Attendance, Home Advantage, and the Effect of a City on its Professional Sports Teams

Jack Willoughby

Dr. Charles Becker, *Guiding Professor*

Duke University
Durham, North Carolina
2014

A rising senior, Jack is on track to graduate Duke University in 2015 with a B.S in Economics and a B.S in Statistics. Prior to his senior year, Jack interned at McKinsey & Company in New York City. Please direct all questions to jjwilloughby95@gmail.com

Acknowledgements

I would like to acknowledge the support and guidance of my Urban Economics professor, Dr. Charles Becker, without whom this paper would likely not have been possible. I also want to thank Duke football fans everywhere for both their enthusiastic support and their role in providing inspiration for this exploration.

Abstract

This paper seeks to determine if geographic location causes variation in different teams' Home Advantages, or the difference between their winning percent in home and away games. OLS and quantile regression indicate that characteristics of a team's home city partially determine attendance rates at home games, which in turn impact a team's Home Advantage. On average, higher home attendance is correlated with higher Home Advantage, but the relationship is nonlinear: teams with a relatively high Home Advantage perform better in front of larger crowds, while teams with a relatively low Home Advantage perform better in front of smaller crowds.

Keywords: Sports, competition, cities, venue locations, spatial distribution, fans

I. Introduction

Why are some cities home to more passionate sports fans than others? Seattle, home to the "12th man," has the loudest professional football fans in the nation, but its baseball fans are not nearly as zealous. St. Louis has arguably the most passionate professional baseball fans in the nation, but its football fans are unenthusiastic by comparison. Is this variation all explained by differences in quality of the teams, or are there other factors that explain why sports fans in some cities are more passionate than others? These factors might include general characteristics of a city, like its population, area, weather, and income per capita, and also sport-specific characteristics, like size and age of a professional sports team's arena, and number of professional sports teams in the city. Linear regression analysis indicates that many of these effects are actually associated with increased attendance at games. In predicting which teams win the most games at home relative to on the road, demographic characteristics of cities have no direct effect, but can have an indirect effect through the medium of attendance. On average, attendance at home games was found to have a significant effect on home team performance, but that effect varies based on how well the team plays at home in general. Teams that play well at home play better with high attendance, while teams who play poorly at home play even worse in front of large crowds.

II. Previous Literature

Previous literature has tried, in part, to identify why home teams consistently outperform their away counterparts. Mizruchi (1985) attempts to determine the cause of the "Home Advantage" of professional basketball teams in the 1981-82 season, where Home Advantage defined as the difference between a team's winning percent in home vs. away games. Mizruchi uses data from the NBA because he believes that by playing in small and often structurally unique arenas, NBA teams will experience a significant Home Advantage professional sport. Using publicly available data on the NBA, Mizruchi finds that professional basketball teams win over 60% of their home games, indicating a significant positive Home Advantage across the entire NBA. Mizruchi hypothesizes why this Home Advantage varies between teams, and then uses regression analysis to test his predictions.

First, Mizruchi speculates that professional basketball teams located in smaller cities will have large Home Advantages. He justifies this claim by arguing that smaller cities have more community pride, and residents of smaller cities have fewer alternatives to attending a basketball game, making these residents more passionate basketball fans and increasing their team's Home Advantage. Additionally, he argues that population growth might be indicative of a lower Home Advantage since a city experiencing growth is filling with people who have not lived locally for a long time and therefore may not have developed loyalties to the local team.

Next, Mizruchi argues that older arenas will be associated with a greater Home Advantage. He suggests that most modern arenas are very nearly identical, but older arenas are more unique, and because fans take pride in their unique arena, older arenas will encourage more enthusiasm from fans. Additionally, older arenas are often smaller with fans sitting closer to the court. The author also argues that the larger the share of the stadium that is occupied for a game, the greater the Home Advantage. Stadiums that are closer to capacity will likely be more engaging for fans and produce more crowd noise, which might predictably increase the Home Advantage. Finally, the longer a team has been located in its current city, the more time the city's fans have had to develop loyalties, and the author argues that this would result in a greater Home Advantage.

Mizruchi used only one year of data from one professional sports league in his analysis, and the covariates that he included in the regression analysis were limited in scope. The regression results loosely confirm his hypotheses but generally without statistical significance, potentially indicating the problems with using such a small sample size of only one sport and one year.

One regressor that Mizruchi failed to include is the distance to the nearest competing professional sports team. At the extreme, residents of city with two professional teams in the same sport will have the option of attending either team's home games at roughly the same travel costs. With equal travel costs, home games for the two teams will be largely substitutable for the fan that attends a sporting event for its entertainment value, rather than just to see his/her favorite team play. On the contrary, close proximity of teams could encourage more passionate rivalries that translate to more loyal fans. Rodney et al. (2004) attempt to determine how the distance to the next closest team affects attendance to professional baseball games. They address this

question in the context of whether or not the MLB should consider expansion, because it is theoretically conceivable that adding a new team geographically close to an existing team would cannibalize the sales of the existing franchise. After collecting public data on attendance and location of MLB stadiums, the authors concluded that two MLB teams that are closer together will have lower attendances than two teams that are farther apart, all else equal. The study controls for the size of a team's home city, the team's win percent and improvement in win percent from the previous year, how recently the team's stadium was built, and whether the team recently changed its home location. Specifically, they found that holding all these variables constant, a one-mile increase in the distance from a reference MLB stadium to the closest MLB stadium results in an increase in attendance by about 1,544 fans per year at the reference stadium. The authors confirm the travel cost theory, which says that all else constant, people will consume goods from the closest seller since travel comes with associated expenses. A complete study of the Home Advantage, therefore, must contain a variable to control for how close the stadium of the nearest competitor is, which Mizruchi (1985) does not do.

Next, researchers should be wary of potential endogeneity in measuring the Home Advantage. One example of possible endogeneity is in the use of income per capita as a covariate. Income per capita is important to include if different income levels in a city result in different behaviors of that city's fan base, but it could be endogenous if successful sports teams with large Home Advantages generate huge revenues that result in higher mean income per capita.

Coates and Humphreys (2003) put this problem into context with their review of the existing literature on the economic effect of sports franchises to cities. The authors were interested in whether or not sports franchises provide economic benefit to their home cities to test the rationale many municipal governments cite when subsidizing the renovation and construction of sports facilities using large sums of public money. The authors explain that, contrary to popular beliefs, economic studies have found that the existence of a sports franchise provides no noticeable financial benefit to a city. The authors found that many government-funded forecasts predict that a large public subsidy for sports facility development will result in net economic gain, but because sporting events cannibalize revenue from other forms of entertainment and the money spent on sports-related infrastructure can often be better spend

elsewhere, economists frequently find in retrospect that governments' predictions were grossly overstated.

These three articles lay the foundation for continued research. Mizruchi (1985) establishes that there is a Home Advantage in which some stadiums and city demographics can predict the increase in a team's winning percent when they play at home vs. away, but his dataset is confined to the NBA and his use of covariates is not comprehensive. Rodney et al. (2004) establish that all else equal, a team's attendance is inversely related to the distance of its closest competing team, indicating that distance to nearest competing stadium ought to be included as a regressor in future studies. Finally, Coates and Humphreys (2003) demonstrate that there is unlikely to be an endogeneity problem in which successful teams will increase the local income per capita, suggesting that income per capita is a valid covariate to include in further analysis. This paper aims to include a more robust dataset and collection of regressors to evaluate the drivers of both attendance and the Home Advantage in sports.

III. Data

This analysis employs data on professional sports teams and the cities in which they play their home games. First, the home city of every professional sports team in the "Big Four" leagues, which are the MLB, NBA, NFL, and NHL, was extracted. For each of these cities, demographic data were taken from the 2008-12 American Community Survey through the Social Explorer platform. The ACS is a subset of the US Census that surveys a small portion of the population every year to provide geocoded demographic data. It is only available for the USA, so cities in Canada with professional sports teams were not used in analysis. The data used refer to the principal city of each professional sports team, rather than the local Metropolitan Statistical Area. Principal cities were used because many MSA's are defined too broadly for this analysis, and some cities of interest would have been combined to other larger, nearby cities, thus reducing the total pool of cities available for study. For example, Newark, NJ, which is the home of the New Jersey Devils and, until the 2013 season, the New Jersey Nets, is included in the same MSA as New York City, while in reality the fan bases from the two areas are segregated. A similar problem arises with Anaheim and Los Angeles, Oakland and San Francisco, and others.

Descriptive measures of demographic information of cities from the ACS (colored blue in Table 3) include measures of population, land area, percentage of population by race and ethnicity, breakdowns of age and educational attainment, average income, unemployment levels, poverty levels, and whether someone has moved in the past year. Not all variables were used in the regression analysis because of multicollinearity problems discussed below. While most of the ACS variables are self-explanatory, some merit explanation. "Pct Other" indicates the percent of the population that does not identify as "White" or "Black," which is composed primarily of people of Latino and Asian descent. "Pct People in Poverty" indicates the share of people whose incomes are less than the amount defined by the poverty line. "Inequality Ratio" equals the share of people making over double of the income poverty line divided by the share of people making under half of the income denoted by the poverty line. Finally, "No Move in Past Year" indicates the share of people who lived in the same county one year ago. The dataset contains one observation per year for each non-Canadian team in the four sports leagues over the past four years, and many of these observations are from the same cities: there are 454 total observations, but only 46 unique cities. Some cities have more sports teams than others, and are therefore more represented in the dataset. The values of demographic statistics of cities with many professional sports teams, like New York and Los Angeles, are therefore weighted more than cities with fewer professional sports teams. The presence of repeated observations could skew the means in Table 3 and also depress the standard deviations. These descriptive statistics do not, however, factor into analysis.

Next, data on each professional sports team from their respective past four regular seasons were collected from www.espn.com (colored red in Table 3). These variables include the win percent of each team in home games, away games, and an aggregate of all games, as well as the attendance percent for each team in both home and away games. The number of teams in each city was found by counting the number of professional teams in each city in all sports leagues for each year. Over the course of 2010-13, the Atlanta Thrashers moved cities and became the Winnipeg Jets, and the New Jersey Nets became the Brooklyn Nets; the data reflect these moves. One may find it curious that the average win percent does not equal precisely 0.500. The reason for this is twofold: first, Canadian teams were excluded from the analysis for the reasons mentioned above, and it is unlikely that their collective win percentage was exactly 0.500; and second, in professional hockey, if a game goes into overtime the winning team is

awarded a win, while the losing team is awarded a tie. Therefore, more wins than losses are observed on the aggregate. Interestingly, the average Home Advantage for all observations is 0.1354, implying that the home team wins on average (0.5 + (0.1354)/2) or 56.77% of the time across the four sports and four years of observations.

Data on the climate of each city were collected from the Information Please Database, made available by Pearson Education, Inc. at http://www.infoplease.com/ipa/A0762183.html (colored green in Table 3). These data specify the average temperature in both January and July in degrees Fahrenheit for each city, as well as the total yearly precipitation. The next piece of the dataset is the distance to the closest competitor in the same sport, (colored yellow in Table 3). These distances between cities were generated using the Geobyte's Distance Calculator at http://www.geobytes.com/citydistance.htm. Finally, the age and capacity of each home stadium or arena were recorded from www.ballparks.com (colored orange in Table 3). The data reflect the few teams who moved stadiums over the course of the past four years.

Table 1: 46 US Cities Included in the Data

	Anaheim	Atlanta	Baltimore	Boston	
Brooklyn	Buffalo	Charlotte	Chicago	Cincinnati	Cleveland
Columbus	Dallas	Denver	Detroit	Green Bay	Houston
Indianapolis	Jacksonville	Kansas City	Los Angeles	Memphis	Miami
Milwaukee	Minneapolis	Nashville	New Orleans	New York	Newark
Oakland	Oklahoma City	Orlando	Philadelphia	Phoenix	Pittsburgh
			Salt Lake		
Portland	Raleigh	Sacramento	City	San Antonio	San Diego
San					
Francisco	San Jose	Seattle	St. Louis	Tampa	Washington

Table 2: Descriptive Statistics of Categorical Variables

Sport	N	Year	N
MLB	116	2010	114
NBA	116	2011	114
NFL	128	2012	113
NHL	94	2013	113
Total Obs	454		

Table 3: Descriptive Statistics of Continuous Variables

Variable	Mean	SD
Avg Temp in January	37.92	13.719
Avg Temp in July	76.3	6.248
Yearly Precipitation (in)	40.32	13.685
Win Pct in Home Games	0.5851	0.167
Win Pct in Away Games	0.4497	0.162
Home Advantage	0.1354	0.161
Total Win Pct	0.5175	0.143
Home Attendance Pct	87.26	15.188
Away Attendance Pct	87.54	10.994
Num of Teams in Home City	3.295	1.514
Distance to Closest Competitor (mi)	178.3	147.7
Stadium Age	19.46	19.181
Stadium Capacity	1.264	0.176
Population	1,392,815	1971897
Population Density (ppl/sq mi)	7517.4	6264
Area (sq. mi)	202.2	169.1
Pct Male	0.4871	0.011
Pct of Pop. Under 18 Years Old	0.2234	0.031
Pct Ages 18 to 44	0.4336	0.036
Pct Ages 45 to 64	0.2352	0.015
Pct White	0.5451	0.163
Pct Black	0.2951	0.200
Pct Other	0.1521	0.105
Pct HS Dropouts	0.1847	0.050
Pct HS Graduates	0.2484	0.053
Pct with Some College	0.2548	0.038
Pct College Graduates	0.3045	0.093
Pct Unemployed	0.07805	0.019
Avg Household Income	$44,091	8355
Avg Income per Capita	$26,602	5605
Median House Value	$221,477	143250
Pct People in Poverty	0.2272	0.052
Inequality Ratio	5.71	1.857
No Move in Past Year	0.9408	0.027

N = 454

V. Predicting Home Attendance

The goal of the study is to determine what factors of a city are associated with higher Home Advantages. In a hypothetical scenario in which the owner of a sports franchise is trying to re-locate his team, the owner would choose to move to the city that maximizes his utility, which would be a function of many different things. To simplify, assume that the owner is moving solely to maximize the value of his investment, which is the sports team that he owns. Furthermore, the value of his investment is a function of the number of people who attend games and his team's winning percent. This simplistic model fails to consider the owner's personal locational preferences and also does not consider things like how being in a larger market might help TV deals and attract more profitable players, but ignoring idiosyncratic features of cities will simplify and aid analysis.

The owner's location decision should therefore jointly optimize both the attendance at home games and the effect that different cities have on how well a team plays at home. Addressing these desires sequentially using an instrumental variable, the first goal is to understand what traits of cities are associated with sports fans that are eager to attend games. To do so, home attendance is predicted using OLS regression by various factors of a city, including demographic and other information. The results of this regression are summarized in Table 4.

Table 4: Results of Regression 1- Predicting Home Attendance

Variable	Estimate	P-Value	95% Confidence Interval
(Intercept)	111.6	0.0406	[4.7828, 218.5125]
2011	0.08263	0.947	[-2.3542, 2.5194]
2012	0.1529	0.904	[-2.3275, 2.6332]
2013	0.2517	0.844	[-2.2606, 2.7641]
NBA	-16.49	2.02E-4	[-25.143, -7.8459]
NFL	-9.09	0.0731	[-19.0348, 0.8555]
NHL	-9.364	0.0756	[-19.6995, 0.9708]
Ave Temp in January	-0.4785	4.32E-11	[-0.6174, -0.3396]
Yearly Precipitation (in)	0.04814	0.370	[-0.0572, 0.1535]
Distance to Closest Competitor	-0.003071	0.467	[-0.0114, 0.0052]
Total Win Pct	18.12	1.09E-7	[11.531, 24.7157]
Away Attendance	0.9129	2.99E-10	[0.6349, 1.191]
Num of Teams in Home City	1.616	4.42E-3	[0.5061, 2.7257]
Stadium Age	0.0314	0.202	[-0.0169, 0.0797]
Population/100,000	-0.2434	3.62E-3	[-0.4069, -0.0799]
Area (sq. mi)	0.01846	1.66E-3	[0.007, 0.0299]
Pct Male	149.5	0.0353	[10.3716, 288.6016]
Pct Under 18 Years Old	-61.12	0.0351	[-117.968, -4.2761]
Pct Ages 18 to 44	-103.7	2.01E-4	[-158.1165, -49.3692]
Pct White	-53.1	2.38E-7	[-72.9725, -33.2304]
Pct Black	-37.5	1.17E-4	[-56.4504, -18.5404]
Pct College Graduates	68.9	1.747E-3	[25.9112, 111.8976]
Pct Unemployed	-36.26	0.4684	[-134.4743, 61.95]
(Income per Capita)/$10,000	-11.2	6.403E-3	[-19.2401, -3.1666]
Inequality Ratio	1.332	5.318E-3	[0.3975, 2.2668]
No Move in Past Year	-80.83	0.0286	[-153.1568, -8.5105]
NBA x Temperature in January	0.462	8.46E-7	[0.2803, 0.6436]
NFL x Temperature in January	0.2892	1.41E-3	[0.1123, 0.466]
NHL x Temperature in January	0.2238	0.0196	[0.0361, 0.4116]

$R^2 = 0.6475$; N = 454

The regression yielded many interesting results, most of which aligned with intuition. First, there are factors that influence attendance that cannot be controlled by an owner's location decision. A 10% increase in a team's total winning percent increases the team's average home attendance by 1.8% (p-value = 1.09E-7). It makes intuitive sense that just as theatergoers prefer a high quality show to a lower quality one, fans value quality of play, or higher win percentage. Moreover, fans like to be a part of victories and might be more eager to attend games when their team is more likely to win.

Some teams are enjoyable to watch because they provide high entertainment value, regardless of whether they win or lose. These teams usually consist of interesting players who make exciting or surprising plays. To control for attendance prompted by entertainment alone, away attendance was included in the regression. Holding the quality of the team constant, fans would be more likely to attend games with exciting opponents rather than uninteresting ones, so controlling for away attendance will piece out the effect on attendance that is independent of its home city and instead due to the excitement value that the players on the team generate. Away attendance also picks up the effect of a nearby rival on home attendance. Away attendance was found to be a significant predictor of home attendance (p-value = 2.99E-10). Next, the age of the stadium was intended to be used as a proxy for the tradition of the franchise, for often old stadiums are home to teams with a long history in their city, but it was found that stadium age has no significant effect on attendance (p-value = 0.202).

The rest of the analysis involves covariates that the owner could manipulate by his choice of home city location. The first interesting connection was between attendance at sporting events and weather. Colder cities, as measured by average temperature in January, are associated with significantly lower attendance levels in all sports. A one degree drop in the average temperature in January decreases the average home attendance by nearly 0.5%, indicating that people who live in warmer areas are more likely to attend sporting events. People in colder weather may prefer other leisure activities when they have the time and money to do so. The effect of the cold is not constant for all sports leagues, however. Compared to Major League Baseball, the attendance of all other sports is less negatively affected by the cold. This makes sense intuitively, since attending baseball games involves standing outside for over three hours, and with less constant on-field action than other sports, weather can be very important to the baseball viewing experience. Hockey and Basketball, both of which are played inside, are much less sensitive to the weather. Football is also played outside, but NFL teams only play eight home games per year, so fans may be more likely to fight through the cold than in baseball, when a team plays 81 home games per year. Furthermore, watching football games in the bitter cold seems to have become a source of pride for fans in cold-weather cities like Buffalo, Green Bay, and Chicago, further explaining the smaller negative effect that cold weather has on NFL attendance rates. Unlike temperature, precipitation was found to have no significant effect on attendance (p-value

= 0.370). Average temperature in July was not included because it was too correlated to the average temperature in January.

Next, the regression indicates that an increase in the proximity to the next closest city home to a team in the same professional sports league, contrary to what Rodney et. al (2004) predict, has no significant effect on the team's attendance (p-value = 0.467). This could be biased by the cities with more than one sports team in a professional league, like Los Angeles and New York, if those teams have high attendance and are precisely 0 miles from their nearest competitors. On the other hand, it is conceivable that fans form passionate rivalries when there are other teams nearby and become more committed to attending sporting events. Furthermore, if a team is closer to its competitors, more fans of the away team may be in attendance on an average night. Relatedly, the presence of each additional sports team in a city increases attendance by an average of 1.6% (p-value = 4.42E-3). This could be a result of the sports culture that some cities with multiple teams have been able to develop. Boston, for example, is home to four sports teams, and has developed a passionate sports culture that somewhere like Orlando, which is home to only one team, may not be able to develop. There could therefore be carryover effects for fans between sports, in which the success of one team in a city might encourage residents of the city to be better sports fans in general and therefore support the city's struggling teams as well.

The remaining covariates are all demographic characteristics of cities. First, despite that there are physically fewer people in smaller cities to fill a stadium, population is inversely associated with attendance at sporting events. An increase in 100,000 in population of a city yields a decrease in average attendance by about .24% (p-value = 3.62E-3). This is confirms Mizruchi's (1985) prediction that smaller city population will result in increased attendance. One explanation might be that residents of smaller cities have fewer alternatives to attending sporting events, or that smaller cities foster increased community pride, which also increases desire to support local teams at sporting events. Slightly contrarily, larger area of a city is associated with increased attendance at sporting events. An increase in land area by one square mile is linked with an increase in attendance rates by about .02% (p-value = 1.66E-3). The connection between the area of a city and the attendance at its hometown sporting events is not intuitively clear, but the relationship is one that an owner looking to relocate should keep in mind. One conjecture is

that holding population constant, an increase in area results in a decrease in population density, and people who live in more spread out cities have fewer alternatives to sporting events.

The types of people who live in a city are also important to attendance rates. The analysis suggests that a larger share of males in a city is associated with, on average, higher attendance rates (p-value = .035). Additionally, racially diverse cities are linked to higher attendance rates at sporting events; cities with larger proportions of whites and blacks relative to other races are connected with significantly lower attendance rates. A one percent shift in population from white to "other" is associated with a .53% increase in attendance rates, while a one percent shift from black to "other" is connected with a .38% increase in attendance (p-values = 2.38E-7 & 1.17E-4, respectively). The primary races that compose the "other" baseline category are Latino and Asian, and the analysis suggest that cities with larger shares of these groups place more value on or derive more enjoyment from attending sporting events, although the reason for this is unknown.

The level of education attainment in a city is also a strong predictor of that city's attendance at sporting events. A ten percent increase in the share of the population that has graduated college is associated with a 6.89% increase in attendance (p-value = .00175). One explanation may be that while in college, people gain an appreciation for sports and become lifelong sports fans. A competing interpretation is that this result includes the effect of education on attendance through the effect of education on income and income on attendance. The share of a city that has graduated from college and that city's income per capita are very closely correlated (cor = 0.925), which could result in a multicollinearity problem that confuses attribution of the effects of income and education level. However, in the regression, income per capita was found to have a significantly negative effect on attendance (p-value = .00640), while in reality it would make sense that higher income levels result in larger amounts of money that people are willing to spend on entertainment, of which sporting events is included. Conversely, it is plausible that cities with lower mean income bond over sports, and individuals choose to spend a much larger share of their entertainment budgets on attending sporting events, or substitute sporting events for more expensive alternatives. More research may be needed to determine the true relationship between income, education level, and the passion of sports fans.

Multicollinearity may also be influencing the predicted effect of unemployment on attendance, which was found through regression to be insignificant (p-value = 0.468). Like education, unemployment level is closely correlated with income per capita (cor= -0.645). Intuitively, larger share of unemployment in a city can open up more time for people to attend sporting events, but it can also lead to people substituting in favor of less expensive means of entertainment, like watching a movie or exercising. Again, further research is needed to discover the true effect of unemployment on attendance at sporting events. Next, higher income inequality within a city is associated with a significant increase in attendance at sporting events (p-value = .00532). Given mean income level, higher income inequality would indicate a larger tail on the income distribution. While the exact mechanism by which inequality increases attendance is unknown, a long tail would imply a larger share of very wealthy people, and these people may drive attendance up.

Finally, a larger share of the population who also lived in the city the year before decreases the overall attendance levels (p-value = .0286). This result contradicts Mizruchi's prediction that the longer people had lived in a city, the more loyalty they had built toward local sports teams and the more games they would attend. Contrarily, the results indicate that cities with more population mobility will have higher attendance levels. One explanation might be that people who migrate to a city are excited about being in the new city and want to immerse themselves in that city's culture and sports. Another idea is that there exists self-selection, and people strategically move to the city that is home to their favorite sports teams so they can attend games. Given all of the other important factors that weigh into the decision on where to live, however, this second explanation seems less plausible.

VI. Predicting Home Advantage

With the determinants of attendance defined, the next question of interest is whether different cities can actually make teams perform better. Here, the outcome variable of interest will be the Home Advantage, or the difference between a team's winning percent at home and on the road. Better teams in general will have a higher win percent at home, but they will have a similarly higher win percent on the road. Therefore, the Home Advantage is a comprehensive metric that determines the effect of playing a game at home, rather than on the road.

The demographic qualities of a city can have both direct and indirect effects on a team's Home Advantage. Their indirect effects arise if different city qualities are associated with higher attendance rates, and high attendance rates affect Home Advantage. These indirect effects can be modeled as follows:

$ATT = C_0 + C_1 * X + C_3 * Z$ (Regression 1, above)
$HA = B_0 + B_1 * ATT + B_3 * Y$ (Regression 2, below)

Where:
ATT is home attendance percent
HA is Home Advantage
X is a vector of city characteristics
Y, Z are vectors of other factors

The indirect effect of X on HA = $(dHA/dX) = (dHA/dATT) * (dATT/dX) = B_1 * C_1$

The direct effects are that different traits of a city may be directly associated with a better Home Advantage. This could be if they result in fans who cheer more loudly or influentially at games, or if different qualities of a city are more intimidating for opponents. OLS regression was used to predict the Home Advantage two separate times, first with baseline parameters and no city demographic parameters and the second time with baseline parameters and all parameters from Regression 1. ANOVA was conducted to determine if the two were statistically different, indicating that the two regressions were not statistically different (p-value > .05). Therefore, there were no significant direct results of the city characteristics, and the primary effects of city demographic parameters are indirect, through the mechanism of attendance. In other words, differences in the demographics and sizes of cities affect the outcomes of games through their effect on attendance, so once fans are in the stadium or arena, their impact on the game is not correlated with the demographic information of their hometown. To streamline the output and prevent multicollinearity with the attendance parameter, the more simple instrumental variable regression will be used for analysis. The results of the regression used to estimate the effect of attendance on the Home Advantage are displayed in Table 5.

Table 5: Results of Regression 2- Predicting Home Advantage

Variable	Estimate	P-Value	95% Confidence Interval
(Intercept)	0.0653	0.454	[-0.106, 0.237]
Distance to Closest Competitor	1.06E-04	0.036	[6.888E-6, 2.053E-4]
Stadium Age	-5.14E-04	0.189	[-1.281E-3, 2.537E-4]
Stadium Capacity	-0.0676	0.118	[-0.152, 0.0172]
Attendance (Predicted from Regression 1)	0.00174	0.005	[5.172E-4, 2.971E-3]
2011	-0.0338	0.105	[-0.0747, 0.00712]
2012	-0.0118	0.572	[-0.0529, 0.0293]
2013	0.0239	0.255	[-0.0173, 0.0650]

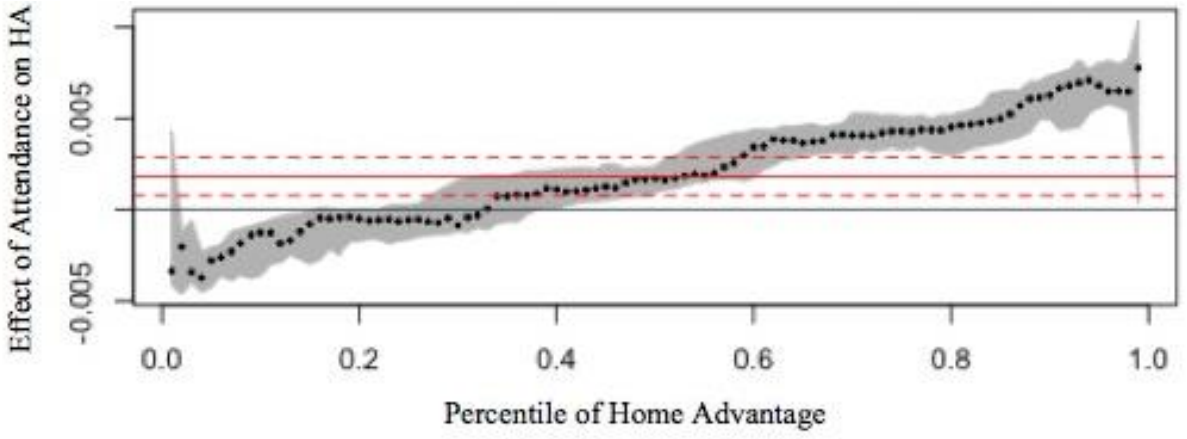
$R^{2 =}$ 0.0604; N = 454

This regression represents the mean effect of different variables on the Home Advantage. The predictors explain only about 6% of variation in Home Advantage, indicating that different teams will play better at home than on the road for a variety of unquantifiable reasons, like the quality of their coach and player leadership.

The results suggest that teams that are farther from their nearest competitor play significantly better in home games (p-value=.036). One explanation could be that when teams are farther from their competitors, there are fewer of the opponents' fans in the crowd, thus increasing the cheering power of the home fans and, therefore, the Home Advantage. The only other significant predictor of Home Advantage is the home attendance percent that was predicted using Regression 1 (i.e., the instrumental variable). For the team with the average Home Advantage, increased attendance has a significantly positive effect on their Home Advantage (p-value = .005).

This effect is not consistent, however, as evidenced by analysis with quantile regression. Quantile regression yields the effect of regressors on the outcome variable at different percentile values of the outcome variable, and therefore illustrates the effect of attendance on teams with varying levels of Home Advantage. For teams with Home Advantage in the bottom third of all teams, increased attendance is actually detrimental to their performance. As teams Home Advantage improves, the effect of attendance on Home Advantage increases linearly. For teams with a Home Advantage above the fiftieth percentile, increased attendance significantly and increasingly helps home performance. The relationship between Home Advantage and the effect of attendance is represented by figure 1.

Figure 1: Relationship between Home Advantage the effect of attendance on Home Advantage



The approximate relationship between the effect of attendance level on Home Advantage and the percentile of home attendance level can be modeled linearly by equation (1):

(Effect of attendance on Home Advantage) =
0.00981*(Percentile of Home Advantage) - 0.00295
(1)

Incorporating previous results yields the effect of city characteristics on Home Advantage:

(Effect of covariate X on Home Advantage) =
(Effect of X on attendance)*[0.00981*(Percentile of Home Advantage) - 0.00295]

Inputting the effects of city demographics on attendance from Table 4 and the percentile of Home Advantage, the owner could calculate the estimated impact of moving to different cities on his Home Advantage. The standard errors vary with the percentile of Home Advantage, and those could be calculated in a statistical package. The owner could then construct a loss function that would weigh the relative benefits of high attendance and high Home Advantage, and from there pick the ideal city to which he wanted to relocate. Equation (1) equals 0 when the percentile of home attendance = 0.3, yielding one intuitive equilibrium: if the owner believes that after the move his Home Advantage will be greater than more than 30% of other professional sports teams, he would be minimizing his loss function by choosing the city that maximized his home attendance rate. However, if the owner believes that after the move his Home Advantage will be less than 70% of other professional sports teams, he would need to weigh the benefit of increased attendance against the detrimental effect increased attendance would have on Home Advantage to determine his optimal destination.

VII. Limitations

The initial motivation for the study was to explore why Durham and Duke University are home to such passionate basketball fans but lackluster football fans, while neighbor NC State in Raleigh is home to passionate football fans but less zealous basketball fans. Having only looked at data from professional sports teams, that initial question has still not fully been answered; the reason for this is that the question may not be possible to answer given the high degree of endogeneity among college sports fans. Many of the fans of college sports teams are alumni or otherwise members of the college, and their association with that college is self-selected. Similarly, many of the fans of a professional sports team are residents of the local city, and those people also chose their residence. The set of determinants of where to go to school versus where to live are very different, however. People seek permanent housing in an area that has jobs, or in which they have familial connections, rather than to be closer to the teams for which they root. Sports play a much larger role in college decisions, however. Anecdotally, I have met many people who chose to attend Duke for the basketball culture, or because they grew up Duke basketball fans. My experience is supported by the literature of Pope & Pope (2009), which finds that both the number and quality of college applicants increase after a college has experienced success in basketball or football. Therefore, students self-select to colleges based on their rooting alliance, so different demographic characteristics of college towns would not breed good fans as much as the tradition of a program would attract good fans. Endogeneity occurs when large Home Advantage draws a certain demographic of person to a college, rather than the characteristics of a city promoting passion among sports fans that then influences the Home Advantage. Therefore, measuring significant contributors to the Home Advantage in college sports is a daunting task that requires further research and expanded use of instrumental variables.

The next limitation is that there is no good way to measure the passion of fans once they are inside a stadium. Some fans go to a game to be social, and they think of attending sports similar to attending a movie or theatrical performance, while other, more passionate fans might cheer loudly behind the belief that their level of energy can impact the home team's performance. The decibel level, number of chants started at a game, or average number of times a fan checks his/her cell phone would all be interesting ways to measure fan involvement and would likely predict Home Advantage, but these data are not available. The only way that this

effect is accounted for in this analysis is by controlling for both the win percentage and the percentage of fans attending away games, which are intended to control for a team's quality and excitement levels, as discussed above. This method is admittedly lacking, and the unmeasured passion of a fan could present an omitted variable bias if it is correlated with outcome variables and also correlated with at least one predictor.

Another limitation of the study is that it did not use any observations from Canada, so the conclusions cannot be interpreted for Canadian cities. The reason for this is that the US Census data did not include any data regarding Canada. Additionally, given the Canada's extreme climate, large area, low population density, and unique preference for hockey, observations from Canada might have proved atypical and difficult to model using regression. Regardless, an owner considering moving his team ought to include Canadian cities in his analysis for completeness.

The final limitation in the study is the repetition of use of different cities. Having observations from a city appear multiple times in regression could bias the results to favor the cities that appear most. While sport and year fixed effects were included in the model, each city had only one set of population parameters, so controlling for city fixed effects would have yielded perfect multicollinearity and not provided results for the effect of any city covariates. The first remedy for this problem would be if the census offered data for each year of observation, instead of requiring the use of the same data for every year of the survey. The presence of four unique years of ACS data would have allowed for the implementation of city fixed effects. Another idea is to run the regression with each city as one datapoint. This methodology would have yielded about 45 observations per year, where each city's Home Advantage would have been a composite score of all sports. The downside with this methodology, and reason why this analysis ultimately opted away from it, is that it would consider the Home Advantage within a city equal across all sports. I think the interactions between sports and factors such as attendance and weather are interesting and would have been unjustifiably lost if Home Advantages were synthesized into composite scores for each city.

VIII. Implications and Conclusion

Analysis indicates that many characteristics of a city are, in fact, good predictors of whether the residents of that city will attend sporting events. Increased attendance, however, has mixed effects on Home Advantage, depending on the value of the Home Advantage for the team in question. Teams that perform poorly at home relative to away actually perform worse as their home stadium fills with fans, while teams that perform relatively well at home perform better in front of many fans. The most interesting extension of the above model might be to attempt to calculate the welfare maximizing spatial distribution of professional sports teams. High attendance rates in general suggest that a home city values having a hometown team, and given that average attendance level is inversely related to population, one implication of the model might be to suggest the expansion of professional sports leagues to smaller cities. Small cities with limited access to teams might benefit considerably more from the addition of a professional sports team than larger cities that already have professional sports franchises in close proximity. To pursue the welfare- rather than profit-maximizing spatial distribution of professional sports teams, some teams would likely need to relocate to smaller markets and make less profit, as revenue from broadcasting is positively correlated with home market size. To make this outcome realistic in context of profit-seeking owners, leagues might need to enhance their revenue sharing programs. These types of interesting questions can only be answered after evaluating the relationship between professional sports franchises and their home cities using analyses like the one above.

References

Coates, Dennis and Humphreys, Brad. "Professional Sports Facilities, Franchises and Urban Economic Development." Public Finance and Management 3.3 (2003): 335-357.

Fort, Rodney, et al. "Location and attendance in major league baseball." Applied Economics (2004): 2117-2124.

Mizruchi, Mark S. "Local Sports Teams and Celebration of Community: A Comparative Analysis of the Home Advantage." The Sociological Quarterly 26.4 (1985): 507-518.

Zimbalist, Andrew and Siegfried, John. "The Economics of Sports Facilities and Their Communities." The Journal of Economic Perspectives (2000): 95-114.

Data:

"Ballparks." Munsey & Suppes. Accessed 3 April 2014. <http://www.ballparks.com>

"City Distance Calculator." Geobytes. Accessed 3 April 2014. <http://www.geobytes.com/citydistance.htm>

"Climate of 100 selected US cities." Information Please Database by Pearson Education, Inc. Accessed 3 April 2014. <http://www.infoplease.com/ipa/A0762183.html>.

ESPN Stats and Information from <www.espn.com>.

U.S. Census Bureau. American Community Survey, 2008-12. Prepared by Social Explorer. Accessed 3 April 2014.