The Worth of an MLB All-Star: Are MLB All-Star Players the Key to Wins, the Playoffs, and the World Series?

Cassidy Mickelson-Carter

Southern Utah University ('18)

Abstract

All-Stars are given the title of being the best a team has to offer but are they helping their teams get more wins, make it to the playoffs and win the World Series? Players are more likely to make the MLB all-star team in the first years of their career, thus a team should identify and purchase these players if it is found that these players could help a team be successful. It is also in the team's best interest to know how many players they should purchase with an "all-star" status. This study found that while holding home game attendance, salary, earned run average (ERA) and fielding percentage constant, a team should try to have 8-9 all-stars on their team to increase wins and probability of making the playoffs, while 6-7 all-stars will help a team to win the World Series.

Southern Utah University Working Paper Series

I. Introduction

After each Major League Baseball season, a team's management team must sit down and reevaluate the last season. They evaluate short falls and successes. For a team, successes could be defined as wins, making the playoffs, or winning the World Series. Through this study we consider if a single player can make a difference, specifically if an "all-star" can bring more wins, and help a team make the playoffs or win the World Series. MLB all-stars are picked first by the general public; they decide the starting nine players. The next sixteen players are selected by the players and managers, then the mangers select the next 9 and then the general public again selects the final player to the roster. The final player was not introduced until 2010. Looking at a recent article (Reiter, 2014) it discusses how the Houston Astros went from being one of the worst teams in baseball to one of the best, simply by of looking at the statistics and data. The Houston Astros increased their All-Stars from two to five from 2016 to 2017; and in 2017 the Houston Astros won the World Series. The winner of the World Series "should" be the best of the best team that year; however the seven game series might not be the best measurement of success for a baseball team, which is why this study looks at both the World Series, number of wins and the playoffs (Mlodinow, 2009). Number all-stars on a team is something, theoretically, a team could control, and therefore it would be in the best interest of the team to have 8-9 all-stars to increase wins and probability of making the playoffs, while 6-7 all-stars is ideal for winning the World Series based upon this study.

II. Data

The data used is from SeanLahman.com. Sean Lahman is a database journalist and author who brings together baseball data and makes it available to the public. The data he collected is from the research he did for the book, *Total Baseball: The Biographical Encyclopedia* (Pietrusza et al., 2000). The original data set was broken down into per player statistics per year. In this study, the data was merged into a per team per year basis, thus summing all of the all-star players. The summary statistics produced by this data is as follows:

	Number of Allstars	*World Series*	*Playoffs*	*Number of Wins*
Mean	2.40	0.03	0.29	80.97
Standard Error	0.07	0.01	0.020	0.50
Median	2	0.00	0.00	82
Standard Deviation	1.56	0.18	0.45	11.37
Minimum	1	0	0	43
Maximum	9	1	1	116
Count	510	510	510	510

The years 2000-2016 were selected because of the expansion teams, the Arizona Diamondbacks and Tampa Bay Rays, which came into existence in the year 1998. Only looking at the data with all teams present for all years decreased the likelihood for any holes in the data.

Drawbacks of the study are that teams cannot always control which players are selected to the all-star team. A player could be selected by a team, and he could be injured or have a less than promising season. Also, players that are not the "best" could also be selected, simply because of that team or player's fan support. This could be controlled by taking into account the city's size, which could be considered the fan base, but not every fan would be included in that city, thus creating a bias in the selection process that cannot be accounted for by this study.

III. Method

This project uses a linear regression model to determine the expected number of wins based upon the number of all-stars on a team. Next, a linear probability model is used to look at the probability of getting into the playoffs and making it to the World Series is based off of the amount of all star players between the years 2000-2016. The model used for this data is:

1) $$(Number\ of\ wins)_{it} = \beta_o + \beta_1(allstar)_{it} + \beta' X + u$$

$$(Playoff)_{it} = \beta_o + \beta_1(allstar)_{it} + \beta' X + u$$

$$(World\ Series)_{it} = \beta_o + \beta_1(allstar)_{it} + \beta' X + u$$

2) $$(Number\ of\ wins)_{it} = \beta_o + \beta_1(allstar)_{it} + \beta_2(allstar)^2{}_{it} + \beta' X + u$$

$$(Playoff)_{it} = \beta_o + \beta_1(allstar)_{it} + \beta_2(allstar)^2{}_{it} + \beta' X + u$$

$$(World\ Series)_{it} = \beta_o + \beta_1(allstar)_{it} + \beta_2(allstar)^2{}_{it} + \beta' X + u$$

3) $$(Number\ of\ wins)_{it} = \beta_o + \beta_1(2-3\ allstar)_{it} + \beta_2(4-5\ allstar)_{it} + \beta_3(6-7\ allstar)_{it} + \beta_4(8-9\ allstar)_{it} + \beta' X + u$$

$$(Playoffs)_{it} = \beta_o + \beta_1(2-3\ allstar)_{it} + \beta_2(4-5\ allstar)_{it} + \beta_3(6-7\ allstar)_{it} + \beta_4(8-9\ allstar)_{it} + \beta' X + u$$

4) $$(World\ Series)_{it} = \beta_o + \beta_1(2-3\ allstar)_{it} + \beta_2(4-5\ allstar)_{it} + \beta_3(6-7\ allstar)_{it} + \beta_4(8-9\ allstar)_{it} + \beta' X + u$$

The dependent variable $Number\ of\ wins$ is measured by the number of wins in a normal season. The dependent variable *Playoff* is measured by a binary term. A team receives a 1 if they made the playoffs that year and a 0 if not. There are ten teams that make the playoffs each year out of thirty. Six teams win their division and are automatically placed into the playoffs; four other teams are "Wild Cards" and are placed into the playoffs for having the best win-loss ratio behind the leaders of the division. The second equation's dependent variable *World Series* is similar to the *Playoff* variable in that it is binary. If a team won the World Series for that year that would be a 1, and every other team for that year would receive a 0. All of these dependent terms are for time *t* for the years 2000-2016 for team *i*.

For the first regression ran (Model 1) the variable *allstar* is a count term which measures how many members of that team were voted onto the MLB All-Star game that year. The All-Star team is voted on by the general public. Each league (American and National) gets a team comprised of the thirty-four "best" players. The second regression (Model 2) *allstar* and $allstar^2$ are both count variables determining if *allstar* is linear or quadratic. The third regression (Model 3) *allstar* is a step variable comparing each category to one all-star. In this case, each category for all-stars is a binary term. For all three regressions this it is for time *t* and team *i*.

The controls for this study are *attendance* which is the overall home game attendance, *salaries* which is the overall team salary, *fielding percentage (FP)* which is an average of how many time a team fields a ball hit or thrown properly, and *earned run average* (*ERA)* which is the average amount of runs let in by the pitchers of that team. These four variables are contained in X. All four of these variables are for each team *i* for time *t*. These variables are the controls because it lets us see if even if a team is extremely popular, pays their players very well, have low ERA and high FP if an all-star can make an effect on a team.

IV. Results

This model's coefficients will give team's management the necessary statistics to make decisions about purchasing all-star players. The β_o in regressions (1) and (2) are not relevant because it is the intercept for when there are no all-stars and all constants are 0 which is an impossible scenario which is why in all cases for all tables the coefficient was negative. A team could not logically have negative wins or a negative probability of making it to the playoffs, but there are not any teams with no all-stars, no fielding percentage, no fans, no ERA, and no salary. Thus this makes β_o irrelevant having OLS still a viable method of evaluating the models.

The β_1 coefficient for regression (1) and (2) for *Number of Wins* is the expected increase/decrease of number of wins for one additional all-star, all else held constant. For *Playoffs* and *World Series* regressions β_1 is the expected percentage point increase of making it to the playoffs or winning the World Series for one additional all-star, all other variables held constant. The results of regression 1 found that *Number of Wins* would increase by $2.471(\beta_1)$, probability of making the *Playoffs* would increase by $8.7(\beta_1)$ percentage points for each all-star added to the team, and probability of winning the *World Series* would increase by $1(\beta_1)$ percentage point. Referring to **Table 1** both *Number of Wins* and *Playoffs* both have the coefficient for *allstars* significant at the 99% confidence level, whereas *World Series* is only significant at the 90% confidence level. This could be because of the lack of observations for winning the *World Series*, compared to *Playoffs*. This regression's limitation assumes that each interval of adding one more all-star has the same amount of increase in wins or probability. The next two regressions address this limitation.

Now β_1 is expanded on by β_2 for regression (2); for all three, if the β_2 is significant then all-stars is a quadratic function. The *allstar^2* coefficient indicates that *Number of Wins* is a quadratic function and that as more all-stars are added to a team the increase of wins diminishes (**Table 2**). This coefficient is significant at the 99% confidence level. Next the *allstar^2* coefficient indicates that *Playoffs* follow similarly, however this is not as strong at a 90% confidence level. This means that having 19 all-stars (the most all-stars per team possible) will not increase probability of making the P*layoffs* or *Number of Wins* as much as the first few all-stars added to a team. Lastly, the *World Series allstar^2* coefficient is not significant indicating that *allstars* is not quadratic.

The third regression set (3) takes the previous two regressions and expands ever further to see if there is a specific interval that will increase wins, playoffs, or World Series percentage. The β_o coefficient is the *Number of Winnings* for a team per season with 1 all-star, all else held constant. For the second two equations, it is the expected percentage for making the *Playoffs* or winning the *World Series* with one all-star on the team, all other variables held constant. The β_1 coefficient for regression (3) is the expected increase or decrease in *Number of Wins* for 2-3 all-stars on the team compared to having one all-star on the team, all else held constant. Now for the *Playoff* and *World Series* dependent variables, the β_1 will be the expected percentage point increase of making the playoffs or winning the world series for having 2-3 all-stars compared to having one all-star, all else held constant. This follows similarly forβ_2,β_3, and β_4 for their respective intervals, in each equation.. Logically, 8-9 all-stars increased the *Number of Wins* and *Playoff* percentage the most (**Table 3**). The only coefficient that was significant for *World Series* was the 6-7 all-star range, which could indicate that most World Series wins came with 6-7 all-stars on a team

These results are limited, however, if a team can identify the all-star players before drafting or purchasing their contract. Also, a team is potentially limited by funds. Some all-stars are not as expensive because they are rookies, however once a player has stood out as an all-star and re-signed for a non-rookie contract that is where the players could get expensive. For example, 2017 all-star, Aaron Judge, makes about $544,500 per year but has only been playing in the MLB for one year. Alternatively, 2017 all-star Jose Altuve has been playing for six years, made the all-star team five times and his average salary is $3,125,000; thus putting a financial strain on a team to purchase Altuve's contract (Spotrac.com, 2017).

Also, based upon this study there are no observations of any teams having more than nine all-star players. A team could have up to potentially 19 all-stars on a team, however unlikely this would ever happen there are no data points to reference that occurrence in the third regression.

IV. Conclusion

Winning the World Series is supposed to be the ultimate glory for a team; however based upon this study, the correlation between *number of allstars* and *World Series* wins is not as great as the correlation between *allstars* and *Number of Wins* and *Playoffs*. Overall, it would make sense that better players help teams get wins and wins get teams to the playoffs and the playoffs are then the first step to the World Series. This study did identify that players with this upgraded status of "all-star" did have a significant correlation with teams doing well. The recommendation based upon this study would be to purchase 8-9 all-star player contracts if financial constraints were not a factor, and if it was known which players were going to be named to the upcoming year's team. More analysis would need to be done to determine if there was a way to identify these players with confidence for the upcoming year.

V. References

Mlodinow, L. (2009) The Drunkards Walk: How Randomness Rules Our Lives. New York, NY: Pantheon Books.

McNeal, S. (2007). They're so money. Sporting News, 231(21), 22-27.

Pietrusza, Silverman & Gershman (Eds.). (2000) Baseball: The Biographical Encycolpedia. Kingston, New York: Total Sports Illustrated.

Reiter, B. (2014). Astro-Matic BASEBALL HOUSTON'S GRAND EXPERIMENT. (Cover story). Sports Illustrated, 120(26), 30-38.

Schmidt, C. W. (2013). EXPLAINING THE BASEBALL REVOLUTION. Arizona State Law Journal, 45(4), 1471-1535.

Spotrac.com. (2017). Aaron Judge. [online] Available at: http://www.spotrac.com/mlb/new-york-yankees/aaron-judge-18499/ [Accessed 4 Dec. 2017].

VI. Appendix

Table 1

	Number of Wins	***Playoffs***	***World Series***
Intercept	-346.820(132.05)***	-2.281(6.77)	-0.635(3.25)
allstar	2.471(0.24)***	0.087(0.01)***	0.010(0.01)*
Constant			
ERA	-9.339(0.68)***	-0.257(0.03)***	-0.025(0.02)
FP	460.833(133.62)***	3.227(6.85)	0.716(3.29)
Salary	-4.16E-08(0.00)***	-7.85E-10(0.00)	1.56E-10(0.00)
Attendance	4.82E-06(0.00)***	1.39E-07(0.00)***	1.24E-08(0.00)

* Significant at the 90% confidence level
** Significant at the 95% confidence level
*** Significant at the 99% confidence level

Table 2

	Number of Wins	***Playoffs***	***World Series***
Intercept	-310.876(132.32)**	-2.245(6.83)	-0.302(3.28)
allstar	4.101(0.73)***	0.088(0.04)**	0.025(0.02)
allstar^2	-0.227(0.10)**	-0.0002(0.01)	-0.002(0.00)
Constant			
ERA	-9.171(0.68)***	-0.257(0.04)***	-0.023(0.02)
FP	421.710(134.02)***	3.187(6.91)	0.353(3.32)
Salary	-3.92E-08(0.00)***	-7.83E-10(0.00)	1.79E-10(0.00)
Attendance	4.64E-06(0.00)***	1.39E-07(0.00)***	1.07E-08(0.00)

* Significant at the 90% confidence level
** Significant at the 95% confidence level
*** Significant at the 99% confidence level

Table 3

	Number of Wins	***Playoffs***	***World Series***
Intercept	-317.025(135.40)**	-3.604(6.94)	-0.954(3.30)
2-3 allstars	5.256(0.79)***	0.108(0.04)***	0.016(0.02)
4-5 allstars	7.392(1.12)***	0.202(0.06)***	0.024(0.03)
6-7 allstars	13.084(1.75)***	0.469(0.09)***	0.093(0.04)**
8-9 allstars	16.954(3.14)***	0.624(0.16)***	-0.055(0.08)
Constant			
ERA	-9.379(0.69)***	-0.267(0.04)***	-0.023(0.02)
FP	432.153(137.06)***	4.690(7.03)	1.037(3.34)
Salary	-3.94E-08(0.00)***	-7.85E-10(0.00)	1.72E-10(0.00)
Attendance	4.90E-06(0.00)***	1.53E-07(0.00)***	1.50E-08(0.00)

* Significant at the 90% confidence level

** Significant at the 95% confidence level

*** Significant at the 99% confidence level