

# Parametric modeling and analysis of NFL run plays

Preston Biro\* and Stephen G. Walker

*Department of Statistics & Data Science University of Texas at Austin, USA*

Received 6 July 2022

Accepted 3 February 2023

Pre-press 12 June 2023

Published 9 November 2023

**Abstract.** The paper is concerned with the modeling of run plays from data obtained from the NFL. Using a parametric regression model based on the skew- $t$  distribution we estimate the shifts from overall league averages for each team within the NFL. From the interpretation of the parameters we can investigate what the best teams are specifically doing to achieve better performance according to the criterion of average yards per play.

**Keywords:** Skew- $t$  distribution, Regression model, Heavy tail, NFL

## 1. Introduction

The NFL has seen offensive development explode over the last several decades, seeing yards per game increase from 295 for an average team in the 1970s to 347 in the 2010s (Gough, 2022). With this much change, it is hard to imagine further room for improvement, and therefore identifying factors that can fundamentally improve a team's performance should prove to be a difficult task. The question of how to improve a team no longer has surface level answers and the use of high level data analysis to inform decisions is, we argue, a necessary next step to seriously consider.

In Biro and Walker (2022), a step towards a data driven decision theoretic approach to play calling was made by establishing a method for an optimal play calling policy. The techniques relied heavily on the probability transition function that dictates how states transition from one to another given a specified action. A state here is represented by a down, a yardage to go to secure a first down, and the line of scrimmage.

The probabilities are estimated from data. At an average level; i.e. over all football teams in the NFL, a nonparametric estimator works. There is sufficient data for all the states to get a good estimation of the probabilities of outcomes following choices of play. However, when we wish to look at team specific outcomes, data is more sparse, and hence parametric models become essential. Provided the model parameters are interpretable, the change in estimated parameters for different teams can lead to an understanding of what activities/decisions the teams are making which determines how well a team does. This is an ambitious endeavour and yet we believe we have found some significant criteria in team choices which explains to some extent team offensive proficiencies, specifically relating to run plays.

Consequently, this paper details the modeling methods used for creating accurate probability transition distributions for run plays. Additionally, we provide insight on some of the interpretable outcomes of the parameters. We attempt to answer the question: "What makes a good running football team?" using information that can be gleaned from the parameters identified in the modeling efforts.

Much of the current literature regarding the modeling of run plays originated in the 2020 Big Data

---

\*Corresponding author: Preston Biro, Department of Statistics & Data Science University of Texas at Austin, USA. E-mail: prestonbiro@gmail.com.

Bowl, where individuals were tasked with predicting the amount of yards gained using player tracking data. These projects proved fruitful, identifying many features of a successful running game, such as a ball carrier's effective acceleration (Ploenzke, 2019), the initial open space generated by offensive linemen (Stern, 2019), and field control generated by each team's positioning (Brighenti, 2019). However, each of these projects rely on tracking data and therefore provide limited insight on the ability to advise decisions in real time, due to the fact that data is not available until the end of the play (or more realistically until after the game). Moreover, NFL and NCAA teams (at the time of writing) are prohibited from accessing computers during the game, and although it may be possible that this restriction is lifted in the future, tracking data still would not be a particularly helpful pre-play predictive tool because of the uncontrollable nature of the opponents' actions. Therefore, the measures created are better suited for descriptive evaluation metrics, while the methods proposed in this paper will focus on generating probability densities that can inform in game decisions using available online information. Additionally, these aforementioned projects tended to focus on evaluation at the player level. The results of this paper will take a higher level approach to improving the run game of a team.

Previous work using NFL play by play data has focused on two topics: creating descriptive statistics and predicting play types. For the former, research into expected points and win probabilities have been exceptionally popular, implemented primarily in the work of Yurko et al. (2018) and Burke (2010a,b). These metrics have been widely accepted in the analytics community, using them as tools to understand decision making and to evaluate players. The latter has been explored by multiple groups (Lefort et al., 2022; Joash Fernandes et al., 2020), attempting to find models that successfully predict play calls. The scope of the work is directed towards adversarial strategy, attempting to find trends in decisions that can be used to help a defense prevent an offense from advancing.

The closest work to our own, we believe, was conducted in Lutz, and Kassarnig (2016), where play by play data was used to predict the number of yards gained (amongst other variables). These authors explored several different machine learning techniques, such as regression trees, support vector regression, and artificial neural networks, to predict a number of outcomes of interest using play by play data. However, their methods used information not

available at the start of the play (such as pass length and side of the field) in their predictions, whereas our research will focus on using only presnap information in order to provide prescriptive decisions. Additionally, their work created point estimators for yards gained, whereas our results will provide distributions of all possible events.

Central to our paper is the modeling of the yards gained outcome following the choice to run for an NFL team. This will be formulated using a regression model where the features are line of scrimmage, yards to first down, and down number. These were successfully assumed to be the key features from Biro and Walker (2022). The key model is the skew- $t$  distribution, which we claim provides a good fit for the data and allows us to separately model important aspects of a distribution; i.e. the mean, the variance, the skewness and the heaviness of tail. The skew- $t$  model is described in detail in section 2. In section 3 we indicate how we analysed the data and present the raw results. Our analysis of the results is to be found in section 4 and section 5 concludes with a discussion.

## 2. Modeling

At each state, which we have indicated is the triple down, line of scrimmage and yards to a first down, we adopt the idea that there are four actions a team can choose from: a run, a pass, a punt, or a field goal attempt. For each state-action pairing, our goal is to model the probability of outcomes which is tantamount to estimating the transition probabilities from state to state. In this paper, we will focus exclusively on the modeling of run plays. First we will begin by modeling the population; i.e. all the teams in the league, and then model the team adjustments. Such a model would be known in the statistical modeling community as a random effects model.

### 2.1. Data

The data used is NFL play-by-play data spanning the 2017 and 2018 regular seasons, obtained via the nflscrapR package (Horowitz et al., 2017). Specifically, the data used for modeling purposes are the plays marked as run plays in the *play\_type* column. We will primarily use the *down*, *ydstogo*, *yardline\_100*, *posteam*, and *yards\_gained* columns; however, other markings will be used in the analysis section. To ensure data consistency, we removed plays that do not constitute standard run plays,

namely penalized plays, quarterback scrambles, and quarterback kneels. Additionally, we removed plays where the pre-play run probability is near one or zero, as these scenarios would result in the defense being over- or under-prepared to stop the run and therefore would not be closely comparable to the rest of the data. Specifically, we will exclude run plays that fall into 2-minute drill, 4-minute drill, extreme score differential scenarios, 3rd and long, and 4th and long situations.

After filtering, the data contains 19881 run plays, with each team represented by a minimum of 518 plays and a maximum of 739 plays. The values of the *down* column span 1 to 4, *yardline\_100* column span 1 to 99, *yards\_gained* column span 1 to 99, which represent their theoretical spans. The theoretical span for the *ydstogo* column is 1 to 99, however, the observed span is much more limited, spanning 1 to 40, with 99% of the observations having values less than 20. We will refer to the values of *down*, *ydstogo*, and *yardline\_100* on a particular play as  $(x_i) = (DOWN_i, DIST_i, LOS_i)$  and *yards\_gained* as  $(y_i)$ . In addition to the data used for modeling, we obtained data for analysis purposes. This included the play-by-play data used for modeling, with the addition of several columns for analysis such as *run\_gap*, *run\_location*, *defteam*, *qb\_scramble*, *shotgun*, *rusher\_player\_id*, *score\_differential*, and *game\_seconds\_remaining*.

### 2.2. The Skew-t distribution

The family of distributions used to model the data is provided by the skew- $t$  distribution, as defined in Arellano-Valle and Azzalini (2013). The density function is given by:

$$f(y; \xi, \omega, \alpha, \nu) = \frac{2}{\omega} t(z; \nu) T \left( \alpha z \sqrt{\frac{\nu+1}{\nu+z^2}}; \nu+1 \right), \quad \text{where } z = \frac{y-\xi}{\omega}. \quad (1)$$

Here  $t$  and  $T$  represent the Student- $t$  density and distribution, specifically

$$t(z; \tau) = \frac{\Gamma(\frac{\tau+1}{2})}{\sqrt{\pi\tau}\Gamma(\frac{\tau}{2})} \left( 1 + \frac{z^2}{\tau} \right)^{-\frac{\tau+1}{2}} \quad (2)$$

$$T(z; \tau) = \frac{1}{2} + x\Gamma \left( \frac{\tau+1}{2} \right) \times \frac{{}_2F_1(\frac{1}{2}, \frac{\tau+1}{2}; \frac{3}{2}; -\frac{z^2}{\tau})}{\sqrt{\pi\tau}\Gamma(\frac{\tau}{2})}, \quad (3)$$

where  $\Gamma(\cdot)$  is the gamma function and  ${}_2F_1(\cdot)$  is the hypergeometric function.

The skew- $t$  distribution has four parameters:  $(\xi, \omega, \alpha, \nu)$ . These parameters can be understood

broadly as (but not fully separately from a mathematical point of view) as mean, variance, skewness, and degrees of freedom, which controls the heaviness of the tails. Figure 1 shows a few examples of the skew- $t$  density evaluated for changing values of each parameter. As can be seen in the images, the  $\xi$  parameter shifts the location of the mode, the  $\omega$  parameter has a large influence over the spread of the distribution, the  $\alpha$  parameter influences the symmetry and skewness of the distribution, and the  $\nu$  parameter controls the size of the tails. Each of the parameters also has some influence over other effects; for example, increasing  $\omega$  changes the spread but also slightly shifts the mode. However, we will use the terms mean, variance, skewness, and degrees of freedom, aka kurtosis, interchangeably with their respective parameters for ease of discussion.

This distribution displays smooth, unimodal behavior, with parameters that allow for relative control over the occurrence of extreme and one-sided events. We observe similar structure in the distribution of run plays, and therefore believe the skew- $t$  distribution is a good choice as a parametric model for the data.

### 2.3. Goodness of fit

Following the aforementioned theme, here we statistically assess the fit of a skew- $t$  distribution for the data. While a visual inspection shows the model fits the data well, a statistical test for fit is necessary. However, it is infeasible to evaluate the goodness of fit for all states, and therefore we will consider the most

common state; namely a 1st and 10 on the offensive team's own 25 yard line; i.e.  $DOWN = 1, DIST = 10, LOS = 75$ . This state represents where a team obtains the ball after a kickoff touchback (amongst other options), and therefore is the most common state, with 1555 plays. We remove the six touchdown outcomes, for reasons that will be elaborated on in proceeding sections, leaving  $n = 1549$  non-touchdown plays with the amount of yards gained spanning  $-9$  to 70 yards. Figure 2 shows this data fit with the corresponding skew- $t$  model. The specific density is obtained by minimizing the objective

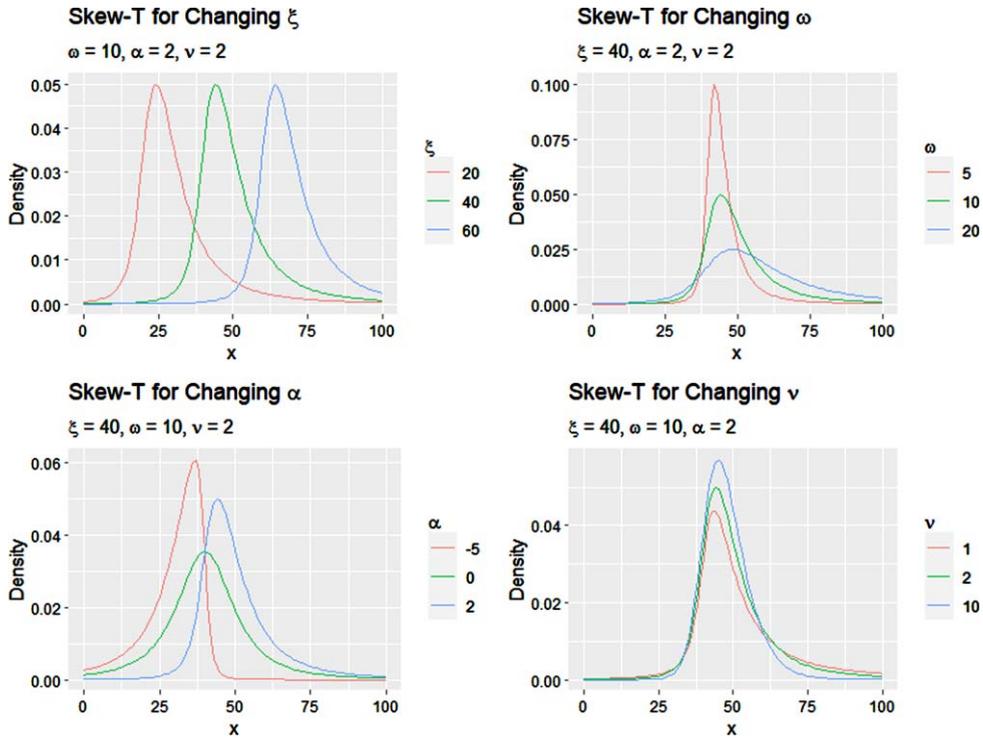


Fig. 1. Plots of skew- $t$  densities for changing parameter values.

function over  $(\xi, \omega, \alpha, \nu)$ ;

$$\sum_{y=-9}^{70} \frac{\{F_n(y) - F(y|\xi, \omega, \alpha, \nu)\}^2}{F_n(y)},$$

where  $F_n(\cdot)$  is the empirical cumulative distribution function for yards gained evaluated at  $y$  and  $F(\cdot)$  represents the cumulative distribution function of the skew- $t$ .

Using Pearson's  $\chi$ -squared goodness of fit test (Pearson, 1900), we obtain a test statistic value of 83.46. For  $df = 80$ , this corresponds to a  $p$ -value of 0.374, and hence there is no reason to reject the hypothesis that the data arose from a skew- $t$  distribution. Performing a test for all states is not appropriate; smaller sample sizes and the play of chance in the testing procedure would not lead to a 100% fit. Nothing seen or checked would suggest the skew- $t$  is not suited for the modeling of the data.

#### 2.4. League average run modeling

The nature of run plays is smooth, with yards gained typically having a unimodal distribution with

a single heavy tail on the positive part. We choose to build a Bayesian model around a skew- $t$  likelihood. The skew- $t$  distribution is truncated at each end to represent the constraints of the game limiting the yards gained to a touchdown on the positive part and a safety on the negative part. Additionally, the skew- $t$  is mixed with a point mass probability of scoring a touchdown; since touchdown scoring events tend to happen with higher frequency than other heavy tail events. Not including this point mass for touchdowns would result in an overcorrection of the parameters to fit the heavy tail of the skew- $t$ , and therefore separating these points allows for more accurate fits. We do not observe the same behavior in the negative part, and therefore do not include a point mass for safety plays.

To incorporate state information, the parameters of the skew- $t$  distribution and the point mass probability are formed as linear combinations of the down, distance, and line of scrimmage with parameters estimated via Markov chain Monte Carlo simulation. The following set of equations represents the skew- $t$  mixture model used for the modeling of yards gained, denoted by  $y_i$  for play  $i$  and with state  $x_i$ :

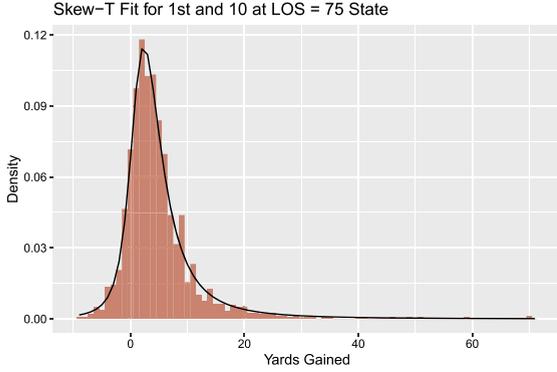


Fig. 2. Skew- $t$  fit for most common state, with the histogram bars indicating the amount of yards gained on observed plays and the black line indicating the predictive density.

$\{\xi, \omega, \alpha, \nu\}$ , along with the point mass weight parameters  $w_i$ .

We consider the  $d_i$  to be latent labels of each of the plays as coming from either the skew- $t$  distribution or the point mass distribution. For all plays that do not result in touchdowns,  $d_i$  is deterministic as a standard play (ST). However, for touchdown plays, there is a high (but not 100%) probability that  $d_i = TD$ . This mixing gives the interpretation of allowing for the total touchdown probability to be a combination of gaining exactly  $LOS_i$  yards coming from the skew- $t$  distribution plus the probability of more yards that could be gained had the yards gained not been truncated by the touchdown, represented by  $1 - w_i$ . Additionally, we have chosen to smooth the  $(w_i)$  values using a rolling average filter

$$p(y_i|x_i, \beta, w_i) = w_i F(y_i|\xi_i, \omega_i, \alpha_i, \nu_i)\mathbb{1}[y_i \in (LOS_i - 100, LOS_i]] + (1 - w_i)\mathbb{1}[y_i = LOS_i] \quad (4)$$

$$\xi_i = \beta_{INT}^{\xi} + \sum_{d=1}^4 \beta_{DOWN=d}^{\xi} \mathbb{1}[DOWN_i = d] + \beta_{DIST}^{\xi} DIST_i + \beta_{LOS}^{\xi} LOS_i \quad (5)$$

$$\log \omega_i = \beta_{INT}^{\omega} + \sum_{d=1}^4 \beta_{DOWN=d}^{\omega} \mathbb{1}[DOWN_i = d] + \beta_{DIST}^{\omega} DIST_i + \beta_{LOS}^{\omega} LOS_i \quad (6)$$

$$\alpha_i = \beta_{INT}^{\alpha} + \sum_{d=1}^4 \beta_{DOWN=d}^{\alpha} \mathbb{1}[DOWN_i = d] + \beta_{DIST}^{\alpha} DIST_i + \beta_{LOS}^{\alpha} LOS_i \quad (7)$$

$$\log \nu_i = \beta_{INT}^{\nu} + \sum_{d=1}^4 \beta_{DOWN=d}^{\nu} \mathbb{1}[DOWN_i = d] + \beta_{DIST}^{\nu} DIST_i + \beta_{LOS}^{\nu} LOS_i \quad (8)$$

with the prior for all the  $\beta$  parameters are independent normal with zero mean and a large variance, and  $w_i \sim \text{Beta}(19 + n_{i,ST}, 1 + n_{i,TD})$ . Here

$$n_{i,j} = \sum_{i=1}^N \mathbb{1}[d_i = j],$$

where  $d_i = ST$ ,  $ST =$  coming from the skew- $t$ , with probability

$$P(d_i = ST) = \frac{w_i F(y_i|\xi_i, \omega_i, \alpha_i, \nu_i)\mathbb{1}[y_i \in (LOS_i - 100, LOS_i]]}{w_i F(y_i|\xi_i, \omega_i, \alpha_i, \nu_i)\mathbb{1}[y_i \in (LOS_i - 100, LOS_i]] + (1 - w_i)\mathbb{1}[y_i = LOS_i]}$$

and  $d_i = TD$  with probability  $1 - P(d_i = ST)$ . The  $\omega_i$  and  $\nu_i$  parameters are modeled on log scale to ensure positivity. Thus, the task of identifying the probability distribution for yards gained on a run play comes down to estimating the parameters  $\beta = \beta_{COV}^{\theta}$ , for each combination of  $COV \in \{INT, DOWN = 1, DOWN = 2, DOWN = 3, DOWN = 4, DIST, LOS\}$  and  $\theta \in$

$(K = 2, \text{two-sided})$  to create a smooth representation of the touchdown probabilities, removing the effects of outliers at specific yard lines. Figure 3 shows the touchdown probabilities (for 1st and goal or 1st and 10 for  $LOS > 10$ ) plotted against the respective frequency at that yard line.

This model yields results both accurate and interpretable. With  $n = 19881$  plays and 28 parameters to fit (with 99  $(w_i)$  parameters that are nearly inde-

pendently fit from the  $\beta_{COV}^{\theta}$ ), each parameter has a sufficient amount of plays that can be used to inform its value. The components of this model are a composition of standard approaches for Bayesian mixture models, with the added novel component of the skew- $t$  distribution to account for the nature of run plays. The model assumes that the distribu-

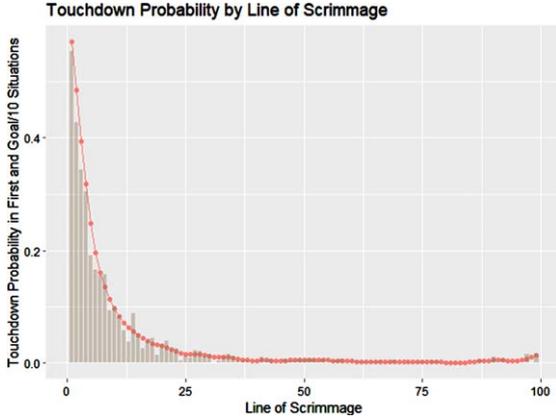


Fig. 3. Modeled probability of scoring a touchdown versus empirically observed touchdown probability for 1st and 10 scenarios, or 1st and goal scenarios for  $LOS$  less than 10.

tion of yards gained on a run play is dictated by the  $\beta_{COV}^\theta$  terms. Each of the individual  $\beta_{COV}^\theta$  parameters (excluding the  $\beta_{DOWN}^\theta$  terms) are modeled assuming a linear relationship between the skew- $t$  parameters and the state information. For example, incrementing  $DIST_i$  by one implies the  $\xi$  parameter will increment by  $\beta_{DIST}^\xi$ . We assumed the relationship between  $DOWN_i$  and the skew- $t$  parameters would likely be far from linear, as the difference in the distributions between 1st and 2nd plays is not similar to the difference between 2nd and 3rd down plays. Treating this relationship as linear would imply an equal shift in parameters for all values of  $DOWN_i$ , therefore we instead chose to treat  $DOWN_i$  as a categorical variable, giving each value its own  $\beta_{DOWN}^\theta$ . We assumed no interaction effects between covariates, assuming a change in one covariate does not affect another covariate's relationship with the parameters. This assumption tends to seem fair for most situations, as the observed data only changes slightly from state to state and therefore a more complex model would only provide marginal accuracy improvements at the cost of computational burdens and intuitive explanations.

### 2.5. Team adjusted run modeling

Given the estimated parameters for the league average data, we can now identify trends at the team level. To make team adjustments, the model is an extension which allows for team additive shifts from the average values as follows:

$$p(y_i|x_i, \beta_{Team}^\theta, \beta_{COV}^\theta, w_i) = w_i F(y_i|\xi_i^j, \omega_i^j, \alpha_i^j, \nu_i^j) \mathbb{1}[y_i \in (LOS_i - 100, LOS_i]] + (1 - w_i) \mathbb{1}[y_i = LOS_i]$$

with

$$\begin{aligned} \xi_i^j &= \xi_i^* + \beta_{Team=j}^\xi, \\ \omega_i^j &= \omega_i^* + \beta_{Team=j}^\omega, \\ \alpha_i^j &= \alpha_i^* + \beta_{Team=j}^\alpha, \\ \nu_i^j &= \nu_i^* + \beta_{Team=j}^\nu, \end{aligned}$$

where  $j$  is an indicator for the team. Here, the  $(\xi_i^*, \omega_i^*, \alpha_i^*, \nu_i^*)$  represent the average parameters of the league model. Additionally, the  $(x_i)$  now contain information indicating the team.

These values were identified using a weighted MLE approach, where the weights were assigned according to the team's run frequency for each state component relative to the league average frequency. This weighting was used to counter the effects of team specific trends for running in situations significantly more or less than the league average and to force the parameters to be a function of the team's proficiency rather than their usage rate in a particular area. Team effects were computed over the entire two-year sample, acknowledging that coaching and roster turnover would naturally result in differing parameters from year to year, but are likely small changes in comparison to the effects of controlled covariates. Note that the values of  $\beta_{COV}^\theta$  and  $w_i$  do not need to be recomputed.

## 3. Results

### 3.1. League average model results

Table 2 shows the average parameter values obtained for the league average run modeling methods. Using these values, one can obtain an accurate distribution for a league average run play for any state. For example, for a 1st and 10 scenario from the teams own 25 yard line, the distribution of yards gained would be modeled using a mixture of a skew- $t$  distribution with mean parameter  $\xi = 0.221$ , variance parameter  $\omega = 4.150$ , skewness parameter  $\alpha = 1.811$ , and degrees of freedom parameter  $\nu = 2.505$ , mixed with the touchdown probability point mass with mixing parameter  $w = 0.997$ .

Examining these values can provide some fundamental insights as to how the model controls the shape of the distribution. For example, as the value of  $DIST$

Table 1

Table of  $\beta_{COV}^{\theta}$  values after MCMC optimization. Values of  $COV$  are shown going down and values of  $\theta$  are shown going across

$\beta_{COV}^{\theta}$	$\xi$	$\omega$	$\alpha$	$\nu$
INT	-.777	1.508	2.909	1.221
DOWN = 1	.151	-.447	-1.291	-.104
DOWN = 2	.417	-.393	-1.423	-.022
DOWN = 3	.439	-.615	-1.221	-.336
DOWN = 4	.833	-.979	-1.664	-.363
DIST	.0553	.0257	-.0115	.0288
LOS	.00392	.00140	.00410	-0.00649

decreases (holding all other values constant), the mean parameter increases, while the skewness parameter will decrease. Hence, comparing the distribution of a 2nd and 10 run play to a 2nd and 1 run play at the same yardline, one would expect the center of the distribution to be slightly larger (shifted right) for the 2nd and 1, but the probability of gaining a large amount of yards to be lower as the skewness decreases. This is likely due to the defense’s response to the offense in each situation, with the defense more worried about giving up a larger play on a 2nd and long play than on a 2nd and short, and thus probably would have lighter personnel on the field (more defensive backs, less defensive linemen and linebackers) to prevent a pass, making them more vulnerable to allowing small gains, but less likely to give up a big gain.

The following set of plots shows the league average predictive posterior distributions for run plays shown for several states, along with corresponding histograms of observed plays coming from the respective states (or closely similar scenarios).

### 3.2. Team model results

Table 1 shows the parameter values obtained for the team adjustments from the league average run model to create a team specific run model.

Similar to the coefficients in the league average model, we can interpret the individual effects of the team adjustment coefficients as well. For example, a negative  $\beta_{Team}^{\xi}$  value, as seen for teams like the Detroit Lions (DET), implies that the center of the skew- $t$  has been shifted negatively from league average. This implies that holding all other parameters constant, the predicted mean yards per carry on comparable states would be lower for Detroit than the average team. However, this team also has a positive  $\beta_{Team}^{\omega}$  adjustment, implying their spread of results is higher than the average team. This (holding  $\beta_{Team}^{\alpha}$  and  $\beta_{Team}^{\nu}$  constant) indicates the team likely has a fairly boom or bust running game, perhaps with a small proportion of 4 to 6 yard gains, but more extreme results such as a 5 yard loss or 15 yard gain. This aligns with their results over the sample period. Interpretations such as these can be continued for the other parameters, providing insight into a team’s trends and proficiencies. To ensure the validity of these estimates, we created several plots of the predictive distributions, indicating a strong fit for each of the individual teams. Several of these plots have been included in Figure 5. Furthermore, we assessed goodness of fit under the skew- $t$  distribution for the 1st and 10 from the team’s own 25 yardline state for each team. Figure 6 shows the distribution of  $p$ -values, showing no unexpected behavior as indicated by distribution

Table 2

Table of  $\beta_{Team}^{\theta}$  values. Values of  $Team$  are shown going down and values of  $\theta$  are shown going across

$\beta_{Team}^{\theta}$	$\xi$	$\omega$	$\alpha$	$\nu$	$\beta_{Team}^{\theta}$	$\xi$	$\omega$	$\alpha$	$\nu$
ARI	0.221	-0.727	-0.2087	-0.174	LA	0.184	1.200	0.3438	0.083
ATL	-0.540	0.464	0.2189	-0.582	LAC	-0.320	-0.081	0.3193	-0.820
BAL	0.124	0.164	0.0369	-0.085	MIA	-0.424	0.243	0.1106	-0.232
BUF	-0.349	0.034	-0.3103	0.225	MIN	0.113	-0.588	-0.2633	-0.671
CAR	0.071	0.287	0.5128	-0.471	NE	0.086	0.897	1.0385	0.178
CHI	0.173	-0.174	-0.2609	-0.558	NO	0.377	-0.604	0.3960	-1.140
CIN	-0.304	0.201	0.1803	-0.345	NYG	0.333	-0.711	-0.4769	-1.044
CLE	0.021	-0.135	0.1540	-0.762	NYJ	-0.704	-0.254	0.4933	-0.508
DAL	-0.024	0.451	0.4805	-0.045	OAK	-0.208	0.209	0.4227	-0.151
DEN	0.160	-0.058	0.2322	-0.327	PHI	-0.013	0.268	-0.0043	-0.209
DET	-0.615	0.302	0.6383	-0.180	PIT	0.417	-0.283	-0.0489	-0.479
GB	-0.250	0.611	0.6774	0.260	SEA	-0.414	0.351	0.4147	-0.041
HOU	-0.091	-0.070	-0.0066	-0.142	SF	-0.668	0.483	0.6205	-0.210
IND	0.136	-0.136	-0.0745	0.036	TB	-0.066	-0.225	-0.1695	0.539
JAX	0.241	-0.368	-0.0148	-0.116	TEN	-0.332	0.113	0.1412	-0.312
KC	-0.200	0.311	0.7544	-0.676	WAS	-0.251	-0.132	-0.1239	-0.300

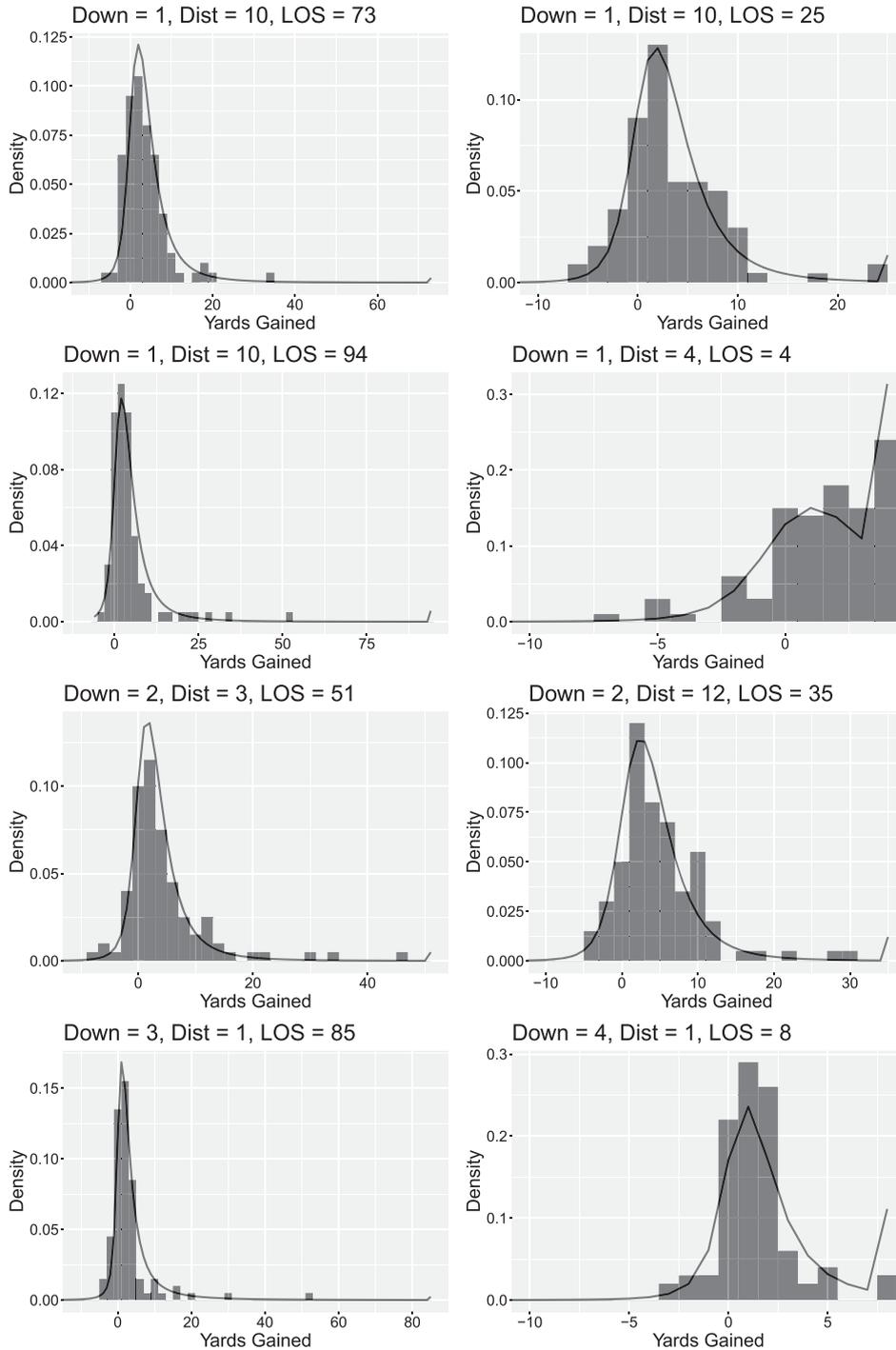


Fig. 4. Several examples of randomly chosen predictive distributions for run plays for a specific state, with observed data from similar states shown in histograms. Distributions have been truncated at -10 yards gained to expand images.

appearing relatively uniform, as is assumed under the null hypothesis.

We also found it to be a fruitful task to compare the predictive distribution of run plays across the differ-

ent teams to identify that we are able to distinguish between different team proficiencies and weaknesses. Figure 7 show a few different team predictive run distributions plotted against one another for identical

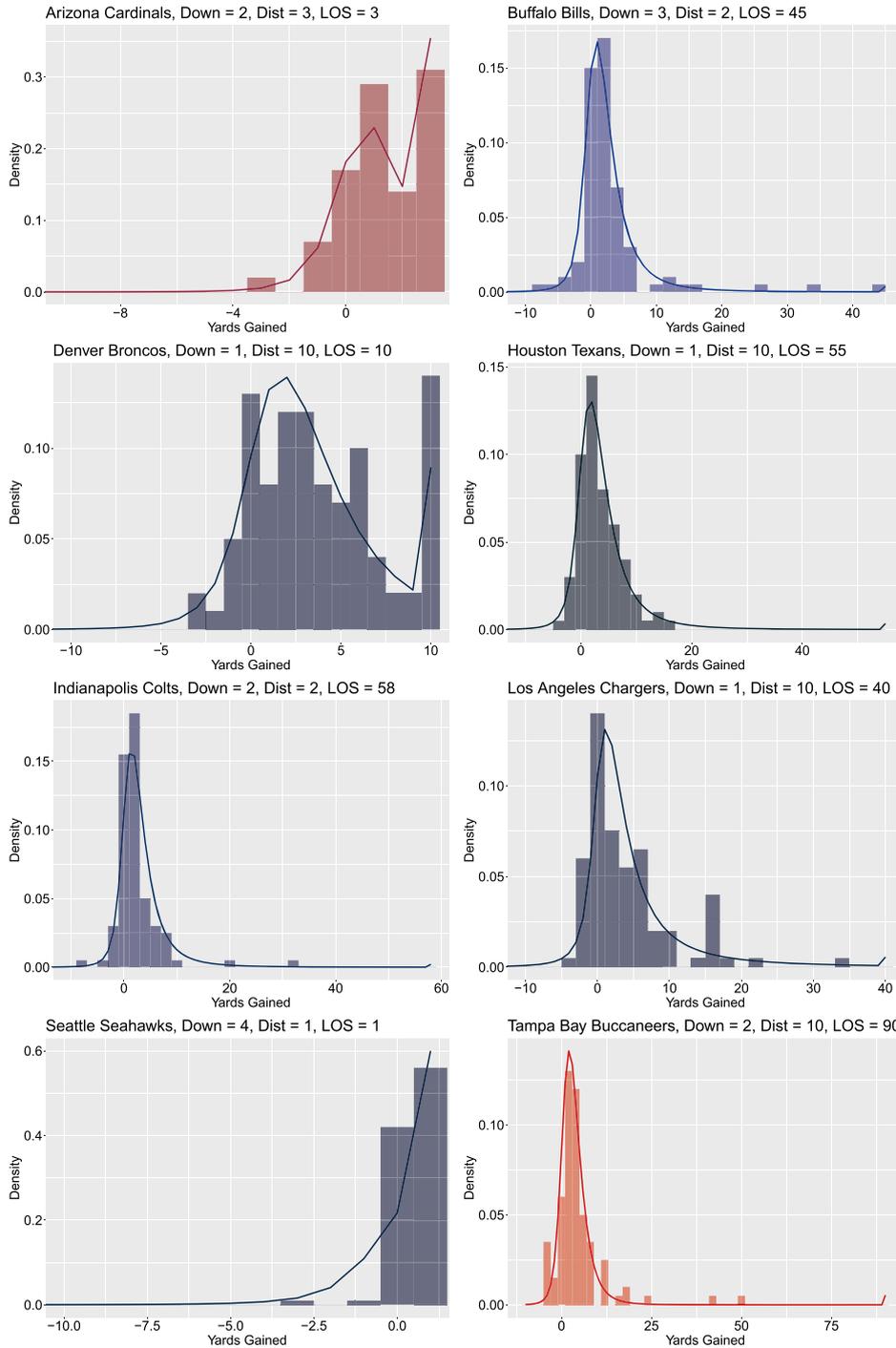


Fig. 5. Several examples of randomly chosen predictive distributions for run plays for a specific team for a specific state, with observed data from similar states for the team shown in histograms.

scenarios. In these plots, one is able to compare the different properties of the predictive distributions for different teams in a visual manner. For example, one can see that the Los Angeles Rams (LA) has a heavier

positive tail than the their divisional opponents (the NFC West), indicating they are more likely to have a larger gain than the Cardinals, Seahawks, and 49ers (ARZ, SEA, and SF). However the positive heavy tail

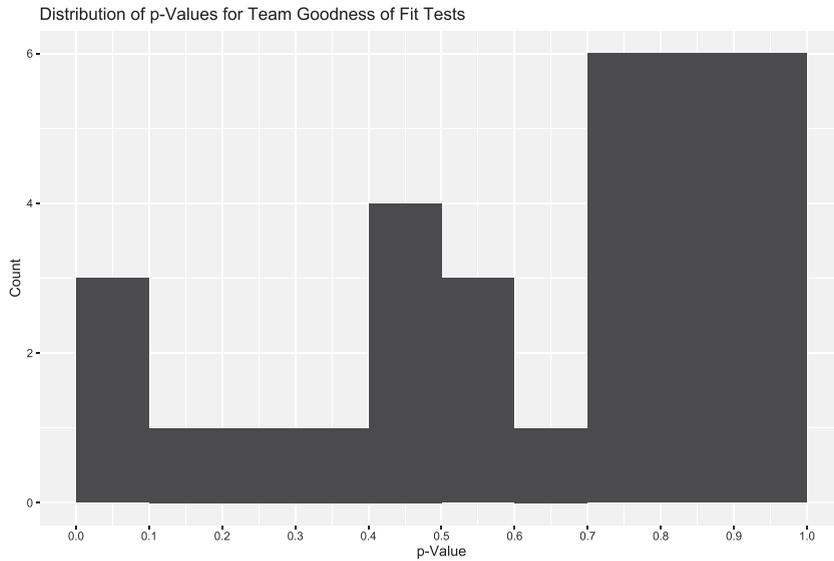


Fig. 6. Goodness of fit  $p$ -values for team parameters for 1st and 10 at their own 25 yardline state. Uniformity implies the skew- $t$  remains a reasonable distribution with which to model to data.

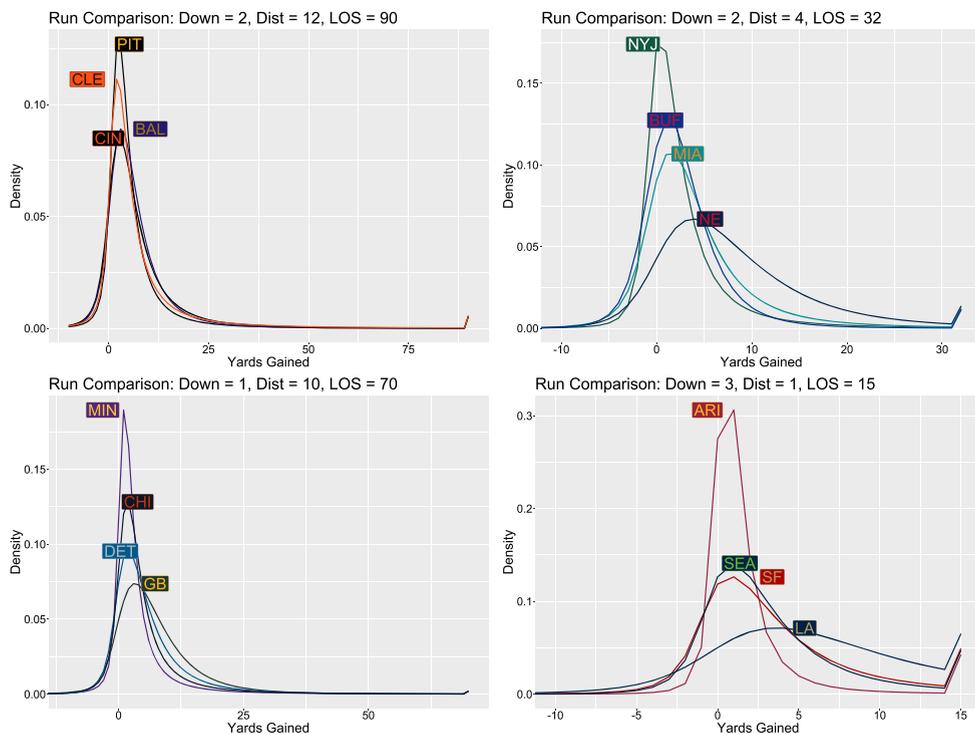


Fig. 7. Comparison plots showing differences in predictive run distributions for different teams for a common state. Colors provided via R's *teamcolors* package (Baumer and Matthews, 2020).

has a cost, specifically by having a larger negative tail as well, indicating they also are more likely to have a run play for a loss. It is apparent that the predictive

distributions are both proper fits for the individual teams and that they are useful in distinguishing trends at the team level.

### 4. Analysis

With these team adjustment parameters identified, we can examine them to identify what trends exist and correlate with successful run games. To do this, we created plots for each of the four sets of parameters against the team cumulative yards per carry over the two year dataset period. The choice to use yards per carry as the metric for effective rushing offense was chosen due to its simplicity in interpretation as well as its ability to avoid volume issues. Other metrics were checked to ensure consistency, such as EPA per run (Yurko et al., 2018), percentage of successful runs (Carroll et al., 1988), and each team’s ranking amongst the league for each of these statistics. All of these metrics have similar relationships with the parameters, and therefore we will proceed using yards per attempt. We will discuss a potential analysis that could be conducted using EPA/attempt, along with appropriate caveats, in Section 4.1.

Figure 8 shows these plots of the parameters against yards per attempt, where a relationship can clearly be seen between  $\beta_{Team}^{\omega}$  (the variance parameter) and  $\beta_{Team}^{\xi}$  (the skewness parameter) with the average yards per carry for each team in the period (correlations of -0.432 and -0.515 respectively). A test for significance for these correlations finds significant  $p$ -values for both parameters ( $p = 0.0135$  and

0.00257 respectively). The relationships with  $\beta_{Team}^{\xi}$  (the center parameters) and  $\beta_{Team}^{\nu}$  (the degrees of freedom parameters) were negligible, and therefore determined to have no meaningful relationship with running success.

It is a usual phenomenon that high variance, and as a consequence higher skewness, while the mean is held fixed, indicates greater risk. Hence, the natural interpretation is to assume that the better rushing teams were willing to attempt more “high risk” run plays, implying that they are more likely to run the ball in a way that increase the probability of a non-successful or even negative play in order to maximize their ability to generate a large run. It is well known that high variance has long been associated with high risk (Winterhalder, 2007; Coombs and Pruitt, 1960), and thus we will henceforth refer to high variance and skewness runs as high risk runs.

Consequently, there is something fundamental about the style of offense run by the successful teams that can be tied to high risk runs. One conjecture is that outside runs (regardless of the actual play scheme) are considered high risk runs. The idea of outside runs being more effective was partially studied in Seth (2020), but we plan on extending this analysis with a more directed approach employing the parametric regression skew- $t$  model. The nature of an outside run requires the ball carrier to run towards

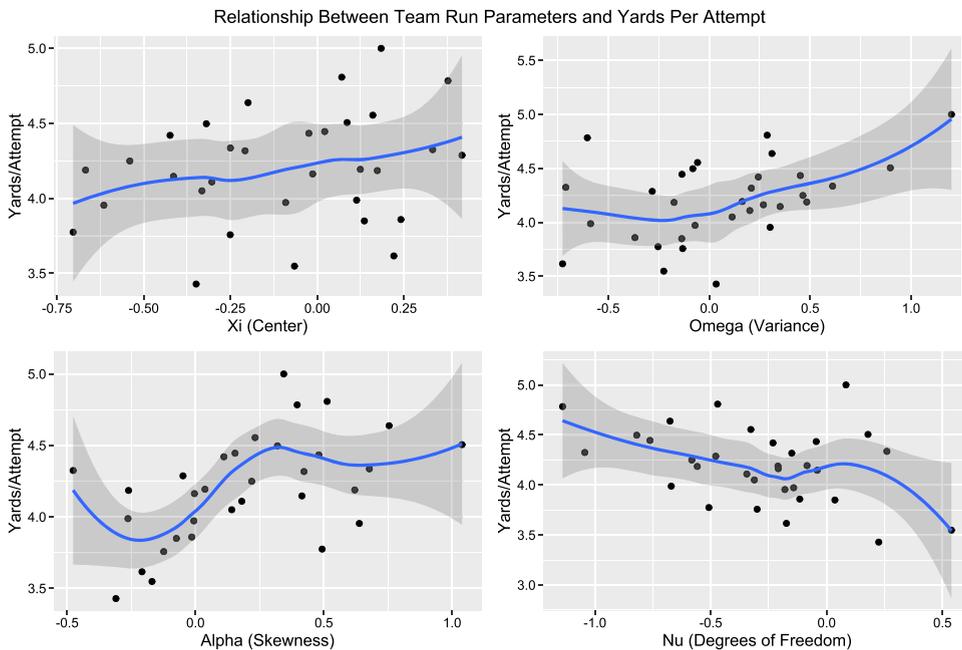


Fig. 8. Parameters from team adjusted run model plotted against team yards per attempt, indicating a correlation between the high risk parameters and successful running teams.

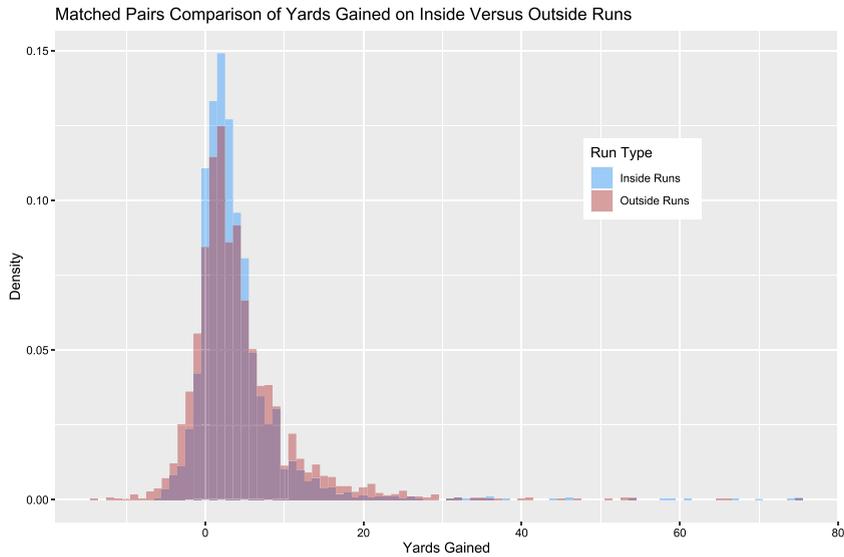


Fig. 9. Overlapping histograms of yards gained on run plays, with blue bars indicating inside runs and red bars indicating outside runs. Plays are matched using one-to-one nearest neighbor matching based on Mahalanobis distance, with exact matching enforced on some factors.

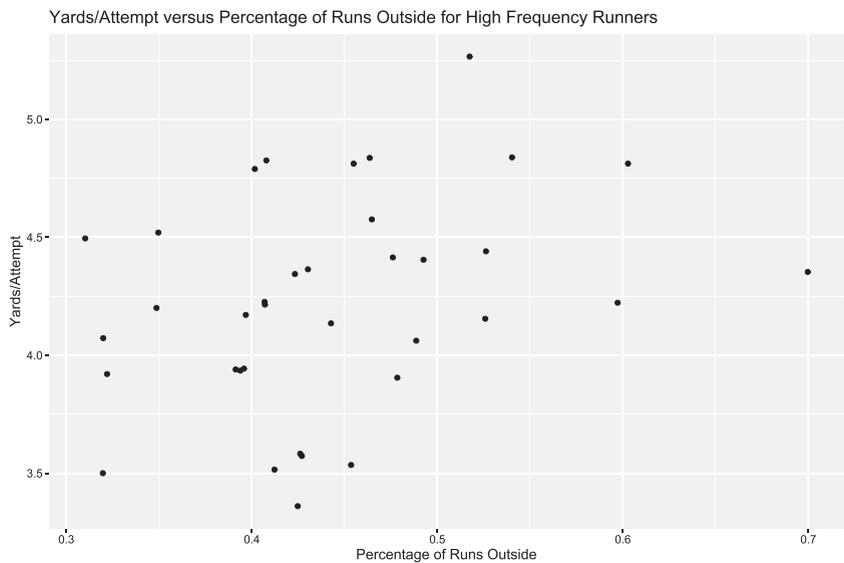


Fig. 10. Scatterplot of yards per attempt versus percentage of runs to the outside for individual runners, specifically for each team's primary runner, showing a positive correlation in between the variables of interest.

the perimeter of the field prior to moving downfield to gain yards, implying that the risk of losing yards may be greater than a play where the ball carrier runs directly downfield. This higher risk of losing yards should also come with a higher chance of creating large gains, as moving outside may allow the back to avoid the larger potential tacklers (typically defensive linemen and linebackers) and may be more likely to induce a broken tackle from a smaller player (defensive backs). Therefore, we expect an outside

run to have an association with high risk run activity.

We should make a clarification here of how we identify an inside or outside run. For the purposes of this paper, we will classify "outside" runs as plays where the *run\_gap* column is marked as "tackle" or "end" according to the *nflfastR* labeling scheme (Carl and Baldwin, 2022). A visual inspection of the film of these runs show that these typically align with runs that are run to a gap located at least outside the

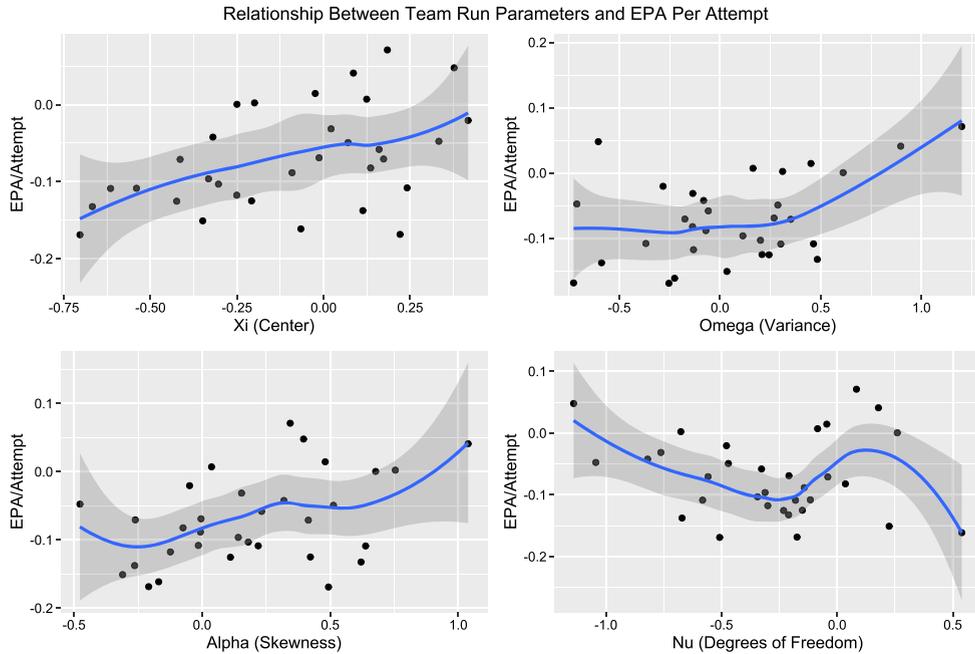


Fig. 11. Parameters from team adjusted run model plotted against team EPA/attempt.

guard (commonly referred to as the B gap), and are more commonly restricted to being outside the tackle (C gap) or even end (D gap). Additionally, the “run location” is marked as “left” or “right”, as opposed to “middle”, which seems to correspond nearly perfectly with “run gap” being labeled “NA”, implying that the run is always “middle”, “guard”, “tackle” or “end”. These runs being labeled as “outside” for our purposes encompass about 47% of all runs, with all other runs being labeled as “inside” runs.

It is also important to note that we recognize that a coach does not call “inside” or “outside” run as their play. While inside and outside zone run concepts exist, these concepts do not fully correspond to what we are labeling as inside and outside runs. The run is classified by an external source that is agnostic to the actual play calling scheme and is determined (by what it seems) solely by the final location of where the back runs. Any run scheme can be used in accordance with the findings of this paper, and therefore the results should be taken into consideration properly in order to determine how an individual team can adjust their scheme to be in line with more successful run behavior.

With these caveats in mind, we calculated the correlation between a team’s percentage of outside run plays called with the parameters of interest. Finding a correlation of .448 with the team variance parameter

(and a slightly less interesting .323 correlation with the team skewness parameter), our hypothesis was supported by the moderate relationship. Furthermore, we conducted a test of significance to determine whether the relationship is statistically significant. Using a significance level of 5%, we performed a hypothesis test for the regression slope of percentage of outside runs. To account for the bounded nature of the percentage of outside runs statistic, we converted the value to the log odds of percentage of outside runs and used it as a covariate in a linear regression model for  $\beta_{Team}^{\omega}$ . This model creates a regression coefficient of 0.594 (standard error of 0.216), resulting in a  $p$ -value of 0.00991, implying there is evidence to suggest that increasing the percentage of outside runs has a significant relationship with the variance parameter of this skew- $t$  run model. Since we are aware that this variance parameter also has a positive relationship with the overall team offensive running proficiency, we are inclined to believe that this means that teams that run the ball outside more should experience higher levels of rushing success.

We will attempt to back this claim via two methods. First, if outside rushes are truly linked to positive results in the running game, we should expect a causal relationship to exist between rushing outside and yards gained on carries. To test this, we conducted a matched pairs observational study, using inside ver-

sus outside run as the treatment variable and yards gained as the outcome. We matched plays (Ho et al., 2011) according to several variables, requiring exact matches for offensive team, defensive team, season (2017 or 2018 regular season), and early/late down, and relatively close matches (using Mahalanobis distance) for distance to 1st down, line of scrimmage, and score differential. After matching, we had 2628 pairs of plays. After confirming balance amongst covariates, we calculated the average treatment affect (Rubin, 1974) of running outside on yards gained. We observe an expected difference of 0.503 yards gained (95% confidence interval: (0.160, 0.846)), implying that running outside has a significant causal effect on yards gained. Figure 9 shows the distribution of each play in the matched pairs sample, with the color of the histogram bars indicating inside versus outside runs. It again becomes apparent in this plot that outside runs are associated with high risk behavior, as large losses and gains happen at a much more frequent rate than what is observed for inside runs. But it also becomes clear how outside runs would generate more yards on average, with a much larger proportion of big gains occurring, bringing the overall average up.

Furthermore, we would expect the same trend to exist at the individual level. Figure 10 shows the relationship between yards gained and percentage of outside runs. Each point represents a player who was the primary rusher for their team during the sample period, removing one outlier (as determined by a standard Cook's distance outlier test (Cook, 1977)) and all non-primary runners to avoid players with lower sample sizes from clouding the trend. It becomes clear that the trend exists on the individual level, as seen by the positive correlation ( $r = 0.40$ ) between the two variables. From this, we conclude that there is evidence to suggest that running outside more will generally lead to a more efficient rushing offense, as measured by yards per attempt.

#### 4.1. Comparison with EPA

One might suggest that it would be helpful to compare the run parameters with expected points added (EPA) per attempt, as EPA has a direct relationship with points scored in the game. One could obtain a similar plot to Figure 8 for EPA/attempt and notice similar correlations ( $r = 0.411$ ,  $p = 0.0195$  for variance,  $r = 0.418$ ,  $p = 0.0173$  for skewness), as seen in Figure 11. The high risk parameters still are associated with an increase in the success metric. One

may even wish to explore the relationship between the  $\beta_{Team}^{\xi}$  parameter, which now seems to have a more meaningful relationship.

However, an analysis of this type should be done with caution. First, the team adjustment parameters were derived using a model built around yards gained as the output. While correlations likely exist between other metrics related to successful offenses, we should not expect the relationship to necessarily be meaningful or replicable unless the modelling process were adapted to the metric of interest. To do so would require a different density for representation, as EPA does not have the same unimodal, smooth features as seen in the skew- $t$ . Moreover, we find that yards are a better tool for discussion, as EPA is not easy for those with a non-technical background to comprehend, as it can not be directly translated to the game as easily as a physical quantity. When possible, tools such as these should be tailored to the target audience.

## 5. Conclusions

The skew- $t$  mixture model allows for an accurate and efficient calculation of the probability transition distribution for run plays that can be used for the play recommendation system described in Biro and Walker (2022). In addition, the parametric nature created reveals trends that allow us to understand underlying factors related to successful rushing. The variance and skewness shifting parameters led us to finding a preference towards high risk run plays, specifically revealing the benefits of outside running. Exploring this directly led us to finding clear evidence that outside runs are associated with an increased yards per attempt.

The benefits of the parameterized approach are clear, not only providing accurate and consistent distributions, but also allowing for interpretable results. This marries the needs of both the data scientist and the coach, where other models may prove too simplistic to incorporate all necessary features or may be too difficult to be accessible by non-technical audiences.

This work opens up room for several followup projects. The most natural would be the extension of similar parametric models for pass plays. However, the structure of pass plays is dissimilar to run plays, with additional point masses needed for incomplete passes and interceptions, along with a non-unimodal curve that would make a skew- $t$  likelihood unsuit-

able. Therefore, another parametric approach would be necessary for pass plays.

Furthermore, there are many ways a team could utilize these results in game preparation. This model provides insight into individual team proficiencies and deficiencies in the run game, and therefore can help a team choose which situations are more appropriate for run plays. A team with a larger positive tail in their predictive distribution would generally be better at generating large runs, and therefore running on 2nd and long may be more appropriate than a team with a larger mean but smaller tail, which would be more apt to pick up short yardage situations. Moreover, a team could adversarially influence an opponent to run in situations which they struggle in by properly aligning their defenders. Finally, a team could examine these results thoroughly to determine which run situations they should spend more time practicing.

In addition, there is room for more expansions to the modeling approach here. For example, the model assumes no dependency between plays, and therefore the sequential nature of plays in a football game are lost, as well as the trends that may be captured within a season. For example, over the course of an NFL season, the main runner of a team may become weary or injured, causing the relative success of the team's running game to decline from week to week. Conversely, at the individual game level, a team may use a run (or pass play) early in the game that tests the defense's response. How the defense responds could be an indicator to the offensive play caller, and may lead them to use a certain run play for a similar look later in the game that may have more success due to the information revealed in the prior plays. While features like this may be present, it would be difficult to represent them in the current model without well labeled data regarding the nature of the effects in question, which is currently not available. Additionally, adding in dependency effects would reduce the ratio of the number of plays to estimated parameters, implying a lower fidelity in the values obtained.

Finally, adversarial tactics could be considered in future approaches. For example, defensive minds such as Vic Fangio and Brandon Staley have recently been lauded by the analytics community in their ability to invite teams to run the ball in situations typically viewed as detrimental for the offense (Pizuta, 2021). Thus, using just the offensive team and state information to model run plays (or pass plays) will lack information regarding the defensive structure. However, it still remains necessary to refrain

from including post-snap or even at-snap information (such as safety alignment or box counts) in these modeling efforts, as the ability to create prescriptive recommendations would be lost. There are alternatives to these measures, such as using metrics such as defensive trend statistics to inform models, such as expected defensive structure in response to certain offensive alignments. This would require additional data to supplement the current methods such as offensive personnel, formations, and defensive response to each feature, which is not currently publicly available and therefore not explored.

### Acknowledgements

The authors are grateful for the helpful discussions with Coach Kevin Kelley which greatly assisted the research. The authors would also like to thank the Editor, Associate Editor and three referees for comments and suggestions on an earlier version of the paper.

### References

- Arellano-Valle, R.B., and Azzalini, A., 2013, The centred parameterization and related quantities of the skew-t distribution, *Journal of Multivariate Analysis* 113, 73-90.
- Baumer, B.S., and Matthews, G.J., 2020, teamcolors: Color palettes for pro sports teams, R package version 0.0.4.
- Biro, P., and Walker, S.G., 2022, A reinforcement learning based approach to play calling in football, *Journal of Quantitative Analysis in Sports* 18(2), 97-112.
- Burke, B., 2010a, Expected points and expected points added explained, Advanced Football Analytics, <https://web.archive.org/web/20210310003124/http://archive.advancedfootballanalytics.com/2010/01/expected-points-ep-and-expected-points.html>.
- Burke, B., 2010b, Win probability added (WPA) explained, Advanced Football Analytics, <https://web.archive.org/web/20210310001449/http://archive.advancedfootballanalytics.com/2010/01/winprobability-added-wpa-explained.html>.
- Brighenti, C., 2019, Modelling run outcomes by field control, NFL Big Data Bowl 2020. <https://operations.nfl.com/media/4201/bdbbrighenti.pdf>
- Carl, S., and Baldwin, B., 2022, nflfastR: Functions to efficiently access NFL play by play data, R package version 4.5.1.
- Carroll, B., Palmer, P., and Thorn, J., 1988, The hidden game of football, Warner Books. New York, NY.
- Cook, D., 1977, Detection of influential observation in linear regression, *Technometrics* 19(1), 15-18.
- Coombs, C.H., and Pruitt, D.G., 1960, Components of risk in decision making: probability and variance preferences, *Journal of Experimental Psychology* 60(5), 265-277.

- Gough, C., 2022, Development of average offensive yards per game (rushing/passing) in the NFL from 1950 to 2021, Statista <https://www.statista.com/statistics/1005862/nfl-average-yards-per-game-development-over-time/>.
- Ho, D., Imai, K., King, G., and Stuart, E., 2011, MatchIt: nonparametric preprocessing for parametric causal inference, *Journal of Statistical Software* 42(8), 1-28.
- Horowitz, M., Yurko, R., and Ventura, S. L., 2017, nflscrapR: Compiling the NFL play-by-play API for easy use in R, R package version 1.4.0.
- Joash Fernandes, C., Yakubov, R., Li, Y., Prasad, A.K., and Chan, T.C.Y., 2020, Predicting plays in the National Football League, *Journal of Sports Analytics* 6(1), 35-43.
- Lefort, R., Guo, W., and Leung, N., 2022, A machine learning approach to predicting NFL plays, NFLPlayPredictions.com.
- Pearson, K., 1900, X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50(302), 157-175.
- Pizzuta, D., 2021, What NFL defenses need to learn from Brandon Staley & Vic Fangio as two-high looks inevitably rise, Sharp Football Analytics <https://www.sharpfootballanalysis.com/analysis/twohigh-shell-staley-fangio-nfl-defense-2021/>.
- Ploenzke, M., 2019, NFL Big Data Bowl sub-contest, NFL Big Data Bowl 2020 <https://operations.nfl.com/media/4204/bdbploenzke.pdf>.
- Rubin, D.B., 1974, Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology* 66(5), 688-701.
- Seth, T., 2020, The NFL's run gap secret, The Michigan Football Analytics Society <https://mfootballanalytics.com/2020/07/15/the-nfls-run-gap-secret/>.
- Stern, A., 2019, Practical applications of space creation for the modern NFL franchise, NFL Big Data Bowl2020. <https://operations.nfl.com/media/4202/bdbstern.pdf>
- Teich, B., Lutz, R., and Kassarnig, V., 2016, NFL play prediction, arXiv preprint, arXiv:1601.00574.
- Winterhalder, B., 2007, Oxford handbook of evolutionary psychology: risk and decision-making, *Oxford University Press*, 433-445.
- Yurko, R., Ventura, S., and Horowitz, M., 2018, nflWAR: A reproducible method for offensive player evaluation in football, *Journal of Quantitative Analysis in Sports* 15(3), 163-183.