Optimizing MLB Player Value with a Performance and Financial Efficiency Rating

Ben Borovinsky
May 1, 2020

Abstract

Major League Baseball (MLB) has been a pioneer in the sports analytics realm. Baseball research has produced sophisticated performance metrics that measure player value, such as Wins Above Replacement (WAR). Recent blockbuster trades continue to break salary records while the inclusion of the Competitive Balance Tax (CBT) keeps MLB organizations cognizant of their payroll. The Performance and Financial Efficiency Rating (PFER) combines the performance and financial value for every player and assigns them a standardized value. This trade-value metric is used in a model that determines the most talented and financially efficient trade options at struggling positions for every team. MLB roster and performance data as of May 1, 2020 are provided courtesy of Baseball-Reference.com, and salary information is provided courtesy of Spotrac. All analysis is done in the statistical computing language R.

1 Introduction

1.1 Evolution of Baseball

In 1869, the Cincinnati Base Ball Club was the first team to become professional by paying its players [1]. The team traveled the country to play other opponents and showcase the thrill of professional baseball. After going undefeated that season and well into the next season, many people were exposed to the sport of baseball.

Baseball fever struck the United States as more teams were formed to compete in the national sensation. Leagues were created to keep rules consistent and eventually Major League Baseball (MLB) was established as the most popular sport in America [2]. Iconic players like Babe Ruth and Ted Williams helped their teams build a strong fanbase and become worldwide brands like the New York Yankees and the Boston Red Sox, respectively. Dynasties brought more fans to the sport, including a Yankees dynasty that lasted roughly three decades and fielded some of the most iconic players to play the game. Everyone knew who these players were on the dominant 1927 and 1939 teams, regardless if they were fans of the sport or not [3]. In fact, many modern dynasties in other sports are compared to the 1939 Yankees, as they are regarded the greatest team of all-time [4].

As the decades passed, some players felt trapped to their teams due to the MLB's controversial reserve clause. One player who challenged the MLB was Curt Flood, a centerfielder for the St. Louis Cardinals in the 1950s and 1960s [5]. He believed many freedoms were taken away from team contracts and thus suffered unfair pay: "A well-paid slave is nonetheless a slave" [6]. Flood went to court against the MLB arguing for a system where players could have the option to play for other teams, as this would alleviate any depressed wages [7]. While the Supreme Court sided with Major League Baseball, his movement on the matter caused other baseball players to follow suit in a collective bargaining effort to demand free agency in lieu of unfair contracts. In 1976, the Major League Baseball Players Association (MLBPA) got the free agency it wanted, with salaries increasing along with profits [8].

1.2 Growth of Baseball Statistics

Baseball is considered by many to be a numbers game, and this is largely attributed to its early adoption of statistics as a means of storytelling. Henry Chadwick created the first box score in a newspaper in 1859, which was able to tell the story of the game in a way that took up minimal real estate [9]. The box score brought fans to the sport who were unable to see the games because they felt they could understand the details of the game through this simple innovation in their newspapers.

Statistics had been kept for most games throughout much of baseball's long history. However, analytics had not significantly impacted the game until a couple major events occurred in the 1970s. In August 1971, L. Robert Davids and 15 other baseball researchers created the Society for American Baseball Research (SABR), a membership-based organization that engages in baseball research and history archival more colloquially known as sabermetrics [10]. Several years later, a former boiler-room attendant named Bill James began writing about baseball from a data and analytical perspective [11]. He then published his writings in 1977 officially called the *1977 Baseball Abstract: Featuring 18 Categories of Statistical Information That You Just Can't Find Anywhere*

Else and continued with updated versions of the *Baseball Abstract* annually through the 1980s as he rose to become a prominent figure in the baseball research community [12].

The creation of SABR and the work of Bill James both have established and reinforced a community for baseball analysis. Baseball researchers and statisticians continue to investigate the game with hopes of discovering something new. One significant innovation these researchers have tried to accomplish is a single metric to signify player value. If used correctly by general managers, this metric can have profound impacts on their organization.

In 2001, the Oakland Athletics had the lowest payroll in baseball and were losing its three best players to trades [12]. General manager Billy Beane could not simply acquire the next best player due to the financial limitations of a small-market organization, so he tried to find value in players that other teams were not looking for. After studying and analyzing the trends at the time, he concluded that getting players on base was conducive to scoring runs and winning games. Beane therefore rebuilt his team around players with a high on-base percentage (OBP), leading his team to the playoffs and an MLB-record 20-game win streak the next season [12]. Baseball has evolved since, but this sparked a revolution for a greater focus on data analytics in the sport.

2 Problem Statement

2.1 Salary Inflation

Within the past decade, baseball organizations have been more willing to pay record-breaking salaries for top players. Before the 2019 baseball season, San Diego Padres third baseman Manny Machado and Philadelphia Phillies rightfielder Bryce Harper each collected record contracts worth $300 million and $330 million, respectively [13]. One year later, starting pitcher Gerrit Cole signed a $324 million contract with the New York Yankees, the most ever for a pitcher [14]. Some players can even extend their contract with their current organization; centerfielder Mike Trout extended his contract with the Los Angeles Angels of Anaheim for $430 million [15].

The best players are getting paid more than ever, but they still have the same shelf life. While some players still perform well under their new lucrative contracts, other players may begin their decline earlier than expected. One organization that has been victim to the latter situation on multiple recent occasions is the Los Angeles Angels of Anaheim. Within a few years, the Angels offered contracts that failed to net a positive gain for the organization. From outfielders Vernon Wells and Josh Hamilton to first baseman Albert Pujols, the team seldom produced more wins with them, and the two outfielders were off the team within two years [16].

The demand for premier talent should be relatively constant as roster and lineup constraints do not change often. However, increased salaries for athletes is a trend across all major sports. In basketball, teams are assigning worth to players who may not deserve it; in other words, they are overpaying for value that is not returned [17]. Organizations in professional leagues are making risky financial decisions without adequate payout, and there needs to be a solution to account for this disparity in value and worth.

2.2 Competitive Balance Tax

The 30 Major League Clubs and the MLBPA have a Collective Bargaining Agreement (CBA) that covers the layout for employment, financial structure, and rules of the game [18]. Unlike other

sports, the MLB does not have a salary cap; instead, there is a tiered penalty system for going over a certain threshold based on the average annual value (AAV) of salaries on the 40-man roster [19]. Organizations that surpass this threshold must pay the league a luxury tax, or Competitive Balance Tax (CBT). The CBT rate that an organization must pay is determined by Table 1.

Table 1: A breakdown of the CBT rates for first-, second-, and third-time offenders, according to the 2017-2021 CBA.

Tax Rate	Excess Payroll	1st Offense	2nd Offense	3rd+ Offense
Base	≤$20M	20%	30%	50%
Base + 1st Surcharge	$20M-$40M	32%	42%	62%
Base + 2nd Surcharge	>$40M	62.5%	75%	95%

A team that exceeds the CBT threshold for the first, second, and third consecutive time is considered a First-Time CBT Payor, Second-Time CBT Payor, and Third-Time CBT Payor, respectively [19]. The rates after the third consecutive time remain the same. If a team falls below the CBT threshold the next season, the penalties reset. For example, a team that goes $50 million over the CBT threshold and is a First-Time CBT Payor must pay 20% on the first $20 million, 32% on the second $20 million, and 62.5% on the remaining $10 million according to Table 1. This forces a traditionally high-spending organization to be cognizant of their spending in relation to their financial situation with the CBT.

2.3 Incorporating Analytics in Player Valuation

The best teams in sports tend to be top-heavy with talent; Major League Baseball is no exception [20]. It is typical for big-market organizations to consistently have competitive teams every season. Teams like the Yankees have a strong fanbase, history, and ownership that result in financial and competitive longevity. Meanwhile, small-market organizations generally have regional fanbases and a much more limited amount of capital that make deep playoff runs more rare. Due to this lack of financial freedom, small-market organizations must get creative with how they invest in their talent.

Bill James popularized baseball analytics as a form of research and Billy Beane was the first to take action primarily motivated by analytics. MLB organizations have now taken analytics more seriously as a way to improve their teams, but it is often hard to determine which statistics apply to overall team improvement and which do not. Using the wrong statistic can cause trades to have potentially lopsided outcomes. The most infamous trade in the history of Major League Baseball was when the Boston Red Sox sold Babe Ruth to the New York Yankees [21]. Not only did Ruth become the greatest player to play the game, but the Red Sox went 86 years without winning another World Series due to the "Curse of the Bambino" [22, 23]. With the rise in baseball analytics, teams have used this particular trade as a cautionary tale to avoid making an embarrassing and possibly catastrophic decision.

One popular metric that aims to determine player value is Wins Above Replacement (WAR). The statistic attempts to calculate "how many more wins he's worth than a replacement-level player at his same position" [24]. Another way to interpret WAR is how many wins that player is responsible for providing to his team. More simply, a positive-WAR player is producing wins for his team while a negative-WAR player is costing his team wins. The higher the WAR, the more impactful that player is to his team.

Front offices are still going to pay top dollar for the best players, but there should be a way to value that risk. Organizations only have so much capital to spend on players and there are rules and penalties set by the MLB to deter overspending for talent. Matt Swartz, a baseball economist and contributor at Hardball Times and FanGraphs, has used a Dollar-per-WAR value to study trends in player salaries and analyze how they change based on several different factors [25]. While players change, the goals for each team remain the same: win more games. If WAR is the accepted metric that captures player value, then teams should build their roster to maximize their WAR. More importantly, teams should minimize their spending to get these high-WAR players by analyzing how WAR impacts salary. This research attempts to create a new metric that combines a player's performance value with his financial value and incorporate this metric in a model that chooses the most efficient player a team can add to their roster.

3 Data Collection

3.1 Performance Data

The primary source for performance data is Baseball-Reference.com. The website serves as a database and an online encyclopedia that provides information on every player and team in MLB history, including statistics, advanced metrics, and salaries. It is a credible public resource for all baseball data.

Relevant information from Baseball-Reference.com that applies to this research include the player's name, age, positions, team, and WAR for a given season. For those unfamiliar with baseball positions, the position labels and numbers are displayed in Figure 1.

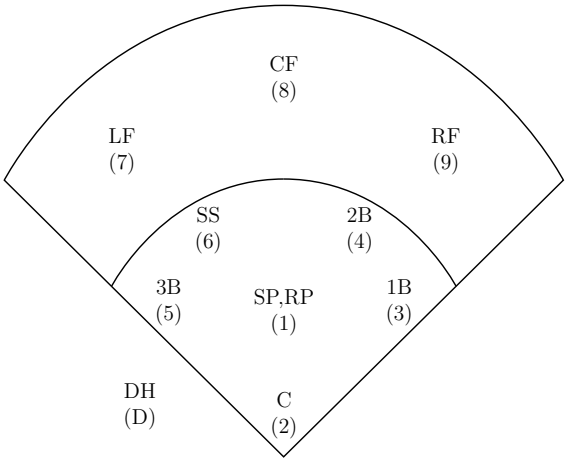


Figure 1: Position locations on a traditional baseball field.

Some players are equipped to play multiple positions during the regular season; for the purpose of this research, these versatile players will be assigned a position they play that most benefits their team with regards to maximizing WAR. The formula for calculating WAR is not published because there are many ways to measure player value; per Baseball-Reference.com,

> There is no one way to determine WAR. There are hundreds of steps to make this calculation, and dozens of places where reasonable people can disagree on the best way to implement a particular part of the framework [26].

Many experts and websites have created their own version of WAR that has its own nuances, so it is important to remain consistent when measuring and comparing player value. With that said, the version of WAR used throughout this research is the one calculated by Baseball-Reference.com.

While the focus is on the current player data, previous data must be considered to provide perspective on the current data. It is therefore necessary to use a player's WAR from every season of their career in order to determine their expected WAR with deviation. The longest active tenured player in the MLB is Albert Pujols who began his career in 2001, so every value of WAR since 2001 must be considered [27].

3.2 Financial Data

The 2017-2021 CBA provides many of the financial guidelines followed by MLB organizations. For the 2020 MLB season, the luxury tax threshold is $208 million and the minimum salary is $555,000 plus a cost of living adjustment [19]. According to the U.S. Bureau of Labor Statistics, the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W) increased from October 2018 to October 2019 by 1.56%, making the MLB minimum salary $563,500 [28].

Financial data for both players and teams are required for this research. MLB players can have a wide range of salaries based on the type of contract they have. These categories that determine a player's salary status are listed in Table 2.

Table 2: Eligibility requirements for the four salary categories for MLB players, according to the 2017-2021 CBA.

Player Status	Qualifications
Pre-Arbitration	MLB Service $<$ 3 years
Arbitration	3 years $\leq$ MLB Service $<$ 6 years
"Super Two"	2 years $\leq$ MLB Service $<$ 3 years, Top 22%
Free Agent	MLB Service $\geq$ 6 years

The AAV for pre-arbitration players tends to be around the league minimum because the team has complete control of the salary; there may be slight variations in AAV depending on if a player received a signing bonus. Players eligible for salary arbitration may request an arbitration hearing for a higher salary based on their contributions to the organization on and off the field [19]. "Super Two" players are arbitration eligible if they rank in the top 22% for amount of service among players who have had at least two but less than three years of MLB service [19]. The remaining veteran players are free agents who have salaries dictated by the open market.

Spotrac is an online database that provides detailed and up-to-date payroll information for every MLB team. This becomes a great resource during the offseason when salaries change and trades occur. There are two types of financial data from Spotrac that are required for this research: player salaries and team payroll. A player's name, age, and team are very important because these three fields are unique identifiers that help merge player information between Baseball-Reference.com and Spotrac. Their contract type, adjusted salary, and AAV are essential for identifying their status in MLB service as well as their financial value. Note that the adjusted salary is how much a player

is making in a particular year while AAV is the average salary a player earns every year. Relevant team information includes the total AAV for its 40-man roster and its Competitive Balance payroll. The difference between the two salary values is the summation of player benefits and Minor League contracts, which are irrelevant for this research. In other words, the main focus for the organization is to analyze the effects of changing rosters with regards to the CBT threshold.

4 Methodology

4.1 Validation Layout

The validation process is about confirming the belief that teams do try to improve at their weakest positions. Every team has a spot on their roster that can be improved. For some teams, the weakest position can be on the offensive side while other teams may have issues with their pitching rotation. The validation step for this research is confirming basic intuition that organizations work to remedy the weakest positions on their team from season to season, regardless of their current win production.

The weakest positions are determined by the three lowest $\mu_{\text{WAR},n}$ values on a team's best starting roster, where $\mu_{\text{WAR},n}$ is the average WAR over n seasons. This value of n should be chosen carefully; a single-season WAR ($n = 1$) may include values that are abnormally high or low from their true worth while a career average may not be indicative of that player's current abilities. A recent example of the former situation is Corey Kluber, a pitcher for the Cleveland Indians who posted all-star WAR values above 5.0 from 2016 to 2018 but suffered an effectively season-long injury in 2019 resulting in a -0.4 WAR [29]. Alternatively, Albert Pujols has had a Hall-of-Fame worthy career primarily as a first baseman, but his production with age has followed the same trend as in Figure 2 [27]. Due to the significance of both recent performance and variation, the value of n has been chosen to be equal to 3 seasons for this research.

A team's starting roster will consist of all offensive positions shown in Figure 1, four starting pitchers, and one relief pitcher. Note that teams in the American League have a designated hitter while teams in the National League do not. Most good teams have four consistent pitchers in their rotation throughout the season, but they will have many more pitchers in their bullpen. Relief pitchers generally do not contribute much value due to the small amount of innings they play. There is usually only one relief pitcher per team who sees more time and thus adds more value to his team. In other words, if more relief pitchers were included in this validation process, nearly every team would have at least one of their three weakest positions be their relief pitcher.

It is considered an improvement if a position is in the bottom three one season and not in the bottom three the next season. There are two ways of measuring the strength of this validation: calculating the percentage of teams that have at least one improvement, and calculating the percentage of improvements across the MLB. Note that an improvement for a team can have multiple interpretations, from trades and player development to injuries. For example, some teams may hold onto their weaker positions in belief that their highly-talented prospect takes over the position upon entering the league. While this validation grossly summarizes the nuances in each team's roster, allowing teams three opportunities to improve at their weakest positions be enough to balance the situation.

4.2 WAR and AAV

Players are analogous to assets for their organization: they can be traded, sold, and bought with their own set of risks. Because these assets are athletes, they tend to have a peak in performance followed by a decline until retirement. Many players have different arcs in performance, but they generally follow a path that is closely related to their age [30]. Therefore, the lifespan for these assets is roughly defined and relatively short.

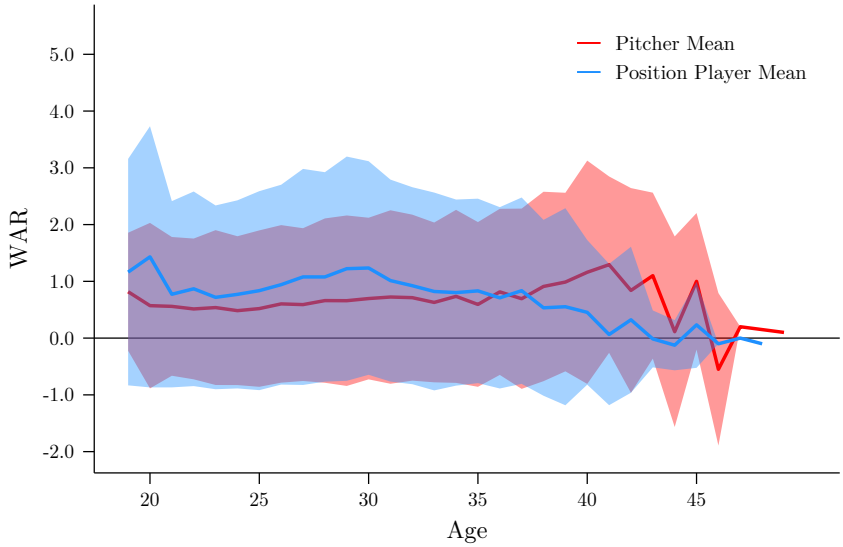


Figure 2: Expected WAR for each age from 2001 to 2019. The bleed around both types of players represents one standard deviation from the mean at each age. Data is provided courtesy of Baseball-Reference.com.

Figure 2 makes some intuitive sense in that a position player should expect to peak at around 30 years old. This is because the player has had enough experience competing with MLB-level talent while still in good enough shape to perform at a high quality. Understandably, a player will not be able to contribute as much to his team if he is not as fit as younger prospects on the team. However, it appears that the expected WAR for a pitcher increases until around 40 years old, a peak that surpasses that of a position player by a decade. Pitchers are not expected to be as fit as their position player counterparts, for their job is to get outs. Some of the best pitchers will know how to pitch to certain batters after facing them many times, thus building their experience and possibly their WAR. For example, 44-year-old pitcher Nolan Ryan threw his seventh career no-hitter, striking out 16 Blue Jays on May 1, 1991 [31]. While not every baseball player who makes it to the Major Leagues will be as successful as Nolan Ryan, age should be considered when evaluating potential future production for a given player.

4.3 Performance and Financial Efficiency Rating

The foundation of this analysis is based on a player's WAR and AAV. The ratio of these two values is the Cost Per WAR (CPW), as shown in Equation (1).

$$\text{CPW} = \frac{\text{AAV}}{\mu_{\text{WAR},n}} \tag{1}$$

A player's CPW denotes how much that player is worth in AAV for every unit of WAR. Players with a higher CPW are less financially efficient for their performance while players with a lower CPW are more efficient. The expected CPW and its standard deviation are calculated using Equations (2) and (3):

$$\mu_Y = 10^{\mu_X + \frac{1}{2}\sigma_X^2} \tag{2}$$

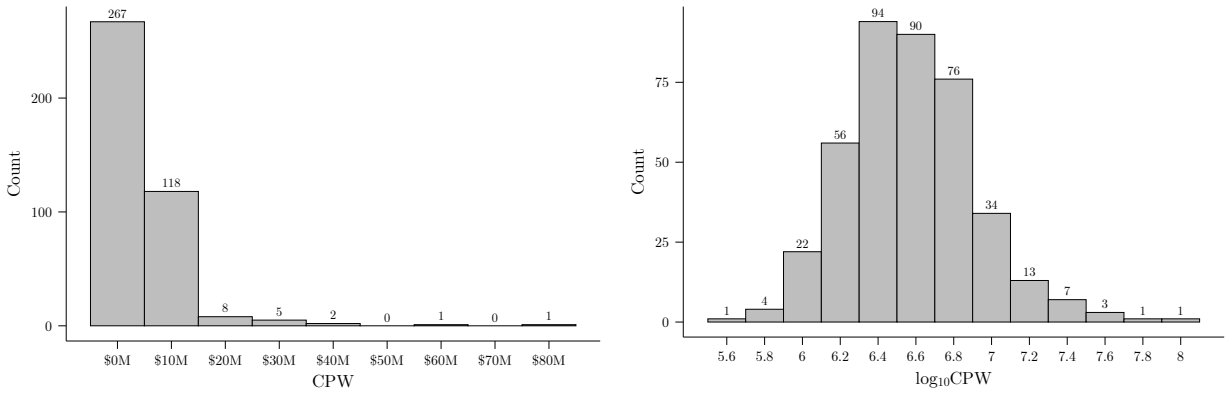
$$\sigma_Y = 10^{\mu_X + \frac{1}{2}\sigma_X^2} \sqrt{10^{\sigma_X^2} - 1} \tag{3}$$

where

$$\mu_X = \mathbb{E}\left[\log_{10} \mathrm{CPW}\right] \tag{4}$$

$$\sigma_X = \sqrt{\mathrm{Var}\left(\log_{10} \mathrm{CPW}\right)} \tag{5}$$

In this context, X denotes the distribution of CPW and Y denotes the log-normal distribution of CPW for all players. Due to the wide spread of salaries in the MLB, the distribution of CPW should be investigated before proceeding. The relevance for Equations (4) and (5) becomes apparent upon viewing the distribution of CPW for every player in Major League Baseball.



(a) Right-skewed distribution.

(b) Log-normal distribution.

Figure 3: Two types of distributions for every player's CPW in Major League Baseball. Data courtesy of Baseball-Reference.com and Spotrac.

Note that σ_Y as calculated in Equation (3) represents the standard deviation for CPW mapped to its logarithm. As displayed by Figure 3a, the standard deviation of CPW would be quite large, surpassing the mean and thus resulting in a coefficient of variation greater than 1. Having a standard deviation greater than the mean while having a strong sample size is generally bad practice, which is why the log-normal distribution is preferred in this situation.

This mapping of means and standard deviations between distributions can be quite confusing to comprehend. Instead of using this rather involved metric that requires a deeper understanding behind the value itself, a standardization is performed on the log-normal distribution of CPW to create the Performance and Financial Efficiency Rating (PFER).

$$\mathrm{PFER} = \frac{\log_{10} \mathrm{CPW} - \mu_X}{\sigma_X} \tag{6}$$

The reasoning for the standardization shown in Equation (6) is to create a more intuitive zero-mean normal distribution with a standard deviation of 1. The negative sign is used to yield a more favorable PFER for those players with efficient financial value relative to their performance. Note that PFER is only assigned to players with a positive $\mu_{\text{WAR},n}$ because those are the only players who organizations should theoretically consider acquiring in this situation. Players who are in pre-arbitration are also not considered for PFER because all salaries for them are around the league minimum and thus have minimal variation in AAV; top pre-arbitration players would be unfairly recommended as valid trade options when they are being paid far less than their performance value would suggest.

5 Results

5.1 Validation Process

As explained in Section 4.1, the validation process involves analyzing the worst positions on every team's starting roster based on $\mu_{\text{WAR},3}$. The number of positions improved for each team is counted from one season to the next season. Table 3 shows this validation process using the 2017 and 2018 seasons.

All but two teams, the Pittsburgh Pirates and the St. Louis Cardinals, had at least one improvement from 2017 to 2018. Looking more closely, 44 out of 90, or 48.9%, possible positions improved throughout all teams in the MLB. This means that there is roughly a coin-flip probability that a position in the bottom three of win production for any given team will not be in the bottom three the following season.

The 93.3% rate that a team will make an improvement on at least one position is very strong, signaling that organizations are motivated to make changes and take risks with the intention of building their WAR and ultimately winning more games. The 48.9% rate translates to an expected number of improvements of 1.47 per team, which further supports the notion that organizations will attempt to fix multiple positions. Both of these rates validate the research because MLB teams are showing a desire to improve, regardless of their current talent or win-loss record.

5.2 PFER Illustrations

From Equations (4) and (5), the mean and standard deviation for the log-normal distribution of CPW are 6.577 and 0.341, respectively. These two statistics determined by the log-normal distribution in Figure 3b are much more favorable than the mean of \$5.420 million and standard deviation of \$6.944 million from the right-skewed distribution in Figure 3a. The PFER for every player is calculated using Equation (6) and analyzed for each team. Rather than display the results for all 30 MLB organizations, two teams have been chosen for illustration purposes: the Boston Red Sox and the Oakland Athletics.

5.2.1 Boston Red Sox

The Red Sox are a big-market organization with a rich history and a strong fanbase that covers New England [32]. The CBT has forced the front office to make difficult decisions in choosing whether to keep paying for top talent or try to develop more from within to save some spending and get under the threshold. One recent example is when the Red Sox traded Mookie Betts to the Los Angeles Dodgers, who was widely regarded as the best player in the MLB behind Mike Trout

Table 3: The worst positions for all 30 MLB teams in 2017 and 2018 along with their improvement values. Data provided courtesy of Baseball-Reference.com.

Team	Worst Positions (2017)	Worst Positions (2018)	Improved (Number)	Improved (Boolean)
ARI	2B, C, SS	SP, C, RP	2	TRUE
ATL	LF, RP, 3B	SS, RP, SP	2	TRUE
BAL	SP, DH, SS	C, 1B, SP	2	TRUE
BOS	C, DH, 3B	3B, C, 1B	1	TRUE
CHC	CF, 2B, LF	RP, CF, RF	2	TRUE
CHW	C, DH, 2B	CF, LF, 2B	2	TRUE
CIN	SP, SP, RF	SP, SP, SS	1	TRUE
CLE	C, RF, CF	RF, C, 1B	1	TRUE
COL	1B, LF, RP	1B, C, LF	1	TRUE
DET	CF, SP, DH	2B, C, LF	3	TRUE
HOU	SP, SP, C	C, RP, DH	2	TRUE
KCR	SS, RF, DH	SS, DH, 1B	1	TRUE
LAA	SP, SP, C	1B, SP, SP	1	TRUE
LAD	1B, RF, SP	SP, SP, 1B	1	TRUE
MIA	SP, SP, SP	SP, SP, RP	1	TRUE
MIL	RF, SP, CF	C, SP, SP	2	TRUE
MIN	SS, 3B, RP	C, SS, SP	2	TRUE
NYM	C, 3B, SP	C, 1B, SP	1	TRUE
NYY	2B, 1B, DH	1B, 3B, RP	2	TRUE
OAK	CF, 1B, RF	SP, SP, C	3	TRUE
PHI	RF, SP, C	RF, SS, SP	1	TRUE
PIT	3B, SS, SP	SS, 3B, SP	0	FALSE
SDP	C, SP, SP	SP, SP, SP	1	TRUE
SFG	LF, CF, DH	DH, LF, 2B	1	TRUE
SEA	RP, 1B, C	LF, SP, SP	3	TRUE
STL	RP, SP, SP	RP, SP, SP	0	FALSE
TBR	2B, SP, RP	SP, SP, SP	2	TRUE
TEX	LF, 1B, CF	SP, LF, CF	1	TRUE
TOR	1B, LF, DH	2B, DH, 1B	1	TRUE
WSN	C, LF, 1B	C, 1B, RP	1	TRUE

[33]. This notion is confirmed by their $\mu_{\text{WAR},3}$ values, as Mike Trout and Mookie Betts sit at the top of the MLB with 8.37 and 7.93, respectively.

Had the Red Sox gone over the CBT threshold, they would be a Third-Time CBT Payor and have to pay the tax rates listed in Table 1. The front office decided to reset their CBT status and face the consequences of trading players to reduce payroll. Even with a player like Mookie Betts gone, the Red Sox still have a talented starting roster according to Table 4.

Only two players on the starting roster are pre-arbitration: Michael Chavis and Rafael Devers. This is fairly standard for a team that is typically at the top of the MLB in payroll spending because they have the financial freedom to acquire elite and established players. Table 5 shows the three positions the Red Sox should focus on improving.

There is only one pre-arbitration player who is included in the worst three positions on the team while the other two players have veteran status. In general, this can be financially inefficient for

Table 4: Detailed information on the optimal starting roster for the Boston Red Sox in 2020. Data provided courtesy of Baseball-Reference.com and Spotrac.

Position	Name	$\mu_{\text{WAR},3}$	$\sigma_{\text{WAR},3}$	Contract	AAV ($M)	CPW ($M)	PFER
SP	Chris Sale	5.10	2.38	Vet	25.667	5.033	−0.36
SP	Eduardo Rodriguez	3.53	2.12	Arb 3	8.300	2.349	0.61
SP	Nathan Eovaldi	0.65	0.92	Vet	17.000	26.154	−2.46
SP	Martin Perez	0.53	1.72	Vet	6.000	11.250	−1.39
RP	Brandon Workman	1.53	1.46	Arb 3	3.500	2.283	0.64
C	Christian Vazquez	0.73	1.66	Arb 3	4.517	6.159	−0.62
1B	Mitch Moreland	1.37	0.38	Vet	3.000	2.195	0.69
2B	Michael Chavis	0.60	*NA*	Pre-Arb	0.571	*NA*	*NA*
3B	Rafael Devers	1.90	2.65	Pre-Arb	0.692	*NA*	*NA*
SS	Xander Bogaerts	4.10	1.91	Vet	20.000	4.878	−0.32
LF	Andrew Benintendi	3.07	1.36	Arb 1	5.000	1.630	1.07
CF	Jackie Bradley Jr.	2.53	0.85	Arb 4	11.000	4.342	−0.17
RF	J.D. Martinez	4.60	1.66	Vet	23.750	5.163	−0.40
DH	Kevin Pillar	1.97	1.01	Arb 3	4.250	2.161	0.71

Table 5: Detailed information on the worst positions for the Boston Red Sox in 2020. Data provided courtesy of Baseball-Reference.com and Spotrac.

Position	Name	$\mu_{\text{WAR},3}$	$\sigma_{\text{WAR},3}$	Contract	AAV ($M)	CPW ($M)	PFER
SP	Martin Perez	0.53	1.72	Vet	6.000	11.250	−1.39
2B	Michael Chavis	0.60	*NA*	Pre-Arb	0.571	*NA*	*NA*
SP	Nathan Eovaldi	0.65	0.92	Vet	17.000	26.154	−2.46

the organization if they are spending a lot for these players. Meanwhile, Michael Chavis is not considered much of a financial burden to the team due to his low AAV, which can motivating for both parties: Chavis will continue to work hard and improve his win production in order to raise his salary when he becomes eligible for arbitration, and the organization has more freedom to experiment with him and his position. The top acquisition options for a starting pitcher and second baseman are listed in Table 6.

Table 6: Detailed information on potential targets for the Boston Red Sox in 2020. Data provided courtesy of Baseball-Reference.com and Spotrac.

Positions	Name	Team	$\mu_{\text{WAR},3}$	$\sigma_{\text{WAR},3}$	Contract	AAV ($M)	CPW ($M)	PFER
S	Mike Clevinger	CLE	4.07	1.05	Arb 1	4.100	1.008	1.68
4,9,8,D	Whit Merrifield	KCR	3.97	1.12	Arb 1	4.062	1.024	1.66
8,4,6	Ketel Marte	ARI	4.20	2.91	Arb 2	4.800	1.143	1.52
S	Jose Berrios	MIN	2.87	0.84	Arb 1	4.025	1.404	1.26
S	Mike Minor	TEX	4.73	2.66	Vet	8.333	1.761	0.97

These players are considered to be financially efficient options based on their win production. Four of the five players in Table 6 are arbitration eligible while one is a veteran. Even though his value is derived from the open market, Mike Minor is included in this group of high-PFER players because he has pitched rather well in the past few seasons, cemented by the achieving the top WAR for a pitcher in 2019 [34]. The 32-year-old starting pitcher may be a sound investment for the Red Sox if his win production continues to follow the trend in Figure 2. Note that the Red Sox have a tax

space of around $12 million, so any of these players in Table 6 can be added to the roster without pushing the Red Sox payroll over the CBT threshold.

5.2.2 Oakland Athletics

Like the Red Sox, the Oakland Athletics also carry a rich history, but they have struggled to achieve playoff success in the 21st century. The A's are a small-market organization that competes for regional baseball attention with the San Francisco Giants, an organization that has seen tremendous success within the past decade [35]. As stated before, this organization does not have the seemingly unlimited capital that its big-market competitors have, so they must be more financially cautious when considering adding more talent.

Table 7: Detailed information on the optimal starting roster for the Oakland Athletics in 2020. Data provided courtesy of Baseball-Reference.com and Spotrac.

Position	Name	$\mu_{\text{WAR},3}$	$\sigma_{\text{WAR},3}$	Contract	AAV ($M)	CPW ($M)	PFER
SP	Mike Fiers	2.17	2.36	Vet	7.050	3.254	0.19
SP	Sean Manaea	1.87	0.55	Arb 2	3.750	2.009	0.81
SP	Chris Bassitt	1.20	0.99	Arb 1	2.250	1.875	0.89
SP	Frankie Montas	0.73	1.50	Pre-Arb	0.564	*NA*	*NA*
RP	Yusmeiro Petit	1.83	0.40	Vet	5.500	3.000	0.30
C	Sean Murphy	0.60	*NA*	Pre-Arb	0.564	*NA*	*NA*
1B	Matt Olson	4.13	1.30	Pre-Arb	0.604	*NA*	*NA*
2B	Chad Pinder	1.57	0.86	Arb 1	2.025	1.293	1.37
3B	Matt Chapman	6.60	2.94	Pre-Arb	0.624	*NA*	*NA*
SS	Marcus Semien	5.10	3.62	Arb 3	13.000	2.549	0.50
LF	Robbie Grossman	1.20	0.44	Arb 3	3.725	3.104	0.25
CF	Ramon Laureano	2.95	0.92	Pre-Arb	0.564	*NA*	*NA*
RF	Mark Canha	1.77	2.55	Arb 2	4.800	2.717	0.42
DH	Khris Davis	1.77	1.79	Vet	16.750	9.481	−1.17

It is apparent that the A's do not have the luxury of buying their wins based on the number of pre-arbitration players in Table 7. There are only three veterans on their starting lineup, allowing the organization to focus more on player development. The most interesting players on this list are Matt Chapman, Marcus Semien, and Matt Olson. These three players not only have the best $\mu_{\text{WAR},3}$ on the team, but Chapman and Olson are also among the cheapest due to their pre-arbitration status. It is players like Chapman and Olson who would have a big impact on the PFER distribution due to their elite financial efficiency.

Table 8: Detailed information on the worst positions for the Oakland Athletics in 2020. Data provided courtesy of Baseball-Reference.com and Spotrac.

Position	Name	$\mu_{\text{WAR},3}$	$\sigma_{\text{WAR},3}$	Contract	AAV ($M)	CPW ($M)	PFER
C	Sean Murphy	0.60	*NA*	Pre-Arb	0.564	*NA*	*NA*
SP	Frankie Montas	0.73	1.50	Pre-Arb	0.564	*NA*	*NA*
SP	Chris Bassitt	1.20	0.99	Arb 1	2.250	1.875	0.89

The A's have run their organization quite well from a financial efficiency perspective upon analyzing Table 8. Their weakest positions are either pre-arbitration or early in arbitration eligibility. In other words, the A's have been great at not risking much to compensate for their worst-performing positions. The summation of AAV for this sample of players is under $3.5 million, which is highly favorable for a team with a consistently low payroll.

Table 9: Detailed information on potential targets for the Oakland Athletics in 2020. Data provided courtesy of Baseball-Reference.com and Spotrac.

Positions	Name	Team	$\mu_{\text{WAR},3}$	$\sigma_{\text{WAR},3}$	Contract	AAV ($M)	CPW ($M)	PFER
S	Mike Clevinger	CLE	4.07	1.05	Arb 1	4.100	1.008	1.68
2	Willson Contreras	CHC	3.37	0.46	Arb 1	4.500	1.337	1.32
2,D	Omar Narvaez	MIL	1.73	0.60	Arb 1	2.725	1.572	1.12
S	Aaron Nola	PHI	6.17	3.50	Arb 2	11.250	1.824	0.93
2,D	Gary Sanchez	NYY	2.70	1.44	Arb 1	5.000	1.852	0.91

All players in Table 9 are arbitration eligible, which is great for a team that must be careful on payroll spending. While the A's have nearly $100 million in tax space, they do not have enough capital to make the CBT threshold their main concern. The small amount of AAV sacrificed for their struggling positions from Table 8 means that they can focus on potentially acquiring a strong starting pitcher in Mike Clevinger or try to fix their catcher situation with Willson Contreras. The PFER metric and this model can be very helpful for the A's as they focus on maximizing there wins per payroll, something they are particularly good at [36].

6 Conclusion

As with most aspects of life, the cost of failure should determine whether or not to make a change. This research has shown that a player's performance and financial value can be quantified and combined to create a metric that can help MLB organizations find other financially efficient players to improve their roster. Teams can theoretically sacrifice less risk and still add win production. The fact that the weakest positions are changing for nearly every team season to season validates the notion that teams may be receptive to PFER and how it selects the most financially efficient players in Major League Baseball.

6.1 Key Notes

The model serves a relatively broad purpose and does not go too deep into certain nuances in the MLB. For starters, contracts are much more complex and possibly obscure in reality, such as the 1999-retired Bobby Bonilla's deferred contract where he gets paid $1.19 million every year until 2035 [37]. Teams may even want to keep a win-costing "core" player as his career is nearing its end because it can allow prospects to cycle in and gain some experience while the he is living his MLB swan song. Integrating core players and promising prospects in the Minor Leagues with the model can make it much more powerful. While both clubs and players agree that WAR is a consistent determinant of value, there may be situations when using different metrics for measuring player value are more advantageous, such as Weighted On-Base Average (wOBA), Weighted Runs Created Plus (wRC+), and On-base Plus Slugging Plus (OPS+). Needless to say, there is a lot of opportunity for this model to grow.

This research was supposed to test and strengthen both the PFER and the model during the MLB regular season, but the COVID-19 pandemic resulted in canceling much of the 2020 season. In other seasons, it may be interesting to merge this research with a season-dependent trade metric like FiveThirtyEight's Doyle Number, which is "the rate at which it should be willing to trade talent in the future for talent in the current season in order to maximize the franchise's total World Series wins" [38]. Incorporating other player value metrics and trade analytics in the model can be very beneficial in future seasons.

6.2 Final Words

Nate Silver, editor-in-chief of the online publication FiveThirtyEight, has some wise words regarding the usage of data as a tool for making decisions: "Data-driven predictions can succeed—and they can fail. It is when we deny our role in the process that the odds of failure rise" [39]. It is important to remember that MLB players are people, and decisions on another player's career should not be made exclusively by a computer with an analytical model. The greater amount of and access to data is exciting for those who desire that information and knowledge, especially as the technology that can process this data becomes more commonplace. Still, as Bill James recently said, "All I know for sure is we will get it wrong several times before we get it right" [40]. Analytical models should not be making decisions, but rather it should warn against the bad decisions and assist in making the good ones.

7 References

[1] "The 1869 Cincinnati Red Stockings Vintage Base Ball Team." *1869 Cincinnati Red Stockings*. Web. 10 Jan. 2020.

[2] Norman, Jim. "Football Still Americans' Favorite Sport to Watch." *Gallup.com*, Gallup, 4 Sept. 2019. Web. 10 Jan. 2020.

[3] Fischer-Baum, Reuben. "The Best MLB Teams Of All-Time, According To Elo." *FiveThirtyEight*, FiveThirtyEight, 9 May 2016. Web. 16 Apr. 2020.

[4] Paine, Neil. "The Warriors Are Great, But Are They Better Than The 1939 Yankees?" *FiveThirtyEight*, FiveThirtyEight, 14 June 2017. Web. 16 Apr. 2020.

[5] Sports Reference LLC. "Curt Flood." *Baseball-Reference.com*, Major League Statistics and Information. Web. 10 Jan. 2020.

[6] Margolick, David. "Fielder's Choice." *The New York Times*, The New York Times, 8 Oct. 2006.

[7] *Flood v. Kuhn*, 407 U.S. 258, 281. 19 June 1972.

[8] Barra, Allen. "How Curt Flood Changed Baseball and Killed His Career in the Process." *The Atlantic*, Atlantic Media Company, 12 July 2011.

[9] Pesca, Mike. "The Man Who Made Baseball's Box Score a Hit." *National Public Radio*, 30 July 2009. Web. 10 Jan. 2020.

[10] "The SABR Story." Society of American Baseball Research. Web. 10 Jan. 2020.

[11] McGrath, Ben. "The Professor of Baseball." *The New Yorker*, The New Yorker, 14 July 2013. Web. 10 Jan. 2020.

[12] Lewis, Michael. *Moneyball: The Art of Winning an Unfair Game*. W.W. Norton, 2013. Print.

[13] Venook, Jeremy. "Two Superstars Gamble on Record-Breaking Deals." *The Atlantic*, Atlantic Media Company, 13 Mar. 2019. Web. 10 Jan. 2020.

[14] Adler, David. "Where Do Strasburg, Cole, Rendon Contracts Rank?" *MLB.com*, 16 Dec. 2019. Web. 10 Jan. 2020.

[15] O'Connell, Robert. "Baseball's Best Player Deserves More Than $430 Million." *The Atlantic*, Atlantic Media Company, 21 Mar. 2019. Web. 10 Jan. 2020.

[16] Page, Vincent. "LA Angels: Ranking the Five Worst Contracts of the Last Decade." *Halo Hangout*, FanSided, 31 Jan. 2019. Web. 10 Jan. 2020.

[17] Crawford, Setrige. "2016 Is A Great Summer To Be A Mediocre NBA Free Agent." *FiveThirtyEight*, FiveThirtyEight, 19 July 2016. Web. 12 Jan. 2020.

[18] Sports Reference LLC. "Collective Bargaining Agreement." *Baseball-Reference.com*, Major League Statistics and Information. Web. 17 Apr. 2020.

[19] "2017-2021 Basic Agreement." *Major League Baseball Players*, Major League Baseball Players Association. Web. 17 Apr. 2020.

[20] Paine, Neil. "Baseball's Best Teams Are Too Damn Good." *FiveThirtyEight*, FiveThirtyEight, 21 July 2017. Web. 16 Apr. 2020.

[21] "New York Yankees announce purchase of Babe Ruth." *History.com*, HISTORY, 16 Nov. 2009. Web. 18 Apr. 2020.

[22] Sports Reference LLC. "Career Leaders & Records for Wins Above Replacement." *Baseball-Reference.com*, Major League Statistics and Information. Web. 18 Apr. 2020.

[23] Sports Reference LLC. "Boston Red Sox Team History & Encyclopedia." *Baseball-Reference.com*, Major League Statistics and Information. Web. 18 Apr. 2020.

[24] "Wins Above Replacement (WAR)." MLB.com. Web. 10 Jan. 2020.

[25] Swartz, Matt. "Trends in Free Agent Spending on Hitters." *FanGraphs Baseball*, Fangraphs Inc., 28 July 2017. Web. 18 Apr. 2020.

[26] Sports Reference LLC. "Baseball-Reference.com WAR Explained." *Baseball-Reference.com*, Major League Statistics and Information. Web. 16 Apr. 2020.

[27] Sports Reference LLC. "Albert Pujols." *Baseball-Reference.com*, Major League Statistics and Information. Web. 17 Apr. 2020.

[28] U.S. Bureau of Labor Statistics. "Consumer Price Index Historical Tables for U.S. City Average." United States Department of Labor. Web. 17 Apr. 2020.

[29] Sports Reference LLC. "Corey Kluber." *Baseball-Reference.com*, Major League Statistics and Information. Web. 17 Apr. 2020.

[30] Weinberg, Neil. "The Beginner's Guide To Aging Curves." *FanGraphs Baseball*, Fangraphs Inc., 10 Dec. 2015. Web. 10 Jan. 2020.

[31] Adler, David. "What We Learned Watching Ryan's 7th No-No." *MLB.com*, 24 Mar. 2020. Web. 15 Apr. 2020.

[32] Giratikanon, Tom, et al. "Up Close on Baseball's Borders." *The Upshot*, The New York Times, 24 Apr. 2014. Web. 19 Apr. 2020.

[33] Gurnick, Ken. "Mookie Betts is officially on his way to LA." MLB.com, 11 Feb. 2020. Web. 16 Apr. 2020.

[34] Sports Reference LLC. "2019 MLB Value." *Baseball-Reference.com*, Major League Statistics and Information. Web. 18 Apr. 2020.

[35] Sports Reference LLC. "San Francisco Giants Team History & Encyclopedia." *Baseball-Reference.com*, Major League Statistics and Information. Web. 19 Apr. 2020.

[36] Morris, Benjamin. "Billion-Dollar Billy Beane." *FiveThirtyEight*, FiveThirtyEight, 24 July 2014. Web. 18 Apr. 2020.

[37] Paine, Neil. "Bobby Bonilla Was More Than The Patron Saint Of Bad Contracts." *FiveThirtyEight*, FiveThirtyEight, 1 July 2017. Web. 18 Apr. 2020.

[38] Paine, Neil. "Should Your Team Buy Or Sell At The MLB Trade Deadline?" *FiveThirtyEight*, FiveThirtyEight, 25 July 2016. Web. 18 Apr. 2020.

[39] Silver, Nate. *The Signal and the Noise: Why so Many Predictions Fail—but Some Don't.* New York: Penguin Press, 2012. Print.

[40] James, Bill. "Moneyball to Homerball with Bill James and Tom Tango." MIT Sloan Sports Analytics Conference, 7 Mar. 2020, Boston Convention and Exhibition Center, Boston, MA. Panel Discussion.