Relationship Between MLB Payroll and Performance

Is it worth it to spend more money on your players?

DATA/IDS 490-B Sports Analytics

Abstract

Major League Baseball (MLB) is a 10 billion-dollar industry with billions of dollars going to players each year. The best players receive the most money. There is a preconceived notion that more money translates to more wins, and therefore more championships. However, there have been an increasing number of individuals who believe that all 30 teams have a chance to win their respective games regardless of the amount of money spent on players. The objective of this research was to explore the relationship between team payroll and team wins. Independent t-test and regression analyses were conducted using data for the 1995-2019 time period. The results show that teams with the top 10 highest payrolls had a better chance of winning the world series than teams with the lowest payrolls. This finding supports the claim that payroll is a predictor of success, but the causal factors are yet to be explored and may be a topic for future research. The focus of this research was professional baseball, so future research may also be extended to explore the implications of compensating college athletes.

Literature Review

In a comprehensive study examining winning percentage and payroll, it was found that during the regular season there was a positive and statistically significant relationship between the two variables across all four major league sports, which was not found during the postseason. The dependent variable (payroll) was transformed by normalizing it as a z-value on a normal curve. Only the NBA and NHL continued this relationship during the postseason (Shorin, 2017).

In a similar study, teams winning percentage and payroll were examined for a sixteen-year period. Payroll was transformed by dividing a team's payroll by the average league payroll. The transformed variable was used in a simple linear regression model. Regression equations were built for each team; less than half were statistically significant. The authors divided teams into groups depending on above/below average payrolls and winning records and used the groupings to conduct another regression analysis. Strong evidence was found to show that payroll was a significant positive indicator of performance (Hasan, 2011).

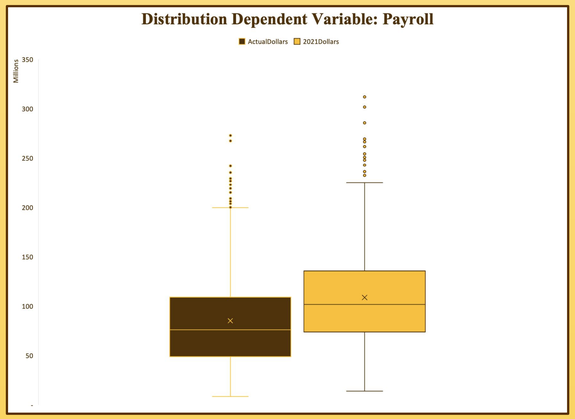
Another researcher looked at whether teams pay to win or win to pay. Graph theoretic methods were used to identify dependence relationships and the direction of those dependencies among winning and payroll. After establishing the dependencies between variables, the researchers used panel regression and impulse responses from system estimation to measure the magnitude of the relationships. It was shown that spending more on payroll helps temporarily to boost team wins, but this model is not sustainable. A team cannot continuously win more by spending more (Stimel, 2011).

Our research was different from prior research, because we looked at payroll in its raw form, using a log transformation and in terms of 2021 dollars. Although prior research used regression, there were differences in the variables used. We also used t-tests to compare means between groups.

Methods

Our primary objective was to look at the relationship between team payroll and performance. To measure performance, we gathered data on team wins, losses, runs scored, runs allowed, and whether a team made the playoffs and/or World Series. Once the data was cleaned, we calculated each team's win percentage and whether or not they had a higher than average win percentage. We studied the period from 1995 to 2019, purposely excluding COVID years, yielding 744 observations for approximately 30 major league teams.

We used visualizations and descriptive statistics to explore variables and their relationships. A series of independent two-tailed t-tests were used to compare the payroll of those who won more than the average number of games versus those who won less than average, to compare the payroll of those who made the playoffs versus those who did not make the playoffs, and to compare the payroll of those who won the World Series versus those who did not win the World Series. Lastly, multiple linear regression using backward elimination was performed to find statistically significant predictors of team payroll.



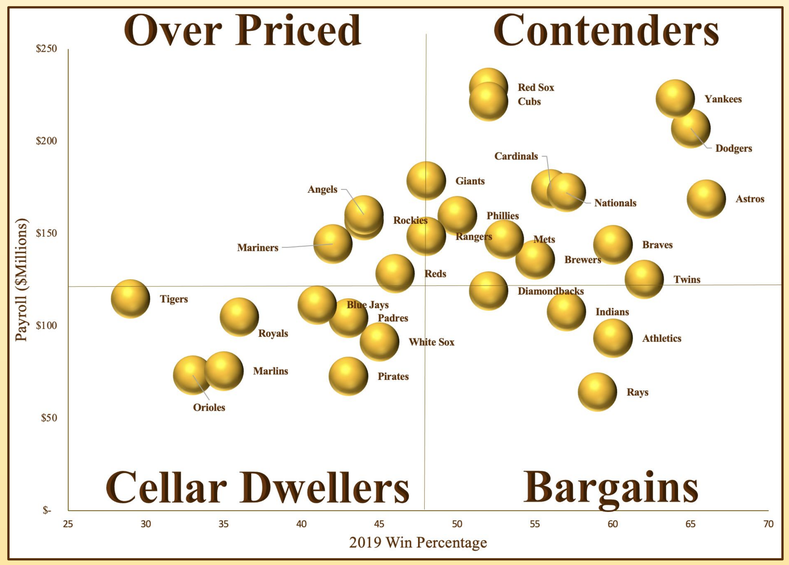


Figure 1

Results

Descriptive statistics is provided in Table 1 for payroll data in actual US dollars. Since a 25-year time period was used, payroll values were also converted to 2021 dollars using the consumer price index (CPI). After running a linear regression model and a log-linear regression model on both the standard US dollar and the Consumer Price Index (CPI) dollar, we concluded that increased fit in the log-linear model was not sufficient to justify the added level of complexity (see Table 2). Our analysis, therefore, focused on actual US dollar as the dependent variable.

In Figure 1, we see that payroll is positively correlated with winning percentages. Teams have been classified into quadrants based on their respective payrolls and winning percentages. Teams with higher payrolls and higher winning percentages are considered "Contenders" teams. These teams have a greater chance of going to the playoffs or world series. Other teams are classified accordingly: "Over Priced" paid more for less wins, "Cellar Dwellers" paid the right amount for less wins, and "Bargains" paid less for more wins. The positive relationship between these two variables is one of correlation and not causation; increasing payroll alone does not guarantee an increase in win percentage.

Our findings showed that there was a statistically significant difference between the payrolls of teams in the playoffs and those who were not ($p = 0.003$). We also found a statistically significant difference between the payrolls of teams in the World Series and those who were not ($p < 0.001$). To evaluate statistical significance based on the number of seasonal wins, teams were placed into two groups: teams at or above average and those below average. Our findings also showed that there was a statistically significant difference between the payrolls of above and below average teams based on their win records ($p < 0.001$).

After running a multiple regression test and using backward elimination, the statistically significant predictors for the payroll of a team are date, number of losses, and number of runs scored. The best predictive model for payroll consists of the date (D) and the number of losses (L) is $\hat{y}$ = -1064373.5L + 4663781.5D - 9189563629. The R value of 0.7405 indicates a strong positive correlation, while the adjusted R^2 of 0.5471 indicates that 57.41% of the variation in payroll is explained by the date and number of losses.

We built an alternative model by selecting variables based on their correlation. This approach resulted in predictors of year (Y) and wins (W). We also looked at log transformations for payroll to address its skewness and used payroll values all expressed in 2021 values. The model with the highest explanatory value (R^2 = 60%) used a log transformation with the payroll values: $\hat{y}$ = 0.059Y + 0.014W - 101. The impact on payroll for each win is: $e^{0.014}$ - 1 = 1.4%. In other words, a team will need to increase its payroll by 1.4% for each additional win.

Conclusion

Increasing the number of wins will cost a team approximately $1 million in payroll (or a 1% increase) to get better players. Alternatively, an additional loss will save the team approximately $1 million by spending less on players. Spending more money improves likelihood of making it to the playoffs or the World Series.

Table 1

Payroll			
Mean	**Minimum**	**Maximum**	**Skewness**
$85,241,537	$8,317,000	$272,789,040	0.9948

Table 2

Correlation Table Models: Actual US Dollars vs 2021 Dollars (CPI)						
Correlation	*Payroll*	*LNPay*	*2021Dollars*	*LN2021*	*Year*	*Wins*
Year	0.691874	0.720382836	0.532191	0.55059685	1	
Wins	0.293909	0.319770213	0.3532934	0.36848881	0.0498	1
R^2	0.546	0.600	0.390	0.420		

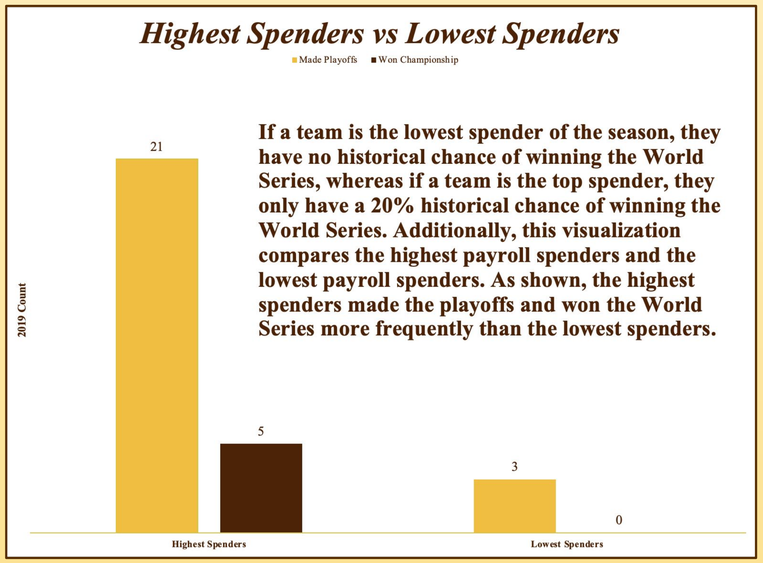


Figure 2

Limitations/Future Research

Our biggest limitation was the time we had to complete our research (i.e. one semester). Also, backward elimination regression, a methodology we used, does not consider all variable combinations; once a variable is removed, it is not re-entered. Our research also was limited to the variables that we provided to generate the "best" model. More combinations and variables should be explored. For example, lagging performance with payroll is a reasonable candidate to add to the list of variables. Tony La Russa, however, questioned this thinking saying that in the past the current year's performance determined the following year's salary, but today players focus on individual performance in place of winning games for the team. He also acknowledged that World Series bonuses did little to motivate players already earning eight figure salaries. These revelations give reason to further explore the causal relationships between performance and salary.

Future research would also benefit from employing other techniques such as clustering. The two-by-two visual analysis used herein to "cluster" teams could be compared to clusters obtained by statistical methodologies such as k-means clustering.

Lastly, the focus of this research was professional baseball, so future research may also be extended to explore the implications of compensating college athletes.

References

Bissinger, Buzz (2005). 3 Nights in August: Strategy, Heartbreak, and Joy Inside the Mind of a Manager. Houghton Mifflin Harcourt.

Hasan, S. (2011). Can Money Buy Success?: A Study of Team Payroll and Performance in the MLB. *Journal of Global Business Management*.

Shorin, G. (2017). Team Payroll Versus Performance in Professional Sports: Is Increased Spending Associated with Greater Success? *Honors Thesis*. Durham, NC: Duke University.

Stimel, D. S. (2011). Dependence Relationships between On Field Performance, Wins, and Payroll in Major League Baseball. *Journal of Quantitative Analysis in Sports, 7*(2).

Faculty Advisor: Dr. Coleen Wilder

Jonathan Bledsoe, Alex Conlin, Grace Edwards, Madison Gawlinski, Mark Lorenz