Fan and Player Alignment

An Analysis of the Effect of Local Demographics on Major League Baseball Attendance

Thomas Vollaro

Abstract:
Literature on fan attendance in sports has shown evidence of discrimination against minority players that has gradually dissipated over the last several decades. In the same timeframe, the population of different minorities in the United States have grown and now amass a significant portion of the total population. While many papers have looked at the rise in both minority populations in the U.S. and minority players in sports, only a small amount of research has been done to look at the potential effects of the interaction between the two. This paper explores the interaction between the racial-ethnic characteristics of the players of a Major League Baseball (MLB) team and its surrounding local population. The primary focus of the research is to determine to what extent alignment between fan and player demographics exists and how any alignment affects game attendance. Results indicate that demographic alignment is a significant determinant of attendance and that the more demographically aligned a team is with its local population, the higher their game attendance will be. Additionally, results find that while fans seemingly do not discriminate against their own players, there is potential for discrimination against players of the visiting team and this discrimination may be more severe in areas with a larger white population, though the results are not significant at conventional levels. This work contributes a new determinant of fan attendance to the literature while also providing a look at the current state of fan discrimination in the MLB.

Acknowledgements:
I would like to thank my senior thesis advisor Professor Anne Preston, Professor Richard Ball, and Norm Medeiros for their help and support throughout the process of completing this project.

Table of Contents:

I. Introduction

Fan attendance in professional sports has prompted a strong body of literature that has looked at a variety of different factors of attendance as well as fan discrimination. Previous research identifies factors such as team performance, team prestige, and local population among others that are significant determinants of game attendance. Fan discrimination in sport game attendance has been a popular topic of research but more recently there has been a shift in the literature. Similar to issues of racial discrimination in American society, the literature in this area has shifted from looking exclusively at fan discrimination when there are more black players present on a team to studying the effects of having more Hispanic and international players on a team. In baseball in particular, there has been a definitive shift in the literature due to the growing changes in player demographics. From when the first black players entered Major League Baseball (MLB) in the late 1940's and into the 1980's, a large majority of players were white or black. From the 1990's to today, there has been a tremendous influx of Hispanic and other international players in the MLB, as well as a decline in black players, that has caused the literature to adapt. With a particularly large increase in Hispanic and international populations both on MLB rosters and in the United States, an intriguing topic of study is how these populations interact with each other and with the existing majority white and minority black populations.

The topic of my study is the effect of local demographics on Major League Baseball attendance. I intend to determine whether alignment exists between fan and player demographics, in a racial-ethnic context, in a local area (Metropolitan Statistical Area or MSA) and if that alignment has any effect on game attendance. There are three main research questions I aim to answer: 1) To what extent does alignment exist between fan and player demographics?

2) Does alignment or lack of alignment have an effect on MLB game attendance? 3) Does fan attendance differ according to the race of the starting pitcher?

In order to isolate the effect that any demographic alignment between players and fans has on attendance, other factors affecting attendance such as team performance, alternative sports teams in the area, and MSA demographic makeup must be taken into account. Integral to this research will be isolating the effect of demographic alignment among so many other attendance factors. In order to do this, regression analysis will be performed with attendance as the dependent variable with a number of different control and dummy variables. Two alignment variables, one for home team alignment and one for visiting team alignment, and two starting pitcher dummy variables, one for whether the home starting pitcher is non-white and one for whether the visiting starting pitcher is non-white, will be the primary variables of interest in the analysis. Alignment will be calculated using the Duncan Index for Occupational Segregation which will look at the racial makeup of an MLB team-MSA pairing and determine how much change would be necessary for all racial groups to be distributed comparably between the two.

My motivation for this study is a combination of the high number of minority players in the MLB and the always-hot topic of minority concentrations in urban areas of the U.S. Each MLB team falls into the MSA of a highly populated city thus likely has a high chance of having a strong concentration of minority fans. Certain areas have different minority groups and I want to look at whether certain MLB teams mimic the demographics of the particular population in their MSA. I intend to establish the effect, if there is any, of racial-ethnic alignment on attendance with the goal of improving the literature on factors of attendance in sports by adding another explanatory variable and by looking at an entire MLB season, game by game.

The significance of my research is that, if any effects are established, or if conversely there are no effects at all, another determinant (or non-determinant) of game attendance will be contributed to the growing literature and results will give a glimpse into the current state of fan discrimination in sport game attendance. Does demographic alignment have any effect on attendance? Is racial discrimination evident in fan attendance in the MLB? These are the questions to which my research intends to provide additional explanation in tandem with alternative explanations explored in the existing literature.

II. Background and Literature Review

Prior literature in the realm of analysis of MLB game attendance focuses primarily on two ideas that are similar in research methods but have different motivations. The most relevant topic that is often studied is whether or not discrimination exists in fan attendance. Do fans attend games more or less if there is an increase in minority players or visibility? While some of the earlier research leans towards the idea that discrimination does exist, some of the more recent research finds inconclusive results. The research in this area provides solid bases in analysis of game attendance and minority players that I will use in the production of this study. The second idea that is more frequently studied is the factors that play into fans' decisions to attend games. What factors or variables have the most influence on game attendance? The research in this area of study has produced consistent results with certain factors such as team performance having significant effects on attendance across multiple studies. This area of research is particularly important in my own study as I will need to account for significant factors of game attendance if I want to accurately isolate the effect of demographic alignment on attendance. While I intend to use the frameworks of previous works to aid in my research, my work will differ in the research questions answered and analysis methods. I hope to contribute to these literatures by creating and

measuring a demographic alignment explanatory variable and by providing a closer analysis of attendance factors down to the individual game level.

In order to understand the dynamics of fan discrimination, it is important to first distinguish the economic relationship involved when discussing sport game attendance. In the market place, an employer hires employees who produce or deliver products to consumers. The sports market is no different. In sports, the employer is the owner of the team, the employees are the players, the consumers are the fans, and the product is the game to which a customer has purchased a ticket. According to Becker (1971), there are many different forces that determine discrimination in the labor force. There is employer discrimination, employee discrimination, consumer discrimination, and market discrimination among others that can have great effects on the market place. Consumer discrimination is of particular relevance to my topic as this is, in fact, fan discrimination. Becker writes that consumer preferences dictate the marketability of any particular output. Simply put, a product's success will be determined entirely on the system of consumer preferences. If a consumer's preferences include tastes for discrimination, two seemingly identical products might have different marketability due to the consumer's desire to discriminate. In sports, a fan with tastes for discrimination might be less likely to attend games in which certain races of players participate. Fan discrimination has the power to affect the marketability of the product (the game) and thus it must be taken into account when discussing sport game attendance.

The seminal piece of literature in studying fan discrimination in attending games is Scully (1973). This article looked at a number of aspects related to discrimination in baseball from salary differences (employer discrimination) to fan attendance (consumer discrimination) and was particularly topical at the time as black players in the league were becoming a major

presence both in quantity and quality. In terms of game attendance, Scully produced a very basic regression equation that included some controls and independent variables that could explain attendance while looking primarily at whether fans attend games less if the starting pitcher is black as opposed to white. Scully's paper laid the groundwork for this kind of baseball game attendance literature, and his results, though not controlling for all factors of attendance, indicated that fan discrimination towards black players did exist. His findings opened the door for future research that would later improve upon Scully's model, take more controls into account, and produce similar, but more significant results (Baade and Tiehan, 1990). The seminal paper that laid the groundwork for the study of factors that influence sport game attendance is Noll (1973). Noll identified several key factors that influence attendance at MLB games such as team performance, per capita income, stadium age, sports competition in the area, number of star players, recent pennant win, close pennant race, and games behind pennant. In his research, Noll regressed these variables along with a few others against attendance and found the most significant explainers of attendance to be past performance and available alternative sports entertainment in the area. He also found that a larger black population in the area of study resulted in a significant decrease in attendance. Noll's work presented the first look into what variables must be accounted for when looking to isolate an effect on attendance. These pieces of work provided the backbone to fan discrimination literature and factors in sport attendance literature, respectively, and allowed for future research to expand upon the basic concepts with the improved data and technology which have become available over the past four decades.

The literature on racial discrimination by fans in attending sports games has evolved over the last several decades. In baseball particularly, studies have shifted from analyzing whether having additional black players on a roster has an effect on attendance to whether having

additional international players or Hispanics has an effect. To some effect, the literature on the subject has paralleled racial issues in the U.S. which today focus strongly on immigration, particularly Latino immigration, as opposed to a few decades ago when the focus was black versus white. A stronger contributing factor to this shift in research, however, is the change in player demographics. There has been a great influx of Latino and other international players in more recent years paired with a decline in black players that has led the literature to adapt.

Medoff (1986) and Schollaert and Smith (1987) are two articles that looked at the effect of black players on attendance in the MLB and National Basketball Association (NBA), respectively. Each paper was founded in different sociological theories about discrimination and was particularly interested in the performance aspect associated with sport. Medoff analyzed MLB attendance data with propositions that performance is independent of social interaction between players and fans, and because social interaction between players and fans is limited, discrimination should not exist. Schollaert and Smith looked at NBA attendance in the context of a theory that whites wish to avoid contact with blacks, especially those of greater or equal status. This is of particular intrigue in the NBA as they found in their research that black performance is stronger than white performance through analysis of individual player statistics and salaries. Both papers ultimately found that having additional black players on a roster didn’t have a significant effect on attendance which is important given that these results contradicted literature (Scully, 1973; Noll, 1973) written a decade earlier which found a significant negative effect. These results are important in my research because they provide sociological contexts but also show that performance, not only at the team level but also at the individual player level, should be taken into account in my analysis. Beyond these two papers, the literature has shifted to primarily focusing on foreign-born players.

The most contemporary and comprehensive work in this area of study is Tainsky and Winfree (2010). The authors analyze MLB attendance over a 21-season frame from 1985-2005, in which there was a significant increase in foreign-born players, in order to determine if fan attendance is affected by the number of international players on a roster. Their results show that having more foreign-born players had a significant effect on attendance; in 1985, having additional international players contributed to lower game attendance and loss in team revenue equivalent to $731,875 per additional player. As the study went along, the effect gradually became positive to the point that in 2000, each additional international player contributed to an increase in team revenue of $595,632 from attendance. This result is especially important in motivating my topic of study. If an MLB team increases their revenue from game attendance by over $500,000 per additional foreign-born player, one can imagine that number might only increase if the international players had similar demographic backgrounds to the local fan base. In other words, if a team wanted to increase their revenues from game attendance it would be beneficial for them to have additional foreign-born players, but if they wanted to maximize revenues, a team would want additional foreign-born players who also parallel the demographics of their fan base as fans might sympathize and attend games more frequently.

Literature on factors in sports attendance spans across multiple sports and often hits many of the same chords. Across the existing research, factors such as team performance, number of alternative sports teams available in the area, number of star players, and population are taken to be significant determinants of sport game attendance. Literature in this area tends to include the main explanatory factors but focuses specifically on contributing a unique variable or variables of interest and attempts to isolate the effect of those unique variables. The idea behind this is that if new factors that influence attendance can be found and proven significant, then teams can use

knowledge of these factors to their benefit and improve game attendance and increase revenue. My motivations are similar; I want to establish the existence and potential significance of alignment between player and fan demographics on game attendance with the goal of contributing another factor of attendance to the existing literature while also giving a more contemporary look into the current state of fan discrimination in baseball.

The article that most closely relates to my work is Breckenridge and Goldsmith (2009), which studies the effect of minority player visibility on game attendance. The authors use the sociological theories of social distance, spectacle, and group threat to explain the effect that minority player visibility had on game attendance from 1930-1961. Social distance refers to the idea that people interact less with people outside their "in-group" (same racial group) who are distanced from them socially. Spectacle is the theory that "racing" or labeling an athlete by their race can attract more white spectators while group threat is the sociological theory that as a minority or outside group grows in numbers, they become more of a threat to the majority or inside group. This article is especially related to my topic for two reasons. First, it created a variable, the minority player index, that most accurately accounts for the presence of minority players by combining multiple playing statistics that capture minority players who actually play significant time as opposed to some who might never play and therefore have little influence on attendance. This is important for my research as I hope also to capture the effect of players with significant playing time who might actually have an effect on attendance. Second, the paper specifically examines the effect that minority visibility has on attendance in areas of high minority populations. This particular aspect of the study is especially relevant to my topic because I hope to establish and analyze the effect of any alignment between fan and player

demographics (i.e. if players and fans are both similar minority groups), and this paper is a basis for examination of the effect of minority players on potential minority attendance.

Overall, the literature on my topic is well established but I think has yet to account for every possible factor of attendance. Existing literature focuses primarily on issues of fan discrimination of minority players or factors of attendance. I intend to continue with this focuses in mind while also looking at the interaction between these two issues. With my research, I intend to provide further evidence of the main factors of attendance through a more recent study while also contributing a new factor of attendance, fan and player demographic alignment. Determination of alignment, that is the construction of some variable that measures similarities between fan and player demographics, will be vitally important in this research.

Duncan and Duncan (1955) created an index of occupational segregation in order to explore primarily gender differences in occupations. The Duncan Index, now widely used in labor economics research, gives the percentage of male (or female) workers that would have to change their jobs in order for the occupational distribution of men and women to be the same. The index is calculated by taking the absolute value of the difference in percent of men and percent of women in each occupation, summing over all occupations, and dividing by two. This yields a number from 0 to 100 with 0 indicating identical distributions across gender and 100 indicating an occupation distribution totally segregated by sex. This index provides a method of determining how equally distributed men and women are across occupations and will serve as my alignment variable in this study. While I will not be looking at gender distribution, the Duncan Index can also be used to look at racial segregation in an occupation. Instead of looking at the difference between men and women in a particular occupation, I will look at the difference between an MLB team's percentage of players of a certain race and an MSA's percentage of

population of that same race. This variable, calculated by taking one half of the summation of the absolute value of the differences for each race (white, black, Hispanic, other), will yield a number which determines alignment between a team's racial-ethnic characteristics and an MSA's racial-ethnic characteristics. Using the Duncan Index, I am also able to look at alignment between the visiting team and the "home" MSA for each game. These alignment values are the focus of my study.

III. Data

The final data set is a cross-section of MLB games over the course of the 2015 MLB regular season for each of the 30 MLB teams excluding home games for the Toronto Blue Jays (not based in U.S.). In total, there are 2,335 observations each representing the attendance of a single game in the 2015 MLB season. The final data set is constructed with data obtained from a number of different sources as well as manually inputted data. The attendance data, along with the names of the starting pitchers, time of day, and day of week data, are from Retrosheet, a prominent baseball statistics database. Other team-level data including previous performance metrics (wpctlag, divisionranklag, playofflag, worldserieslag) and team prestige information (stadiumage, franchiseage) are from The Baseball Cube. The MSA census data (MSAincome, MSApop, MSAwhitepercent, MSAblackpercent, MSAhispanicpercent) was obtained from the American Community Survey (ACS) data for 2015 from the U.S. Census Bureau and includes demographic information of interest for the 26 MSAs of the 29 MLB teams in the study (Toronto is excluded and 6 teams share an MSA with another team). The rest of the data was manually inputted and merged into the master data set.

Variables in the data set used for analysis include: attendance; wpctlag (a team's winning percentage the previous season); divisionranklag (a team's rank in the division the previous

season); stadiumage (the age of a team's ballpark); numberofothermsateams (the number of other professional sports teams in an MSA); starplayers (the number of superstar players on a team), franchiseage (the number of years a team has existed); whitepercent (the percentage of white players on a team); blackpercent (the percentage of black players on a team); hispanicpercent (the percentage of Hispanic players on a team); MSAincome (median household income of the MSA); MSApop (the total population of an MSA); MSAwhitepercent (the percentage of the MSA population that is white); MSAblackpercent (the percentage of the MSA population that is black); MSAhispanicpercent (the percentage of the MSA population that is Hispanic); playofflagdummy (a dummy variable indicating if a team made the playoffs in the previous season); worldserieslagdummy (a dummy variable indicating if a team made the World Series in the previous season); timeofdaydummy (a dummy variable indicating whether a game was played during the day or at night); weekenddummy (a dummy variable indicating if a game was played on the weekend); awayspdummy (a dummy variable indicating whether the visiting starting pitcher is non-white); homespdummy (a dummy variable indicating whether the home starting pitcher is non-white); alignment (how closely a team's player demographics mirror their surrounding MSA), and away alignment (how closely the visiting team's player demographics mirror the demographics of the "home" MSA). Descriptive statistics are presented in Table 1.

Table 1: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
1. Attendance	2,335	30374.5	9554.079	8701	53518
2. Winning percentage lag	2,335	.4996086	.0592669	.395	.605
3. Division rank lag	2,335	2.999143	1.438246	1	5
4. Stadium age	2,335	24.42099	23.48699	3	103
5. Number of other MSA teams	2,335	2.794004	1.349079	1	6
6. Star players	2,335	1.485653	1.223392	0	5
7. Franchise age	2,335	87.30021	45.14381	17	144
8. Percent of white players	2,335	.5330462	.1248863	.25	.8333333
9. Percent of black players	2,335	.1219363	.0859051	0	.3076923
10. Percent of Hispanic players	2,335	.3222378	.131945	.0833333	.6666667
11. MSA median income	2,335	65064.66	11258.92	48911	93294
12. MSA population	2,335	6173229	4854254	1575747	2.02e+07
13. MSA white pop. percentage	2,335	.5634693	.1531322	.299455	.8578676
14. MSA black pop. percentage	2,335	.1399817	.0733799	.0486086	.3314368
15. MSA Hispanic pop. percent.	2,335	.1875072	.1301707	.0165577	.450951
16. Playoff lag dummy	2,335	.3456103	.4756687	0	1
17. World series lag dummy	2,335	.069379	.254152	0	1
18. Time of day dummy	2,335	.6830835	.4653742	0	1
19. Weekend dummy	2,335	.3199143	.466543	0	1
20. Away pitcher race dummy	2,335	.3010707	.4588216	0	1
21. Home pitcher race dummy	2,335	.2997859	.4582622	0	1
22. Demographic alignment	2,335	.279561	.1324733	.048191	.6238019
23. Away demographic alignment	2,335	.281561	.1266234	.0310349	.6542522

The main variables of interest for the regression analysis in this study are the starting pitcher and alignment variables (awayspdummy, homespdummy, alignment, awayalignment). The starting pitcher variables (awayspdummy, homespdummy) are simply dummy variables whose values are zero if the starting pitcher is white and one if the starting pitcher is non-white (black, Hispanic, or other). Alignment is determined using a modified Duncan Index for Occupational Segregation which gives the percentage of team players of one race that would have to change race in order for the racial distribution of team players and the racial compositions of the MSA to be equal. In practice, alignment is determined by taking the absolute value of the difference between a racial group's team percentage and the same racial group's MSA population percentage, adding up the values of each racial group pairing, and dividing by two. The resulting alignment value is a number from zero to one with numbers closer to zero indicating a strong alignment between the team-MSA pair and numbers closer to one indicating a weak alignment between the team-MSA pair. The away alignment variable is constructed in the same manner except that the visiting team's roster percentages are used instead of the home team. Note that according to rows 22 and 23 of table 1, both mean alignment and mean away alignment are approximately 0.28.

From Table 1, some interesting patterns can be observed in the data. The mean attendance for an MLB game in the 2015 season was 30,375 but there is a good deal of variance in the attendance numbers as the standard deviation is 9,554. Player demographic characteristics and MSA demographic characteristics both show that there is a good deal of variation between MLB teams and MSAs in terms of percent population by race, indicating that many MSAs and even many MLB teams can vary greatly from one another. The mean of superstar players is just about 1.5 star players per team but as the minimum and maximum values show, some teams have

as many as 5 superstars and some have none at all. This presents a significant advantage, in terms of attendance, for the small amount of teams with multiple superstars. Both franchise age and stadium age show a great deal of variance and as will be presented later in Table 3, younger teams and stadiums are outperformed by their elders in terms of game attendance.

In my study, I intend to determine alignment between player and fan racial-ethnic demographics but I do not want my results to include irrelevant data. I do not intend to look at a total team's player demographics but only the demographics of players who actually see significant playing time and who might actually have an effect on fan attendance. In other words, I want to eliminate the potential of "bench players" or relatively unknown players skewing the data. Thus, I chose to "qualify" players for my research based on playing time limitations that I dictated in the following way: a player, to be qualified, must have at least 400 plate appearances, 140 innings pitched, or 30 saves. The MLB uses approximately 500 plate appearances and 160 innings pitched in order to qualify for end-of-year awards so I used that as my base measure. I decided to drop the limit to 400 plate appearances and 140 innings pitched in order to include more observations and capture the effects of more players with significant playing time. I added the 30 save qualification for relief pitchers as closers in the MLB today are very popular players and have significant impacts late in games, despite a limited number of innings pitched. The list of qualified players was obtained from ESPN using sorting tools that allowed for the exclusion of players who did not meet my specific limitations. With this list of qualified players, I manually coded each qualified batter and pitcher as white, black, Hispanic, or other based on their country of birth, images of each player, their names, and my own discretion when necessary. Country of birth data was obtained online from The Baseball Cube and images of players were viewed on Google images. The roster percent variables (whitepercent, blackpercent,

hispanicpercent) were constructed using the number of qualified players of each race for each team. Similarly, the starting pitcher data was constructed by manually coding the race of each starting pitcher for each game from the Retrosheets data as white, black, Hispanic, or other based on country of birth, images, and my own discretion.

The number of superstar players, traded players, and number of other teams in an MSA were also manually coded for each team. I constructed a list of superstar players with the criteria that the players won either the Most Valuable Player (MVP), Cy Young, or Rookie of the Year (ROY) award in the past ten years or appeared on a list of top-selling MLB jerseys in 2012, 2013, or 2014. The list was constructed using award race data from Baseball Reference and jersey sales information from ESPN. One additional list of manually coded data included players who were traded mid-season. From my qualified list of players, I extracted the players who had two teams listed because they had been traded in the middle of the season. With this new list, I manually coded the date of each trade using data from an MLB trade tracker from CBS Sports. With the date of each trade coded, I can be sure the effect of each traded player is attributed to the right team throughout the season. Also manually coded was the number of other major sports teams in the local MSA. For each team, I recorded the number of other sports teams from the MLB, National Football League (NFL), National Basketball Association (NBA), and National Hockey League (NHL) in the same MSA.

Table 2: Demographic Alignment by Team

TEAM	ALIGNMENT
Chicago Cubs	0.048191
Texas Rangers	0.064166
Tampa Bay Rays	0.07409
Arizona Diamondbacks	0.1313234
Minnesota Twins	0.1377658
New York Yankees	0.1462825
Houston Astros	0.1655027
Miami Marlins	0.1869615
Atlanta Braves	0.1893553
New York Mets	0.2188199
Philadelphia Phillies	0.2649647
Colorado Rockies	0.2706921
Boston Red Sox	0.2752628
Milwaukee Brewers	0.2805041
Chicago White Sox	0.2809422
St. Louis Cardinals	0.2865357
San Diego Padres	0.2912719
Washington Nationals	0.2919841
Seattle Mariners	0.301539
Cleveland Indians	0.3088055
San Francisco Giants	0.3413375
Cincinnati Reds	0.3499289
Baltimore Orioles	0.3602956
Pittsburgh Pirates	0.4203676
Los Angeles Angels	0.4278178
Oakland Athletics	0.4398223
Kansas City Royals	0.4447973
Los Angeles Dodgers	0.4819649
Detroit Tigers	0.6238019

Table 2 gives alignment values by team. Some of the teams with the strongest alignment to their local MSA population include the Chicago Cubs, Texas Rangers, and Tampa Bay Rays. The team with the least demographic alignment to their local MSA is the Detroit Tigers. The large variance in alignment values is more likely a result of differences in MSA population characteristics rather than team characteristics. Tampa Bay and Texas are areas of high white and Hispanic concentrations which align more closely with the overall MLB player population while

Detroit has a higher relative black population which is dissimilar to the overall MLB player population.

IV. Methodology

The main purpose of the study is to determine the extent to which demographic alignment between an MLB team roster and the population of the surrounding MSA is a significant determinant of game attendance. In addition to the alignment variables described in the dataset, other factors of game attendance must be included in analysis in order to increase the accuracy of any findings. The dependent variable in the study is the total game attendance of each MLB game in the 2015 MLB season excluding any home games for the Toronto Blue Jays and any game in which attendance was not observed. A number of controls and dummy variables are included in regression analysis to account for alternative factors of game attendance. These variables include factors related to a team's performance in the previous year (rank in the division, whether they made the playoffs, whether they made the World Series), factors related to the history and prestige of a team, in particular franchise age and stadium age, factors related to a team's roster (number of star players, percent of white players, percent of black players, percent of Hispanic players, percent of other players), factors related to the game itself (day of week, time of game), and factors related to a team's local MSA (total population, population broken down by race, income, number of other teams in the MSA). However, the variables of interest in this particular study are demographic alignment (alignment between a home team and its local MSA), visiting team demographic alignment (alignment between the visiting team and the local MSA), home team starting pitcher race (a dummy variable which indicates 1 if non-white), and visiting team starting pitcher race.

Regression analysis was performed using a number of robust OLS regression models. There are two sets of four models of regressions to look at the alignment and pitcher race variables separately. The first set of four models represented by equation 1 below looks specifically at the effect of alignment on attendance with a broader look at other factors of attendance. The first model includes only demographic alignment and visiting team demographic alignment regressed on the natural log of attendance. The second model adds variables for the percent of players of each race on each team. The third model adds the MSA population demographic variables. The fourth model includes the rest of the control variables in the study.

Equation 1:

$$\ln(attendance_i)=\alpha+b(alignment_i) + c(awayalignment_i) + d(controls_i) + \varepsilon_i$$

The second set of four models represented by equation 2 below looks specifically at the effect of the race of the starting pitcher on attendance, particularly if the starting pitcher is non-white. This set of models also looks at whether a stronger interaction between the race of the starting pitcher and the MSA white population indicates more or less discrimination. In other words, this interaction will describe if areas with larger white populations tend to discriminate more against the starting pitcher. The first and third models include only the dummy variables with none and all of the controls, respectively. The second and fourth models include the dummy variables as well as the interaction terms with none and all of the controls, respectively.

Equation 2:

$$\ln(attendance_i)= \alpha+b(homespdummy_i) + c(awayspdummy_i)$$

$$+ d(sphomewhiteinteraction_i) + e(spawaywhiteinteraction_i) + f(controls_i) + \varepsilon_i$$

V. Results

The results of the first set of linear regression models are presented in Table 3 and show that demographic alignment is a significant determinant of attendance. To reiterate, the alignment variables range from zero to one with a zero indicating identical racial distributions between an MLB team-MSA pairing and a one indicating the exact opposite. That is to say an increase in the alignment value actually means a decrease in alignment. In the first three models, both the alignment and visiting team alignment variables show that an increase in the alignment value results in an increase in attendance. In the first model in which only the alignment variables are included (column 1), an increase in the alignment variable (decrease in alignment) results in a significant increase in attendance (row 1) and an increase in the away alignment variable results in a significant but smaller increase in attendance (row 2). In the second model, when the racial characteristics of the team are included (column 2), the coefficient on alignment remains significant and increases in magnitude while the coefficient on away alignment becomes insignificant. In the third model, when the racial characteristics of the team and the local MSA are included (column 3), the coefficient on alignment remains significant and decreases in magnitude while the coefficient on away alignment remains insignificant. However, when the rest of the controls are added for the fourth model (column 4), particularly the team performance variables such as division rank, playoff participation, and World Series participation, the alignment variable changes direction. Results show that a one standard deviation increase in the alignment value results in a 8.25% decrease in attendance. The coefficient for alignment is significant at the 99% level and indicates that teams with greater alignment with their MSA population (a lower alignment value) will receive greater attendance figures. The away alignment variable is significant at the 95% level and remains positively correlated with

attendance such that a one standard deviation increase in the away alignment value results in a 1.25% increase in attendance. These results indicate that once team-specific variables are controlled for, greater alignment between home fans and home players results in an increase in attendance while greater alignment between home fans and the visiting players actually results in a decrease in attendance. This could be a result of home fans sympathizing with the home players of similar demographic backgrounds and attending games more frequently while also attending more frequently when teams of dissimilar demographic backgrounds come into town, perhaps in hopes of seeing their home players defeat the visiting players of a different race.

Aside from the alignment variables, the control and dummy variables, as seen in column 4 of table 3, performed as expected and in line with previous research. The better a team performed the previous season (lower division ranks, participated in the World Series), the greater their attendance numbers were. A team who made the World Series in the previous year (row 11) experienced a 30.1% increase in attendance on average compared to teams that did not make the World Series. Similarly, the more “prestige” a team has (the older their stadium or franchise is), the greater their game attendance was. A one year increase in stadium age (row 12) and franchise age (row 15) resulted in a 2.03% and 2.37% increase in attendance, respectively. For each additional star player (row 14), a team experienced a 10.4% increase in game attendance. For each additional professional sports team in the local MSA (row 13), a team experienced a 6.5% decrease in attendance. This makes sense as it is logical that a consumer will purchase less of the original product when there are more substitutable products available (other teams). Games played on the weekends unsurprisingly experienced a large increase in fan attendance when compared to games played during the week. Games played on the weekend (row 17), on average, experienced almost a 20% increase in attendance compared to those played

during the week. The population controls also performed as previous research indicated they might. As the MSA population and the median household MSA income increase (row 18 and row 19, respectively), teams experienced a significant increase in attendance. All of these control and dummy variables were significant at the 99% level and performed almost identically in the second set of regressions.

Table 3: Demographic Alignment Regression Results

VARIABLES	(1) lnattendance	(2) lnattendance	(3) lnattendance	(4) lnattendance
1. alignment	0.550***	0.800***	0.507***	-0.623***
	(0.0539)	(0.0663)	(0.0628)	(0.119)
2. awayalignment	0.106*	0.00481	0.0660	0.0987**
	(0.0588)	(0.0604)	(0.0591)	(0.0473)
3.whitepercent		-2.125***	-3.123***	-0.924***
		(0.210)	(0.224)	(0.248)
4.blackpercent		-2.016***	-2.406***	-1.041***
		(0.217)	(0.222)	(0.216)
5.hispanicpercent		-2.222***	-2.535***	-0.741***
		(0.215)	(0.216)	(0.267)
6. MSAwhitepercent			-1.220***	2.532***
			(0.145)	(0.253)
7. MSAblackpercent			-1.917***	0.562**
			(0.184)	(0.280)
8. MSAhispanicpercent			-1.005***	2.932***
			(0.173)	(0.308)
9. divisionranklag				-0.213***
				(0.0119)
10. playofflagdummy				0.0437
				(0.0288)
11. worldserieslagdummy				0.301***
				(0.0202)
12. stadiumage				0.00203***
				(0.000311)
13. numberofothermsateams				-0.0654***
				(0.0116)
14. starplayers				0.104***
				(0.00834)
15. franchiseage				0.00237***
				(0.000327)
16. timeofdaydummy				-0.0196
				(0.0128)
17. weekenddummy				0.197***
				(0.0122)
18. lnMSApop				0.0768***
				(0.0276)
19. lnMSAincome				0.920***
				(0.0713)
Constant	10.08***	12.13***	14.02***	-0.276
	(0.0219)	(0.200)	(0.269)	(1.142)
Observations	2,335	2,335	2,335	2,335
R-squared	0.048	0.093	0.163	0.498

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4 presents results from the second set of linear regressions. These results are particularly intriguing and indicate fans do discriminate based on the race of the starting pitcher and that areas of larger white populations tend to discriminate more, though not at a significant level. Again to reiterate, for the starting pitcher dummy variables a value of zero indicated that the starting pitcher was white and a value of one indicated that a pitcher was non-white. In the first model in which only the starting pitcher dummy variables are included (column 1), the coefficient on the away starting pitcher dummy variable is insignificant while the coefficient on the home starting pitcher dummy variable is significant and negative, such that on average if a home pitcher is non-white there is 3.21% decrease in attendance. In the second model, when the interaction terms are added (column 2), the direction on the coefficients of both dummy variables change from negative to positive but are both insignificant. The coefficients for the interaction terms are also insignificant in this model. In the third model with the pitcher dummy variables and all of the controls (column 3), the coefficient on the away pitcher variable is significant and negative such that on average if an away pitcher is non-white there is a 4.19% decrease in attendance. The coefficient on the home pitcher variable is insignificant in this model and the controls performed similarly to the first set of regressions presented in Table 3. In the fourth model with the dummy variables, interaction terms, and all of the controls, the directions changed again on the coefficients of the pitcher dummies but both were insignificant while the coefficients on the interaction terms also remained insignificant. However, the coefficients on the interaction terms between both the home and away starting pitcher race and MSA white population were close to significant at the 90% level and indicated a negative effect on attendance. These results indicate that once the team-specific variables are controlled for, fans are unlikely to discriminate against the home starting pitcher based on his race but are still prone

to discriminate against the visiting starting pitcher. This presents an intriguing result. While it does show that fan discrimination exists, it also indicates that perhaps fan discrimination is adapting such that fans only discriminate against the opposing players and not just all players of minority descent, regardless of team. Additionally, these results show that areas of larger white populations tend to discriminate against the starting pitcher more, though not at conventional levels of significance.

Table 4: Starting Pitcher Discrimination Regression Results

VARIABLES	(1) lnattendance	(2) lnattendance	(3) lnattendance	(4) lnattendance
1. awayspdummy	-0.0265	0.0482	-0.0419***	0.0152
	(0.0163)	(0.0513)	(0.0115)	(0.0387)
2. homespdummy	-0.0321**	0.0613	-0.0169	0.0385
	(0.0162)	(0.0571)	(0.0125)	(0.0464)
3. sphomewhiteinteraction		-0.158		-0.0953
		(0.0977)		(0.0797)
4. spawaywhiteinteraction		-0.133		-0.101
		(0.0873)		(0.0682)
Controls	None	None	All	All
Constant	10.28***	10.28***	-1.079	-1.092
	(0.00982)	(0.00981)	(1.130)	(1.130)
Observations	2,335	2,335	2,335	2,335
R-squared	0.003	0.006	0.495	0.496

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VI. Discussion and Conclusions

There are several important results to take away from this research. First, demographic alignment between MLB players and local fans is a significant determinant of attendance.

Stronger alignment between home team players and local fans results in an increase in game attendance that for MLB teams signifies an increase in revenue from ticket sales. Second, fans attend games more frequently when playing a team with which they are actually less demographically aligned. Third, fans attend games less frequently when the starting pitchers are non-white, though this effect is only significant for the away starting pitchers. Fourth, in areas of larger white populations there appears to be more discrimination against both the home and away starting pitchers though results are insignificant. Lastly, the team-level and MSA-level controls for attendance perform as expected given their effects in the large body of previous literature on fan attendance.

The reasons for the results found in this paper cannot be stated with absolute certainty but there are some assumptions that can be made. Fans attend games more frequently when players are of the same demographic background as themselves. This is a rather intuitive result as fans are more likely to sympathize with players whom they believe to be similar to themselves in their background, in how they were raised, and in their culture. Less intuitive is the result that fans attend games more frequently when there is less demographic alignment with the visiting team. One possible scenario is that fans want to see their own hometown players of similar demographic background beat the visiting team of players with different demographic backgrounds. Like in many large international sports tournaments (World Cup, World Baseball Classic, Olympics), rivalries between certain nationalities might exist and could be affecting how fans attend MLB games. Another intriguing result is that fans do discriminate against the starting pitcher, though it is only significant for the visiting pitchers. This could mean fans do not discriminate against their own players because they are on the hometown team but do not feel the same obligation to visiting players and thus they tend to discriminate. Though not significant, it

is also interesting that areas with larger white populations seem to discriminate more against the starting pitchers. This is not entirely surprising, nor is it in contrast to previous studies on fan discrimination, but it certainly shows that white fans might discriminate more than others.

Future research can further solidify the significance of demographic alignment by most importantly looking at multiple years of data on the subject. This research only looks at one MLB season, but it would be helpful to look at a larger timeframe and to see how the results may have adapted over time. It might also help to research other professional sports leagues in the U.S. or even abroad to see if similar patterns emerge across all sports. Diving even deeper into more specific racial-ethnic backgrounds to determine if certain nationalities or countries feel stronger senses of “connection” than others would also be an intriguing topic of study.

References:

1. Becker, Gary S. 1971. *The Economics of Discrimination*, 2nd ed. Chicago, IL: The University of Chicago Press.
2. Scully, Gerald. 1973. "Discrimination: The Case of Baseball." In *Government and the Sports Business*, edited by Roger G. Noll, 221-273. Washington, D.C.: The Brookings Institution.
3. Baade, Robert A., and Laura J. Tiehan. 1990. "An Analysis of Major League Baseball Attendance, 1969-1987." *Journal of Sport and Social Issues*, 14(1): 14-32.
4. Noll, Roger G. 1973. "Attendance and Price Setting." In *Government and the Sports Business*, edited by Roger G. Noll, 115-131. Washington, D.C.: The Brookings Institution.
5. Medoff, Marshall H. 1986. "Baseball Attendance and Fan Discrimination." *Journal of Behavioral Economics*, 15(1): 149-155.
6. Schollaert, Paul T., and Donald Hugh Smith. 1987. "Team Racial Composition and Sports Attendance." *The Sociological Quarterly*, 28(1): 71-87.
7. Tainsky, Scott, and Jason A. Winfree. 2010. "Discrimination and Demand: The Effect of International Players on Attendance in Major League Baseball." *Social Science Quarterly*, 91(1): 117-128.
8. Wakefield, Kirk L. 1995. "The Pervasive Effects of Social Influence on Sporting Event Attendance." *Journal of Sport and Social Issues*, 19(4): 335-351.
9. Breckenridge, R. Saylor, and Pat Rubio Goldsmith. 2009. "Spectacle, Distance, and Threat: Attendance and Integration of Major League Baseball, 1930-1961." *Sociology of Sport Journal*, 26: 296-319.