Matt Butts
Miranda Edwards

Payroll and Wins in Major League Baseball

Research Questions

1. Is money a deciding factor in a baseball team's success?

By analyzing the ratio of number of wins of a baseball team and the money spent by that team, we found a medium-strong correlation between these two factors across the league.

2. How has the implementation of a 'luxury tax' affected this relationship?

The 2003 implementation of a 'luxury tax' was supposed to fundamentally change the impact of a team's financial influence on its own wins. We found that the luxury tax actually increased the influence, but decreased the effectiveness of that influence.

Motivation and Background

Popular professional sports leagues such as the NFL or NBA are well-known for having highly paid athletes, even to the point of controversy. What is less well-known is that the officials of these leagues attempt to limit the salaries of pro athletes through the implementation of various types of 'salary caps.' These can vary in leniency between each league and sport. The NBA, for example, has a 'soft' cap, allowing some exceptions, whereas leagues such as the NFL or NHL use 'hard' salary caps, which allow no exceptions. Major League Baseball is unique in that, until recently, it had no such salary cap. Teams were (and even now are) essentially limited only by their own financial resources.

The fact that baseball has no real salary cap often leads to debate as to whether every team has the same fair odds of winning. It appears that the more money a team spends, the more successful it is. This was often cited as an issue in the 1990s. For instance, the two contenders for 1999 World Series were the New York Yankees and the Atlanta Braves. At the time, the two teams had the 1st and the 3rd highest payrolls in the country, respectively.

The differences between team payrolls across the country are large enough to potentially have a substantial influence on game outcomes. The New York Yankees have the highest payroll, which is four times the size of the lowest, of the San Diego Padres. Not only that, but the Yankees have 27 World Series titles while the Padres have zero. This simple observation makes it appear that money is a significant factor in a team's chances of winning.

However, in 2003, a luxury tax, or a 'competitive balance tax', was levied on the league. Essentially, there is a 'maximum' payroll set so that a given team whose overall player payroll exceeds this threshold must pay a fine equal to a percentage of the payroll. This is intended to discourage the use of salary to guarantee a team's wins. However, for a team with unlimited resources, this may not be a set-back at all. The only consequence would be to pay a little more.

The total salary of all the players of a team does not necessitate that the team is spending it as effectively as the next team. The total number of the players on the team, as well as the differences in salary per player, can cause the total payroll to remain constant while still affecting the outcome of the team. By separating our calculations into categories of the average salary per player and the overall payroll, we can determine if how the team weights its payroll across the team affects its wins during the season.

Baseball is the only national professional sport where the competing teams have such discrepancies in their comparative payrolls. The correlation between payroll and success becomes an important question because baseball is a national sport and pass-time for Americans, and has billions invested in it as an organization.

Dataset

The dataset was extracted from a set of playing statistics found in the Lahman Baseball Database. This database is a publically available collection of statistics for major league baseball, and can be found at http://seanlahman.com/baseball-archive/statistics/. The data we will analyze is found in the 2012 version folder, in the files titled "Teams.csv" and "Salaries.csv." Here, we will collect the number of wins for each team, and the team salaries per player, respectively.

Methodology

We specifically looked for a pattern between a baseball team's payroll spending and its success throughout the league in the years 1983-2012. The results were graphed in a series of scatterplots, three per decade. By dividing the data into decades, we could easily track the changes over a longer period of time while compensating for inflation. The win-ratio for a team was compared with the total player salary, the average salary per player on the team, and the ratio of the team's payroll over the league average (the payroll percentage of the league).

The success of the team was measured by the ratio of wins to total games played in the regular season. The number of wins for each team will plotted dependently against the overall payroll spending of that team. We calculated the win ratio as opposed to the integral number of wins to account for the varying number of games played each season. We looked for a correlation between the two data sets by drawing a best-fit line to the graph, and from that, calculating the coefficient of determination (r). The coefficient of determination allows a quantitative measurement of how well the line fits the scatter plot. This process was repeated when comparing the average salary per player per team and the team's payroll percentage of the league.

We also considered including in our comparison other variables. However, variables such as batting or pitching averages can't necessarily point towards a team's success against other teams.

The resultant values of r lead us to conclude that over the last 30 years, money has played an increasingly important role in the success of a baseball team. Not only that, but by computing r for the years immediately preceding and following the implementation of the 'competitive balance tax', we have found that the tax did not truly affect the relationship between a team's payroll and success. The equation of the line of best-fit for each set of data draws attention to the rate of change of money spent versus the team's success.

Results

According to our graphs, as time passed from 1983 to 2012, the correlation coefficient *r* grew more significantly, for all three variables graphed versus success (the total player salary, the average salary per player on the team, and the payroll percentage of the league). The *r*-values corresponding to the 1980s indicated almost no relationship. However, in the two subsequent decades, the correlation appeared stronger. The most important factor here seems to be the team's total payroll in comparison to the rest of the league, as the coefficient of correlation was strongest here.

To answer our second question, we looked at the slope of the best-fit lines. In the 1980s, there was almost no change (and also no correlation) between a team's payroll and a team's success. However, the change in success with respect to payroll spiked in the 1990s, but dropped once again in the 2000s, presumably because of the luxury tax. This rate of change can be noted below.
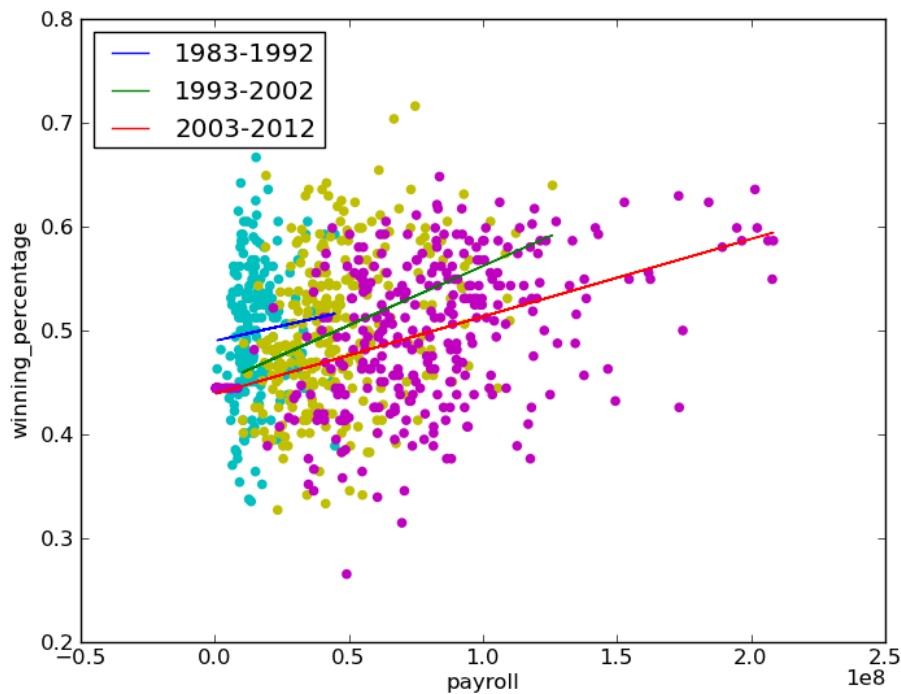
However, the correlation actually appeared to increase in the 2000s, seemingly contrary to the intent of the luxury tax. We therefore concluded that while the tax appears to have lessened the effect of winning based off of money, it also increased the likelihood of being able to 'buy' a win for a team.

Reproducing Our Results

The program can be run from the command line by simply calling it. This assumes that the user has downloaded and stored the two necessary files into the working directory "Teams.csv" and "Salaries.csv."

When the file is run, the equations for the lines of best fit will be printed out, along with the Pearson coefficient for each line shown. The final graph, graphing the data from 2002 and 2004, best shows the direct impact of the luxury tax.

Collaboration

We did not collaborate with anyone else on this assignment.

Reflection

The results were definitely a bit different from what we expected. The results of effects of the luxury tax, especially, were surprising because the changes were immediate (2002 vs. 2004).

Overall, this assignment was cool to work from start to finish. It was definitely difficult to figure out how to best analyze the data. The tough thing was finishing on time, even without procrastinating.