The Factors to effect MLB Position Player’s Salary

Background

In recent years, MLB players’ salaries are getting higher and higher. For example, this year, Max Scherzer’s salary reaches to 43.33 million dollars. He has the highest salary in MLB this year, also is the highest record in MLB history. The average of player’s salary in 2021 was 4.17 million dollars. The top five salary in 2021 were 43.3 million, 36 million, 35.5 million, 35.1 million, and 35 million. There is huge difference between them. This made me want to know more and make research on the relationship between player’s salary and performance. Also, I want to find out what is the most important performance data to effect on player’s salary.

Also, I want to prove if the players have better performance will have higher pay.

Object

The objects for this research are the position players who have higher than 10 million dollars salary. The reason I only collected the players who have higher than 10 million salary is because I want to see what kind of player can get high salary. Also, there are a lot of minimum wage players not consistent to play in the major, sometime will play in minor league for long time until the manger to call them up when the team needs them for major. Because of this, their performance data will nor be complete or not enough to analysis. For those high wage players, I didn’t collect their data one year before they sign the contract with team. Because some of the players got hurt or some reason, they didn’t play a lot of games in big league, I decided to collect the data of how good or bad they would perform in 162 games (a season). Basically, to use their career performance data to divide by how many games they play and time 162, so that can get the average performance data per 162 games. In this research, the main way I used to analysis the relationship between dependent and independent variables is regression model by excel. There are 67 players’ data in my research.

Variable

In this research I used a lot of performance data. I will explain each of the variables in this part. Avgsalary, which represents player’s average salary of contract. Years, means how many years of contract that the player sign with team. Signedage, means the players’ age when they sign the contract with team. Totsea, means how many seasons player has played in the major, for this variable will not count the season they play in the minor league. RBI, represents run batted in, basically it say how many runners that the player send them back to home for a season. WAR, it is a advance data or a combination data. It measures a player’s value in all facets of the game by deciphering how many more wins he is worth than a replacement-level player at his position. It combines a player’s defense, offense, and base performance, this data can evaluate what is the level of the player at. SLG, it is slugging percentage which represents the total number of bases a player records per at-bat. If the player’s SLG is high means he has a good ability to hit extra base hit. SO%, it is strike out percentage, to see how many at-bat the player will get one strike out. HR is homerun. BA is batting average. TB is total bases that player can produce, with a high TB means either the player hit a lot of extra base hit or has a lot of hits. R is run, which is to calculate how many times that the player run back to home and score for the team.

Explanation of Variables & Restricted and Unrestricted model

In my research, I used a lot of variables, because I want to find out which has more correlation with player’s salary. I used two regression model. One is unrestricted model and the other one is restricted model. The purpose of doing this is to see what variables are statistically significant and statistically insignificant. I will take off the variables which have low correlation with salary. I used two regression models. One is unrestricted model includes which includes more variables than restricted model. Year, signedage, totsea, HR, RBI, WAR, BA, H, TB, R, SLG, and SO% are put into unrestricted model. Year, signedage, RBI, and BA are statistically significant, because their t-stat are greater than critical value which is 1.674. For this model, the R square is .681. The restricted model only includes year, signedage, totsea, HR, RBI, WAR, SO%, and TB. I took off some variables from the unrestricted model, due to their t-stat are too small. In restricted model, the statistically significant variables are the same with unrestricted model.

To decide which model I will use, I made the decision depend on F-test. F-test is .42 and critical value is 2.09. Because of F-test is less than critical value, it is failed to reject and means “jointly insignificant”, so it is better to use the restricted model.

Interpret the result

Avgsalary = -28.353 + 1.493year + .867signedage - .459totsea - .164HR + .18 RBI + 1.011WAR

(.249) (.275) (.298) (.171) (.095) (.474)

- .099SO% + .018TB

(.158) (.035)

This is the final equation I used for this research. The variables with yellow colored are statistically significant. There are four variables which are statistically significant, I will interpret them. Year, with the coefficient of 1.493, which means 1.493 million dolor more for each additional year of the contract. The player has long year contract usually have good performance so that the team is willing to use long contract to sign this player and expect to use his contribution to help the team to win the game. Signedage with the coefficient of 0.87 which represents that 0.87 million dollar for each additional player’s age. It seems like not making sense, because usually athlete’s age has high correlation with the performance, higher age usually has worse performance. But in my research, to become a free agent player in MLB must play in major at least 6 years. After becoming a free agent player, player will have a higher chance of earning bigger contract. I believe this is the main reason to make this variable positive but not negative. Next is RBI, 1.8 million dollars for each 10 additional RBIs. The player who has high RBI usually is a slugger. The winner for a baseball game is whoever score more, so if a player can help the team score more, it is the most directly way to help the team win the game. WAR’s coefficient is 1.01, which means 1.01 million dollars for each additional WAR. The player who has higher WAR means he is a higher-level player. 0-1 WAR is scrub player, 6+ WAR is MVP level player.

Also, I made a table to compare the player’s performance between higher pay and lower pay. I put the average salary and some performance data of player into the table, I split them into two group. One group is players who have more than 20 million salary and the other group is players who have more than 10 million salaries. The purpose of doing this is to see if the player who has higher performance will have higher pay. The result supports this assumption.

Conclusion

-The result of the research shows that the player who has higher RBI, longer contract years, higher WAR, and higher age will have higher salary. The player who has high RBI in a year usually is a good hitter. That kind of player usually is the best hitter in their team, because they can help their team to score more in the game so that the team can win the game. The player who signed a long-year contract usually have good performance on the field, so the team is willing to sign with the player for long contract. Because depending on that player’s previous seasons stats, the team believes this player will have good performance in the future. The player who has higher age will also have higher pay. Because Most of the high-pay players have become a free agent (must played at least 6 years in Major), after becoming a free agent, they will have a chance to earn a bigger contract. War can show a player’s combination performance, including offensive, defensive, and base running. The higher WAR the player has means they have more contribution to their team. Overall, the player who has good performance usually will get higher salary, but the player who has higher salary not always have really good performance.

Reference

<https://www.baseball-reference.com/> (Player performance data)

<https://www.spotrac.com/> (Player contract information)