

# Factors Affecting the Probability of Arrests at an NFL Game

Patrick Brown

## 1. Introduction

Every NFL season comes with its fair share of stories about rowdy fans taking things too far and getting themselves arrested. While sports media tends to favor stories of fighting and public intoxication out of Oakland and Buffalo, game time arrests still occur at every NFL stadium. Sure, anytime a crowd of thousands gathers for an event, there's a risk of criminal activity. But is there anything specifically about NFL games that can sway the occurrence of arrestable offenses one way or another? As a football fan, I'll admit a fascination with the varying fanbase culture from team to team, and I often look forward to hearing what some of the rowdier fans are going to do next. But from an economics standpoint, fan culture poses some interesting questions. The question I seek to answer today is whether the probability of arrests during an NFL game depends more on the culture surrounding the NFL stadium or the characteristics and outcome of the game itself.

## 2. Data

In order to understand the probability of arrests at an NFL game, I collected and merged data from six different sources:

- The Washington Post NFL Arrests Dataset
- Pro-Football-Reference's NFL Games Attendance Database
- NFL Stadiums Seating Capacity via Wikipedia
- NFLSavant.com's Game Time Weather

- Bureau of Economic Analysis Data on Real GDP per Metropolitan Statistical Area by Year
- The FBI's Uniform Crime Report Table 6

An aggregated .csv file was constructed in Google Sheets and prepared for import into Stata.

The NFL Arrests Dataset was originally compiled by Kent Babb and Steven Rich of the Washington Post by organizing public record requests from police departments that oversaw NFL stadium security between 2011 and 2015. Of the 31 jurisdictions in which there is an NFL stadium (note that the New York Giants and the New York Jets reside in the same jurisdiction), all but Cleveland and New Orleans submitted data. While most agencies provided complete game-by-game arrest records, Buffalo, Miami, and Oakland provided only partial records and were omitted from this study.<sup>1</sup> Further, St. Louis provided only yearly arrests data, while Detroit, Minneapolis, and Atlanta excluded parking lot arrests.<sup>2</sup> This dataset provided further game-specific information like the season, week, day, and time of day that the game was played, the scores obtained by both the home and away team, and binaries for whether or not the game was a division game or went into overtime. Weeks and days were further categorized into dummy variables.

A level analysis of total arrests per game would not account for proportional variance in crowd and stadium size, so data on game attendance and stadium seating capacity were merged with the Washington Post's dataset as well. Pro-Football-Reference provides a meticulous database of NFL statistics, including attendance per game. These two datasets were merged

---

<sup>1</sup> Anecdotally, Buffalo and Oakland are two fanbases with rowdier reputations, so it's somewhat unfortunate that these teams have been excluded from the data.

<sup>2</sup> One assumption is that the heaviest alcohol consumption and fighting occurs in the parking lot, where police and security supervision is less prevalent. Note that the data does not provide information on area-specific police presence or the type of crime committed that lead to arrest.

using the home team, week number, and season as unique identifiers. Wikipedia provides data on the seating capacity at each stadium, and this data was merged using the home team as a unique identifier. Note that the San Francisco 49ers changed stadiums from Candlestick Park to Levi's Stadium during the period of analysis, and this switch was accounted for in the data merge. The Wikipedia data also contained each stadium's roof status, which was expressed as binaries for *retractable*, *open*, and *fixed*.

NFLSavant.com provides data on the temperature and humidity during several games from 2011-2013, and this was merged using home team, away team, and season as unique identifiers for each game.

In an attempt to categorize what I refer to as the "*you think you're better than me?*" factor, I merged Bureau of Economic Analysis (BEA) data on the per capita real GDP for the home and away team's Metro Statistical Area (MSA). Under the assumption that some arrests will be the result of fighting between fans, I wanted to be able to determine if income inequality was a contributing factor to the probability of arrest. These real GDP per capita numbers were further categorized with a binary describing if the home team's fan base was wealthier than the away team's fanbase, and a calculation of the absolute value of the difference between both fanbase's real GDP per capita.

Finally, crime rate data for the home and away team MSA's were merged from the FBI's Uniform Crime Report (UCR) Table 6. Rates were expressed as violent crimes per 100,000 people and property crimes per 100,000 people and categorized by year.

Within Stata, six additional variables were generated:

- *attend2capac*: total game attendance divided by stadium seating capacity<sup>3</sup>
- *arrest2attend*: total game arrests divided by total game attendance
- *lhomeviocrmrt*: the natural logarithm of the home team's violent crime rate
- *lhomepropcrmrt*: the natural logarithm of the home team's property crime rate
- *ltemp*: the natural logarithm of the average temperature during game play
- *lscorediff*: the natural logarithm of the absolute value of the difference between the home and away team's respective scores

The original Washington Post dataset contained 1006 observations corresponding with 1006 games. After merging the remaining 5 data sets, 295 fully complete records remained. Missing records can be attributed to both the weather and UCR data.<sup>4</sup> The average NFL game in our set is attended by 68,805 fans and results in 6.3 arrests. The average probability of arrests on a per-game basis is .009218%. The highest number of arrests at any particular game occurred during the 2011 season's Week 10 Thursday night game featuring Oakland at San Diego in which 69 people were arrested. With attendance totaling 68,109, the arrest probability during this game was .1013%. Interestingly enough, the second highest number of arrests occurred during the same matchup with Oakland at San Diego during the following season, in which 60 fans were arrested.

---

<sup>3</sup> In some cases, this variable was greater than 1. The Wikipedia data accounted only for seating capacity. Several stadiums having "standing room only" areas as well. In the event of a full sellout at a stadium with the latter feature, the ratio would be greater than 1.

<sup>4</sup> Recall that game time weather was provided only through 2013. The UCR is dependent on reporting from agencies within individual jurisdictions, and as such, some records are incomplete.

### 3. Methodology

The model was intended to describe the probability of arrests during an NFL game by taking into account the outcome of the game, crowd stressors, and regional crime risk. The primary regression equation to be estimated was:

$$arrest2attend = \beta_0 + \beta_1 hometeamwin + \beta_2 gametime + \beta_3 attend2capac + \beta_4 ltemp + \beta_5 lhomeviocrmrt + \beta_6 lhomepropcrmrt$$

My hypotheses regarding the signs of the individual coefficients were as follows:

Variable	Coefficient Sign	Reasoning
<i>hometeamwin</i>	(-)	The home team losing may lead to more crowd aggression, which may lead to more arrests.
<i>gametime</i>	(+)	A game that starts later in the day may lead to larger degrees of intoxication among the crowd.
<i>attend2capac</i>	(+)	A game in which stadium seating capacity is maximized may be indicative of a game with a high degree of fan importance. As such, feelings of rivalry may be more prevalent.
<i>ltemp</i>	(+)	As the temperature increases, the risk of dehydration due to intoxication increases, thus leading to higher incidence of public intoxication.
<i>lhomeviocrmrt</i>	(+)	A higher violent crime rate for the area may lead to higher rates crime within the stadium.
<i>lhomepropcrmrt</i>	(+)	A higher property crime rate for the area may lead to higher rates crime within the stadium.

Further controls were added to the regression model to determine the partial effects of whether or not the game was played outside, if the game was played on a Sunday, seasonality due to the week in which the game was played, the score differential, and whether or not the the home team's fan base was wealthier than the away team's fans. Heteroskedasticity tests were taken into account, and when applicable, heteroskedasticity robust standard errors were used in place of the normal standard errors.

The primary source of estimation bias arose from the data itself. Recall that the crime-specific data was sourced from local police agencies separately by both the Washington Post and the FBI. This practice resulted in inconsistencies in both collection and reporting methods across the 31 jurisdictions, and as such, some data points on arrests and crime were missing. Anecdotal evidence implies that the average probability of arrest per game could be biased downwards with the omission of Buffalo and Oakland, but unfortunately this could not be confirmed without relevant data. It should be noted that Babb and Rich of the Washington Post cleaned their data set by omitting reporting they deemed to be unreliable. Without an efficient method of adjusting for potential collection bias, the data was analyzed as is while noting that an unavoidable bias due to collection methods could have existed.

#### **4. Results**

The primary regression model was found to be statistically significant in all variables to at least the 10% level with an R-Squared of .2885. I found that the home team winning, higher attendance relative to seating capacity, and higher local property crime rate all decreased the probability of arrests at an NFL game, while later start times, higher temperatures, and higher local violent crime rates tended to increase the same probability.

Factors Affecting the Probability of Arrest at an NFL Game

	(1)	(2)	(3)	(4)	(5)	(6)
<b>hometeamwin</b>	<b>-.0000193**</b> <b>(8.91e-06)</b>	<b>-.0000189**</b> <b>(8.84e-06)</b>	<b>-.0000193**</b> <b>(8.74e-06)</b>	<b>-.0000227*</b> <b>(.0000135)</b>	<b>-.0000218</b> <b>(.0000146)</b>	<b>-.000024*</b> <b>(.0000135)</b>
gametime		.0001597*** (.0000381)	.0001823*** (.0000379)	.0002814*** (.0000594)	.0002393*** (.0000655)	.0001295** (.0000606)
attend2capac			-.0003339*** (.0000658)	-.0003834*** (.0000913)	-.0002706*** (.0000994)	-.000265*** (.0000906)
ltemp				3.72e-06 (.0000201)	.0000145 (.0000231)	.0000514** (.0000211)
lhomeviocmrt					-.0000315 (.0000252)	.0000508** (.0000246)
lhomepropcmrt						-.0003217*** (.0000287)
Constant	.0001033*** (6.77e-06)	6.21e-06 (.0000241)	.0003168*** (.0000657)	.000298** (.0001252)	.0003607** (.0001693)	.0023167*** (.0002277)
Observations	1,006	1,006	1,006	510	443	407
R-Squared	0.0046	0.0218	0.0463	0.0742	0.0584	0.2885

Notes: Standard errors in parentheses: \*\*\*p<.01, \*\*p<.05, \*p<.1

Both the Breusch-Pagan Test and the White Test confirmed heteroskedasticity in the model, but all factors remained statistically significant even with heteroskedasticity robust standard errors.

Factors Affecting the Probability of Arrest at an NFL Game (Robust S.E.)

	(1)	(2)	(3)	(4)	(5)	(6)
<b>hometeamwin</b>	<b>-.0000193**</b> <b>[9.25e-06]</b>	<b>-.0000189**</b> <b>[9.15e-06]</b>	<b>-.0000193**</b> <b>[9.06e-06]</b>	<b>-.0000227</b> <b>[.0000143]</b>	<b>-.0000218</b> <b>[.0000157]</b>	<b>-.000024*</b> <b>[.0000145]</b>
gametime		.0001597*** [.0000406]	.0001823*** [.0000405]	.0002814*** [.0000678]	.0002393*** [.0000703]	.0001295** [.0000632]
attend2capac			-.0003339*** [.0000734]	-.0003834*** [.0001051]	-.0002706** [.0001041]	-.000265*** [.0000925]
ltemp				3.72e-06 [.0000177]	.0000145 [.0000212]	.0000514*** [.000019]
lhomeviocmrt					-.0000315 [.0000221]	.0000508** [.0000203]
lhomepropcmrt						-.0003217*** [.0000269]
Constant	.0001033*** [7.67e-06]	6.21e-06 [.0000239]	.0003168*** [.0000712]	.000298** [.0001212]	.0003607* [.0001426]	.0023167*** [.0002397]
Observations	1,006	1,006	1,006	510	443	407
R-Squared	0.0046	0.0218	0.0463	0.0742	0.0584	0.2885

Notes: Standard errors in parentheses: \*\*\*p<.01, \*\*p<.05, \*p<.1

Several additional models containing the variables *open*, *sunday*, dummies for each week, *scorediff*, *lscorediff*, and *homericher*<sup>5</sup> were estimated, but in each case the additional variables lacked statistical significance. Particularly, I had expected that the probability of arrest would be higher on outdoor Sunday games later in the season when the score was close and the home team was less wealthy than the away team<sup>6</sup>, but there was no evidence to indicate that any of these factors were significant within this dataset.

Returning to the primary model, the magnitude of certain factors bore significant economic weight. Note that N decreased as variables were added to the equation. In the full model case where N=407, the average probability of arrests was .00986%. At an average attendance of 68,700 fans, around 7 people were arrested at each game. However, if the home team won, the expected arrests dropped to 5. This was consistent with the idea that the home team winning results in less negative stress among fans, potentially resulting in lower levels of intoxication or aggression. A 1% increase in the temperature during game play increased expected arrests to 10. While it's possible that a hot day may lead to higher levels of intoxication, this could be more indicative of geographic cultures. For instance, games played in Jacksonville will, on average, be warmer than games played in New England. Cultural differences between the two areas may be the causal element, not the heat itself. A 1% increase in the area's violent crime rate also increased the expected number of arrests to 10, but a 1% increase in the property crime rate resulted in a total arrests expectation of -15. While negative arrests are not actually possible, this is indicative of an interesting trend where violent crime rate contributed positively to game arrests and property crime contributed negatively. This could

---

<sup>5</sup> This variable indicated whether or not the real GDP per capita of the home team was greater than that of the away team.

<sup>6</sup> Referred to earlier as the "you think you're better than me?" factor.

have been a result of local police force management. If an area was known to have a higher rate of violent crime, police agencies may have been more likely to allocate officers to high-population events like an NFL game. Consider that property crime includes petty theft, misdemeanor vandalism, and other nuisances that may be considered less important than violent crime. If violence is less of an issue for a locale, the proportion of officers allocated to an NFL game relative to the proportion allocated to police the rest of the MSA on game day may have been less than in violence-prone areas, resulting in lower expected arrests.

## **5. Conclusion**

While there is evidence within this dataset to indicate that an NFL game's outcome, start time, attendance to capacity ratio, weather, and local crime rates have an effect on the probability of arrests during the game, definitive decisions regarding police allocation and local gameday policy will require continued research. Inconsistencies in collections and reporting methods between jurisdictions make this sort of analysis difficult, and future uniformity of methods may be beneficial. Regardless, the results paint an interesting picture of crowd behavior in the presence of football-based stimuli, and continued data collection may lead to interesting research on crowd dynamics in the future.