!
**Expected Points Added (EPA)** estimates a play’s value based on the change in situation, providing a point value

\[ EPA_{\text{play}_i} = EP_{\text{play}_{i+1}} - EP_{\text{play}_i} \]

David Johnson’s 11 yards leading to 4th down? 0.0096 EPA

Le’Veon Bell’s 4 yards converting 1st down?
**Expected Points Added (EPA)** estimates a play’s value based on the change in situation, providing a point value:

\[
EPA_{\text{play}_i} = EP_{\text{play}_{i+1}} - EP_{\text{play}_i}
\]

David Johnson’s 11 yards leading to 4th down? 0.0096 EPA

Le’Veon Bell’s 4 yards converting 1st down? 0.9225 EPA

His 4 yards are almost 100 times more valuable!
Using EPA, can calculate new metrics to evaluate players

The following slides will go through various EPA-based metrics:

- Total EPA
- Dividing by attempts for rate measure
- Introduce new weighted metrics
Quarterbacks: Total EPA

Total EPA = \sum^{Attempts} EPA

Total EPA and Pass Attempts in 2016

Ron Yurko (@Stat_Ron)
NFL Player Evaluation
Great Lakes Analytics, 2017
Quarterbacks: EPA per Attempt

EPA per Attempt = \frac{Total\ EPA}{\#\ of\ Attempts}

EPA per Attempt and Pass Attempts in 2016

[Graph showing quarterback names and EPA per attempt values]

Ron Yurko (@Stat_Ron)

NFL Player Evaluation

Great Lakes Analytics, 2017
Quarterbacks: Success Rate

**Success Rate** = \[\frac{\# \text{ plays with EPA} > 0}{\# \text{ plays}}\] (Berri & Burke, 2012)

EPA per Attempt and Success Rate in 2016
Quarterbacks: Weighted Completion %

Weighted Completion % = \frac{\sum \text{Completions} \cdot \text{EPA}}{\sum \text{Attempts} \cdot |\text{EPA}|}

Weighted Completion % and Completion % in 2016
Receivers: Weighted Reception %

Weighted Reception % = \frac{\sum \text{Receptions} \cdot \text{EPA}}{\sum \text{Targets} \cdot |\text{EPA}|}

Weighted Reception % and Reception % in 2016

Ron Yurko (@Stat_Ron)

NFL Player Evaluation

Great Lakes Analytics, 2017
Rushers: EPA per Carry

EPA per Carry = \frac{Total \ EPA}{\# \ of \ Carries}

EPA per Carry and Carries in 2016

Ron Yurko (@Stat_Ron)
NFL Player Evaluation
Great Lakes Analytics, 2017
Rushing is costly, need to evaluate players relative to average.
Important to understand when stats stabilize to “trust” them

- *e.g.* How many pass attempts does Dak Prescott need before we can believe the results?
Stabilization

Important to understand when stats stabilize to “trust” them

- e.g. How many pass attempts does Dak Prescott need before we can believe the results?

Follow approach by Tom Tango for baseball stats:

- For given number $N$ of attempts (or targets, carries, etc.), find the players with at least that number
- For each player, randomly select two samples of size $\frac{N}{2}$ and calculate the stat of interest, repeat
- Compute the correlation $r$ and repeat the process for each considered $N$, identify when correlation stabilizes
Receiving Stabilization

Stability of Stats over Number of Targets

Stability of Stats over Number of Receptions
Passing Stabilization

Stability of Stats over Number of Pass Attempts

- EPA-based
- Traditional

- Comp %
- EPA per Att
- Weighted Comp %
- INT per Att
- Success Rate
- TD per Att
- Yards per Att

- 25
- 50
- 75
- 100
Rushing Stabilization

Stability of Stats over Number of Carries

- EPA per Carry
- Success Rate
- TDs per Carry
- Yards per Carry

- EPA-based
- Traditional

- 25
- 50
- 100
- 150
Accurate metrics of player ability **should be consistent** over time.

Check the correlations between player seasons in two ways:
- Same team in both seasons
- Different team in each season

Ideally, measure of player’s ability should be independent of team.
Receiving Correlations

High level of consistency for receiving stats
Passing Correlations

Season-to-Season Correlations for Passing Statistics
Different Teams (left) versus Same Team (right)

Drop between 2014-2015 led by Peyton Manning and Tony Romo

Ron Yurko (@Stat_Ron)
Rushing Correlations

Clearly **team dependent** (O-line?), e.g. DeMarco Murray:
- DAL in 2014 - Success Rate = 43.16%
- PHI in 2015 - Success Rate = 34.38%
Recap

- Traditional metrics do not always properly evaluate a play
- Using nflscrapR we can calculate EPA based metrics (as well as WPA) and view player production across the field
- Passing is more efficient than rushing from an EPA view
Recap

- Traditional metrics do not always properly evaluate a play.
- Using `nflscrapR` we can calculate EPA based metrics (as well as WPA) and view player production across the field.
- Passing is more efficient than rushing from an EPA view.
- Receiving:
  - EPA per Rec and Yards per Rec are highly consistent.
- Passing:
  - Success Rate and Comp % are more appropriate than yards.
- Rushing:
  - Success Rate is most consistent.
  - Most difficult to evaluate.
Future Work: Isolate Player Contribution

Obviously one player is not solely responsible for EPA

Need to account for the situation (down, yards to go, etc.) and also the teams and players involved
Future Work: Isolate Player Contribution

Obviously one player is not solely responsible for EPA

Need to account for the situation (down, yards to go, etc.) and also the teams and players involved

**Mixed model** approach (Judge et al., 2015):
- **Fixed effects** for the situation
- **Random effects** for the individual players, teams

This work will be presented at NESSIS 2017
Future Work: Probability Measures

Multinomial logistic regression model generates probabilities for each of the 7 next score events - can create **new statistics for events**

- e.g. **Probability of Touchdown Added (PTDA)**
Website application:
- Developing reports and shiny apps to host on Tartan Sports Analytics Club website
- https://tartansportsanalytics.com/

nflscrapR development on Github:
https://github.com/maksimhorowitz/nflscrapR

Follow us on twitter:
Tartan Sports Analytics - @CMUAnalytics
Ron - @Stat_Ron
Sam - @stat_sam
Max - @bklynmaks
References

[Hastie et al., 2009]. [Carroll et al., 1998].

Football outsiders.

numberfire.

Measuring productivity of nfl players.

Burke, B.
Advanced football analytics.


Expected Points Model

Combined two ideas for weighting plays:
- Score differential - more weight for close score games
- Distance in drives away from next score