Kyle Chmielewski
Prof. Larsen
Econ 253
December 10, 2021

Optimizing Team Offense or Team Pitching: Does Team OPS or Team ERA have a bigger impact on Wins?

Introduction:

The game of baseball has been played for centuries and statistics have been kept for just as long. In this paper, I will be examining statistics from the 1985 season to the 2016 season to determine if team offense or team pitching is more important to winning. The specific stats I will be looking at are team OPS (on-base plus slugging) for offensive performance, and team ERA (earned run average) for pitching performance. This is an incredibly important question to answer because in the modern baseball market player's salaries are lofty investments for the franchise and it essential for General Manager's to optimize their teams performance. Therefore by determining which part of the game increases winning the most, GM's can make the tough decisions regarding their roster.

This has been a question on the minds of researchers for years and many scholarly articles have been written about the subject. For example, Jay, Javier Lopez, Daniel J. Mundfrom, and R. Schaffer wrote the paper, "What Makes a Winning Baseball Team and What Makes a Playoff Team?." where they try and predict if a team will make the playoffs using linear regression of OPS and ERA. Their playoff prediction model, using the same population model as mine, was able to predict playoff teams with 80% accuracy. Another set of researchers looked into just the 2013 season and how pitching, defense, and offense contributed to wins. They concluded that every one point increase in ERA led to about 19 less wins, every additional error per game led to roughly 13 less wins, and finally every additional point of OPS led to 2.165 more wins. All things considered, the question of whether offense or defense is more important has been considered since the advent of the game, and people have tried to quantify it with statistics for decades.

Data:

In this experiment, I will be using the stats from the 1985-2016 MLB seasons, this includes over 918 observations, with one for each team. I am specifically looking at team stats, OPS for their offensive performance, and ERA for each team's pitching performance. The distribution of these two variables is very interesting as this dataset includes some of the best teams in MLB history, such as the regular season win record holding 2001 Seattle Mariners. The lowest ERA in the dataset is of the 1988 Mets with a 2.91, and the highest OPS is of the 2003 Boston Red Sox posting an .853 mark.

Model:

This experiment utilizes linear regression to determine the relationships between team OPS, team ERA, and wins. My basic population model for this experiment is as follows:

$$Wins_i = \beta_0 + \beta_1 teamOPS_i + \beta_2 teamERA_i + \varepsilon_i$$

Since this is a multivariate regression, I will be testing the hypothesis below which tests if our coefficients equal each other under the null.

$$H_0: \beta_1 = \beta_2 = 0$$

$$H_1: \beta_1 \neq \beta_2$$

This is an appropriate hypothesis because if the two coefficients are equal then that implies that both OPS and ERA have the same effect on wins, and therefore defeats the purpose of this experiment. I expect the results of the model to help us reject this hypothesis in favor of the alternative, even though as much as people argue about baseball and what is more important hitting or pitching, one is definitely more important to winning than the other.

Results:

When examining the results of this experiment it is important to start with our simple multivariate regression, which is represented by column one in our table below. This regression gave us a β_1 of 0.159, where this is interpreted as every additional 1 point increase in team OPS leads to 0.159 more wins. Our β_2 was -16.18, where this states that every one point increase in team ERA leads to 16.18 less wins. Secondly, to get a clearer picture of possible bias in the regression I added some controls to the regression, and the controls I chose were other team statistics such as home runs, walks, and strikeouts. This regression, displayed in column 2 below, resulted in a β_1 of .0239, and a β_2 of -15.81, both having the same interpretation as above. This regression shows us that OPS experiences some negative bias as the coefficient gets smaller, and ERA experiences positive bias as the new coefficient is larger than the simple regression. This is likely caused by the importance of strikeouts and defense to helping a team win. After running this regression, I decided to test my hypothesis on this regression with the controls, so I ran a F-test which resulted in a F-stat of 1193.47 and a P-value of 0.0000. These results lead us to reject the null hypothesis, and favor the alternative hypothesis, which was expected.

Lastly after testing our hypothesis, I decided to test how team OPS and ERA affect a team's chances of winning the world series. I did this through a simple linear probability model, where the regression gave us a β_1 of .0000228, where this is interpreted as every additional 1 point increase in team OPS leads to 0.00228 percent increase in the chance of winning the World Series. Our β_2 was -.0538593, which means that for every one point increase in team ERA leads to a 5.38 percent decrease in a team's chances of winning the World Series. The last two regressions I ran included interaction terms to determine the difference in the effects of OPS and ERA between AL and NL teams. First off for OPS, I was able to determine that a one point increase in OPS for AL teams increase wins by .0143159 wins more than it does for NL teams. Secondly, when looking at ERA the regression results say that for every one point increase in ERA for AL teams increases wins by .3557564 wins more than NL teams.

Conclusion:

Based on these results, it becomes painfully clear that pitching is more important to winning than hitting, and that winning teams are good because of their pitching. Not only does one regression confirm this result, but all do. This is very interesting, because it is clear some teams don't believe in this philosophy as power hitters often garner some of the most lucrative deals in the league, even when the math shows that a one point increase in ERA reduces a teams probability of winning the World series by a massive 5%. I don't believe there is a significant amount of possible bias still left to explore, as the controlled regression didn't lead to much change. Ultimately, I think these results could lead to an increase in teams investing in their pitching staffs, which we are already seeing with the contracts of players such as Gerrit Cole and Max Scherzer.

References:

Fullerton, S. L., Fullerton Jr, T. M., & Walke, A. G. (2014). An econometric analysis of the 2013 major league baseball season.

Jay, J. L. D. J. M., & Schaffer, R. What Makes a Winning Baseball Team and What Makes a Playoff Team?. *Multiple Linear Regression Viewpoints*, 23.

Tables and Figures:

Figure 1 (Regression table):

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Simple Multivariate Regression	Multivariate Regression With Controls	Linear Probability of Winning Championship	Interaction Term for OPS	Interaction Term for ERA
AL				-13.81	-1.426
				(10.76)	(2.581)
OPS_points	0.159***	0.0239***	2.28e-05	-0.0674***	
	(0.00783)	(0.00631)	(0.000240)	(0.0123)	
OPS_Interaction				0.0143	
				(0.0145)	
errors_t		-0.0753***	-0.000640**	-0.156***	-0.0828***
		(0.00941)	(0.000305)	(0.0152)	(0.00908)
hit_t		0.0435***	0.000165**	0.0444***	0.0451***
		(0.00192)	(6.84e-05)	(0.00319)	(0.00192)
hr_t		0.0906***	0.000276	0.0732***	0.105***
		(0.00651)	(0.000276)	(0.0122)	(0.00540)
so_t		-0.00658***	-0.000101**	-0.0101***	-0.00819***
		(0.00114)	(4.63e-05)	(0.00227)	(0.00109)
bb_t		0.0315***	7.70e-05	0.0516***	0.0330***
		(0.00253)	(0.000104)	(0.00446)	(0.00245)
era_t	-16.18***	-15.81***	-0.0539***		-15.68***
	(0.444)	(0.329)	(0.0134)		(0.435)
ERA_Interaction					0.356
					(0.613)
Constant	30.31***	50.73***	0.0988	56.77***	64.98***
	(5.169)	(4.615)	(0.174)	(9.641)	(3.161)
Observations	918	918	918	918	918
R-squared	0.586	0.844	0.047	0.403	0.841
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					

Figure 2 (Summary Statistics):

```
                win_t
-------------------------------------------------------------
      Percentiles      Smallest
 1%           52            43
 5%           60            47
10%           65            47       Obs                 918
25%           71            49       Sum of wgt.         918

50%           80                     Mean           79.94336
                      Largest        Std. dev.      11.84022
75%           89           106
90%           95           108       Variance       140.1909
95%           98           114       Skewness       -.197728
```

99%	103	116	Kurtosis	2.664434

era_t

Percentiles		Smallest		
1%	3.09	2.91		
5%	3.36	2.94		
10%	3.5	2.95	Obs	918
25%	3.8	2.96	Sum of wgt.	918
50%	4.17		Mean	4.205
		Largest	Std. dev.	.5565088
75%	4.58	5.71		
90%	4.94	5.76	Variance	.309702
95%	5.17	6.01	Skewness	.3033773
99%	5.53	6.38	Kurtosis	2.867634

teamOPS

Percentiles		Smallest		
1%	.647	.627		
5%	.673	.634		
10%	.688	.636	Obs	918
25%	.71	.637	Sum of wgt.	918
50%	.737		Mean	.7388627
		Largest	Std. dev.	.0410717
75%	.766	.84		
90%	.795	.844	Variance	.0016869
95%	.81	.85	Skewness	.1495002
99%	.837	.851	Kurtosis	2.773862

Figure 3 (Graph of OPS vs. Wins):

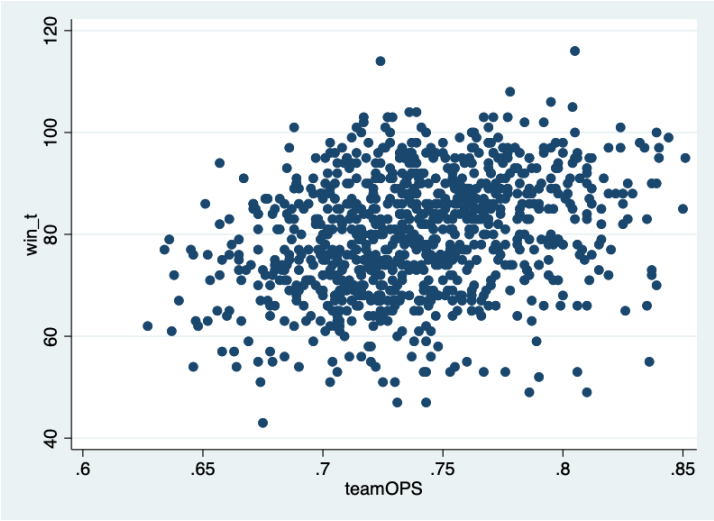
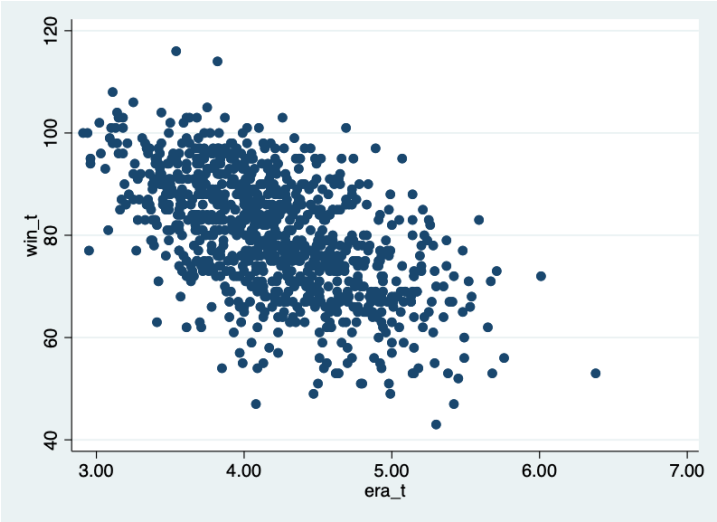


Figure 4 (Graph of ERA vs. Wins):


win_t
120
100
80
60
40
3.00
4.00
5.00
6.00
7.00
era_t