Factors that Impact MLB Attendance

Principal Investigators (Group 21): Garrett Schreffler (schreffler@wisc.edu)

Introduction

Major League Baseball is a unique sport compared to other professional sport leagues in North America. Baseball has by far the largest discrepancies in payroll between teams. In 2019, the Boston Red Sox had an opening day payroll of $225,183,602, which was the highest amount while the lowest, the Tampa Bay Rays, had a mere $64,901,866. The Red Sox had 3.5 times as much money as the Rays! For small market teams like the Rays, attendance to home games can be a great source of revenue to evenly compete against these large market bullies. Since attendance is such an important factor to these types of teams, I studied what factors play an impact on attendance. More specifically, how payroll, home runs hit, and winning percentage play a factor.

All of the data I used was compiled from baseball-reference.com. In my experiment, I took data between 2015 and 2019 to create a large enough sample size to be able to make inferences about the data. I used the more recent years in order to create the most accurate data possible.

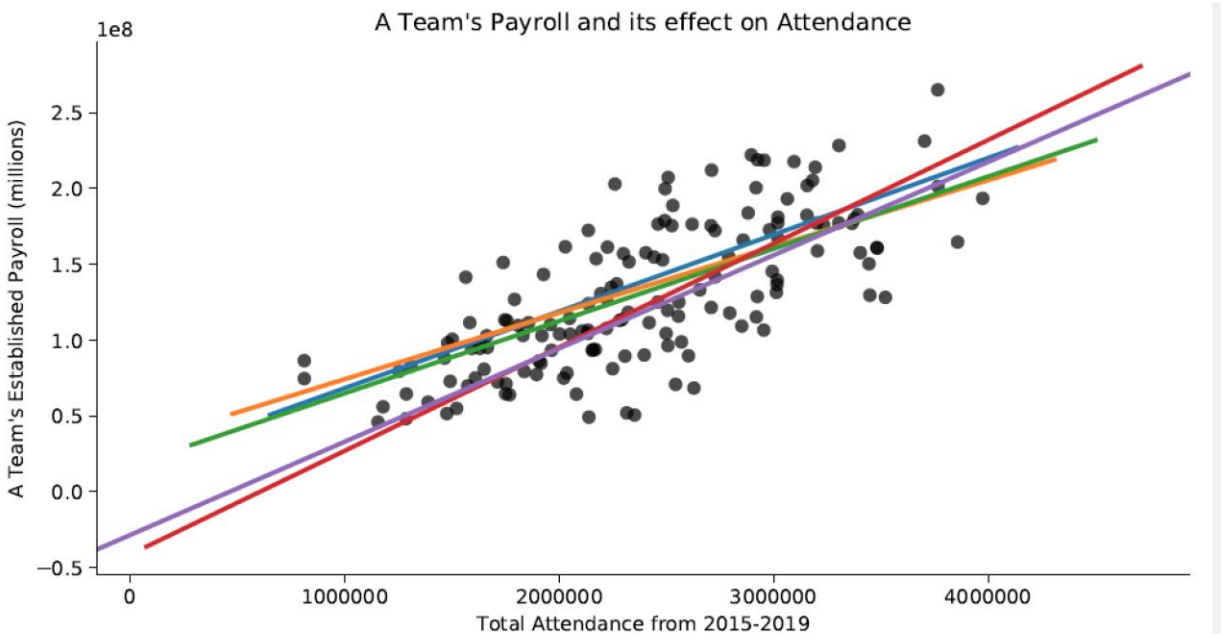
In the first part of my study, I looked at how payroll and attendance play off of each other. Not surprisingly, there was a positive correlation between the two. The higher the payroll, the higher the attendance. For most fans, home runs are one of the most exciting thing about baseball to watch. Because of this, I wanted to see how it impacts attendance. I ran an OLS regression to see exactly how home runs affect it. Although it was minimal, I found there to be a positive correlation between the two. Lastly, I compared how well a team is playing to its attendance. This ended up being the most important factor in increasing attendance. As a team wins more, attendance grows.

Analysis

Payroll and Attendance

To easily show the relationships expressed in my study, I use seaborn. I chose seaborn over matplotlib because I wanted to include the different lines of best fit in order to more easily see trends. Plotting the total attendance per team and the total payroll of that team, figure 1 shows a positive relationship between the two. This is unsurprising because the highest payrolls in baseball are the big market teams like Chicago, New York, Boston, and Los Angeles. These large market teams will never have an issue with poor attendance because of how highly populated the cities are.

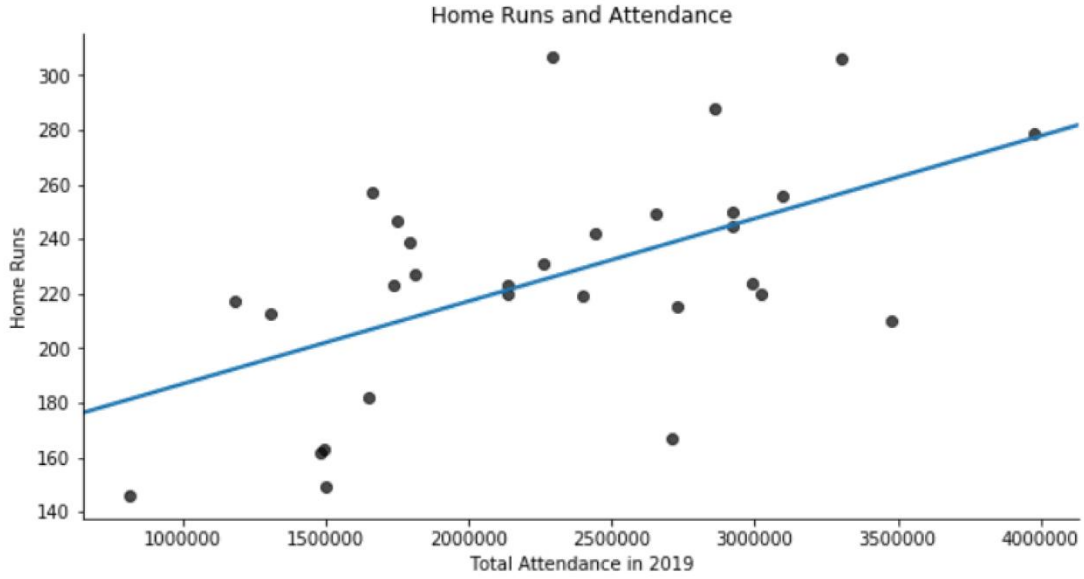
Figure 1



Home Runs on Attendance

In 2019, the league hit a total of 6,776 home runs, crushing the previous record of 6,105 set in 2017. Seeing that this was such a record breaking year, I wanted to see how home runs in 2019 would compare to attendance that year. In figure 2, it can be seen that there is a positive correlation between the two. As the number of home runs rose in 2019, so did attendance.

Figure 2



Although figure 2 shows that there is a positive relationship between home runs hit and attendance, I wanted to know just exactly how big of an impact home runs played. In order to do this, I ran an OLS regression on these two variables, shown in figure 3. As the number of homeruns increases by 1, attendance increases by 10,110 people, according to this regression. Figure 3 also gives us an R-squared (coefficient of determination) value of 0.933. This means that the independent variable (home runs) explains the dependent variable (attendance) well.

Figure 3

Dep. Variable:	Attendance	**R-squared:**	0.933
Model:	OLS	**Adj. R-squared:**	0.930
Method:	Least Squares	**F-statistic:**	401.1
Date:	Wed, 29 Apr 2020	**Prob (F-statistic):**	1.58e-18
Time:	15:00:00	**Log-Likelihood:**	-442.88
No. Observations:	30	**AIC:**	887.8
Df Residuals:	29	**BIC:**	889.2
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
HR	1.011e+04	504.722	20.029	0.000	9076.660	1.11e+04

This regression proved that more home runs hit will increase attendance at games. However, this does not appear to be the case on a yearly basis. As I have said previously, 2019 had a record breaking number of home runs hit. However, looking at figure 4, 2019 only had the 21st highest total attendance in MLB history. 2017 was the second highest home run total ever, and that year only places it at 14th highest attendance. 2007, the highest total attendance ever, ranked 15th in total homeruns ever hit. Comparing figures 3 and 4, I realized homeruns are not the only factor playing an impact on attendance.

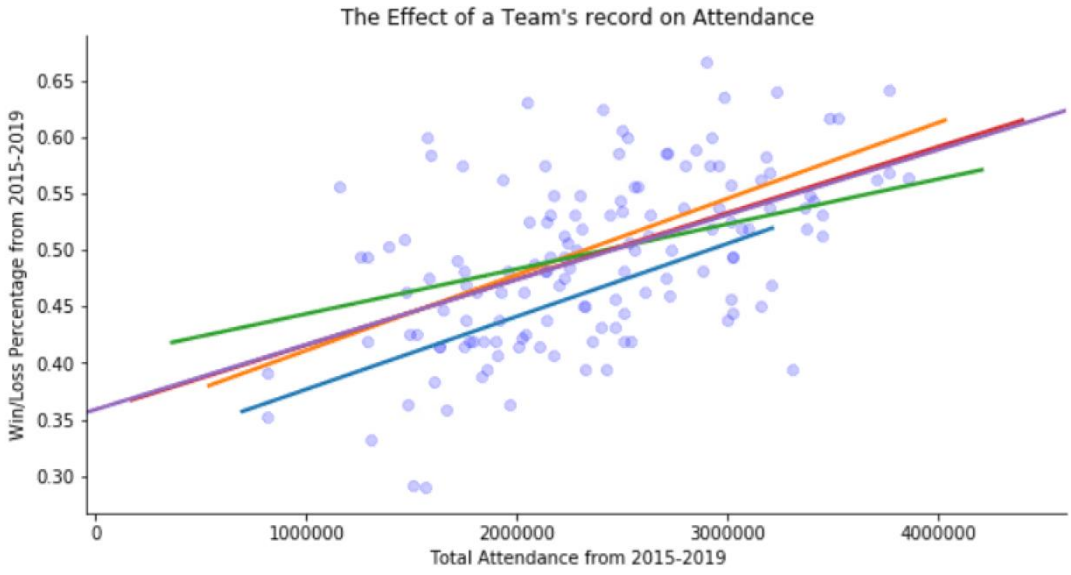
Figure 4

Year	Attendance	Attend/G	Lg Payroll*
2007	79484718.0	32696.0	2.499199e+09
2008	78624315.0	32382.0	2.694090e+09
2006	76043902.0	31306.0	2.337875e+09
2005	74915268.0	30816.0	2.189013e+09
2012	74859268.0	30806.0	2.950093e+09
2013	74027037.0	30451.0	3.150728e+09
2014	73739622.0	30345.0	3.398869e+09
2015	73719340.0	30349.0	3.680887e+09
2009	73430580.0	30218.0	2.791645e+09
2011	73425667.0	30228.0	2.872257e+09
2016	73159044.0	30131.0	3.761012e+09
2010	73061763.0	30066.0	2.757480e+09
2004	73022972.0	30075.0	2.078658e+09
2017	72678797.0	29908.0	3.983893e+09
2001	72581101.0	29881.0	1.969086e+09
2000	71358907.0	29377.0	1.685768e+09
1998	70601147.0	29030.0	1.285345e+09
1993	70257938.0	30964.0	9.031152e+08
1999	70139380.0	28887.0	1.503589e+09
2018	69671272.0	28659.0	3.964097e+09
2019	68494752.0	28198.0	3.999827e+09

Winning Percentage on Attendance

Winning is the most important part about baseball no matter the size of the market or what team you are. The goal of every franchise in every professional sport is to win the championship and to do that, you have to win games. Figure 5 shows the relationship between a team's winning percentage and their total attendance in the last five years. This is where the "fair-weather fan" comes into play. A lot of fans only want to go see their team play if the team is playing well. If they are have a losing year, these fans will not want to pay for a ticket to just go and watch their team lose. Figure 5 shows exactly this.

Figure 5



As discussed previously, it can be assumed larger market teams never have an issue achieving a high attendance every year, given that their populations are so large. Even if they are having a losing season, they will still fill up their stadiums. Given this, I ran an OLS regression on winning percentage and attendance that excluded the top 5 largest markets in order to get more accurate results. Unsurprisingly, the top attendance holders every year are the largest market teams. Running a regression in this way reduces this bias. Figure 6 shows us that as a small market team's winning percentage goes up, so does attendance.

Figure 6

Dep. Variable:	Attendance	**R-squared:**	0.940
Model:	OLS	**Adj. R-squared:**	0.936
Method:	Least Squares	**F-statistic:**	220.2
Date:	Fri, 01 May 2020	**Prob (F-statistic):**	5.87e-10
Time:	12:53:22	**Log-Likelihood:**	-217.20
No. Observations:	15	**AIC:**	436.4
Df Residuals:	14	**BIC:**	437.1
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
W-L%	4.288e+06	2.89e+05	14.838	0.000	3.67e+06	4.91e+06

Conclusions and directions for future research

From the study above, having a higher payroll correlates to having a higher attendance at games. However, having hitting more home runs and winning more games will increase yearly attendance as well. Small market teams like the Brewers and Royals can increase attendance by playing better. You do not have to be a large market team with a higher payroll to get fans to show up to your games.

Given more time and more data, I could expand this research over a longer period of time than just the past 5 years to get even more accurate results. Given more data, I would also look at other factors that could potentially play a role in increasing attendance like average ticket prices and promotional events on certain gamedays.