Security Management at NFL Stadiums

Modeling the number of arrests using an negative binomial distribution

Introduction

Football stadiums attract thousands of people with every game that is played, and with large crowds, security becomes an important issue to manage. Every game, arrests are made. Sometimes, to the degree that stadiums like the Lincoln Financial Field in Philadelphia build their own prisons within the stadium.

Objective & Datasets Utilized

To understand what drives the number of arrests at football games and how to better manage security during NFL games, I investigated a dataset that I sourced from Kaggle.com describing the number of arrests during football games between 2011 and 2015.[1] The objective was to create a negative binomial distribution (NBD) model that accurately describes the number of arrests made at NFL stadiums based on observable attributes of the game and to subsequently use such a model to make inferences that would improve security management at NFL stadiums.

The dataset I used is based on public records from the police departments that monitor the security at each one of the NFL stadiums. Although not all jurisdictions released their data, the dataset contains figures from 25 NFL stadiums and 966 games. In addition to data on the number of arrests made at each games, each game also holds information such as day of the week, time of day, score outcome, season, and participating teams (see Appendix 1 for first 10 rows).

To complement the dataset with information I believed would be relevant to understanding what drove the number of arrests, I sourced two additional datasets. The first dataset is from the Federal Bureau of Investigation and contains information on crime levels per 100,000 people in 2015 in

[1] https://www.kaggle.com/washingtonpost/nfl-arrests

the cities where the NFL stadiums are located.[2] The dataset divides total crime levels into two broad categories: violent crimes and property crimes, both of which are subsequently divided into more specific crimes (see Appendix 2 for first 10 rows). The second dataset is from ESPN and contains information on attendance at NFL stadiums in 2015.[3] The dataset provides data on average and total attendance across the 7-8 games played during the season, both home and away games (see Appendix 3 for first 10 rows). Note that since attendance rates in 2015 were missing for San Diego, I sourced those numbers from Statista.[4]

Concerns regarding the datasets

Since all three datasets come from credible sources such as the Washington Post, the FBI, and ESPN, I believe these numbers to be accurate. My main concern about the two latter datasets would be that the crime and attendance level data lack generalizability for game arrests made before 2015. However, since I believe crime rates and attendance level data to be generally stable over a five year period, I assume that 2015 figures can be applied across all games 2011-2015.

Hypothesis

In considering the possible drivers of the number of arrests made at NFL stadiums, I came up with four hypothesis that I depended on to create an accurate NBD model:

1) **Attendance levels** at NFL stadiums have a positive relationship with the number of arrests made during games played at those NFL stadiums. The more people are present at the stadium, the more likely that one of them misbehaves and is arrested.

[2] https://ucr.fbi.gov/crime-in-the-u.s/2015/crime-in-the-u.s.-2015/tables/table-6
[3] http://www.espn.com/nfl/attendance/_/year/2015
[4] https://www.statista.com/statistics/250047/average-home-attendance-of-the-san-diego-chargers/

2) **Crime levels** in areas surrounding NFL stadiums have a positive relationship with the number of arrests made during games played at those NFL stadiums. The more criminals live in the surrounding areas, the more likely it is that those criminals will attend a game, misbehave, and get arrested.

3) **The outcome of the game** will affect the number of arrests made at each game. If the home team, for example, were to lose, I believe there is a possibility that more arrests will be made since the majority of attendants, in my own experience, are home fans (due to proximity to the stadium). If a majority of the attendants is upset by the outcome, I believe that chances are higher that a larger number of people will misbehave and get arrested.

4) **The day of the week** that the game is played will affect the number of arrests made during a given game, since people generally drink more during weekends than on weekdays and alcohol increases impulsive behavior.[5]

[5] "Impulsive behavior in males increases after periods of heavy drinking," *Science Daily* (2010), https://www.sciencedaily.com/releases/2010/11/101116181945.htm

Methodology

In creating a model that would accurately describe the number of arrests made during individual games, I used the hypothesis above to guide the four models I ran against the training data. The training data consisted of games played during 2011-2013 and represented 60% of all observed games during 2011-2015. For each of the models, I used the Maximum Likelihood Estimation methodology to estimate the r and alpha parameters. I subsequently chose the model with best perceived accuracy, which I inferred using the Chi-Squared Test (binning all observations for which the number of arrests exceeded 10), the Likelihood Ratio Test (LRT) (comparing the given model with a model for which no subgroups had been created in estimating the r and alpha parameters – see figure 1), the mean absolute percentage error (MAPE), and histograms comparing Actual data with Expected data, to validate against a holdout sample of games played 2014-2015 (40% of all observed games).

The first model ("Attendance" - figure 2) divides individual games into four subgroups based on average attendance levels of the home teams (since the home team stadium is the object of interest): very low (below the 25^{th} percentile), low (between the 25^{th} and 75^{th} percentile), high (between the 50^{th} and 75^{th} percentile), and very high (above the 75^{th} percentile).

The second model ("Crime" - figure 3) divides individual games into four subgroups based on the number of violent crimes committed in the city where the NFL stadium is located: very low, low, high, and very high (with the exact same methodology as above for creating the four subgroups / labeling each game, i.e. using percentiles).

The third model (“Game Outcome” - figure 4) divides individual games into two subgroups based on whether the home team wins or loses.

The fourth model (“Day of Week” - figure 5) divides individual games into two subgroups based on whether the game took place on a weekday (Monday, Tuesday, Wednesday, Thursday) or on a weekend (Friday, Saturday, Sunday).

The parameter estimates, log likelihoods, chi-squares, degrees of freedom, p-values, and MAPE for each of the subgroups as well as for the overall models can be found in Table 1. Summary of the Likelihood Ratio Tests can be found in Table 2 for each of the models.

Results

Figure 1: All NBD Model (No subgroups)

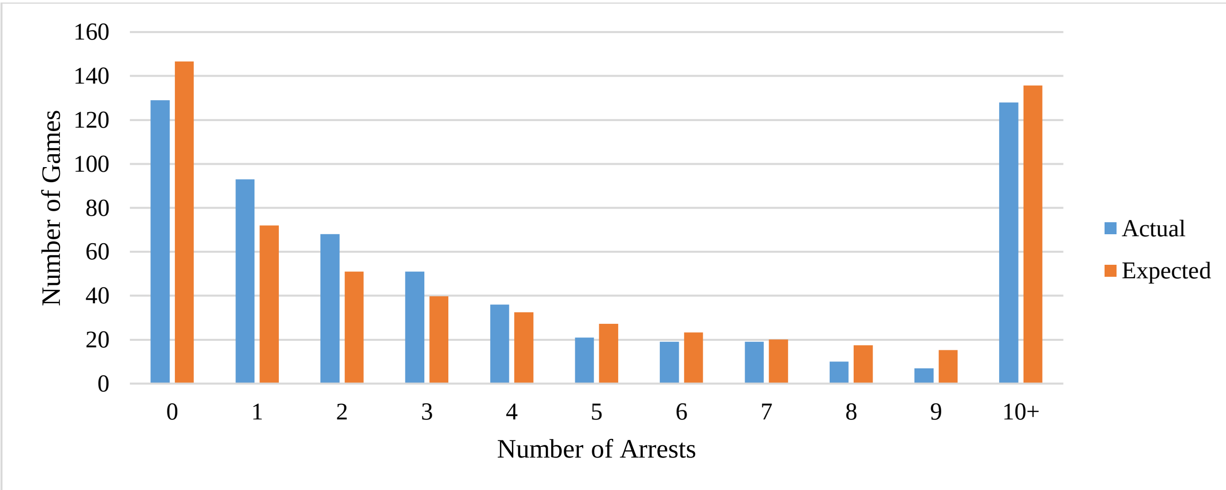


Figure 2: Attendance NBD Model

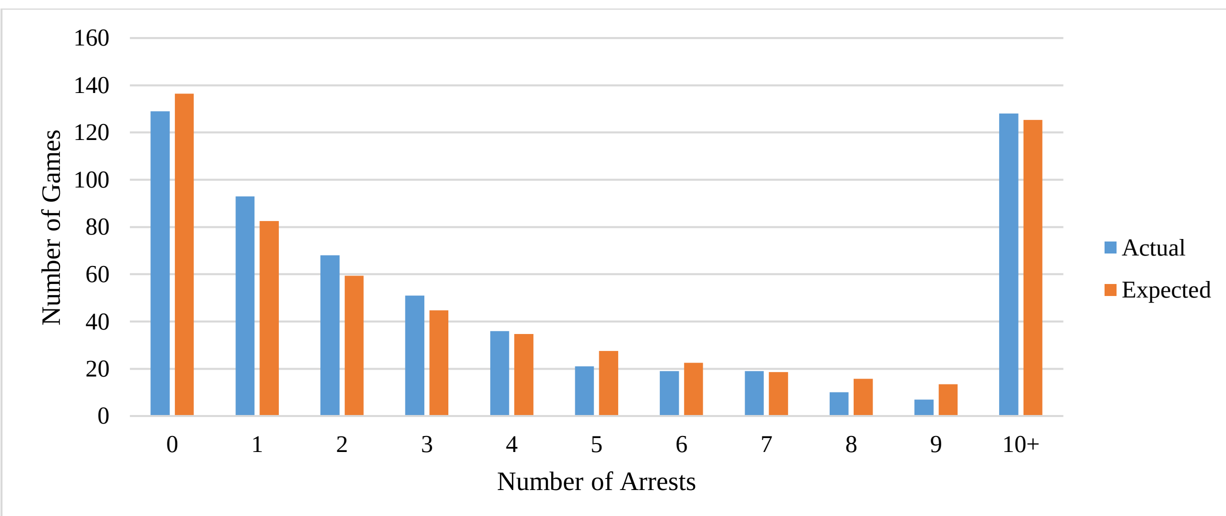


Figure 3: Crime NBD Model

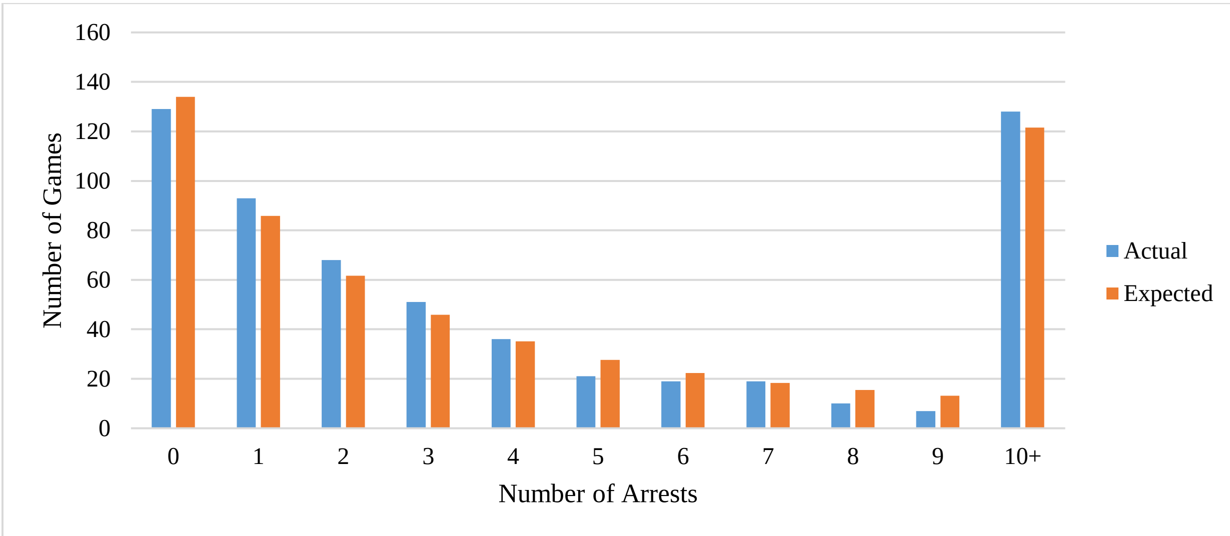


Figure 4: Game Outcome NBD Model

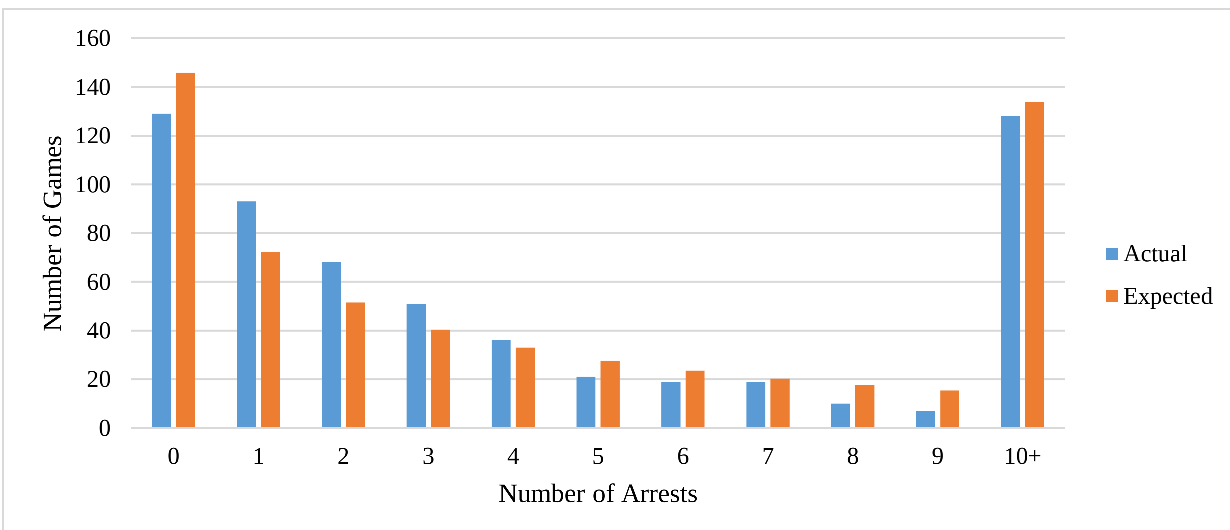


Figure 5: Day of Week NBD Model

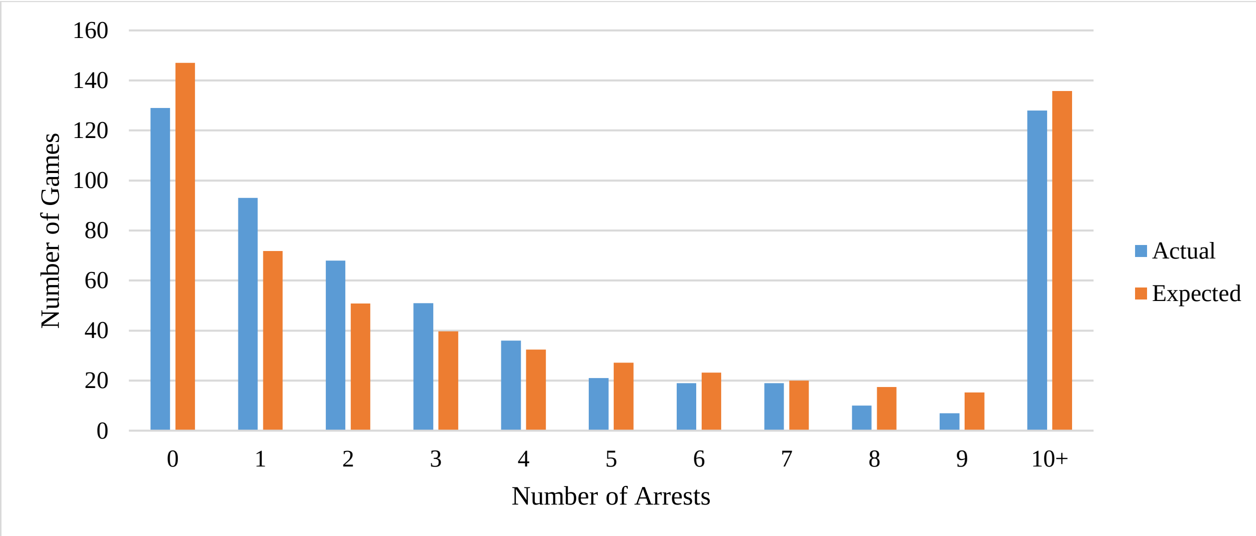


Table 1: Summary Data of Models

NBD Model	Subgroup	Parameter Est.		LL	Chi-Squared Test			MAPE
		r	α		Chi-Sq.	Df	p-value	
All	**None**	**0.53**	**0.08**	**-1667**	**28**	**7**	**0.023%**	**31.9%**
Attendance	**Aggregated**			**-1602**	**11**	**6**	**13%**	**22.7%**
	Very Low	0.50	0.12	-292	25	7	0.089%	-
	Low	0.58	0.06	-523	21	7	0.32%	-
	High	1.09	0.52	-311	7	7	44%	-
	Very High	0.77	0.08	-476	2	7	95%	-
Crime	**Aggregated**	**-**	**-**	**-1593**	**9**	**7**	**23%**	**21.2%**
	Very Low	0.67	0.05	-511	15	7	3.1%	-
	Low	0.60	0.10	-465	9	7	25%	-
	High	1.20	0.63	-300	8	7	31%	-
	Very High	0.67	0.11	-317	6	7	52%	-
Game Outcome	**Aggregated**			**-1662**	**27**	**7**	**0.036%**	**31.8%**
	Home Win	0.62	0.10	-958	18	7	1.4%	-
	Home Lose	0.45	0.06	-704	22	7	0.30%	-
Day of Week	**Aggregated**			**-1663**	**28**	**7**	**0.019%**	**31.9%**
	Weekday	0.70	0.08	-227	12	7	10%	-
	Weekend	0.51	0.08	-1436	25	7	0.088%	-

Table 2: Likelihood Ratio Tests (comparing each model to the model w/o subgroups)

	Attendance	Crime	Game Outcome	Day of Week
2LR	130	148	10	8
Df	6	6	2	2
p-value	1.3E-25	1.7E-29	0.84%	1.8%

Discussion

Selecting a model

According to the results, the most likely and accurate model out of the four is the "Crime" model; it has the highest p-value = 23% (Chi-Squared) and the lowest mean absolute percentage error (MAPE) = 21.2% (see Table 1). Additionally, it generates the lowest p-value = 1.3E-29 (LRT) when compared to the NBD model with no subgroups. Although the p-value is much lower than 0.0001, it does not confirm Hypothesis 2. It does, however, suggest that the data provides strong support for the subgrouping of the games (based on crime levels in the city where the NFL stadium is located) to describe the number of arrests at each game (see Table 2). Indeed, comparing the r parameters of the subgroups with that of the model with no subgroups indicates that the games in the subgroupings are more homogenous in terms of the number of arrests than without the subgroupings ($r_H = 1.20$, $r_{VL,VH} = 0.67$, and $r_L = 0.60 > 0.53$).

The "Attendance" model generates comparable results to that of the "Crime" model with a Chi-Squared p-value slightly below that of the "Crime" model (13%) and a slightly higher MAPE = 22.7% (see Table 1). These results suggests that although grouping games by attendance levels is not as accurate as grouping by crime levels, the model does a decent job at describing the number of arrests at each game. In fact, when compared to a model with no subgrouping, the "Attendance" model produces an LRT p-value of 1.3E-25, suggesting that the underlying data provides strong support for the subgrouping (see Table 2). Again, it does not confirm Hypothesis 1, but does suggest that Attendance Level plays a factor in the number of arrests per game (although not necessarily directly). That being said, the significantly lower Chi-Squared p-values of the "Very Low" and "Low" subgroupings (0.089% and 0.32%) compared to that of the "High" and "Very

High” subgroupings does indicate that the model may not be significant out-of-sample (see Table 1). A similar, but less substantial difference can be seen in the Chi-Squared p-values of the lower subgroupings of the “Crime” model and will be discussed under “Concerns about the model.”

The remaining two models tested – “Game Outcome” and “Day of Week” – failed to improve the accuracy of the model without subgroupings. Both generated mean absolute percentage errors almost identical to that of the model without subgrouping (31.8% and 31.9%) and chi-squared p-values very close to 0% (0.036% and 0.019%), which suggests that Hypothesis 3 and 4 are most likely untrue.

Validating the model

Ultimately, I chose the “Crime” model for describing the number of arrests made at a given NFL stadium given its superior performance and accuracy. To validate its performance, I ran the model against the remaining two years (or 40%) of game data (2014-2015). Figure 6 depicts the histogram generated by the model and compares it to the actual arrest figures.

In analyzing the resulting summary data, the out-of-sample results were not as promising as the in-sample results. With a Chi-Squared p-value of 11% (compared to the in-sample p-value of 23%), the model is less likely to hold true out-of-sample, which is problematic for practical use. The difference in p-value may be a sign of non-stationarity – that is, the lambda distribution of the underlying games changes with time. Even so, it was surprisingly more accurate out-of-sample than in-sample, with a MAPE of only 15.4% (compared to 21.2% in-sample). This discrepancy can either attributed to chance (see Table 3).

Figure 6: Validation of Crime Model against 2014-15 Hold-Out Data

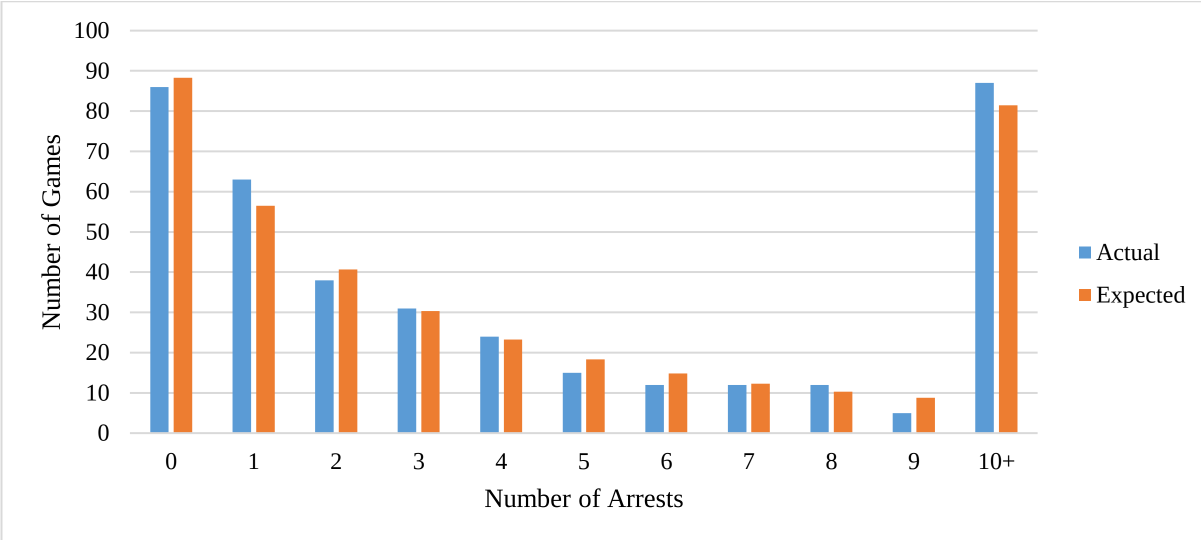


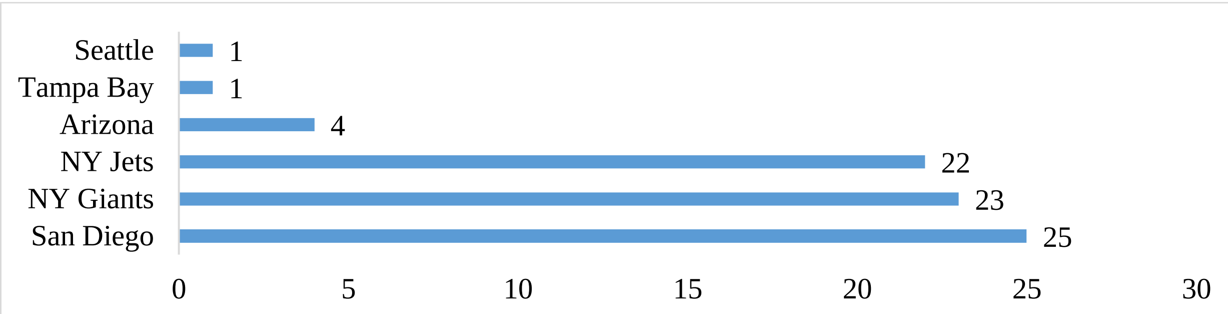
Table 3: Summary Data of the Validation (out-of-sample performance)

	Chi-Squared	Df	p-value	MAPE
Crime Model	4	2	11%	15%

Concerns about the model

As alluded to earlier, the "Crime" model did not provide a strong Chi-Squared p-value for <u>all</u> subgroupings. In fact, the p-value for the "Very Low" subgrouping was only 3% (compared to the other subgrouping, which had p-values of 25%, 31%, and 52% respectively), suggesting that the model did not do a good job describing the number of arrests per game among stadiums with "Very Low" levels of crime and that confounding factors may be at play. When examining the average number of arrests among teams that have "Very Low" levels of crime in the city where the stadium is located, we see that the number of arrests for games played where the six teams are home teams are very different, which may be the reason why the NBD model does not provide as good of a fit compared to those of other crime levels (see Figure 7). To combat the issue, I would consider sampling other covariates on games played by those teams to control for confounding factors such as the number of guards present at those games and the number of attendants at those particular games for a visitors-per-guard ratio.

Figure 7: Average number of arrests for NFL stadiums in cities of "Very Low" levels of crime



Managerial implications

Examining the Lorenz Curves under the "Crime" model (see figure 8), we see that stadiums in cities of all four type of crime levels have a similar degree of concentration in terms of arrests across games. Table 4 shows that 20% of games make up between 49% - 61% of all arrests made with the highest degree of concentration among stadiums in cities of "Low" levels of crime, where 20% of games make up 61% of all arrests. Consequently, police departments supervising the games should carefully manage the number of guards at the games such that most guards are present during those games during which most arrests are made. In fact, the concentration of arrests in games suggests an opportunity to save money by predicting the number of arrests at any given game and allocating resources accordingly. Therefore, police departments supervising the games should considering modelling the number of arrests based on more game-specific factors such as opponent team and weather to see if the variables can improve their current expectation of criminal activity at games. Police departments may, for example, use Bayes Theorem to forecast a conditional expected number of arrests based on past games taking into account the covariates above. Furthermore, in doing so, police departments should control for the number of guards at each game to make sure it is not the number of guards that drive the number of arrests.

Figure 8: Lorenz Curve Comparison for Different Crime Levels

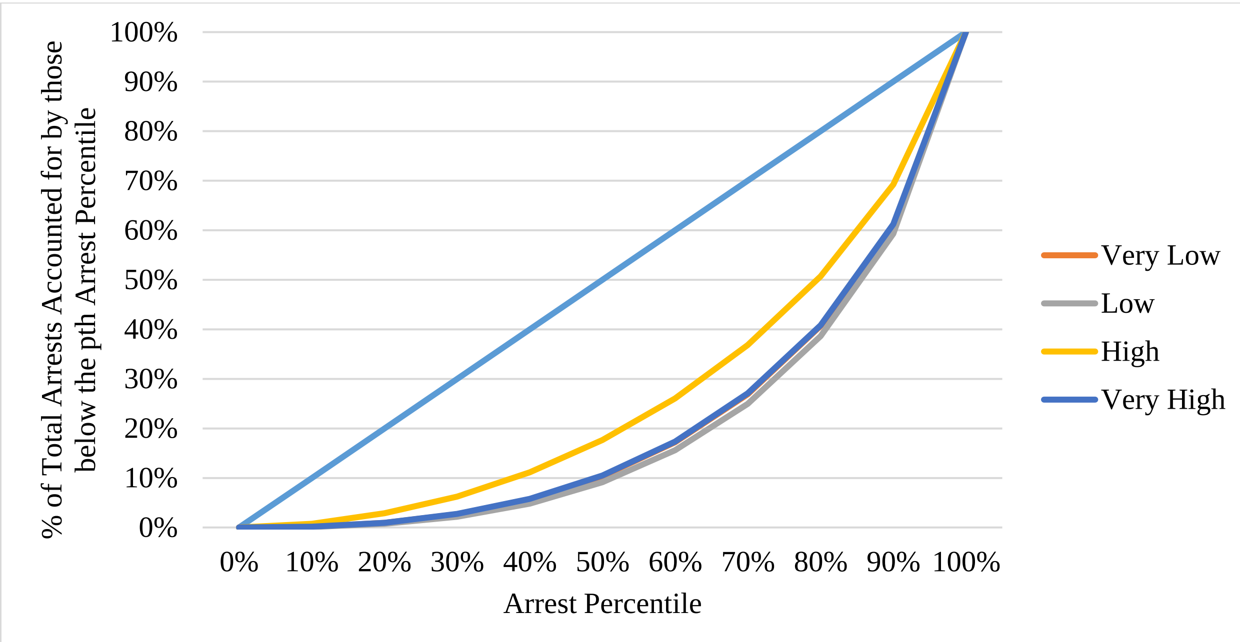


Table 4: Lorenz Curve Comparison for Different Crime Levels

Percentile	Lorenz Curves			
	VL	L	H	VH
0%	0%	0%	0%	0%
10%	0%	0%	1%	0%
20%	1%	1%	3%	1%
30%	3%	2%	6%	3%
40%	6%	5%	11%	6%
50%	10%	9%	18%	10%
60%	17%	16%	26%	17%
70%	27%	25%	37%	27%
80%	41%	39%	51%	41%
90%	61%	59%	69%	61%
100%	100%	100%	100%	100%

Next steps: Improving the model

I see a four important ways in which the model could potentially be improved upon:

1. Improve the accuracy of the crime level variable by separating out on the crimes that are most relevant to stadium arrests (such as vandalism and assault).
2. Obtain data on attendance level for each individual game to improve the accuracy of the model.
3. Create more subgroups based on both crime and attendance levels in modeling the number of arrests and/or properly model for multiple covariate's influence, which we have not yet learned
4. Control for the number of guards present at each game to ensure it is not the presence of more vigilant guards that is driving the number of arrests.

Appendices

Appendix 1: First 10 Rows of NFL Arrests Dataset, 2011-2015

Season	Week	Day	Game Time	Home Team	Away Team	Home Score	Away Score	OT Flag	Arrests	Division Game
2011	1	Sunday	1:15 PM	Arizona	Carolina	28	21		5	n
2011	4	Sunday	1:05 PM	Arizona	NY Giants	27	31		6	n
2011	7	Sunday	1:05 PM	Arizona	Pittsburgh	20	32		9	n
2011	9	Sunday	2:15 PM	Arizona	St. Louis	19	13	OT	6	y
2011	13	Sunday	2:15 PM	Arizona	Dallas	19	13	OT	3	n
2011	14	Sunday	2:05 PM	Arizona	SF	21	19		4	y
2011	15	Sunday	2:15 PM	Arizona	Cleveland	20	17	OT	1	n
2011	17	Sunday	2:15 PM	Arizona	Seattle	23	20	OT	4	y
2012	1	Sunday	1:25 PM	Arizona	Seattle	20	16		0	y
2012	3	Sunday	1:05 PM	Arizona	Philadelphia	27	6		12	n

Appendix 2: First 10 Rows of Crime Levels per 100,000 people Dataset, 2015

State	City	Population	Violent Crime					Property Crime				
			Total	Murder and	Rape	Robbery	Aggravated Assault	**Total**	Burglary	Larceny-Theft	Motor Vehicle Theft	Arson
NM	Albuquerque	559,721	**965.8**	7.7	72.2	301.2	584.8	**6073.2**	1071.2	4076.7	925.3	15.9
CA	Anaheim	349,471	**363.7**	5.2	36.9	125.6	196	**2872.3**	422.4	1972.4	477.6	8
AL	Anchorage	301,239	**1070.9**	8.6	171.6	206.1	684.5	**3917.5**	559.4	2975	383.1	35.2
TX	Arlington	387,565	**502.1**	2.1	53.7	136.5	309.9	**3443.6**	559.9	2657.1	226.5	7.5
GA	Atlanta	464,710	**1119.6**	20.2	36.6	429.3	633.5	**5499.3**	1028.8	3549.1	921.4	10.8
CO	Aurora	360,237	**460.8**	6.7	97.7	124.1	232.3	**2936.7**	467.2	2119.4	350	17.8
TX	Austin	938,728	**372.5**	2.5	51.9	99	219.2	**3771**	532.6	2990	248.3	9.7
CA	Bakersfield	373,887	**484.1**	5.9	19	175.2	284	**4161.4**	1036.9	2484.2	640.3	90.7
MD	Baltimore	621,252	**1535.9**	57.8	46.2	694.2	740.1	**4980.4**	1248.6	2842.3	889.5	41.9
MA	Boston	665,258	**706.8**	5.7	36.1	233.1	431.9	**2316.1**	354.3	1769.5	192.3	N/A

Appendix 3: First 10 Rows of Attendance Levels Dataset, 2015

2015 Attendance		Home				Road				Overall			
RK	TEAM	GMS	TOTAL	AVG	PCT	GMS	TOTAL	AVG	PCT	GMS	TOTAL	AVG	PCT
1	Dallas	8	731,672	91,459	91.5	8	580,857	72,607	98.4	16	1,312,529	82,033	94.4
2	NY Giants	8	632,011	79,001	95.8	8	566,152	70,769	96.4	16	1,198,163	74,885	96.1
3	Green Bay	8	627,308	78,413	107.2	8	520,478	65,059	99	16	1,147,786	71,736	103
4	NY Jets	8	625,280	78,160	94.7	8	582,366	72,795	99.2	16	1,207,646	75,477	96.9
5	Denver	8	615,381	76,922	101	8	527,010	65,876	97.9	16	1,142,391	71,399	99.6
6	Washington	8	609,672	76,209	89.7	8	586,415	73,301	96.2	16	1,196,087	74,755	92.8
7	Kansas City	7	518,604	74,086	96.5	8	528,579	66,072	97.7	16	1,047,183	69,812	97.1
8	Carolina	8	592,454	74,056	100.4	8	565,235	70,654	94.3	16	1,157,689	72,355	97.3
9	New Orleans	8	584,305	73,038	95.5	8	550,722	68,840	97.6	16	1,135,017	70,939	96.5
10	Houston	8	574,159	71,769	101	8	524,655	65,581	95.7	16	1,098,814	68,675	98.4