Persuaded under pressure: evidence from the National Football League

Michael J. Lopez
Assistant Professor of Statistics
Skidmore College
14 Horseshoe Drive
Saratoga Springs NY 12866
518-580-5297
mlopez1@skidmore.edu
February 3, 2016

Abstract

We exploit a natural experiment within each National Football League game, finding the first evidence in professional sports that referees succumb to the pressures of satisfying team personnel in the vicinity of possible violations. Using generalized additive models for binomial outcomes, we show that these sideline-based differences in penalty rates, which are observed on common but influential penalties including pass interference and holding, peak near the centralized location of players and coaches on the sideline. With sizable interests in referee decisions, coaches and players often try to manipulate referee behavior with verbal and nonverbal communications; such actions appear to be persuasive (JEL classification Z0, H3).

Keywords: Referee bias, professional sports leagues, peer pressure, generalized additive models, football

**Acknowledgment: Many thanks to Skidmore College’s Cody Couture, panelists in the 2015 Joint Statistical Meetings session on referee behavior, as well as the anonymous referees and Journal co-Editor, for their contributions and insights.**
Persuaded under pressure: evidence from the National Football League

I 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, Dohmen 2008, Pettersson-Lidbom and Priks 2010, Buraimo et al. 2010, Moskowitz and Wertheim 2011), 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, Pope and Pope 2015), and previous sequences of violations (Lopez and Snyder 2013, Abrevaya and McCulloch 2014, Gift 2015).

One aspect missing in the literature, but a part of the flow of any athletic event, is accounting for the 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.

This paper presents the first evidence of a successful sideline pressure in sports. 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. Finally, using the observed differences in penalty rates, we
explore the practical significance of our findings on play and game outcomes, while comparing our effect sizes to other known referee biases.

II 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 about 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 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 coded as left, middle, or right.\(^1\) We also wish to explore the possibility of a sideline bias on special teams plays, including kickoff and punt returns. However, NFL game reports, and thus the AA data, do not label special teams plays with a direction. As a result, these plays are dropped.

End zone directional information - that is, the direction that the offensive and defensive units are facing - is missing from both AA’s data and traditional NFL play-by-play reports. We used coin toss data 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 tosses, 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.

\(^1\)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.
Sideline information for each of the home and visiting teams was collected manually using team websites and seating guides. 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,

\[
Sideline_i = \text{Direction of play } i, \quad i = 1, \ldots, n_r, n_r + 1, \ldots, n_r + n_p \\
= \{\text{'Offense' if direction of play } i\text{ is the offensive team's sideline, 'Defense' if defense's sideline}\}
\]

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.

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

- \( OHR_i = \{1 \text{ if there was an offensive holding penalty on play } i, 0 \text{ o/w}\}, \quad i = 1, \ldots, n_r \)
- \( DAP_i = \{1 \text{ if there was a defensive aggressive penalty on play } i, 0 \text{ o/w}\}, \quad i = 1, \ldots, n_r + n_s \)
- \( DPI_i = \{1 \text{ if there was a defensive pass interference penalty on play } i, 0 \text{ o/w}\}, \quad i = n_r + 1, \ldots, n_r + n_p \)
- \( OPI_i = \{1 \text{ if there was an offensive pass interference penalty on play } i, 0 \text{ o/w}\}, \quad i = n_r + 1, \ldots, n_r + n_p \)

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 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.


4 From a statistical modeling perspective, we also use to ‘logit’ link in Section III. 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.
II.I Methods

We first analyze the relationship between $Y$ and Sideline. Under an existing sideline pressure, we hypothesize that there will be higher $OHP$ and $OPI$ rates on plays run in the direction of the defensive team’s sideline, relative to the offensive team’s sideline. This can be attributed to either a pressure from the defensive team and coaches to throw a flag on the offensive unit on plays near the defensive sideline, or to a fear of throwing a penalty on the offensive team when in the neighborhood of the offensive sideline. Using a similar logic, we hypothesize that there will be higher $DAP$ and $DPI$ rates on plays run towards the offensive team’s sideline, relative to the defensive team’s sideline. We will 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. $\chi^2$ 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). By rule, the sidelines at each stadium fit the same dimensions. 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 example, one issue with multiple logistic regression is that it requires the associations between line-of-scrimmage and penalty outcomes to be specified as linear, quadratic, and/or piecewise, despite the true functional form being unknown. 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

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

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, \( \logit(p) = \log\left( \frac{p}{1-p} \right) \), and \( f_{\text{Offense}}(\text{LOS}_i) \) and \( f_{\text{Defense}}(\text{LOS}_i) \) are the smoothed functions of the log-odds of a penalty based on plays run at the offensive and defensive teams sidelines, respectively, by \( \text{LOS} \). Separate \( \text{LOS} \) surfaces are estimated for each play direction using penalized thin-plate regression splines \([\text{Gu and Wahba} (1993)]\). As in \([\text{Wood} (2006)]\) and implemented by \([\text{Mills} (2014)]\), 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.

Model (1) allows for plays toward each sideline to have different baseline penalty rates, as measured through the \( \beta_1 \) coefficient, and for line-of-scrimmage effects to vary based on \( \text{Sideline} \).

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

\[
\logit(P(Y_i = 1)) = \beta_1 \times I(\text{Sideline}_i = \text{Offense}) + f_{\text{Overall}}(\text{LOS}_i), \tag{2}
\]

\[
\logit(P(Y_i = 1)) = f_{\text{Overall}}(\text{LOS}_i), \tag{3}
\]

\[
\logit(P(Y_i = 1)) = \beta_1 \times I(\text{Sideline}_i = \text{Offense}). \tag{4}
\]

Models (2) and (3) include an overall surface term for line of scrimmage, \( f_{\text{Overall}}(\text{LOS}_i) \), 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 \( \text{LOS} \) impact does not significantly differ by \( \text{Sideline} \). 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 \( \text{Sideline} \). Models are fit in the \( R \) statistical software \([\text{R Core Team} (2014)]\), and are contrasted using the Akaike Information Criterion (AIC, \([\text{Akaike} (1974)]\)), which includes a penalty for unneeded parameters to discourage overfitting.

### III RESULTS

AAs and FOs 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’s 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

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

*Total plays 42322, 40793

Model estimates are shown in Table 2. After adjusting for LOS using GAMs, there are baseline differences in both DAP and DPI across all model specifications, as judged by significant β₁ estimates. 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 team’s sideline, compared to one run towards the defensive team’s 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 team’s sideline. There are no
Table 2: Model statistics and estimated coefficient

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>( \hat{\beta}<em>1(SE</em>{\hat{\beta}_1}) )</th>
<th>AIC</th>
<th>( \hat{\beta}<em>1(SE</em>{\hat{\beta}_1}) )</th>
<th>AIC</th>
<th>( \hat{\beta}<em>1(SE</em>{\hat{\beta}_1}) )</th>
<th>AIC</th>
<th>( \hat{\beta}<em>1(SE</em>{\hat{\beta}_1}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td><strong>4969</strong></td>
<td>-0.117 (0.086)</td>
<td>6145</td>
<td>0.331 (0.091)**</td>
<td><strong>9519</strong></td>
<td>0.199 (0.069)**</td>
<td>4498</td>
<td>-0.056 (0.107)</td>
</tr>
<tr>
<td>(2)</td>
<td>4976</td>
<td>-0.130 (0.085)</td>
<td><strong>6142</strong></td>
<td>0.334 (0.091)**</td>
<td>9523</td>
<td>0.161 (0.067)*</td>
<td>4495</td>
<td>-0.054 (0.105)</td>
</tr>
<tr>
<td>(3)</td>
<td>4976</td>
<td>N/A</td>
<td>6154</td>
<td>N/A</td>
<td>9527</td>
<td>N/A</td>
<td><strong>4493</strong></td>
<td>N/A</td>
</tr>
<tr>
<td>(4)</td>
<td>4978</td>
<td>-0.130 (0.085)</td>
<td>6142</td>
<td>0.335 (0.091)**</td>
<td>9561</td>
<td>0.162 (0.067)*</td>
<td>4505</td>
<td>-0.054 (0.105)</td>
</tr>
</tbody>
</table>

*OHR taken on run plays, DPI and OPI pass plays, DAP on all plays

Model with lowest AIC in bold

significant differences in the baseline rates of OHR or OPI violations. For each Y, estimates of \( \beta_1 \) are robust between the four model fits.

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, whose lowest AIC’s are achieved via Models [2] and [3], respectively.

While estimates from GAMs can be difficult to interpret, 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 rates, 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.

*Smoothing parameter estimates are available upon request.
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
Offensive Pass Interferences per 1000 pass plays

Figure 4: Offensive pass interference calls by sideline, line of scrimmage, estimated per 1000 pass plays with 95% confidence intervals
Near an offensive team’s 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 appear to subside around the 50-yard line. While this initially seems to violate our hypothesis of a greater sideline effect near midfield, 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 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.

IV DISCUSSION

It is generally assumed that football referees assess each possible violation on its own accord. However, we find multiple penalties whose likelihoods vary significantly 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 IV). 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 \( \beta_1 \) in these adjusted fits was within a one-hundredth of the estimates shown in Table II. Further, we did not observe any significant interactions between the minute of the game and the sideline effect. This perhaps indicates that the sideline bias is more instantaneous, as opposed to a pressure that grows over the course of a contest.

We compared DPI by LOS because the NFL’s play-by-play data does not give precise infor-
mation 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 team’s 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 team’s sideline. On passes thrown near one of the end zones, the ratio of flags by sideline was roughly 1:1 (counts of 201 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. Using the estimated yardline to which the pass was thrown, Figure 5 shows the model-estimated penalty rates by sideline. As in our other penalty outcomes, the largest differences in DPI rates occur on plays likely to have ended near the middle of the sideline.

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.

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. One alternative explanation 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 players 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 teams with relatively higher frequencies of passing.

---

8 Adding 8 yards to the line-of-scrimmage for each short pass, and 25 yards for each deep pass.
Figure 5: Offensive pass interference calls by sideline, imputed line of penalty, estimated per 1000 pass plays with 95% confidence intervals.
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. Nor can we easily infer or identify referees that are more prone to exhibit a bias; referee crews are randomly assigned to games, and, without looking at game film, it is unknown which officials stand on which sideline during a specific contest.

Although the differences in certain penalty rates meet criteria for statistical significance, a final important consideration is that of practical significance. Notably, each of our penalty outcomes carries a hefty punishment against the violating team. DPIs and DAPs each give the offensive unit a new set of downs with which to possess the ball, and DPIs are a spot foul, meaning that the line-of-scrimmage moves to wherever the penalty took place. As a result, while DAPs are usually a 15-yard infraction, many DPIs can be worth additional yardage. Additionally, offensive teams are penalized 10 yards for any holding penalty. Such punishments are relatively severe compared to other infraction types, and can have obvious impacts on the game. Anecdotally, Burke (2015) compared an offensive team’s win probability with and without a potential DPI call from a 2014 game between Detroit and Dallas, finding that an accepted penalty (on one team’s sideline, at roughly the 35 yard line) would have increased the offensive team’s chances of winning from 67% to 79%. It is unlikely that all DPI penalties have this large of an affect on game outcomes, but it is not unrealistic to believe that many DPI calls or non-calls swing game outcomes by more than 5%.

As the majority of offensive plays in an NFL game do not end in OHRs, DPIs, or DAPs, our penalty outcomes are relatively rare. However, given that these infractions are among the game’s most important so far as determining its outcome, the relative fraction of penalty outcomes that can be attributed to a sideline bias is noteworthy. For example, our data set contains 596 DPIs estimated to have occurred near midfield (204 at defensive sideline, 99 down the middle, 293 at offensive sideline). Using the defensive sideline DPI rate as a baseline, roughly 15% of all DPIs in this area can be explained by a sideline pressure. This adds context to Figures 1 through 5 which show absolute differences in rates varying by up to 12 penalties per 1000 plays (OHR), or relative differences as high as 50 (DPI) to 100 (DAP) percent, given the LOS.

Perhaps most importantly, the aggregated effects of a sideline bias dominate any differences in
penalties between home and away teams. For example, our data set of sideline passes contained 60 additional DPI calls when the home team was on offense. This difference represents about two-thirds of the difference in flags that we observed based on Sideline. Moreover, using the same data presented above, home teams on offense were called for more OHRs (we would expect fewer under a home bias) and drew fewer DAPs (we would expect more), relative to the away team being on offense. On average, as far OHR and DAP are concerned, there is no obvious advantage to being a home team, particularly when compared to a Sideline effect. Additionally, 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, as judged by the significance of the interaction term and AIC fit values.

Finally, although there were 842 DAPs in our data set, this only represents about 45% of these penalty outcomes; an additional 1029 occurred on special teams. As mentioned in Section II, directional information for special teams plays is not included in the NFL’s play-by-play reports. Given that several of these missing penalties came on kick and 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.

V 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 behavioral economics and 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; Buraimo et al., 2010). We propose that in several settings, sideline pressure dwarfs that of the home crowd. 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.
References


