Persuaded under pressure: evidence from the National Football League

Michael Lopez

Skidmore Colllege

(mlopez1@skidmore.edu)

August 4, 2015

Abstract

With sizable interests in referee decisions, coaches and players often try to manipulate referee behavior with verbal and nonverbal communications. We exploit a natural experiment within each National Football League game, finding that referees succumb to the pressures of satisfying the personnel in the vicinity of possible violations. Using generalized additive models for binomial outcomes, we find that the penalty rates for certain types of common but influential penalties, including holding and pass interference, vary from one sideline to the next by as much as 50%, depending on play location.

1 Introduction

Referee decision making in sports has been shown to be associated with, among other factors, a social pressure to support the home team [Sutter and Kocher, 2004, Boyko et al., 2007, Pettersson-Lidbom and Priks, 2010, Moskowitz and Wertheim, 2011, Buraimo et al., 2010], changes to both the number of referees [Heckelman and Yates, 2003] and their positioning during play [Kitchens, 2014], player characteristics such as race, size and stature [Price et al., 2007, Mills, 2014, Gift and Rodenberg, 2014], and previous sequences of violations [Lopez and Snyder, 2013, Abrevaya and McCulloch, 2014].

One aspect missing in the literature, but a part of the flow of any athletic event, is accounting for the consistent pressure and monitoring applied on referees by team employees. Formally, coaches and players admit that 'working the referees,' which includes acts of kindness and reverence as well as screaming during times of frustration, is part of a game-plan [Abrams, 2008]. Although there is some evidence that suggests referees can be tricked into favoring either team [Petchesky, 2014], and that certain positions use their stature to gain leverage [Mills, 2014], it us unknown to date if the immense and constant pressure placed on referees by players and coaches in all sports alters or impairs referee judgment. Using data from the National Football League (NFL), we identify four sets of penalties in which referees are forced to make difficult decisions in the presence of one team's sideline. With both standard and advanced modeling strategies, we find significant evidence that in the presence or surrounding of a particular team's coaches and players, referees are more likely to call one of several penalties on that team's opponent. Further, rates of offensive holding, defensive pass interference, and aggressive defensive infractions, including personal fouls and unnecessary roughness, peak near midfield, but only on one team's sideline. Given that this is the most likely spot for team personnel to influence referees one way or the other, we posit that referee decision-making is strongly impacted by a pressure to appease the nearest stake-holders.

2 Data

For the first and third quarters of an NFL game, coin tosses are used to assign teams end zones to defend. At the end of each of these 15-minutes of play, teams flip-flop sides for the second and fourth quarters. Such a set-up ensures that each team plays 30 minutes in each game moving towards and defending both end zones. Assuming that teams call plays independently of sideline location, the first and third quarter side changes make for a natural experiment that happens twice within every game. Because team benches remain on the same sideline for the duration of each contest, the flip-flopping of directions ensures that, on average, each team will run the same number and type of play towards each team's sideline by game's end. As a result, we test if referees are influenced by a pressure to appease the nearest coaches and players by contrasting penalty rates based on finishing sideline.

Armchair Analysis (AA, `www.armchairanalysis.com`) provided play and game-specific characteristics for each regular season NFL play between the 2010 and 2014 seasons. These include each play's offensive unit, defensive unit, line of scrimmage, down, distance, score, and outcome, play directions for both runs and passes, as well as each game's stadium. Directions for runs and passes are labeled as left, middle, or right.[1] Special teams plays, however, which include all kicks and punts, are directionless and dropped.

[1] Pass plays are also categorized as either 'deep' or 'short.' We characterize any run play listed as 'right end' or 'left end' as an outside rush.

End zone directional information is missing from both AA's data and traditional NFL play-by-play reports. We used coin toss information courtesy of Football Outsiders (FO, `www.footballoutsiders.com`), which extracted this information from post-game 'Game Books' that are put out by the NFL. These Game Books, and thus the FO data, are missing 97 halves of coin toss information, roughly 4% of the overall data set. Given that the missing games are scattered throughout the five years, we inferred that this information is missing completely at random [Little, 1988] and performed analysis on the remaining data only.

Sideline information for each of the home and visiting teams was collected manually using team websites and seating guides.[2][3] The stadium, coin toss, and play-by-play data were merged to calibrate our covariate of interest, $Sideline_i$, where for n_r rush plays and n_p pass plays,

$$\begin{aligned} Sideline_i &= \text{Direction of play i}, i = 1, \ldots, n_r, n_{r+1}, \ldots, n_r + n_p \\ &= \{\text{`Offense' if direction of play } i \text{ is the offensive teams sideline, `Defense' if defense's sideline}\} \end{aligned}$$

We drop all middle rushes (those between the tackles) or middle passes from our data; these are by nature different types of play calls than the ones run towards each sideline, and it is less likely for an identifiable sideline bias to impact these plays.[4]

We consider the four outcomes Y_i, where $Y_i \in \{OHR_i, DPI_i, OPI_i, \text{ and } DAP_i\}$, most likely to vary based on pressure to appease team personnel, such that

- $OHR_i = \{1 \text{ if there was an offensive holding penalty on play } i, 0 \text{ o/w}\}, i = 1, \ldots, n_r$[5]
- $DAP_i = \{1 \text{ if there was a defensive aggressive penalty on play } i, 0 \text{ o/w}\}, i = 1, \ldots, n_r + n_s$
- $DPI_i = \{1 \text{ if there was a defensive pass interference penalty on play } i, 0 \text{ o/w}\}, i = n_r{+}1, \ldots, n_r{+}n_p$[6]

[2]Teams using multiple stadiums include Buffalo (also playing home games in Toronto), Minnesota (also at University of Minnesota), and San Francisco (new stadium for 2014). We dropped games played at London's Wembley Stadium given that there is no clear distinction for the game's home and away sidelines

[3]See, for example, `http://images.patriots.com/2014-individualgameprice.jpg`

[4]From a statistical modeling perspective, we also use to 'logit' link in Section 3. This makes the independence of irrelevant alternatives assumption [McFadden et al., 1973], under which the relative odds of a penalty from one sideline versus the other acts independently of plays over the middle of the field.

[5]For offensive holding penalties, we look only at running plays given that on passing plays, most holding violations occur in the center of the field, away from either sideline.

[6]Pass interference penalties cannot be called on rushing penalties.

- $OPI_i = \{1 \text{ if there was an offensive pass interference penalty on play } i, 0 \text{ o/w}\}, i = n_r+1, \ldots, n_r+n_p$

Four violations are included in DAP: unnecessary roughness, personal foul, unsportsmanlike conduct, and horse collar tackle, which each penalize the defensive team 15 yards.

2.1 Methods

We first contrast Y and $Sideline$. Under an existing sideline pressure, we expect higher-than-expected OHP and OPI rates on penalties run in the direction of the defensive team's sideline, and lower-than-expected DAP and DPI rates on penalties run towards the offensive team's sideline. We can also check to determine if there are any systematic differences with $Sideline$ and other play-specific variables, such as down, distance, and the game's score. χ^2 tests are used to contrast each of these associations.

Our second goal is to identify if a referee bias varies by the play's line-of-scrimmage. Coaches, players, and other team staff are required to stand within a 36-yard zone in the center of each sideline, between endpoints of a trapezoid that extends between the field's two marked 32-yard lines [NFL, 2014]. Under a referee bias to appease a sideline, we expect the greatest difference in penalty rates between these two 32-yard line marks (e.g, the middle of the field).

Generalized additive binomial logistic models (GAM, Hastie and Tibshirani [1986]) are used to measure the effect of line-of-scrimmage on Y. GAMs require fewer assumptions and will allow us to more flexibly gauge the association between line-of-scrimmage and penalty likelihood, relative to a purely parametric approach such as multiple logistic regression. For all penalty outcomes Y_i, the effect of LOS_i, the line-of-scrimmage for play i, is measured nonparametrically using the full model

$$\text{logit}(P(Y_i = 1)) = \beta_1 * I(Sideline_i = Offense) + f_{Offense}(LOS_i) + f_{Defense}(LOS_i), \tag{1}$$

where $I(Sideline_i = Offense)$ is an indicator for whether or not the play was run in the direction of the offensive team's sideline, $\text{logit}(p) = \log(\frac{p}{1-p})$, and $f_{Offense}(LOS_i)$ and $f_{Defense}(LOS_i)$ are the smoothed functions of the log-odds of a penalty based on plays run at the offensive and defensive teams sidelines, respectively, by LOS.

Model (1) allows for plays towards each sideline to have different baseline penalty rates, as measured through the β_1 coefficient, and for line-of-scrimmage effects to vary based on *Sideline*. Separate *LOS* surfaces are estimated for each play direction using penalized thin-plate regression splines [Gu and Wahba, 1993]. As in Wood [2006] and implemented by Mills [2014], the model was fitted using penalized iteratively reweighted least squares, and a generalized cross-validation procedure was used to prevent overfitting with the smoothing degrees of freedom.

For each Y, Model (1) is compared is compared to three reduced fits:

$$\text{logit}(P(Y_i = 1)) = \beta_1 * I(Sideline_i = Offense) + f_{Overall}(LOS_i), \quad (2)$$

$$\text{logit}(P(Y_i = 1)) = f_{Overall}(LOS_i), \quad (3)$$

$$\text{logit}(P(Y_i = 1)) = \beta_1 * I(Sideline_i = Offense). \quad (4)$$

Models (2) and (3) include an overall surface term for line of scrimmage, $f_{Overall}(LOS_i)$, implicitly making the assumption that any effect of the play's line-of-scrimmage on penalty likelihood acts independently of play direction. If (2) or (3) provide stronger fits than (1), we would conclude that a *LOS* impact does not significantly differ by *Sideline*. Model (4) assumes no effect of the game's line-of-scrimmage on a penalty outcome, but like Model (2), still allows for differences in the baseline penalty rates by *Sideline*. Models are fit in the *R* statistical software [R Core Team, 2014], and are contrasted using the Akaike Information Criterion (AIC, Akaike [1974]), which includes a penalty for unneeded parameters to discourage overfitting.

3 Results

AA's and FO's data yielded 152,751 offensive plays, including 69,636 which were discarded as middle runs or middle passes. Of the 83,115 outside plays, $n_p = 66,925$ were passes (80.5%), with the remaining plays rushing attempts. Just over half (42,322, 50.9%) were run in the direction of the offensive team's sideline.

There do not appear to be any obvious differences between play calls (run/pass) or play situations (score, down & distance) based on *Sideline* (Table 1). However, there are significantly higher rates of *DAP* (p-value <0.001) and *DPI* (p-value = 0.018) among plays run at the offensive team sideline. This follows our hypothesized association that suggested more defensive penalties in the presence of offensive team personnel. There are not significant differences in the rates of OHR or OPI by $Sideline$.

Table 1: Play characteristics and penalty outcome counts (with %'s) by play direction

		$Sideline$		
Covariate	Level	$Offense$	$Defense$	p-value*
Play Type	Rush	8326 (19.6)	7864 (19.3)	0.152
	Pass	32929 (80.7)	33996 (80.4)	
Score (offense)	Behind 2+ possessions	9695 (22.9)	9425 (23.1)	0.456
	Behind 1 possession	11253 (26.6)	10624 (26.0)	
	Tied	7748 (18.3)	7502 (18.4)	
	Up 1 possession	8163 (19.3)	7984 (19.6)	
	Up 2+ possessions	5463 (12.9)	5258 (12.9)	
Down/Distance	First and 10	17436 (41.2)	16935 (41.5)	0.750
	Second and long	10747 (25.4)	10411 (25.5)	
	Second and short	3450 (8.2)	3273 (8.0)	
	Third/fourth and long	6229 (14.7)	5911 (14.5)	
	Third/fourth and short	4460 (10.5)	4263 (10.5)	
Penalty Outcomesˆ	OHR	278 (3.3)	298 (3.8)	0.133
	DAP	298 (0.7)	205 (0.5)	< 0.001
	DPI	494 (1.5)	408 (1.2)	0.018
	OPI	179 (0.5)	183 (0.6)	0.643
Total plays		42322	40793	

ˆOHR taken on run plays, DPI and OPI pass plays, DAP on all plays

*Calculated using χ^2 tests of association

After adjusting for LOS using GAMs, there are baseline differences in both DAP and DPI across all model specifications, as judged by significant β_1 estimates (Table (2)). There is an estimated 39% increase (95% confidence interval (CI), 1.16 to 1.66) in the odds of a DAP when the play is run towards the offensive teams sideline, compared to one run towards the defensive teams sideline. The overall odds of a defensive pass interference are 22% higher (95% CI, 1.06 to 1.40) on plays run towards the offensive teams sideline. There are no significant differences in the baseline rates of OHR or OPI violations. For each Y, estimates of β_1 are robust between the four models.

As implied by a noticeably lower AIC, Model (1) produces the best fit for OHR and DPI outcomes, implying that the effect of LOS on penalty likelihood significantly differs by $Sideline$ for these penalties. There do not appear to be unique sideline effects based on LOS for DAP or OPI,

Table 2: Model statistics and estimated coefficient

	OHR		DAP		DPI		OPI	
Model	AIC	$\hat{\beta}_1(SE_{\hat{\beta}_1})$	AIC	$\hat{\beta}_1(SE_{\hat{\beta}_1})$	AIC	$\hat{\beta}_1(SE_{\hat{\beta}_1})$	AIC	$\hat{\beta}_1(SE_{\hat{\beta}_1})$
(1)	**4969**	-0.117 (0.086)	6145	0.331 (0.091)**	**9519**	0.199 (0.069)**	4498	-0.056 (0.107)
(2)	4976	-0.130 (0.085)	**6142**	0.334 (0.091)**	9523	0.161 (0.067)*	4495	-0.054 (0.105)
(3)	4976	N/A	6154	N/A	9527	N/A	**4493**	N/A
(4)	4978	-0.130 (0.085)	6142	0.335 (0.091)**	9561	0.162 (0.067)*	4505	-0.054 (0.105)

ˆOHR taken on run plays, DPI and OPI pass plays, DAP on all plays

*(**) p-value <0.05 (0.01)

Model with lowest AIC in **bold**

whose lowest AIC's are achieved via Models (2) and (3), respectively.

While estimates from GAMs can be difficult to interpret,[7] one advantage to this framework is that estimated probabilities can be used to depict the relationship between LOS and penalty likelihood. Figures 1 through 4 show estimated penalty probabilities, along with 95% confidence intervals, based on each play's LOS and using model (1) for each penalty outcome. The x-axis in each figure represents the LOS, with the offensive team moving out of its own end zone from left to right. The 36-yard area containing team personnel lies between yard markers representing 32 and 68 yards from the offensive team's end zone.

While OHRs are relatively consistent on plays run towards the offensive team's sideline, there is a noticeable spike on plays run towards the defensive team's sideline between roughly each of the 30-yard markers (Figure 1). At the 50-yard line, the fraction of OHRs is significantly higher - a difference of about 12 penalties per 1000 plays - on play's run at the defensive team's sideline. This suggests a successful pressure from defensive coaches and sideline players to draw violations on the offense, but only in locations where personnel are located during the run of play.

DAPs are higher on plays run at the offensive team's sideline (Figure (2)). In the middle of the field, infractions are about 50% more likely to occur in front of the offensive team's sideline. The differences in DAPs by $Sideline$ are smaller closer to each end zone.

Near an offensive teams own end zone, DPIs are called significantly more often on plays run in the direction of the offensive team sideline. In contrast to OHRs and DAPs, however, differences in DPI rates by $Sideline$ subside around the 50-yard line. This initially seems to violate our hypothesis of a greater sideline effect near midfield, but keep in mind that most DPIs occur well beyond the LOS, at the yard line to which the ball was thrown. As a result, passes thrown from

[7] Smoothing parameter estimates are available upon request.

near the offensive team's end zone may yield a violation at midfield, and in the presence of a sideline pressure, but passes thrown from midfield may land well beyond where teams are allowed to stand on each sideline.

Unlike our other outcomes, there are no noticeable differences in OPI rates by LOS or sideline.

4 Discussion

It is generally assumed that football referees assess each possible violation on its own accord. However, we find multiple penalties whose likelihoods significantly vary based on whether or not the violation occurred in the presence of one team's sideline. Further, model estimates suggest that the sideline locations including team personnel result in the largest differences in call rates. The highest ratios of the estimated differences range from 1.3 (offensive holding) to 2 (defensive pass interference) times as many violations, comparing one sideline to the other.

Models (1) through (4) do not adjust for any variables or functions of variables besides $Sideline$ and LOS. However, our covariate of interest, $Sideline$, appears to act independently of other play and game-specific covariates (Table 1). As a result, problems such as omitted variable bias, in which variables that effect both the treatment ($Sideline$) and the outcome (Y) can produce confounded associations, should not impact our estimated $Sideline$ effects. To check this assumption, we also adjusted each model for point differential, down and distance, game minute, a season-specific factor, and an indicator for whether or not the home team was the offensive team. Each estimated β_1 in these adjusted fits was within a one-hundreth of the estimates shown in Table 2.

While the effect size for DAP is relatively small, 45% of the roughly 1700 DAPs since 2010 were unaccounted for in our data set because they occurred on special teams.[8] Given that several of these missing calls came on kick or punt returns, the importance of a sideline pressure on DAP calls is potentially higher than what we were able to find. Other than the 4% of halves missing coin toss data, all sideline OPIs, DPIs, and OHRs are accounted for.

In principal, each of our penalty outcomes should result in enough punishment on the offending team as to deter players from committing the foul to begin with.[9] One alternative explanation

[8] As mentioned in Section 2, directional information for special teams plays is not included in the NFL's play-by-play reports.

[9] DPI is a spot foul, where the ball is placed where the foul occurred. Accepted OHR and OPI calls each penalize the offensive team ten yards, while DAP calls are worth 15-yards against the defense.

for a referee sideline bias, however, is that players adapt their behavior in the presence of coaches and players on the other team, perhaps acting more aggressively near opponents. However, such an explanation would not justify the observed differences in OHRs, given that most offensive personnel first engage in contact close to the line-of-scrimmage, away from referees and the sideline.

It does not appear that any team or unit has taken advantage of the varying referee behaviors. For example, we checked to see if there were any offense's with relatively higher frequencies of passing plays towards their own sideline, with the intent of drawing defensive pass interferences. The highest fraction of team-specific sideline calls is Jacksonville (53.1% towards its own sideline), and the lowest is Oakland (48.8%), with other team rates symmetrically distributed in between those two cutoffs.

We compared DPI by LOS because the NFL's play-by-play data does not give precise information on where each pass was thrown, other than to identify each throw as either 'deep' (roughly 15 yards or more) or 'short.' Using these labels as proxies for where each pass was thrown to, we also categorized passes by whether or not they were likely to have landed near team locations on the sideline. These plays included all deep throws from before a team reaches the 50-yard line while coming out of its end zone, and any short throw between a teams own 15-yard line and their opponents 30-yard line. Of n_p, 46,225 (69.1%) were plays that more than likely were thrown in vicinity of team personnel on either sideline. The ratio of defensive pass interference calls was nearly 3:2 (raw counts, 293 and 204) in this window, favoring plays towards the offensive teams sideline. On passes thrown near one of the end zones, the ratio of flags by sideline was roughly 1:1 (counts of 101 and 204). Such a drastic difference adds context to Figure 3, which is limited in the fact that it uses only the LOS to contrast DPI frequencies.

While DPI frequencies appear to vary based on the location of the throw, we found no significant differences in OPI rates. However, there are several reasonable explanations for the null finding. First, OPI may require less discretion on behalf of the officials, in which case such decisions would be less likely to vary based on a referee's pressure to appease. Second, there are fewer than half as many OPIs as DPIs. In addition to making it more difficult to discern a statistical difference with OPI, perhaps it is also less likely that coaches and players try to sway referee decisions with such a relatively rare infraction.

5 Conclusion

A home bias due to referee calls has been extensively studied, and one purveying theory is that under duress, referees use crowd noise as a cue with which to inform decisions [Nevill et al., 2002, Sutter and Kocher, 2004, Unkelbach et al., 2010]. This confirms work in psychology, where it has been shown that the most salient cues have the largest effects when people are forced to make decisions under a time pressure [Wallsten and Barton, 1982].

Until now, however, it has been assumed that the primary impetus for uncertain referees has been a crowd noise in favor of the home team [Nevill et al., 2002]. We propose that in several settings, sideline pressure dwarfs that of the home crowd.[10] However, this isn't to say a home bias does not exist in football; instead, we argue that if noise is a cue, a sideline noise is more salient than that of the home crowd. This follows results of Buraimo et al. [2010], who noted effect of noise on referee behavior was proportional to how close noise was to the referees.

Evidence is strong that team sidelines are rent seeking, exerting an influence on NFL referee behavior without reciprocation. Several sports, for example basketball, hockey, and soccer, also similar planned directional changes to the one contrasted here. However, granular data with respect to the location of referees and participants in these sports may be difficult to obtain. If such information is available, future work is warranted to unify and extend the effects presented here.

Acknowledgement

The author would like to thank Cody Couture for his persistence and assistance in this project.

[10] Using interaction terms in Models (1) through (4), we also checked if our sideline effects differed based on whether or not the home team was on offense. There was no evidence that a sideline effect differed based on the offensive team's status.

References

Jonathan Abrams. *Working the Refs Is Part of a Coach's Game Plan*, 2008. URL `http://www.nytimes.com/2008/12/16/sports/basketball/16refs.html?pagewanted=all&_r=0`.

Jason Abrevaya and Robert McCulloch. Reversal of fortune: a statistical analysis of penalty calls in the national hockey league. *Journal of Quantitative Analysis in Sports*, 10(2):207–224, 2014.

Hirotugu Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974.

Ryan H Boyko, Adam R Boyko, and Mark G Boyko. Referee bias contributes to home advantage in english premiership football. *Journal of sports sciences*, 25(11):1185–1194, 2007.

Babatunde Buraimo, David Forrest, and Robert Simmons. The 12th man?: refereeing bias in english and german soccer. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173 (2):431–449, 2010.

Paul Gift and Ryan M Rodenberg. Napoleon complex height bias among national basketball association referees. *Journal of Sports Economics*, page 1527002514535168, 2014.

Chong Gu and Grace Wahba. Smoothing spline anova with component-wise bayesian "confidence intervals". *Journal of Computational and Graphical Statistics*, 2(1):97–117, 1993.

Trevor Hastie and Robert Tibshirani. Generalized additive models. *Statistical science*, pages 297–310, 1986.

Jac C Heckelman and Andrew J Yates. And a hockey game broke out: Crime and punishment in the nhl. *Economic Inquiry*, 41(4):705–712, 2003.

Carl Kitchens. Identifying changes in the spatial distribution of crime: Evidence from a referee experiment in the national football league. *Economic Inquiry*, 52(1):259–268, 2014.

Roderick JA Little. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404):1198–1202, 1988.

Michael Lopez and Kevin Snyder. Biased impartiality among national hockey league referees. *International Journal of Sport Finance*, 8(3):208–223, 2013.

Daniel McFadden et al. Conditional logit analysis of qualitative choice behavior. 1973.

Brian M Mills. Social pressure at the plate: Inequality aversion, status, and mere exposure. *Managerial and Decision Economics*, 35(6):387–403, 2014.

Tobias Moskowitz and L Jon Wertheim. *Scorecasting: The hidden influences behind how sports are played and games are won.* Crown Archetype, 2011.

Alan M Nevill, Nigel J Balmer, and A Mark Williams. The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, 3(4):261–272, 2002.

NFL. *NFL Rulebook*, 2014. URL `http://static.nfl.com/static/content/public/image/rulebook/pdfs/4_2013_Field.pdf`.

Barry Petchesky. *Earl Thomas Says The Seahawks "Are Playing The Referees Too"*, 2014. URL `http://deadspin.com/earl-thomas-says-the-seahawks-are-playing-the-referees-1648391758`.

Per Pettersson-Lidbom and Mikael Priks. Behavior under social pressure: Empty italian stadiums and referee bias. *Economics Letters*, 108(2):212–214, 2010.

Joseph Price, Justin Wolfers, et al. *Racial discrimination among NBA referees.* Centre for Economic Policy Research, 2007.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2014. URL `http://www.R-project.org/`.

Matthias Sutter and Martin G Kocher. Favoritism of agents–the case of referees' home bias. *Journal of Economic Psychology*, 25(4):461–469, 2004.

Christian Unkelbach, Daniel Memmert, et al. Crowd noise as a cue in referee decisions contributes to the home advantage. *Journal of sport & exercise psychology*, 32(4):483–498, 2010.

Thomas S Wallsten and Curtis Barton. Processing probabilistic multidimensional information for decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(5):361, 1982.

Simon Wood. *Generalized additive models: an introduction with R*. CRC press, 2006.

Table 3: Penalty frequencies (%'s) by play direction

Outcome	*Sideline* Offense	*Sideline* Defense	p-value*
OHR	278 (3.3)	298 (3.8)	0.133
DAP	298 (0.7)	205 (0.5)	< 0.001
DPI	494 (1.5)	408 (1.2)	0.018
OPI	179 (0.5)	183 (0.6)	0.643

*Using χ^2 tests

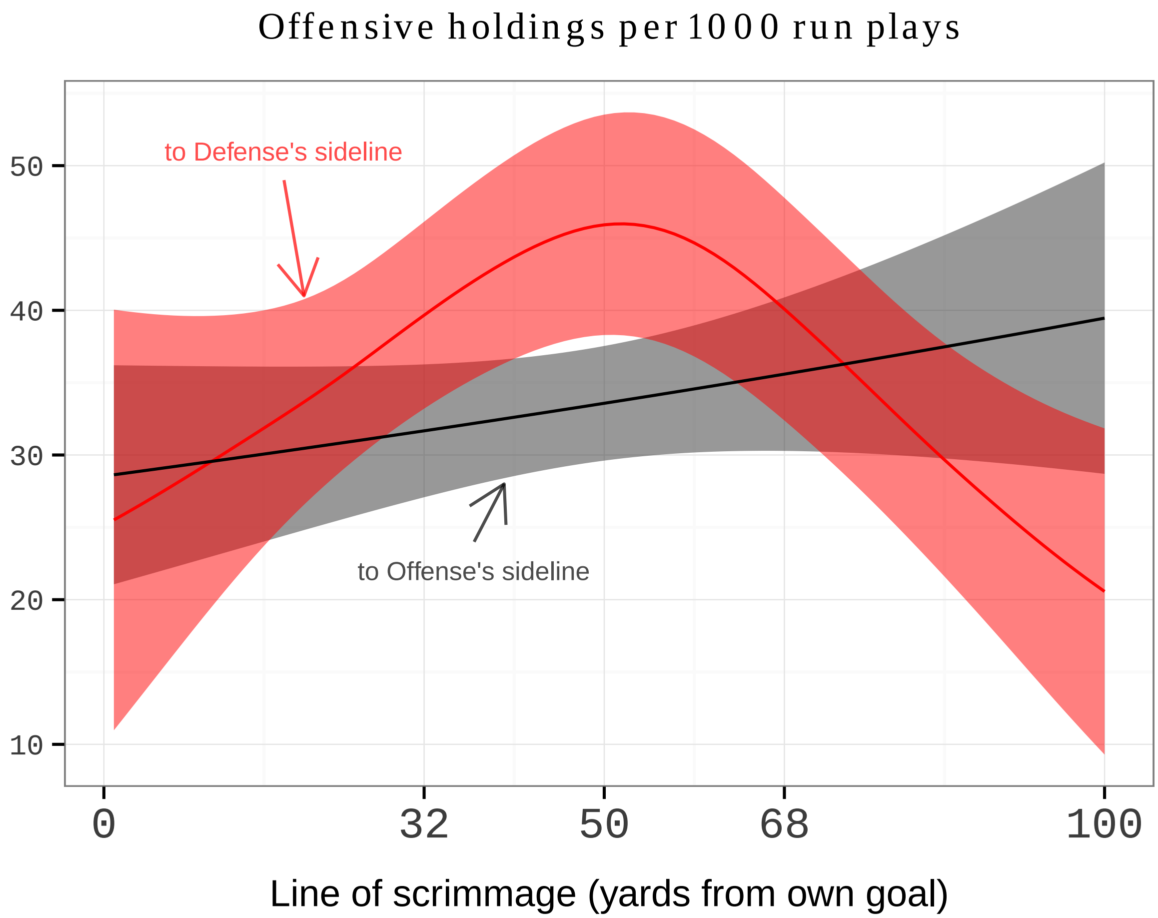

Figure 1: Offensive holding penalties by sideline, line of scrimmage, estimated per 1000 running plays with 95% confidence intervals

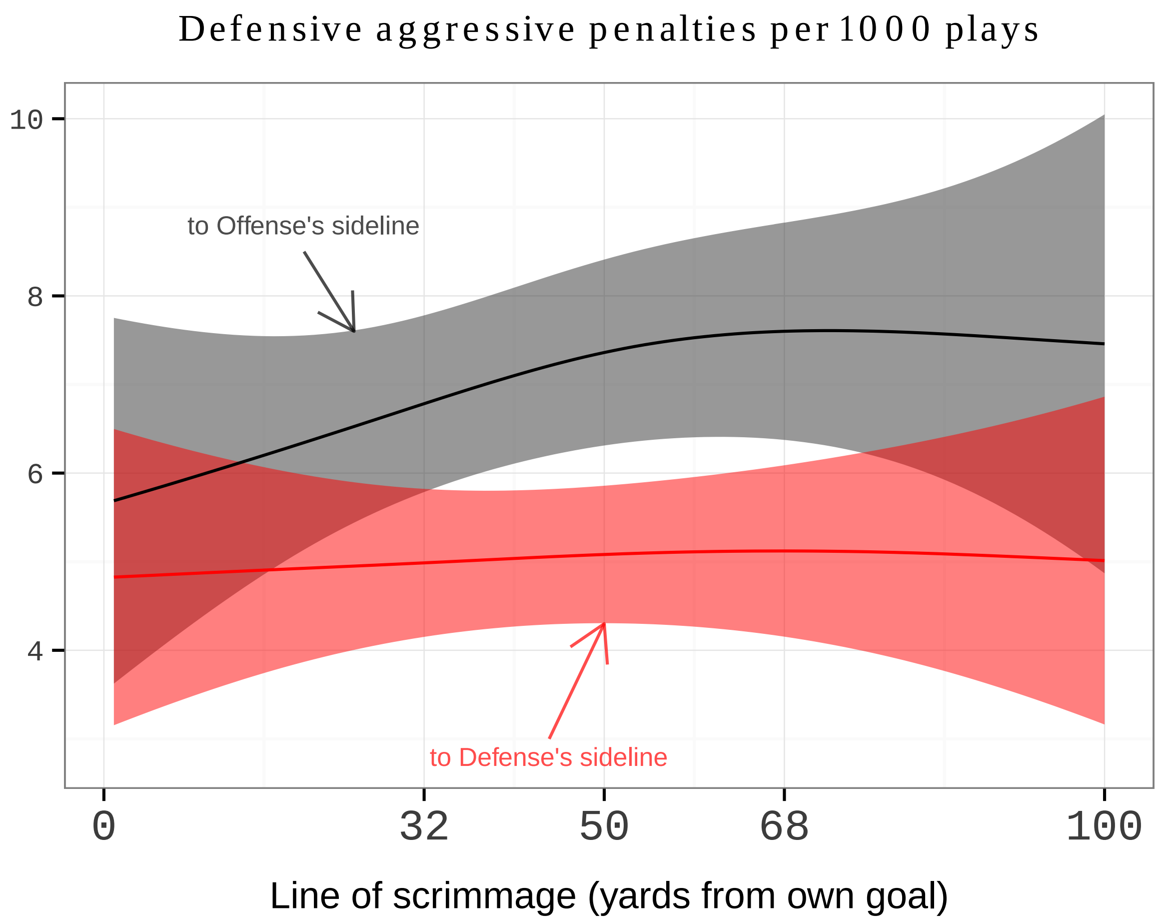

Figure 2: Defensive aggressive penalties by sideline, line of scrimmage, estimated per 1000 plays with 95% confidence intervals

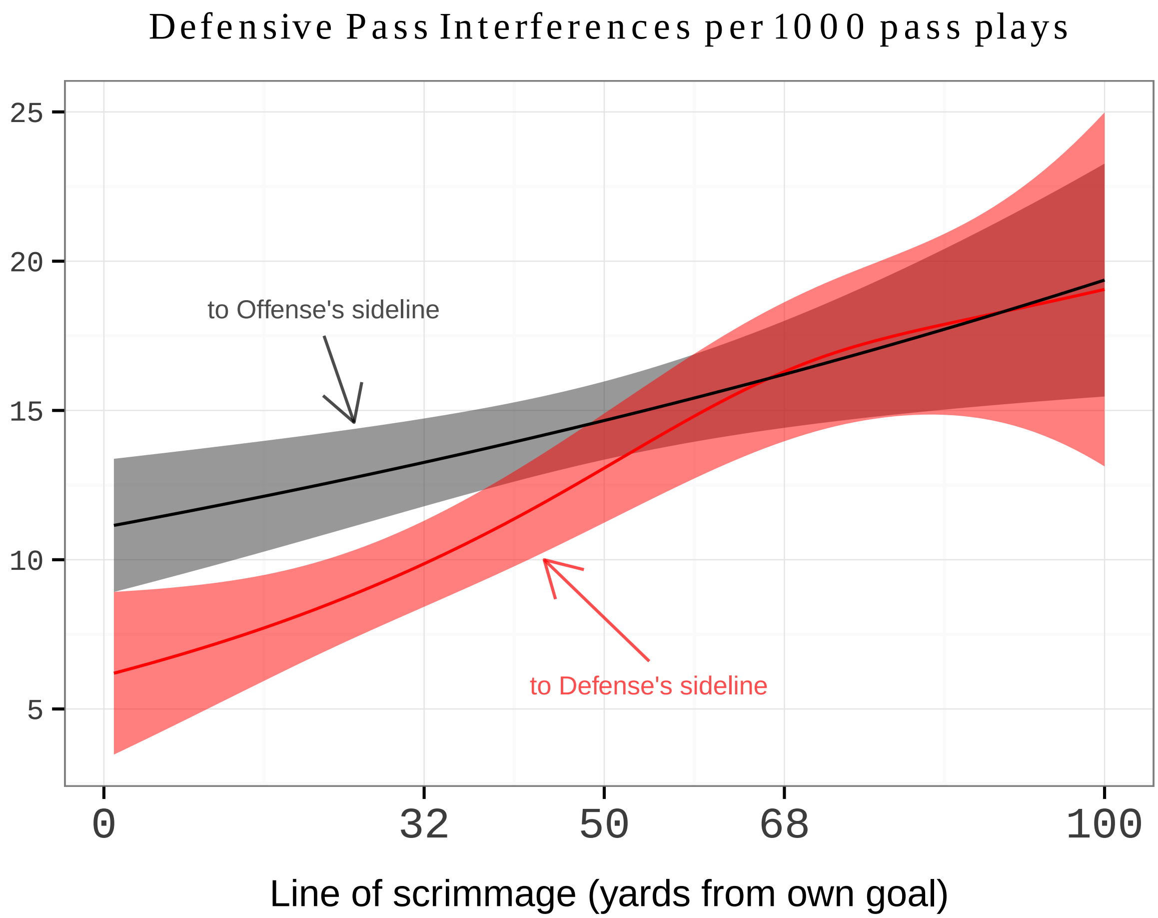

Figure 3: Defensive pass interference calls by sideline, line of scrimmage, estimated per 1000 pass plays with 95% confidence intervals

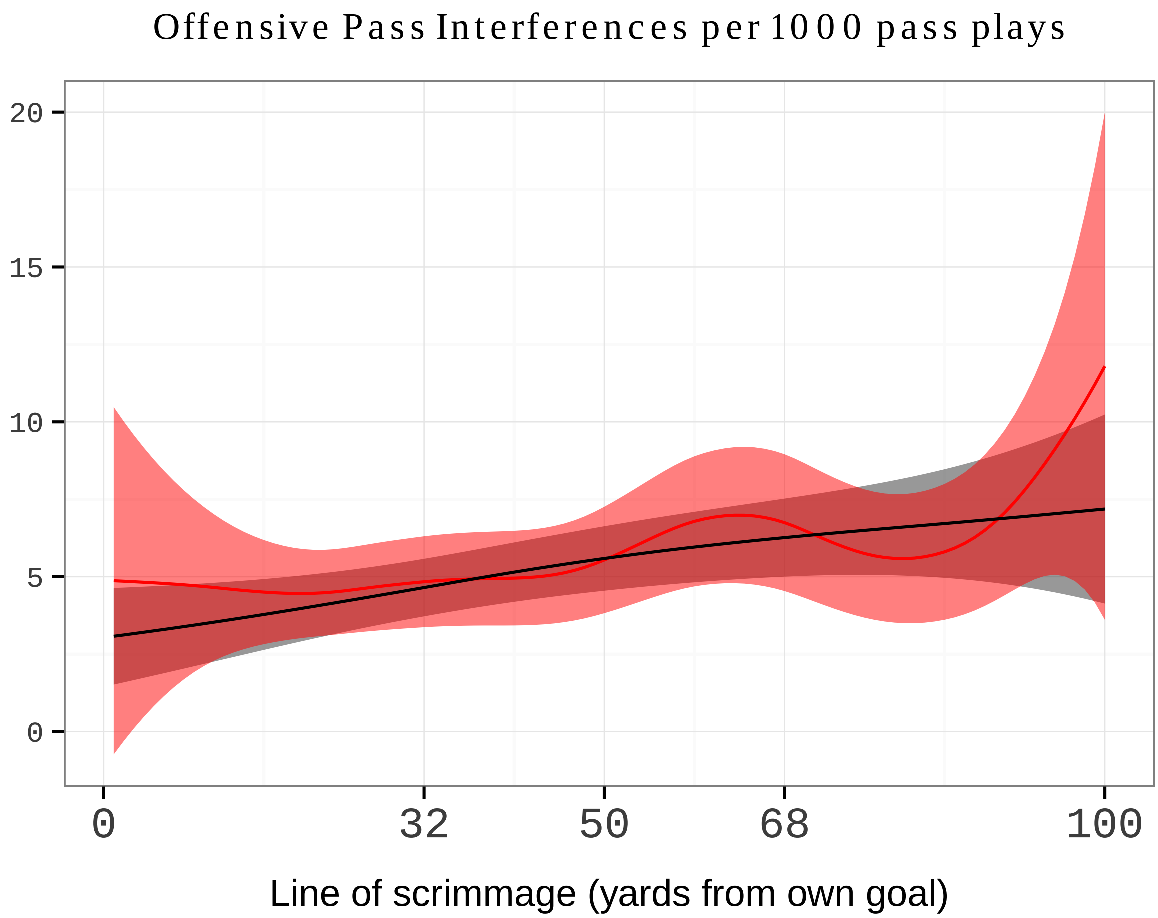

Figure 4: Offensive pass interference calls by sideline, line of scrimmage, estimated per 1000 pass plays with 95% confidence intervals