Sean Cooke
Jeffrey Hrebenach

The Effect of High Salaries on MLB Players

The Problem

Major League Baseball (MLB) is a very turbulent industry that puts a huge amount of pressure on its players to perform. The success of teams in the league is not only astronomically important to the players and coaches, but also to the millions of fans around the world. However, a player's performance is not entirely about fitness and skill, the mental state of each player plays a massive role in such a technical game. We will examine the effect that salary has on player performance in MLB and whether there are certain types of players that respond better to salary increases than others.

Related Work

Michael Lewis' Moneyball emphasizes that a player's on-base percentage and slugging percentage are excellent statistics for measuring batting performance, and that there are undervalued batters who possess high values in these metrics with a low salary.[6] We will only consider Baseball-Reference.com's wins against replacement (bWAR) to measure productivity and will not consider on-base percentage and slugging percentage.[5] bWAR is an all-encompassing statistic that attempts to quantify the amount of wins a player adds to their team over an average replacement player at their position for that year.[7]

Moneyball emphasizes that successful teams can be assembled with inexpensive players relative to the free-agent market We will consider the success of MLB players relative to themselves rather than assembling a successful team with undervalued players. We will run a comparative study that compares a batter's bWAR when their salary is in the upper 10th percentile versus their bWAR when their salary is in the lower 90th percentile. This analysis will help us determine how salary affects a MLB player's performance.

Sean Cooke
Jeffrey Hrebenach

Data Acquisition

The Lahman Database and Baseball Reference's bWAR databases were used as data sources for our analysis.[3][5] These databases provide us with a comprehensive set of attributes, including batting statistics, pitching statistics, bWAR statistics, salary information and demographic data. These databases also provide us with the ample time frame of 1985-2014, which will be sufficient to allow us to make strong conclusions. The master table of the Lahman Database holds demographic data for all players, the salaries table holds salaries by year for players and the batting, pitching, b_war and p_war tables hold batting and pitching statistics for players by year between 1985 and 2014.

Performance Evaluation

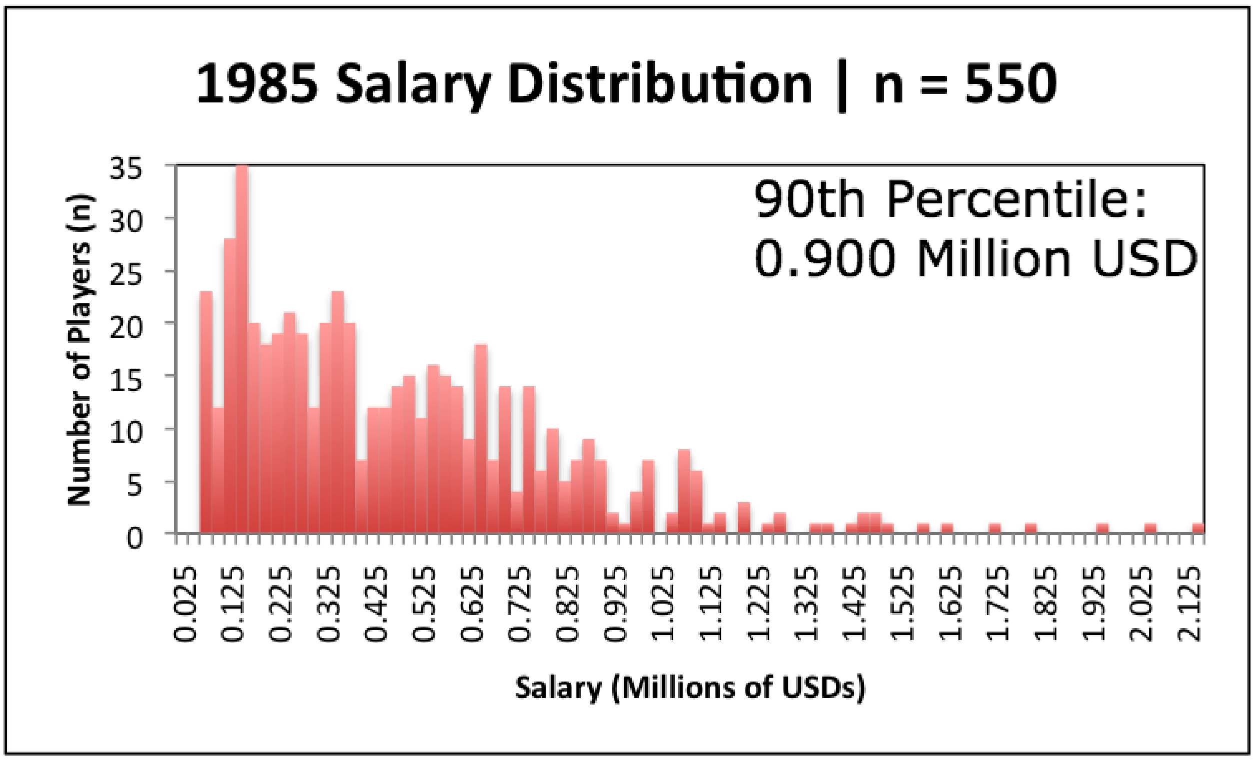

Figure 1a

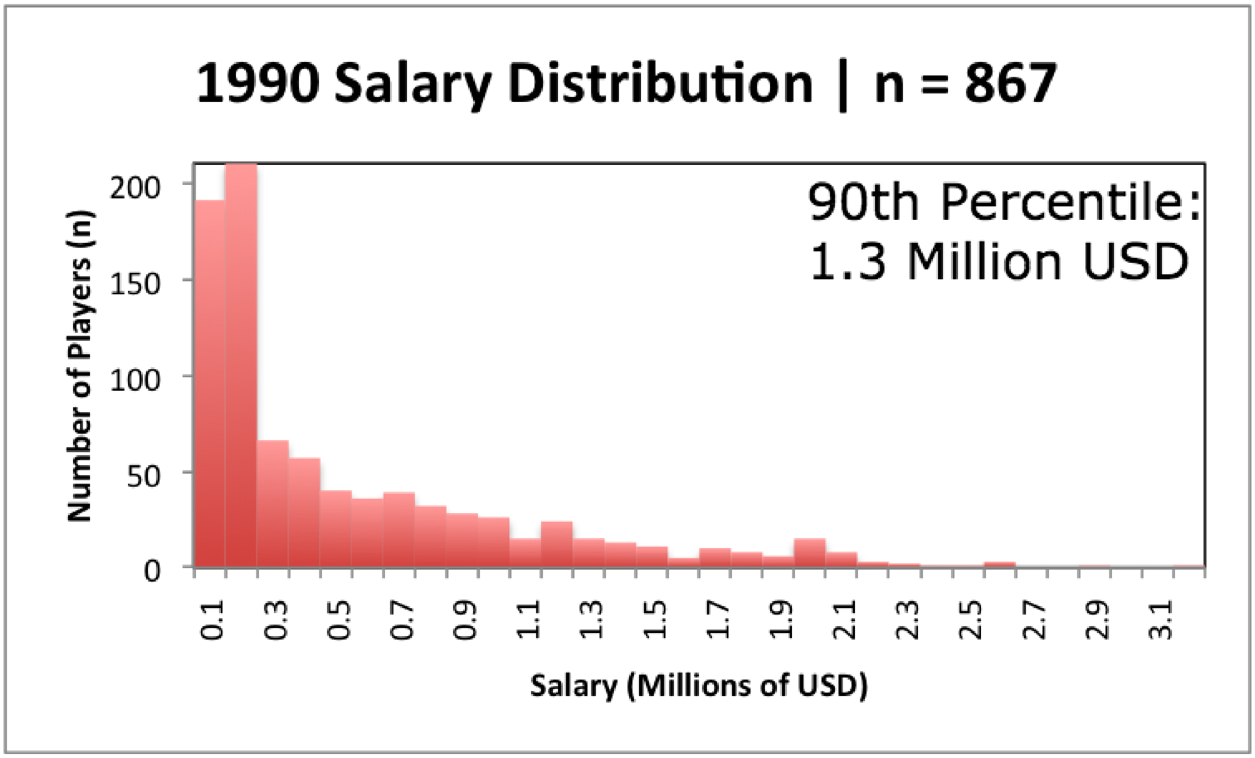

Figure 1b

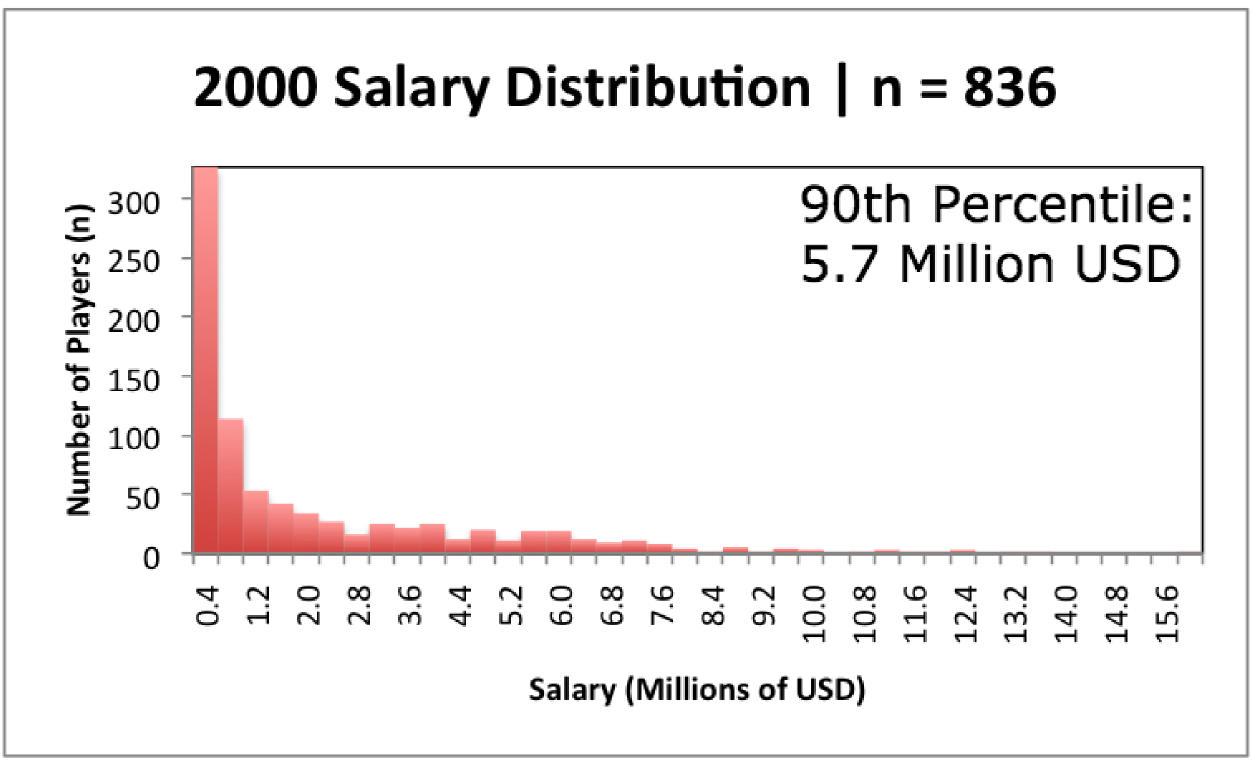

Figure 1c

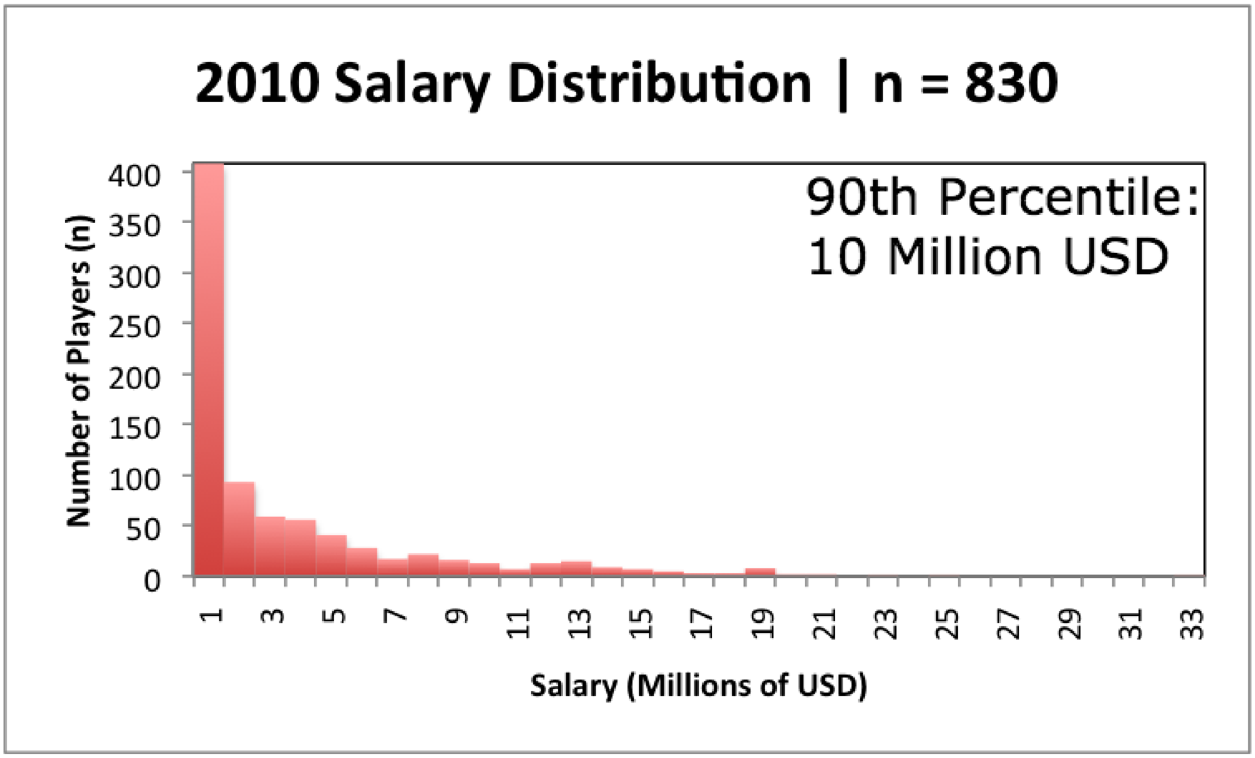

Figure 1d

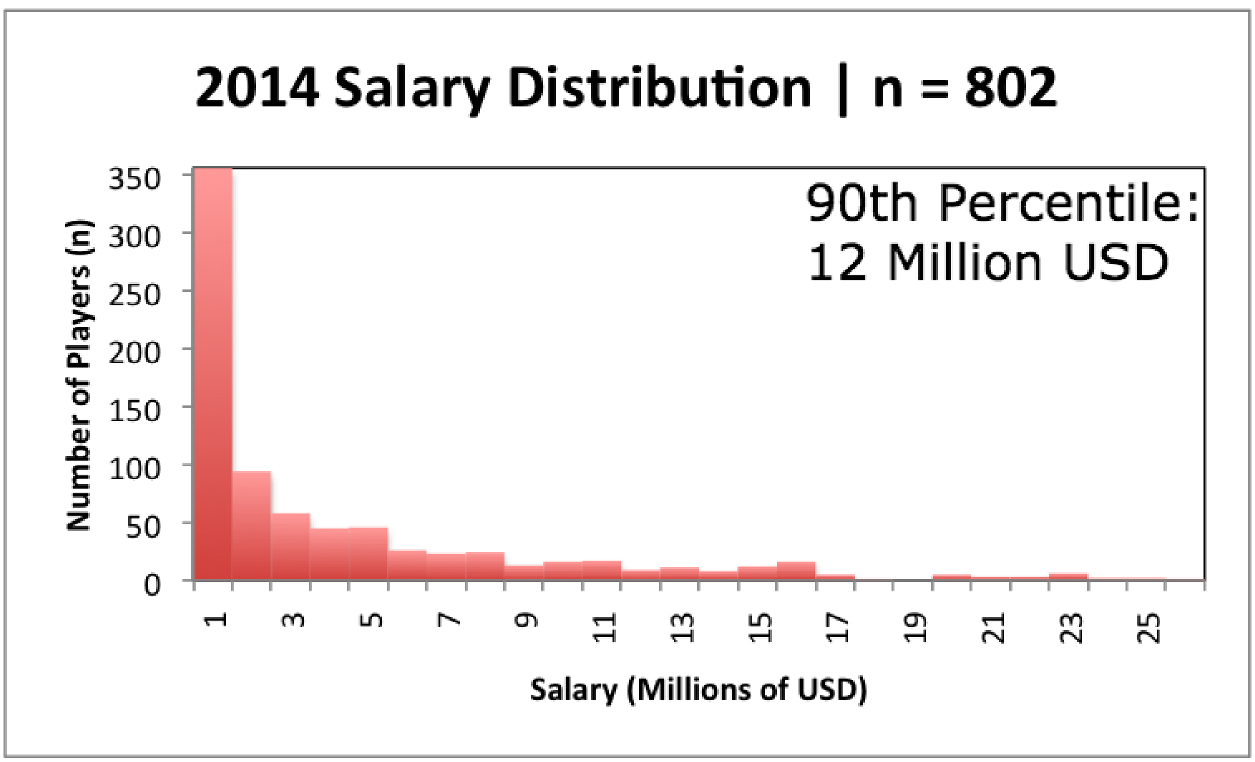

Figure 1e

As seen in figures 1a, 1b, 1c, 1d, and 1e, the concept of a high salary in MLB has changed significantly between 1985 and 2014. Between 1985 and 2014, the salary distributions of MLB players became increasingly positively skewed. This means that over time, there has been an increased percentage of players making salaries relatively close to MLB's minimum, with a smaller amount of players making salaries significantly larger than MLB's average.

For these reasons, we will only consider players who, at one point in their careers, were making 'big money'. We define 'big money' as a salary equal to or above the 90th percentile for a year. By grouping salary into the resilient bins 'top 10th percentile' and 'bottom 90th percentile' we discretize the continuous variable 'salary'. We will evaluate the change in a player's performance from when they are making a salary in the top 10th percentile to when they are making a salary in the bottom 90th percentile. This will allow us to run a comparative study to determine how salary affects a player's performance on an individual basis.

To measure a MLB player's performance, we will first classify MLB players as either batters or pitchers. Second, we will run a clustering algorithm to cluster batters and pitchers based on their statistics. Batters will be clustered into the subgroups 'power hitters' and 'contact hitters' and pitchers will be clustered into the subgroups 'power pitchers' and 'control pitchers'. Using statistics derived from the Lahman Database, we will be able to successfully cluster players from both these positions into their appropriate subgroups. Before running the clustering algorithm, we will normalize our clustering statistics to the range [0, 1]. This will ensure that all used statistics will have equal weight in determining our clusters. All clusters of MLB players (power hitters, contact hitters, power pitchers and control pitchers) can earn a high salary.[4][8]

To measure an MLB player's performance we will use the statistic bWAR. bWAR attempts to quantify the number of wins that a player will provide their team over the average replacement player at their position for that year. According to Baseball-Reference.com 'most sabermetricians agree that comparing players to a general replacement level is the best approach to valuing players'.[5] In economic terms, this approach is known as evaluating the 'opportunity cost' of playing a particular athlete. bWAR can be both positive and negative; a negative bWAR indicates a player performed worse than the average replacement player should have performed over the course of the year.[7] Although bWAR is calculated differently for batters and pitchers, the general goal to quantify the number of wins a player contributes over a replacement player remains the same.

Algorithm Idea

We will take a clustering approach in our study. K-means clustering with two distinct clusters will be used to label batters as either power hitters or contact hitters. Power hitters have a high career home run per at bat rate and a high career strikeout per at bat rate. Contact hitters have a low career home run per at bat rate and a low career strikeout per at bat rate.[2]

K-means clustering with two distinct clusters will also be used to label pitchers as either power pitchers or control pitchers. Power pitchers have a high career strikeout per batter faced rate and a high career walk per batter faced rate. Control pitchers have a low career strikeout per batter faced rate and a low career walk per batter faced rate.[8]

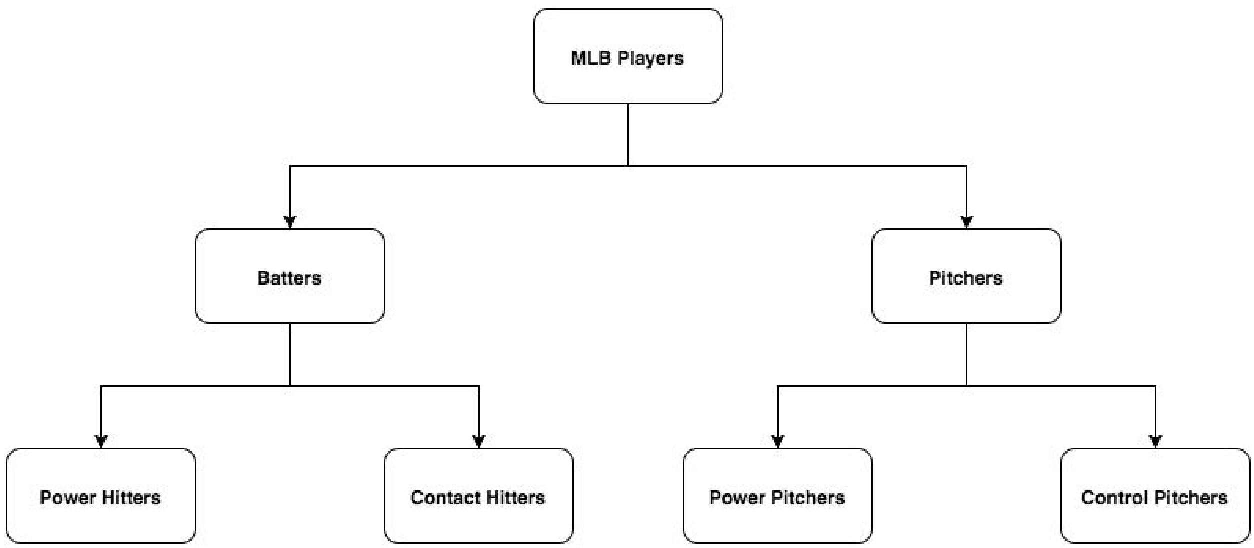

Figure 2

Experiment Design

We will create two clusters of hitters: power hitters and contact hitters. Similarly, we will create two clusters of pitchers: power pitchers and control pitchers. All players considered for this study were in the top 10th percentile of earners at one point in their career, and were in the bottom 90th percentile at another point. This will allow us to run a comparative study; only comparing players to themselves which eliminates any skill bias in our analysis.

To draw conclusions about our findings, we will compare a player's average bWAR from when their salary was in the top 10th percentile to what their average bWAR was when their salary was in the bottom 90th percentile. We will refer to the difference between these two averages as 'WARdifference'. For each given player, 'WARdifference' will either be:

1. Positive - Player performed better when they were in the top 10th percentile of salaries.
2. Negative - Player performed better when they were in the bottom 90th percentile of salaries.
3. Equal to 0 - Salary change did not affect performance.

By taking the average WARdifference across all players throughout our seven groups (MLB players, batters, pitchers, power hitters, contact hitters, power pitchers and control pitchers), we will be able to draw both absolute and relative conclusions about how each group responded to salary change.

Expectation

When our study has been completed, we will have a better idea of how salary affects performance among MLB players. Our hypothesis is that an increase in salary will negatively affect a player's performance. We make this prediction because we believe that increasing a player's salary will diminish their incentive to perform, since earning that high salary is often the goal they were working hard for in the first place. Once a player achieves their target salary, it is likely that they will be more complacent with their skill level and their ambition will decrease. We are expecting to find that the process of setting a baseball player's salary is truly complex, and effectively setting a player's salary relies on finding the balance where players are getting paid enough to be incentivized to perform well, but not well enough to adversely affect their psychological state.

Results

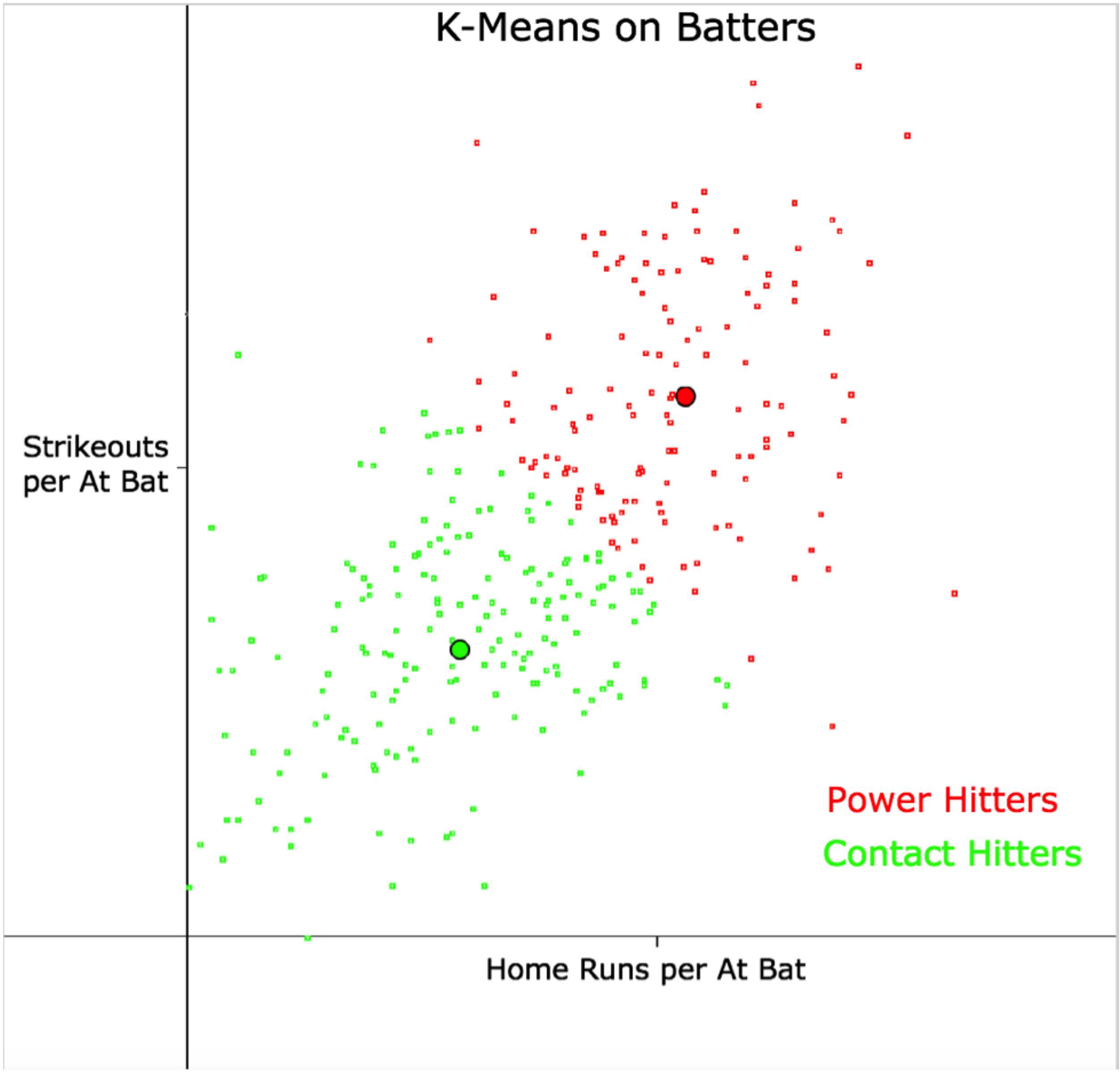

Figure 3a

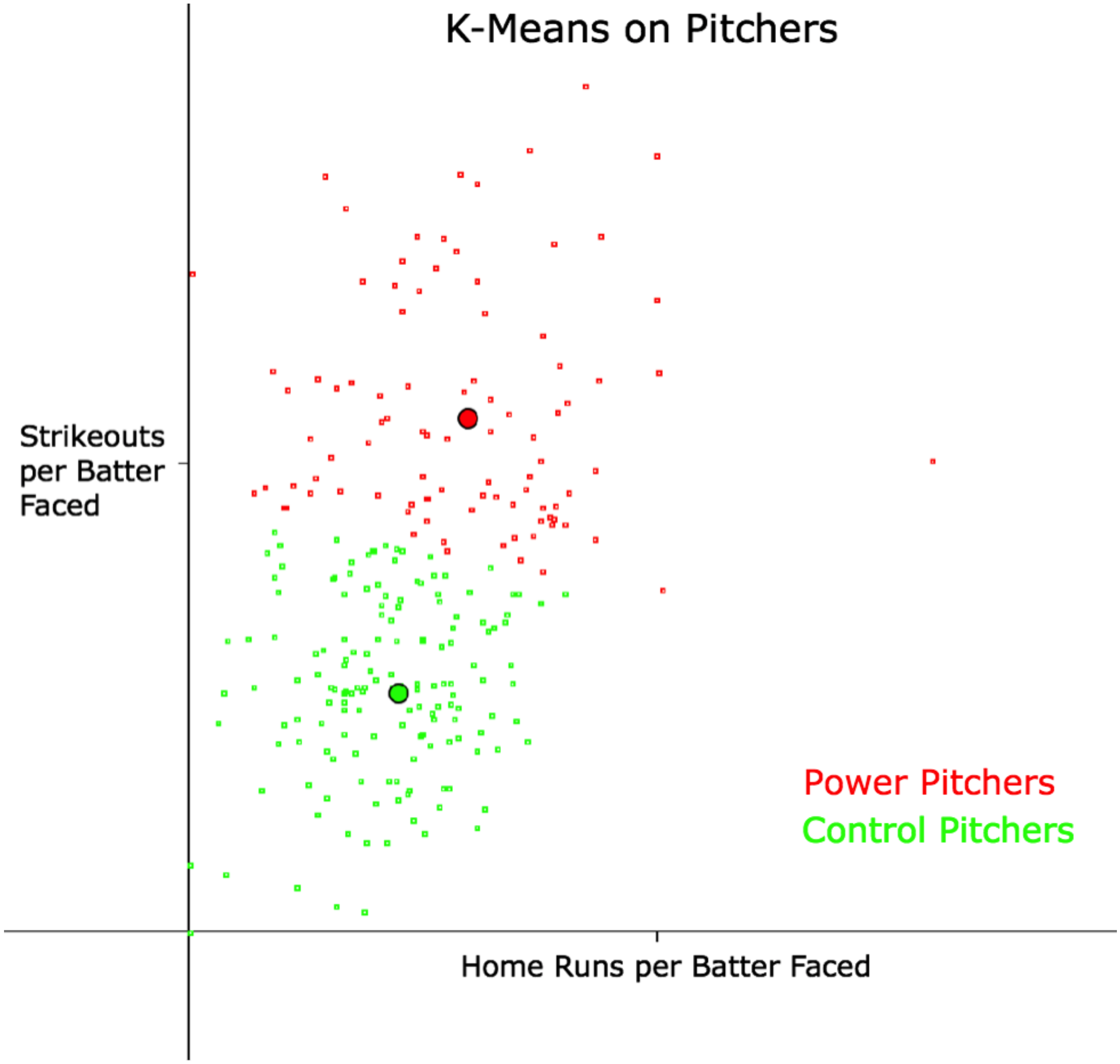
K-Means on Pitchers
Strikeouts per Batter Faced
Home Runs per Batter Faced
Power Pitchers
Control Pitchers

Figure 3b

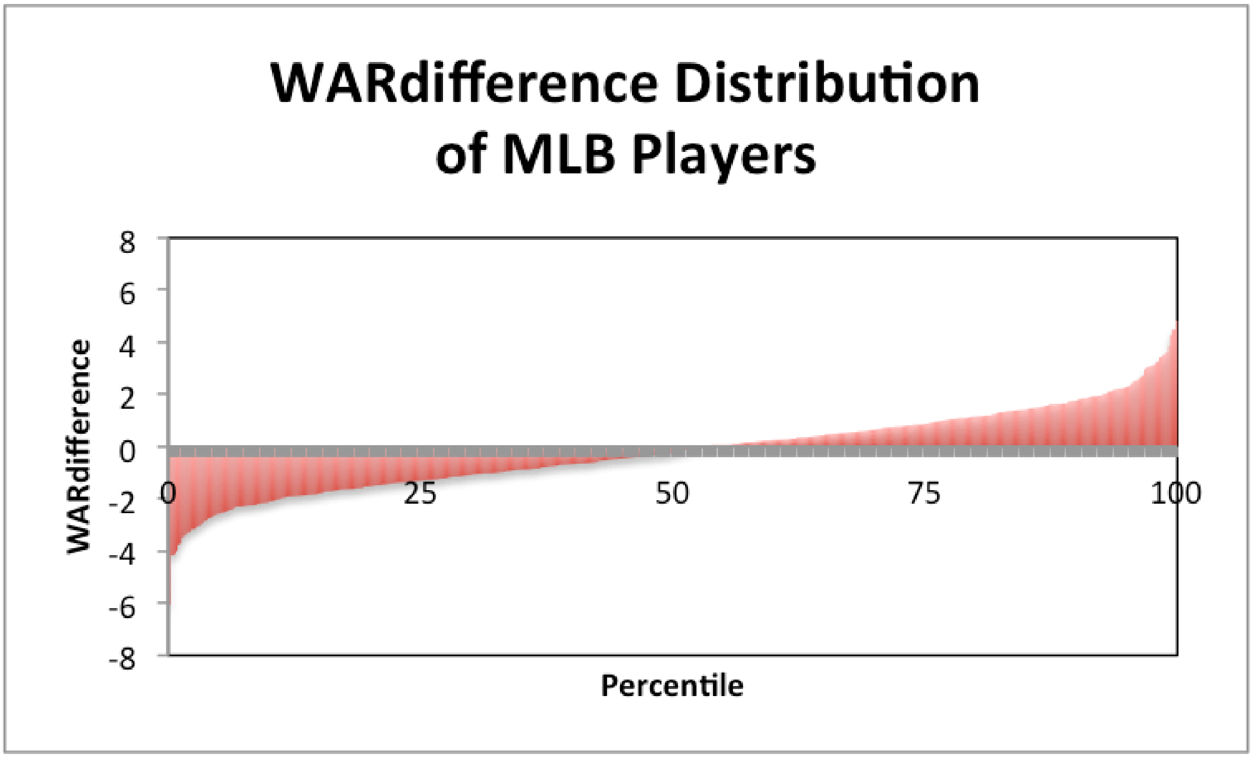

Figure 4a

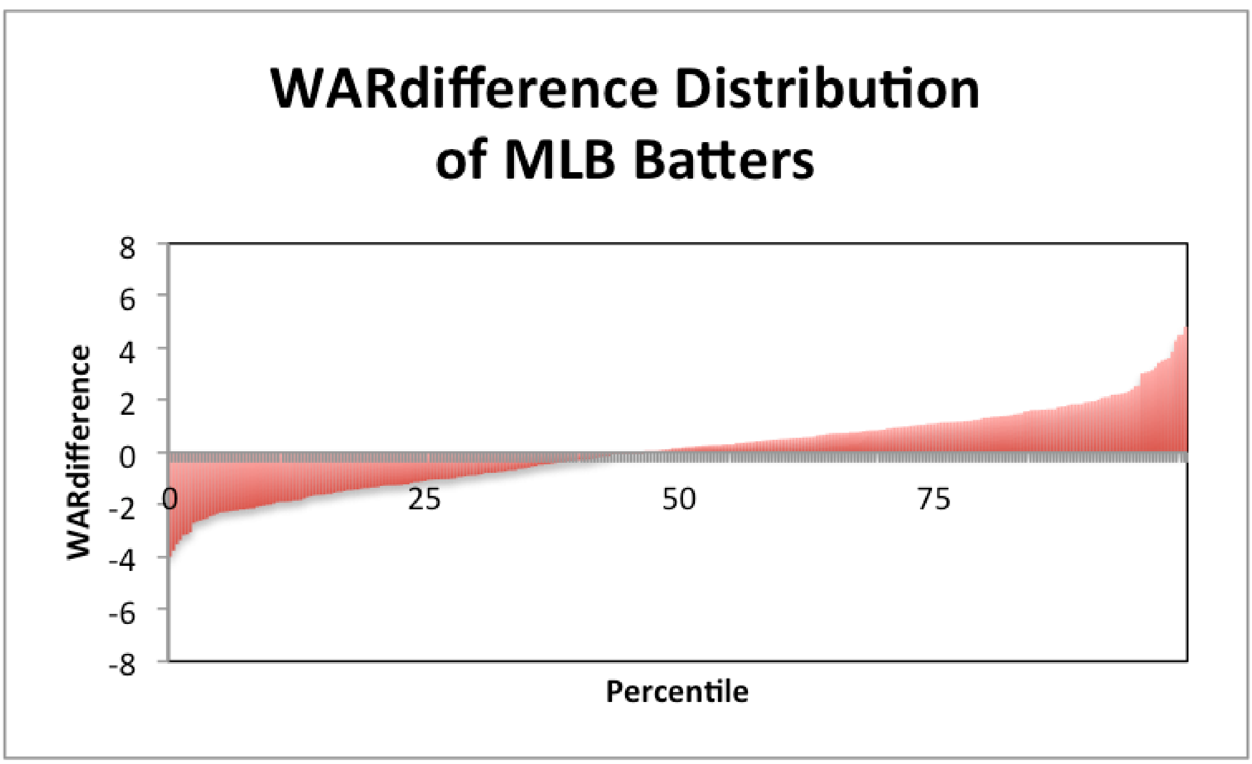

Figure 4b

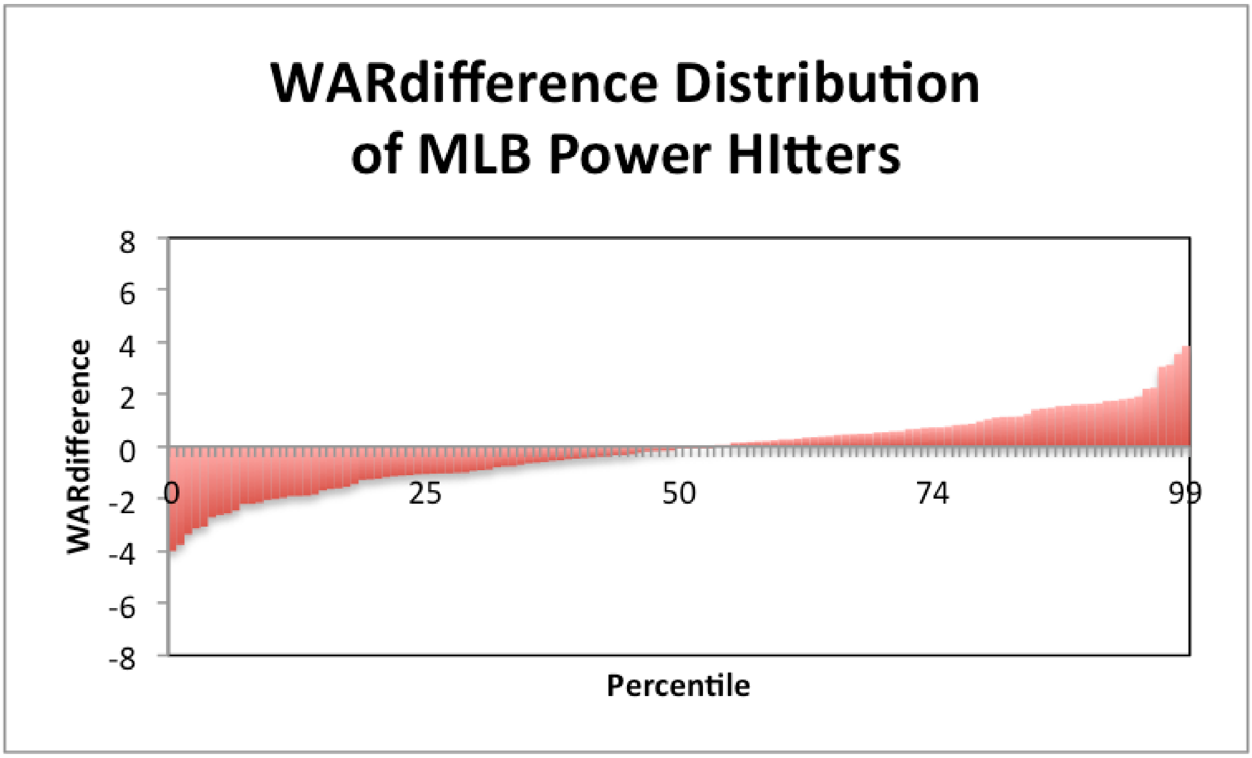

Figure 4c

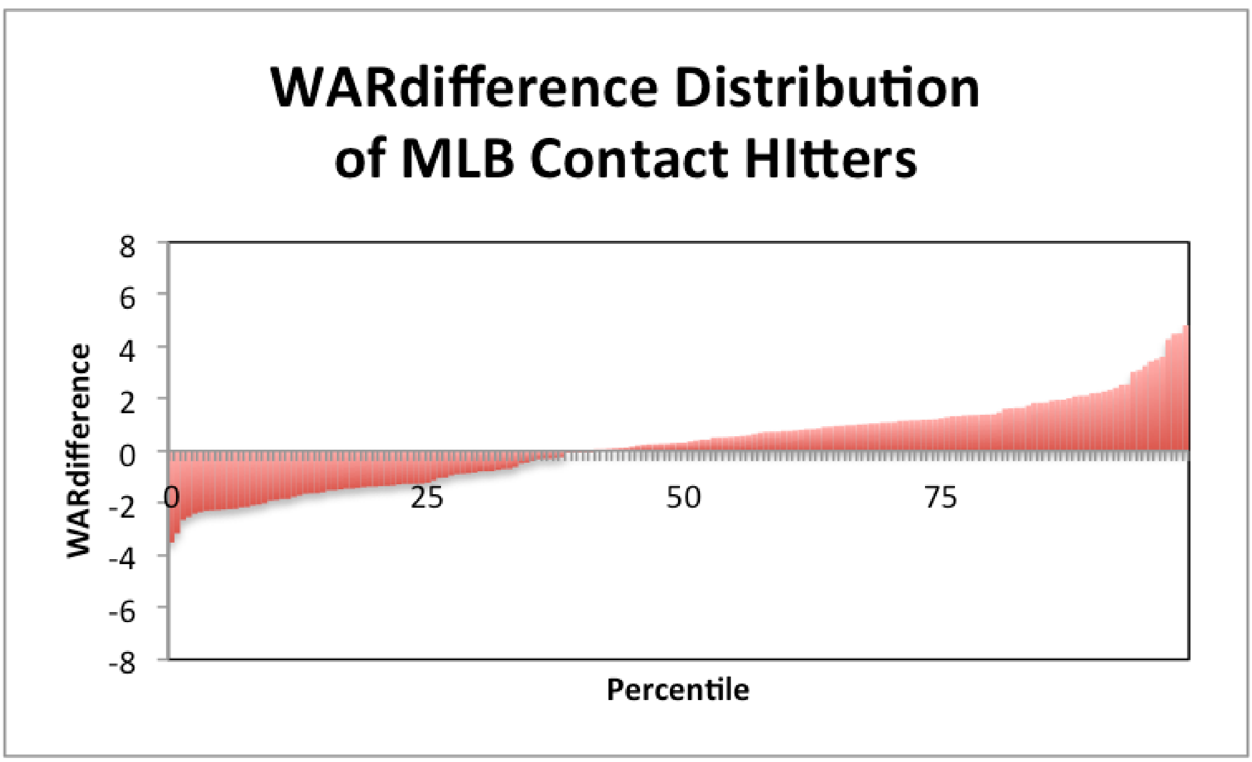

Figure 4d

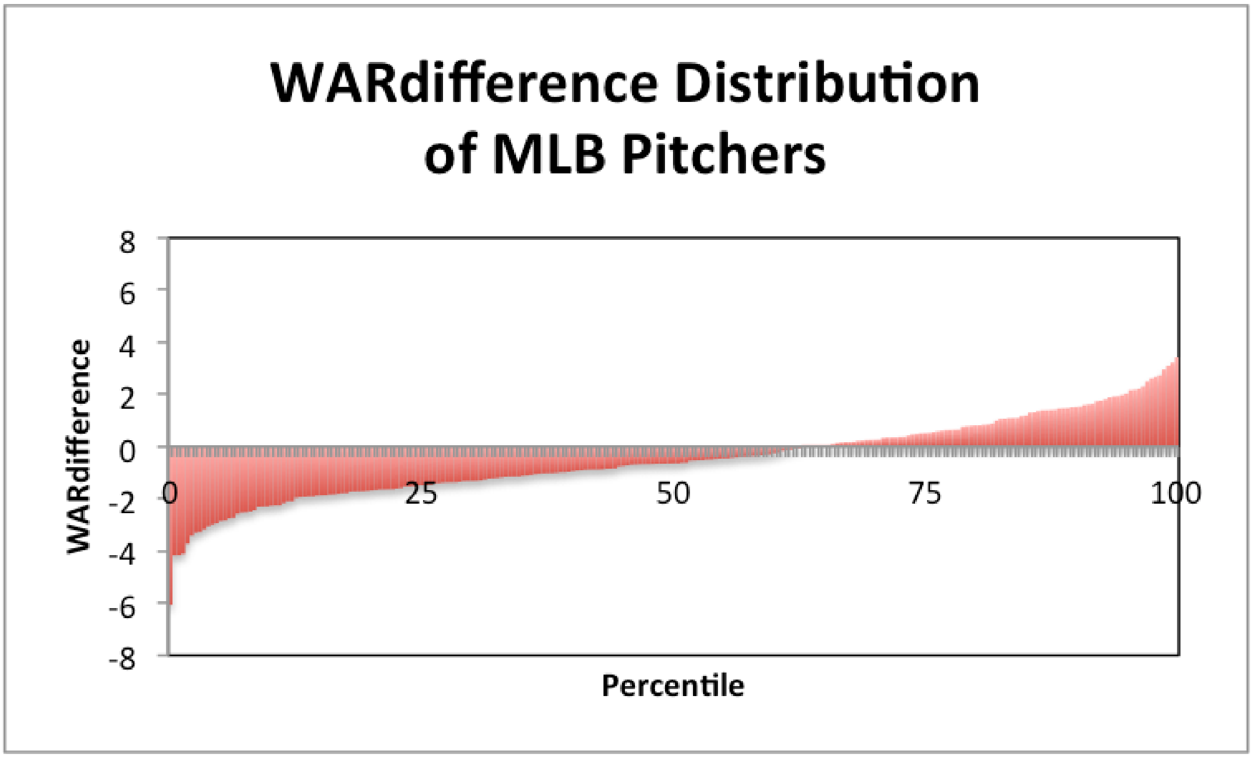

Figure 4e

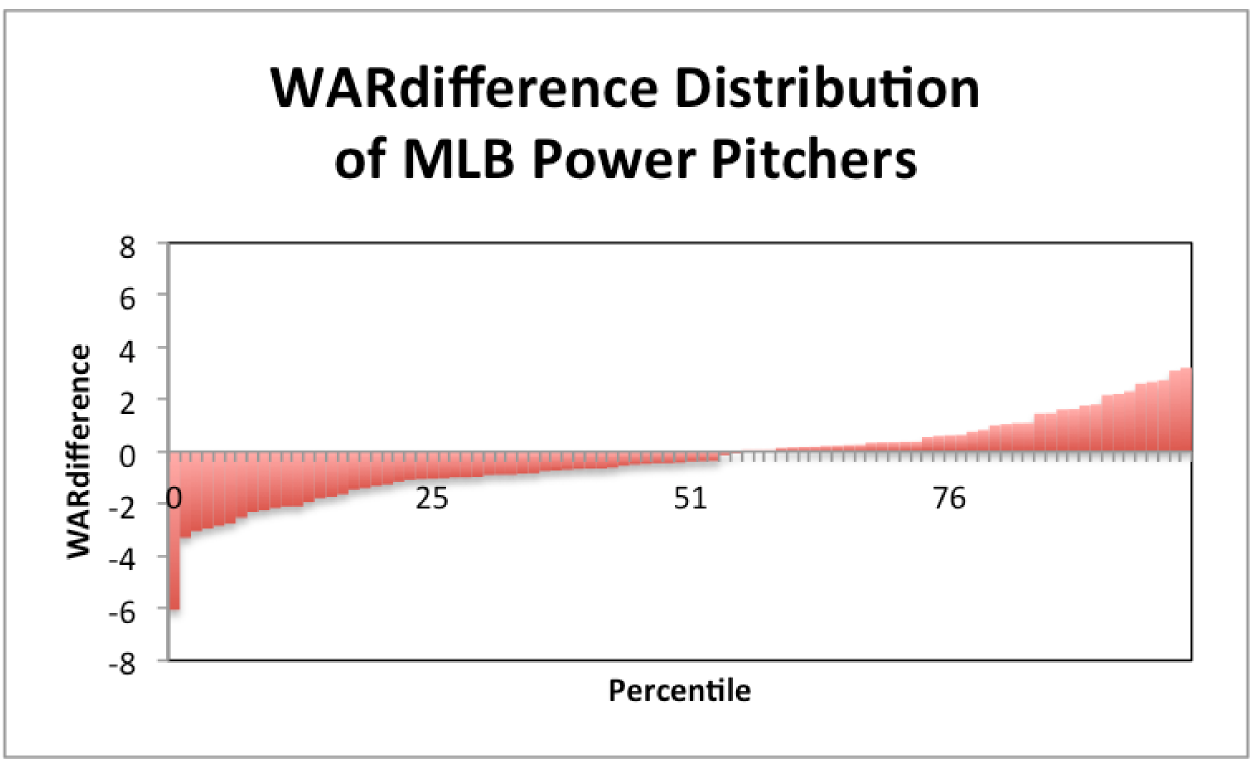

Figure 4f

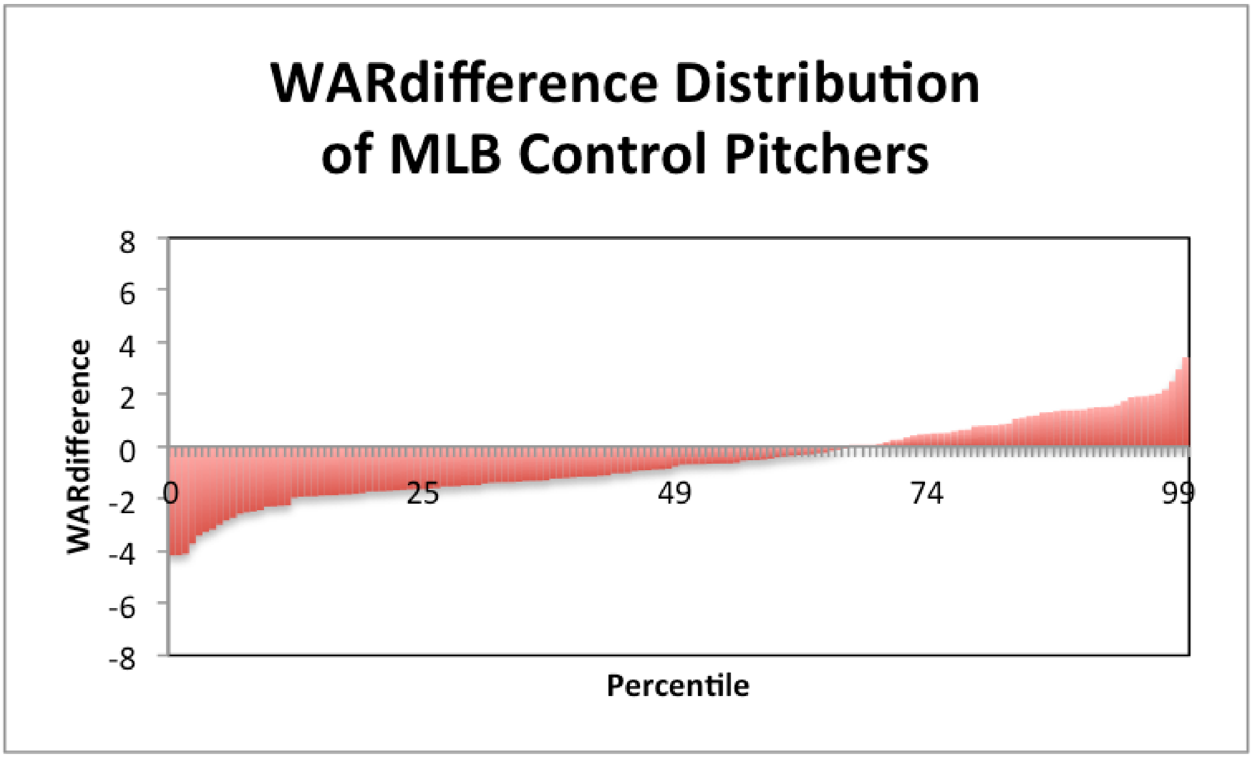

Figure 4g

	Number of Players	Percentage of Players Who Had a Decrease in WARdifference	Average WARdifference	Mean Age	Mean Age bottom 90th Percentile	Mean Age top 10th Percentile
MLB Players	545	53.76	-0.181	30.64	29.61	32.40
Batters	304	46.71	0.058	30.68	29.64	32.28
Power Hitters	129	53.49	-0.157	30.09	28.48	32.17
Contact Hitters	175	41.71	0.216	31.13	30.42	32.37
Pitchers	241	62.66	-0.483	30.58	29.58	32.59
Power Pitchers	91	56.04	-0.283	30.14	28.67	32.69
Control Pitchers	150	66.67	-0.605	30.85	30.10	32.52

Figure 5

Conclusions

Giving a high salary to a player in MLB will not make them a significantly better player. We can conclude this because across MLB Players, the percentage of players who had a decrease in WARdifference was greater than 50%. Through our research, we have concluded that there are two major reasons as to why this is the case. First, higher salaries come with lower incentives to work hard for a pay raise, which is detrimental to performance. Second, players receiving a salary in the top 10th percentile have often peaked in performance and are starting to decline is skill.

Another conclusion we make is different player types respond to salary changes differently. The cluster where the largest percentage of MLB players experienced a positive WARdifference was contact hitters (58.29%). The cluster where the lowest percentage of MLB players experience a positive WARdifference was control pitchers (33.33%). Due to the volume of players examined in our analysis, this is a very substantial difference. This implies that certain player types have either: 1) a significantly higher range in skill, 2) a significantly larger impact on the outcome of a game, 3) substantially different responses to large salaries. Building on this idea, we can conclude that certain player clusters should be awarded very high salaries while other player clusters should be kept at more modest salaries. In particular, from our results we can conclude that the best type of player to give 'big money' to is a contact hitter. Additionally, the worst type of player to give 'big money' to is a control pitcher.

One may assume that players in the top 10th percentile of salary earners are significantly older than players in the bottom 90th percentile of salary earners. On average in MLB, a player making 'big money' is 2.51 years older than he was when he was not making 'big money'. Since the average age of players in our analysis pool is 30.64 with a standard deviation of 4.81 years, this is a relatively small age difference. Although this difference may impact our results in a slight manner, it is not a large enough difference to significantly confound our findings.

Due to lack of precision in the bWAR statistic, we cannot make strong conclusions from WARdifferences that are close to 0. Although many agree that bWAR is the strongest comparison measure available today to evaluate players, it is still an estimate and will never be perfectly precise.[5] Due to the inherent inaccuracy that is ubiquitous throughout athletic performance measures, we can only conclude that giving a big contract to a player in MLB will not make them perform significantly better. Building on this, we can conclude that players in MLB are paid based on their past performance rather than on their expected future performance. This is an interesting find as, due to free-agency, players who sign 'big money' contracts often sign contracts with a new team.

References

1. Katie Stankiewicz. "Length of Contracts and the Effect on Performance of MLB Players" *The Park Place Economist*. Volume XVII. pp. 76-83.

2. "**K% and BB%**" *FanGraphs Sabermetrics Library*. Web. 6 Nov. 2015.

3. Lahman, Sean. "**Download Lahman's Baseball Database**" *http://www.seanlahman.com*. Web. 6 Nov. 2015.

4 "**Types of Hitters**" *Wikipedia*. Wikimedia Foundation. Web. 6 Nov. 2015.

5. "**Baseball-Reference.com WAR Explained | Baseball-Reference.com.**" *Baseball-Reference.com*. Web. 30 Nov. 2015.

6. Lewis, Michael. *Moneyball: The Art of Winning an Unfair Game*. New York: W.W. Norton, 2003. Print.

7. "**What Is WAR?**" *FanGraphs Sabermetrics Library*. Web. 17 Nov. 2015.

8. "**Control Pitcher.**" *Wikipedia*. Wikimedia Foundation. Web. 24 Nov. 2015.