5-2016

Baseball Portfolio Optimization

Keegan Henderson
University of Arkansas, Fayetteville

Baseball Portfolio Optimization

A thesis submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Science in Industrial Engineering

by

Keegan Henderson

May 2016
University of Arkansas

Abstract

Portfolio optimization techniques are methods used to determine the best set of stocks in which to invest. Mean-variance optimization, one method of portfolio optimization, attempts to find the set of portfolios that have the maximum expected return at each level of risk (Jorion, 1992). Another technique, Monte Carlo simulation, uses random number generation to create a probability distribution of potential returns (Kwak & Ingall, 2007). This can be used to determine the risk of potential investments not returning a certain desired amount (Thompson & McLeod, 2009). Though traditionally used in the world of finance, these tools can also be utilized by professional sports teams, such as those in Major League Baseball, to make more efficient investments in personnel and increase their likelihood of reaching the postseason.

This research effort explores strategies to optimize the allocation of a baseball team's resources in the free agent market. In this effort, we use a portfolio optimization approach and explore a variety of baseball performance metrics. A prototype optimization model is created and evaluated. This model is designed to assemble the team with the highest likelihood of making the playoffs while accounting for various budget and roster constraints faced by Major League Baseball teams. The prototype is utilized to create an optimized 2015 roster for three teams: the Boston Red Sox, Kansas City Royals, and San Diego Padres. These optimized rosters are then compared to each team's actual 2015 opening day roster. Several iterations of this model are discussed in an attempt to find the option that returns the most value. After multiple alternatives are analyzed, three different options are identified that compare favorably to the teams' actual opening day rosters with regards to 2015 performance of the players selected. Weaknesses of the model are then discussed, as well as ways in which it can be improved.

Keywords: portfolio optimization, mean-variance optimization, Monte Carlo simulation, expected returns, risk, performance metrics, stochastic optimization, linear optimization

Introduction

Professional sports have always involved high financial stakes, and those stakes have only grown in the modern era of televised sporting events and expanded media coverage. From the beginning of professional sports, teams have faced a large degree of risk in signing athletes to expensive contracts due to the variable and unpredictable nature of athletics. However, as these contracts have continued to grow, they have had a tendency to morph from a slight risk to a dangerous gamble that can lead to the financial success or failure for an organization over the foreseeable future. Major League Baseball, in particular, has seen a rapid increase in money spent by franchises trying to assemble competitive teams in recent years. This increase in spending is illustrated by the change in the average Major League Baseball player's salary, which has increased from $578,930 in 1990 all the way up to $3,440,000 in 2012 (CBS Sports).

With such a large amount being spent on players, it is necessary to use a disciplined and strategic approach when investing money in order to best utilize a team's resources. This problem is further complicated by the large degree of uncertainty in the return on investment of a large baseball contract. Some of this uncertainty is essentially unavoidable due to the inherent unpredictability of athletes and sports in general, which further emphasizes the need for a structured approach that accounts for uncertainty in constructing a MLB roster. Organizations must be careful to balance the risk of a large and expensive contract with the potential benefit of additional wins that a new player can bring to their team. The most efficient way to construct a roster is constantly changing due to the conditions of the free agent market and the needs of specific teams. It is helpful to create a mathematical approach to aid in the process that is less susceptible to bias and can more easily adjust to particular scenarios than the approach of solely relying on personal judgement. One way this scenario can be modeled mathematically is to look at the free agent market as a group of potential stocks and then use a portfolio optimization approach to determine the best combination of players to maximize the likelihood of having a successful team.

The purpose of this research is to describe current portfolio optimization techniques and discuss how they might be used in player investment decisions, as well as the most effective publicly available metrics to determine a baseball player's overall contribution to his team's performance. Finally, we will demonstrate how portfolio optimization can be applied to the construction of a professional baseball roster using examples.

Research Methodology

According to Speidell, Miller, and Ullman, portfolio optimization, at its simplest form, "is a procedure for measuring and controlling portfolio risk and expected return" (Speidell, Miller, & Ullman, 1989). With regards to stock portfolios, risk is defined as the amount of deviation or variance from the expected market return. Investments with a relatively high amount of standard deviation of potential returns are considered to have a large degree of risk (Speidell et al., 1989). One of the basic underlying assumptions of portfolio optimization is that, as the expected return of a portfolio increases, the risk increases as well. The goal of portfolio optimization is to ensure that an investor is achieving the highest amount of expected returns possible at whatever level of risk they deem acceptable (Speidell et al., 1989). At each level of risk, there is one portfolio with the highest expected return. Together, this set of portfolios make up what is known as the efficient set, which represents the highest level of expected returns that can be obtained at any particular risk level (Jorion, 1992).

This approach, invented by Harry Markowitz, is known as mean-variance optimization (Jorion, 1992). Because traditional mean-variance optimization creates an optimal portfolio at each level of risk, it is reliant on the optimizer to determine the ideal level of risk based on their personal preference (Speidell et al., 1989). As the investor does not always know the best level of risk for a particular scenario, this can sometimes cause problems as the investor may make a subjective or poor decision that leads to worse returns than could have been obtained at other levels of variance.

One technique that can be used to help address the issue of ideal risk is that of Monte Carlo simulation. At its most basic form, Monte Carlo simulation is simply the use of random number generation to determine a range of outcomes from a system of probability distributions (Kwak & Ingall, 2007). The concept is described in more detail by Young Hoon Kwak and Lisa Ingall:

"A model or a real-life system or situation is developed, and this model contains certain variables. These variables have different possible values, represented by a probability distribution function of the values for each variable. The Monte Carlo method simulates the full system many times (hundreds or even thousands of times), each time randomly choosing a value for each variable from its probability distribution. The outcome is a probability distribution of the overall value of the system calculated through the iterations of the model" (Kwak & Ingall, 2007).

With regards to portfolio selection, this method can be used to determine the likelihood that a particular set of securities loses money or returns less than a certain specified benchmark (Thompson & McLeod, 2009). With only a probability density function of each security in a potential portfolio, an investor can use Monte Carlo simulation to receive a distribution of the different potential returns, along with the likelihood of each scenario occurring (Kelliher & Mahoney, 2000).

Analysis of Applicable Research

In order to determine how to best model the selection of a professional baseball team, several articles were analyzed to see if they could be applied to this particular problem. The search for articles was focused on papers that included the phrases "Portfolio Optimization", "Mean-Variance Optimization", or "Monte Carlo Simulation." 19 total articles regarding methods of portfolio optimization were read. The title and author of all these articles are listed in Appendix A. Together, they are able to capture both the key concepts in portfolio selection as well as a wide variety of creative and interesting attempts to create improved models.

In order to effectively model the creation of a baseball team as a portfolio model, it is necessary to determine how the returns on an investment will be measured. While the goal of a traditional investment is to return monetary value, the goal of a baseball team is ultimately to win games. Therefore it makes sense to determine a baseball player's return on investment by his contributions towards helping the team win games. Unfortunately, there is not a large degree of academic research with regards to the use of portfolio optimization in baseball roster creation. The baseball blog "Beyond the Box Score" did post an article in May 2015 discussing how the principles of Modern Portfolio Theory, particularly investment diversification, can be used by Major League Baseball teams. However, the article does not describe the creation of an actual mathematical model or metrics that could be used to determine return on investment (Lampe, 2015).

Fortunately, the baseball website "Fangraphs" has a large amount of publicly available research discussing performance metrics that can be used to describe a player's contribution to his team both in terms of runs and wins contributed. Among these are several statistics that break down a player's impact by batting, base running, fielding, and pitching, using a scale of runs above or below average (Fangraphs, c).

Literature Review Analysis

Of the 19 portfolio optimization articles analyzed, 13 proposed specific models to aid in the problem of portfolio selection, while the others described the overarching concepts used in portfolio modeling. These papers ranged in date from as recently as 2012 all the way back to 1952. Out of the 19 papers, 11 were found using ProQuest and the other 8 were found from JSTOR. One interesting note is that the ProQuest articles analyzed tended to be more recent than those viewed from JSTOR. This point is demonstrated in Figure 1 below, which shows the breakdown of articles by year and location where they were found.

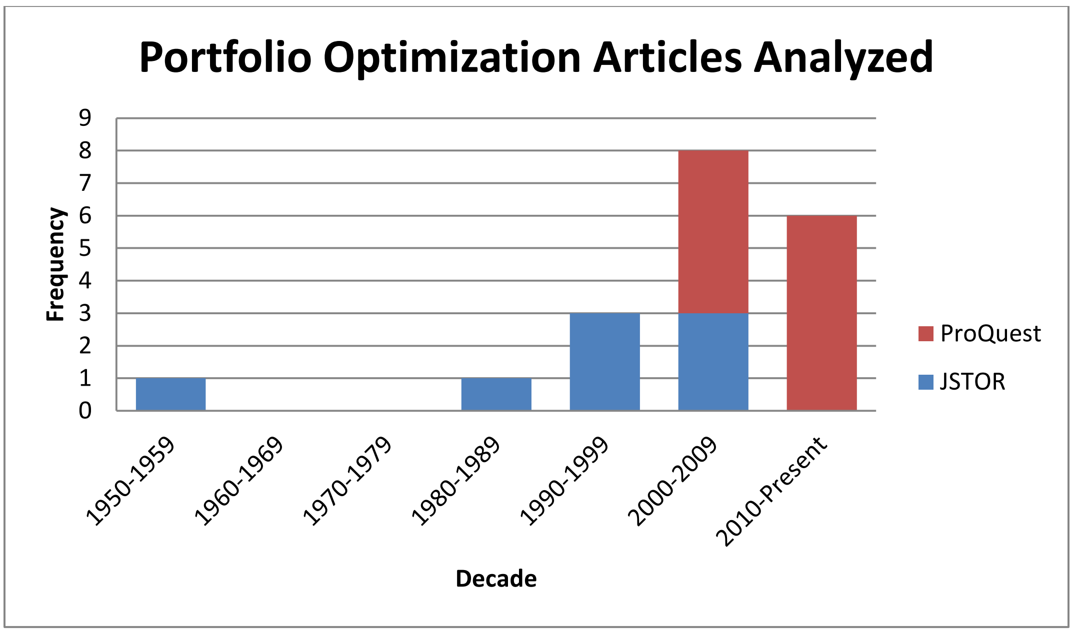


Figure 1: Portfolio Optimization Articles Analyzed

One extremely interesting work that was read is "Portfolio Selection" by Harry Markowitz. In this paper, Markowitz introduces the concept of an efficient set of portfolios based on maximizing the expected returns and minimizing the variance (Markowitz, 1952). This paper would become the foundation of mean-variance optimization, a concept that is still integral to the process of portfolio optimization today (Kritzman, 2011). The importance of "Portfolio Selection" by Markowitz is best illustrated by its usage in other academic literature. According to Google Scholar, it has been cited a total of 22,017 times, while all other portfolio optimization articles analyzed for this paper have been cited a combined total of 1,648 times (Google Scholar).

Another relevant and interesting article is "Using Monte Carlo Simulation to Improve Long-Term Investment Decisions" by Kelliher and Mahoney, which discusses how Monte Carlo simulation can be leveraged to make decisions about future investments, using any type of probability distribution to describe expected returns. Although the paper uses a real estate investment as its example, the principles discussed can be applied to essentially any decision where capital is being invested with a variable distribution of potential returns (Kelliher & Mahoney, 2000).

Several different sites were utilized to determine the necessary statistics to calculate the expected value of baseball players and the variance of that expected value. In total, 21 different articles were analyzed in order to gain an accurate view of the different metrics that should be included to capture a player's value in all aspects of the game. The majority of these articles were found on Fangraphs, with the site accounting for 16 of the 21 resources used. The other 5 articles are from a variety of sources, including several baseball blogs and one book. The title and author of each paper is listed in Appendix B. The sources are also broken down by origin in Figure 2.

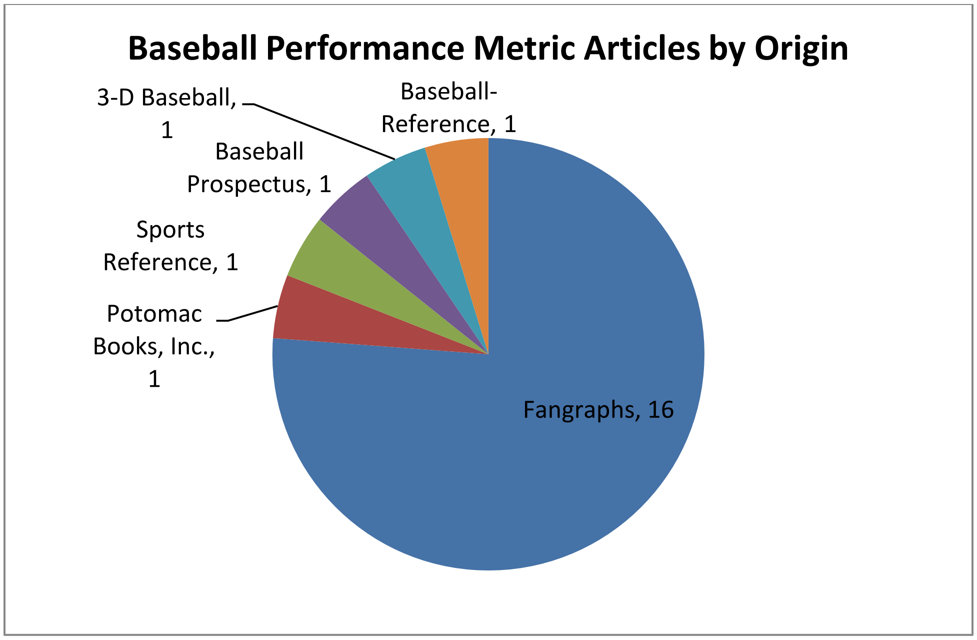


Figure 2: Baseball Performance Metric Articles by Origin

Fangraphs does calculate an overarching statistic designed to compile a player's total value in one number, called Wins Above Replacement (WAR). For position players, WAR essentially attempts to add up contributions from hitting, base running, and fielding into one number that expresses the player's total value. For pitchers, WAR is calculated solely by looking at their pitching contributions to the team (Fangraphs, h). Because WAR is composed of several different categories with different sample sizes and potential ranges of outcomes, WAR can be conceptually challenging to understand. While WAR values are given relative to a replacement level player, WAR component categories are analyzed on a runs above or below average scale. Fortunately, each athlete's component category statistics are also available on Fangraphs (Fangraphs, c). These component statistics are briefly described below in an attempt to make WAR slightly easier to understand.

Batting contributions are calculated from a statistic called Weighted Runs Above Average (wRAA). wRAA is based on a metric known as Weighted On-Base Average (wOBA)

(Fangraphs, j). According to Fangraphs, "*Weighted On-Base Average (wOBA) is a rate statistic which attempts to credit a hitter for the value of each outcome (single, double, etc) rather than treating all hits or times on base equally*" (Fangraphs, i).

Base running value is based on three different statistics: Weighted Stolen Base Runs (wSB), Ultimate Base Running (UBR), and Weighted Grounded Into Double Play Runs (wGDP). wSB simply calculates the amount of runs a player adds to his team relative to an average player by stealing bases or being caught stealing. UBR computes the amount of runs a player contributes above or below average based on their ability or failure to advance bases compared to the average. wGDP determines the amount of runs a player costs or saves his team based on the frequency in which he hits into double plays (Fangraphs, a).

Because of their vastly different role than other defensive positions, defensive value is calculated differently for catchers than it is for other defensive positions. For non-catchers, defensive runs are computed by adding a player's Ultimate Zone Rating (UZR) to a positional adjustment. The positional adjustment attempts to credit athletes who play a more demanding position, while debiting those who play an easier position (Fangraphs, g). UZR is calculated by adding Outfield Arm Runs (ARM), Double-Play Runs (DPR), Range Runs (RngR), and Error Runs (ErrR). Outfield Arm Runs are the amount of runs an outfielder saves relative to the average by preventing base runners from advancing. Double-Play Runs, as the name implies, are the amount of runs an infielder contributes above or below average by turning double plays. Range Runs are the amount of runs saved or cost by a fielder's ability to reach more or less batted balls than the average fielder. Error runs is simply a run value based on the amount of errors that the fielder commits (Fangraphs, e). The defensive value of catchers is computed by adding their Stolen Base Runs (rSB) and Runs Saved on Passed Pitches (RPP) to their positional adjustment (Fangraphs, g). Stolen Base Runs measures the amount of runs added by a catcher throwing out base runners who attempt to steal a base or keeping runners from attempting to steal. Runs Saved on Passed Pitches calculates the amount of runs relative to the average that a catcher contributes by blocking potential passed balls (Fangraphs, b).

As the various component metrics that sum up to encompass a player's total value can be confusing and difficult to remember, the following table has been included to summarize the different statistics and the aspect of the game they are designed to measure.

Statistic Name	Abbreviation	Aspect of Game Measured
Weighted Runs Above Average	wRAA	Batting
Weighted Stolen Base Runs	wSB	Base Running
Ultimate Base Running	UBR	Base Running
Weighted Grounded Into Double Play Runs	wGDP	Base Running
Positional Adjustment	Pos	Defense
Outfield Arm Runs	ARM	Non-Catcher Defense
Double-Play Runs	DPR	Non-Catcher Defense
Range Runs	RngR	Non-Catcher Defense
Error Runs	ErrR	Non-Catcher Defense
Stolen Base Runs	rSB	Catcher Defense
Runs Saved on Passed Pitches	RPP	Catcher Defense

Figure 3: Positon Player Component Statistics (Fangraphs, g)

Determining pitcher value is much easier to compute than the method used for position players. Pitchers' contributions to their teams are based on a single statistic known as Fielding Independent Pitching (FIP) (Fangraphs, f). FIP is best described in the following paragraph from Fangraphs:

"Fielding Independent Pitching (FIP) is a statistic that estimates a pitcher's run prevention independent of the performance of their defense. FIP is based on outcomes that do not involve defense; strikeouts, walks, hit by pitches, and home runs allowed. FIP uses those statistics and approximates a pitcher's ERA assuming average outcomes on balls in play. While

it is not a complete accounting of pitcher performance, it is generally a better representation of performance that ERA" (Fangraphs, d).

While it is useful to understand the context and methodology behind the various performance metrics, they are all already calculated by Fangraphs and therefore do not to be computed individually (Fangraphs, c).

Discussion

Reviewing the articles and papers describing portfolio optimization and baseball performance metrics has led to several key discoveries regarding the creation of an optimization model to describe baseball roster construction. The first of these findings is that, though portfolio optimization research has been around for over 60 years, it is still a relevant research topic. In fact, ten of the nineteen articles analyzed have been written in the past ten years. Another key observation was made specifically regarding the application of optimization techniques to professional baseball. Though there was an article from "Beyond the Box Score" stating that treating player contracts like portfolio investment decisions could be used to reduce risk, there were no articles found that described the creation of a specific model to analyze the optimal roster construction of a Major League Baseball team. Of the models that were analyzed, all measured expected returns using the measurement unit of dollars. The proposed baseball optimization model utilizes wins contributed relative to a replacement player as the unit of expected returns instead of a monetary amount, which is unique from the models that were studied. One other key observation was that there are already a wide variety of publicly available metrics that assess a player's value to his team. This vastly simplifies the modeling process, as these values can be used to determine expected returns of potential investments.

Future Research

One of the most basic factors that make optimizing a baseball lineup different than a financial portfolio is the goal of the investor. In financial portfolio optimization, the goal is to maximize your expected returns while minimizing the risk of losing money. In creating a baseball lineup, the goal is to create a roster that wins enough games to win the team's division, or at least make the playoffs. While this seems like a relatively simple and arbitrary difference, it has a major factor on the method used to model each scenario. It does not matter if the baseball team fails to reach the postseason by one game or by twenty; the result of each season is essentially the same. Therefore, it makes sense that a model analyzing the construction of a baseball roster should vary slightly from that of a financial model. The proposed baseball model should use mean-variance optimization to determine an efficient set of potential rosters, and then Monte Carlo simulation to determine which of the options has the highest probability of generating enough wins to reach the postseason. The roster with the highest probability of reaching the postseason should then be selected. This varies from Monte Carlo simulation use for financial modeling where, in most of the articles analyzed, Monte Carlo simulation was utilized to determine the risk of losing money, or the probability of certain worst case scenarios. In the proposed baseball model, Monte Carlo simulation will be used to calculate the probability of certain best case scenarios.

Another potential area of future research that will not be analyzed in this paper is the improvement of the metrics used to determine player value. Though the current metrics are generally very accurate and effective at measuring each player's contributions, they are certainly not perfect. Improving the metrics used to value players would vastly improve the quality of any model using the metrics. Kritzman discusses this concept in his defense of mean-variance optimization, where he states that it is not possible to create an accurate model if the expected returns used in the model are inaccurate (Kritzman, 2011). That being said, the current statistics should be effective enough so as to not significantly limit the accuracy of the model.

Performance Metric Utilized

As previously discussed, there are a large number of metrics that can be used to assess the performance of baseball players. After looking at all the different statistics available, we eventually made the decision to solely use Wins Above Replacement (WAR) in the optimization model. WAR attempts to combine a player's total impact into one metric that represents the total amount of wins a player contributes to his team. This win value is given relative to that of a replacement player who can easily be acquired for a negligible cost. This theoretical replacement player is expected to be worth a winning percentage of 0.294, which means a team composed solely of replacement players would win 47.7 games over the course of a MLB season. Therefore, just as the name implies, WAR represents the amount of wins a player contributes to his team above those that would have been provided by a replacement level player (Fangraphs, h). The fact that WAR is an all-encompassing statistic that attempts to factor in almost every aspect of a baseball game makes it an appealing choice to use as the metric in our model. It also helpful that WAR is measured in wins, as wins are the values upon which the success of a baseball team's season is determined. Though multiple different metrics could have been included to factor in various aspects of the game, this would have made the model more difficult to understand and likely increased the necessary computing time by making the model more complicated. WAR, though imperfect, is able to describe virtually the same skills as a set of multiple metrics, but is able to do so while streamlining the model and vastly simplifying the process.

After making this decision, WAR values were recorded for every player in the model going back five seasons. Because we were looking at the 2014-2015 offseason, WAR was recorded for the 2010-2014 seasons. Statistics were recorded for five seasons in order to get a more complete picture of a player's ability rather than just looking at one or two seasons, but still keeping the seasons considered limited to those that are fairly recent. In addition to WAR values, plate appearances (PA) for position players and innings pitched (IP) for pitchers were recorded for each season under consideration. After determining the metric to be utilized, the next step was to calculate the expected WAR and standard deviation of expected WAR for each player in the model. Because WAR is a counting statistic, not a rate statistic, athletes with more playing time in a particular season are likely to have accumulated more WAR, even if their performance was not as good relative to the average as players with less plate appearances or innings pitched. This can be both a good thing and a bad thing, as durable players who are able to stay healthy throughout an entire season and accumulate a large amount of WAR should be rewarded in their expected WAR values. However, some players included in the model were called up from the minor leagues towards the end of the season, or simply stuck as a backup behind a very good player. They should not be penalized for aspects outside of their control. In order to account for this, the equation for expected WAR values utilized a weighted mean, weighting by plate appearances in each season for position players and innings pitched for pitchers, respectively. The equation for the weighted mean is shown below in Figure 4.

$$\overline{x_w} = \frac{\sum_{i=1}^{n}(w_i x_i)}{\sum_{i=1}^{n} w_i}$$

Figure 4: Weighted Mean Formula (Mathematics Stack Exchange)

To create a probability distribution of Wins Above Replacement, expected WAR values were assumed to be normally distributed. In order to test this assumption, a Kolmogorov-

Smirnov test was performed on WAR values for five different players with a relatively large amount of Major League Baseball playing experience. For all five of these tests, we failed to reject the null hypothesis of a normal distribution at a significance level of α=0.01. At a significance level of α=0.05, we failed to reject the null hypothesis of a normal distribution for four of the five samples (see Appendices C-G). The following formula was used to calculate the standard deviation of the weighted mean:

$$sd_w = \sqrt{\frac{\sum_{i=1}^{n} w_i (x_i - \bar{x}_w)^2}{\frac{(n' - 1) \sum_{i=1}^{n} w_i}{n'}}}$$

Figure 5: Standard Deviation of Weighted Mean (Dataplot Reference Manual, 1996)

Because it is not possible to compute a standard deviation value for a player with only one year of experience, a sample standard deviation had to be used. For those players with only a year of experience, a random sample of 15 players at that particular athlete's position were taken, and the average standard deviation from the random set of players was used as the standard deviation.

Pool of Players

In order to create a pool of potential free agents to use in the model, MLB Trade Rumors' Free Agent Tracker was used to create a list of free agents for the 2014-2015 offseason (MLB Trade Rumors). The main source to determine each player's 2015 salary was the website Cot's Baseball contracts (Baseball Prospectus). In cases where there was confusion over what a particular player was being paid, the website Baseball Reference was used as well (Baseball Reference). Players who were unsigned or signed to minor league contracts were assumed to be making the major league minimum of $507,500 if signed to a major league contract (Major

League Baseball Players Association). The website Fangraphs was primarily used to classify the various positions that each player is capable of playing (Fangraphs, c).

Model Concepts

The primary concept of the optimization model was to use Monte Carlo simulation and each player's WAR probability distribution to choose the roster with the highest probability of making the playoffs. In order to achieve this, 1000 "seasons" were simulated, with the objective to maximize the percent of seasons with WAR totals above an arbitrary value that was deemed necessary to make the playoffs.

In order to determine whether or not a season reached the necessary win total, each player was modeled as a binary integer variable, and given a value of "1" if they were selected for the roster and a value of "0" if not. For each season simulated, a random number generation was run for each player using their expected WAR mean and standard deviation using the assumption that WAR follows a normal distribution. Each of these seasons were then multiplied by the binary variables representing which players were on the roster, so that only the selected players were included in the team's win total. The individual WAR values for each season were then added together, giving 1000 different team WAR values. The number 47.7 was then added to each team WAR to represent the replacement level win total. This essentially changed the scale from Wins Above Replacement simply to wins. The final step was to calculate the proportion of the 1000 seasons that had a win total above the predetermined amount considered necessary to qualify for the postseason. This final proportion represented the theoretical likelihood that the specific roster would make the playoffs.

In addition to maximizing the objective function value, there were several constraints that each potential roster was required to meet. The first of these was a budget constraint. In order to determine a team's budget, their 2015 opening day payroll was first considered. The 2015 salaries of players already under contract with the team were considered a sunk cost, and therefore subtracted from the team's actual 2015 opening day payroll. The remaining amount was determined to be the team's available budget to spend on free agents. The sum of 2015 salaries for all free agents signed by the team was required to be less than or equal to the available budget. There were also several roster constraints, based on the number of players at each position, included in order to ensure that the team was able to field a valid lineup. The complete mathematical formulation for the model is shown on the following pages.

<u>Sets</u>
P = set of players

<u>Parameters</u>
W_p = wins above replacement contributed by player p
μ_p = expected wins above replacement contributed by player p
σ_p = standard deviation of expected wins above replacement contributed by player p
T = total wins above replacement needed to qualify for postseason
s_p = salary of player p
B = team budget to spend on free agents
SP_p = 1 if player p is a starting pitcher; 0 otherwise
RP_p = 1 if player p is a relief pitcher; 0 otherwise
P_p = 1 if player p is a pitcher; 0 otherwise
C_p = 1 if player p is a catcher; 0 otherwise
$1B_p$ = 1 if player p is a first baseman; 0 otherwise
$2B_p$ = 1 if player p is a second baseman; 0 otherwise
SS_p = 1 if player p is a shortstop; 0 otherwise
$3B_p$ = 1 if player p is a third baseman; 0 otherwise
LF_p = 1 if player p is a left fielder; 0 otherwise
CF_p = 1 if player p is a center fielder; 0 otherwise
OF_p = 1 if player p is an outfielder; 0 otherwise
IF_p = 1 if player p is an infielder; 0 otherwise

<u>Variables</u>
x_p = 1 if player p is selected for roster; 0 otherwise

Figure 6: Optimization Model Sets, Parameters, and Variables

Objective	Goal
Max $P(\sum_{p\in P}[[W_p \sim N\left(\mu_p, \sigma_p{}^2\right)]x_p] \geq T)$	Maximize probability of sum of wins being greater than or equal to amount needed to qualify for postseason with each player's wins contributed following a normal distribution
S.T.	Rules
$\sum_{p\in P} x_p = 25$	Roster must have exactly 25 players
$\sum_{p\in P} s_p x_p \leq B$	Sum of each player's salary must be less than or equal to the team budget
$\sum_{p\in P} SP_p x_p \geq 5$	Sum of starting pitchers must be greater than or equal to 5
$\sum_{p\in P} RP_p x_p \geq 7$	Sum of relief pitchers must be greater than of equal to 7
$\sum_{p\in P} P_p x_p = 12$	Sum of pitchers must equal 12
$\sum_{p\in P} C_p x_p \geq 2$	Sum of catchers must be greater than or equal to 2
$\sum_{p\in P} 1B_p x_p \geq 1$	Sum of first basemen must be greater than or equal to 1
$\sum_{p\in P} 2B_p x_p \geq 1$	Sum of second basemen must be greater than or equal to 1
$\sum_{p\in P} SS_p x_p \geq 1$	Sum of shortstops must be greater than or equal to 1
$\sum_{p\in P} 3B_p x_p \geq 1$	Sum of third basemen must be greater than or equal to 1
$\sum_{p\in P} LF_p x_p \geq 1$	Sum of left fielders must be greater than or equal to 1
$\sum_{p\in P} CF_p x_p \geq 1$	Sum of center fielders must be greater than or equal to 1
$\sum_{p\in P} RF_p x_p \geq 1$	Sum of right fielders must be greater than or equal to 1
$\sum_{p\in P} OF_p x_p \geq 5$	Sum of outfielders must be greater than or equal to 5
$\sum_{p\in P} IF_p x_p \geq 6$	Sum of infielders must be greater than or equal to 6
$x_p \in \{0,1\}\ \forall\ p \in P$	Binary variable constraints

Figure 7: Optimization Model Mathematical Formulation

Because the objective function is not linear, the simplex method or other traditional linear optimization techniques could not be used to solve the problem. Instead, an Evolutionary

Algorithm was utilized to find a solution. The Evolutionary Algorithm is a genetic algorithm that uses random sampling in an attempt to improve the current best solution. If the current solution cannot be improved in a set amount of time, it is accepted as the final solution (Frontline Solvers). For this particular problem, the amount of time allotted to improve the current solution was set to 15 minutes. Because the algorithm is just accepting a solution after a certain time period, it cannot guarantee that the solution is actually optimal. It can only guarantee that it was the best solution found in the allotted time (Frontline Solvers).

First Modeling Attempt

In order to test the capability and accuracy of the model, a prototype was built based on the 2014-2015 offseason for the Royals, Red Sox, and Padres. A combination of the each team's offseason outlook on the site MLB Trade Rumors and their actual opening day 25 man roster was used to project players already available on their payroll (Adams, 2014; Glaser, 2015; Links, 2014; McAdam, 2015; Polishuk, 2014; Rieper, 2015). Because these players' salaries were considered sunk costs and already subtracted from the available budget, they were given a 2015 salary of zero dollars in the model. A list of 206 free agents was also created, primarily based on MLB Trade Rumors' free agent tracker, with the sites Fangraphs and Baseball Reference used to validate the information. The Excel solver engine only allows for 200 decision variables, however, so the free agents with the lowest expected performance were removed until only 200 players remained in the model.

In the initial optimization attempt, the solver engine built into Excel was used with the evolutionary solving method selected. The model took a large amount of computing time, though, and the built in solver engine was unable to find an effective solution in the time allotted. It was determined that, in order to improve the quality of the solution, a faster solving engine than that already available in Excel needed to be utilized. In order to solve this problem, the Frontline Solvers Analytic Solver Platform was downloaded and used instead. While the program still limited the amount of decision variables to 200, it was able to find a much better solution while reducing the necessary computing time. The initial objective function value used was the proportion of simulated seasons with 92 or more wins. The results of this first optimization, along with the actual WAR of the each team's opening day roster, are shown in Figure 8.

Team	Probability of 92 or More Wins	Expected WAR of Optimized Roster	Actual 2015 WAR of Optimized Roster	Actual 2015 Opening Day Roster WAR
Padres	88.8%	53.91	16.6	23.4
Red Sox	100%	63.79	28.4	19.1
Royals	96.7%	58.13	31.7	35.7

Figure 8: Initial Optimization Results

As evidenced by the fact that the difference between the optimized roster's expected 2015 WAR and actual WAR was over 25 wins for all teams, the initial values over-projected performance by a significant amount and were generally inaccurate. Only one of the three optimized rosters, the Red Sox roster, was able to beat their corresponding opening day roster with regards to actual WAR.

There are multiple reasons why this first optimization attempt was ineffective. First, it did not penalize players who had missed full seasons due to injury in the past. These players should have been considered injury risks that were likely to miss significant chunks of time again, but instead the model did nothing to account for this. As a byproduct of weighting all seasons based solely on playing time without including how recently the season occurred, the model also over-projected players who had performed worse in more recent seasons and were trending downward. Likewise, it under-projected players who had performed better in more recent seasons and were trending upward.

Modeling Adjustments

Several adjustments were added in order to account for these inadequacies. The first of these was to penalize players who had missed full seasons due to injury. In order to account for this, players were given a WAR value of zero for any season in which they did not play following their debut during the five year period we considered. In order to determine the amount of plate appearances or innings pitched to weight these seasons by, the random sample of 15 players at each position described earlier was again considered. The average plate appearances or innings pitched of the 15 players at the same position as the one missing time was used as the sample size for missed seasons. This is because the amount of playing time received differs based on the position being analyzed. Relief pitchers will not throw as many innings in a season as a starter, so the weight of a missed season by a reliever should be reflected accordingly. Another key adjustment to the model was that, in addition to weighting by the number of PA or IP in each season, recent seasons were weighted more heavily. Several different weighting iterations were used in order to find the best possible projections. A summary of these different iterations is given in Figure 9. The figure describes the amount by which each season was weighted in addition to PA or IP, along with the percentage of players whose actual 2015 performance fell within, above, or below one standard deviation of their expected performance at each weight.

	Weights					Percentage of Players at Least 1 Standard Deviation Below Expected Performance	Percentage of Players at Least 1 Standard Deviation Above Expected Performance	Percentage of Players Within 1 Standard Deviation of Expected Performance
	2014	2013	2012	2011	2010			
Attempt 1	1	1	1	1	1	40.5%	9.9%	49.6%
Attempt 2	3	3	2	2	1	34.7%	10.6%	54.7%
Attempt 3	3	2	2	1	1	32.5%	10.6%	56.9%
Attempt 4	4	3	2	1	1	30.3%	12.0%	57.7%
Attempt 5	3	2	1	0	0	29.2%	18.2%	52.6%
Attempt 6	5	3	1	0	0	29.9%	17.9%	52.2%

Figure 9: Analysis of Weighting Alternatives

The alternative eventually chosen was Attempt 5, which weighted 2014 by three, 2013 by two, and 2012 by one, but did not consider 2010 or 2011. Though this weighting system did not have the highest percentage of players within one standard deviation of expected performance, it had a more even distribution of players above and below the projections than any of the options with more players within one standard deviation. This should lead to this method being less likely to systematically over-project performance. Another key factor in the selection of Attempt 5 is that, under the initial constraints, it performed better than any other weighting system with regards to actual 2015 WAR, as demonstrated in Figure 10 below.

2015 WAR of Optimized Rosters						
Team	Attempt 1	Attempt 2	Attempt 3	Attempt 4	Attempt 5	Attempt 6
Padres	16.6	18.8	18.8	21.5	24	19.8
Red Sox	28.4	34.7	40.5	38.6	37.5	33.7
Royals	31.7	24.2	33.7	29.7	32.6	34
Total	76.7	77.7	93	89.8	94.1	87.5

Figure 10: Performance of Weighting Alternatives Based on 2015 WAR

It is worth noting that there is still an 11% difference between players above one standard deviation and those below one standard deviation, meaning the likelihood of over-projection remains somewhat high. However, it should still be able to provide relatively accurate and balanced projections, and was therefore the weighting system utilized for the remainder of the project.

In addition to changing the weighting system in order to improve the model projections, the objective function was changed from the proportion of seasons with greater than or equal to 92 wins to the proportion of seasons with greater than or equal to 85 wins. Though somewhat arbitrary in nature, the thought process was that 85 wins was a more reasonable expectation for an opening day roster. It is likely that a team will need to win more than 85 games in order to make the playoffs, but the opening day 25 man roster shouldn't be expected to provide all of those wins. Injuries are inevitable, along with midseason trades or players being called up from the minor leagues, meaning that it will actually be more than 25 players that have a role in contributing the necessary wins to make the playoffs. The Royals, for instance, ended up winning the World Series in 2015 and accrued a total of 35.7 WAR from their opening day roster. Combined with the 47.7 wins that a replacement level team provides, their opening day roster would reach approximately 83.4 wins based on the theory used in this model. Thus, it seems reasonable to conclude that a team whose opening day 25 man roster could generate the WAR necessary to reach 85 wins would be highly likely to make the playoffs.

Updated Stochastic Optimization Model

After changing both the projection system and the objective the function, the rosters were again optimized, using the same sets of players already on the payroll and the best available free agents until the limit of 200 decision variables was reached. The results of these optimization runs are shown in Figure 11.

Team	Probability of 85 or More Wins	Expected WAR of Optimized Roster	Actual 2015 WAR of Optimized Roster	Actual 2015 Opening Day Roster WAR
Padres	79.5%	41.16	25.5	23.4
Red Sox	99.4%	48.81	34.9	19.1
Royals	99.1%	48	36.9	35.7

Figure 11: Updated Stochastic Optimization Results

This time, the model generates a roster whose 2015 WAR is actually better than that of the corresponding opening day roster for all three teams. This represents a vast improvement over the previous attempt. However, it is still worth noting that the actual 2015 WAR of the optimized roster falls significantly short of the expected value in every instance. Also, none of the teams actually meet their objective of 85 wins. Adding the 36.9 WAR of the Royals' optimized roster, the best of the three analyzed, to the 47.7 wins generated by a replacement level roster gives us a total of 84.6 wins contributed by the optimized roster. Though a very good result, it still falls short of the 85 wins that the roster was supposed to have a 99.1% chance of reaching. In spite of these deficiencies, the model still represents a significant success in its ability to beat the actual opening day roster WAR of all three teams, including the Royals, who finished the 2015 season with the best record in the American League (MLB.com, a).

Other Optimization Methods

In addition to the stochastic optimization method described throughout this paper, three attempts using a linear objective function were also made in order to see how they compared to the stochastic results. Because they had a linear objective function, these methods utilized the simplex method to reach an optimal solution instead of the evolutionary method used in the stochastic model. In the first of these methods, the same projections included in the stochastic optimization described above were used. The same constraints were used as in the previous model. The only difference was that the objective function was changed from maximizing the amount of seasons with at least 85 wins to simply maximizing the expected Wins Above Replacement. The results of this optimization run are shown below, with the results of the previous stochastic method included for comparison.

Team	Expected WAR of Linear Optimization Roster	Actual 2015 WAR of Linear Optimization Roster	Actual 2015 WAR of Stochastic Optimization Roster	Actual 2015 Opening Day Roster WAR
Padres	44.14	20.8	25.5	23.4
Red Sox	58.38	39.9	34.9	19.1
Royals	52.98	34.5	36.9	35.7
Total	155.5	95.2	97.3	78.2

Figure 12: Linear Optimization

The linear optimization method fails to beat the stochastic method on the aggregate with regards to actual WAR. While this is certainly favorable evidence towards the benefits of a stochastic method, it does not guarantee that the stochastic model is actually any better. As discussed previously, the evolutionary algorithm used in the stochastic method cannot guarantee an optimal solution. It simply provides the best answer it can find in the amount of time allotted (Frontline Solvers). Therefore, it is very possible that, given more time to find an improved solution, the stochastic method would have reached the same solution as the one given by the linear method.

In addition to utilizing linear optimization with the preexisting projections, linear regression was also used to create new and ideally more accurate projections. This method was driven by the theory that the model was still biased towards older players on the decline of their careers. If age could be factored into the model, then possibly better answers could be found. In order to create the equation to test this theory, the weighting system factoring in 3 seasons was scaled back one year and used to find expected WAR values for the 2014 season. Ages for each player in 2014 were also recorded. A multivariate linear regression was then run with actual 2014 WAR as the response variable and 2014 expected WAR and 2014 age as the predictor variables. The results of this regression are displayed in Figure 13.

	Coefficients	Standard Error	t Statistic	P-value
Intercept	1.368779077	0.642274299	2.13114409	0.034030211
Expected Wins	0.602299409	0.064322956	9.363677331	4.13487E-18
Age	-0.045708801	0.020862512	-2.1909538	0.029356876

Figure 13: Linear Regression Output

The coefficients from this equation were then applied to each player's 2015 age and 2015 expected WAR in order to generate a new set of projections. With these new projections, another optimization run was made with the objective of maximizing the expected wins. The same constraints as used previously were also included. The outcome of this particular run is as follows, with the results of the previous stochastic optimization again included for comparison.

Team	Expected WAR of Linear Optimization Roster	Actual 2015 WAR of Linear Optimization Roster	Actual 2015 WAR of Stochastic Optimization Roster	Actual 2015 Opening Day Roster WAR
Padres	25.94	24.1	25.5	23.4
Red Sox	33.18	41.4	34.9	19.1
Royals	32.47	37.8	36.9	35.7
Total	91.59	103.3	97.3	78.2

Figure 14: Linear Optimization with Projections via Linear Regression

The optimization with projections from linear regression is higher than both the stochastic method and the opening day rosters with regards to total 2015 WAR generated. It is also the only method that did not over-project 2015 performance. In fact, the model significantly under-projected the 2015 WAR for the Red Sox and the Royals' rosters. The primary weakness of this method is that it gives you a single estimate for performance, while the stochastic method gives you a distribution of potential outcomes. However, the projections utilizing linear regression seem to be significantly more accurate than those that rely solely on weighting previous seasons.

In an effort to build off the success of the initial regression model, a second regression attempt was made with the goal of increasing the accuracy of the projections. This attempt used the same methodology as the first, with age and expected WAR included as the predictor variables. The only difference is that players were separated by position group, with separate regression equations created for each group instead of simply creating one equation to project all players, as was done previously. The thought process behind this model was that players age differently depending on their position, and different projections should be used to account for the varied impact of age. For instance, it is likely that aging will have a greater impact on an outfielder who has to frequently run down fly balls than it will a relief pitcher, who is often only in the game for one or two innings at a time.

In order to implement this concept, players were split into one of six position groups: starting pitchers, relief pitchers, catchers, corner infielders (first and third basemen), middle infielders (second basemen and shortstops), and outfielders. Linear regression was then used to project 2014 WAR for each player, with 2014 age and expected WAR as the predictor variables. The results of these regressions can be seen in Appendices H-M. The coefficients for each position group were then utilized, along with 2015 age and expected WAR values, to create WAR projections for every player. After these new projections were created, the same constraints as described in the previous model were used to create optimized rosters for each team. The results of these optimizations are shown below.

Team	Expected WAR of Optimized Roster	Actual 2015 WAR of Optimized Roster	Actual 2015 Opening Day Roster WAR
Padres	28.28	28.5	23.4
Red Sox	37.17	49	19.1
Royals	34.11	31.4	35.7
Total	99.56	108.9	78.2

Figure 15: Linear Optimization with Separate Linear Regression Projections by Position

The results of this particular set of optimizations are very promising, but also somewhat inconsistent. The total 2015 WAR of 108.9 is the highest of any of the optimization attempts, and the Red Sox 2015 WAR of 49 is the highest of any individual team. However, the Royals' optimized roster failed to beat their opening day roster by a significant margin, while both the initial regression and stochastic methods had no trouble surpassing the WAR of the Royals' actual roster. Overall, though, there does seem to be potential for a substantial amount of value from separating the projections by position group.

Weaknesses of Model

Overall, both the stochastic optimization and regression methods seem to provide very reasonable answers, as all three were able to beat the opening day rosters with regards to total 2015 WAR. This does not mean they are not without weaknesses, though. The stochastic method consistently over-projects future performance, which significantly hinders its usefulness. In fact, the primary way to improve the model is likely to improve the quality of the projections used. It does not matter how good the model is if the input data that it uses is inaccurate. The model also fails to factor in the correlation of performance between players. If you have two above average position players at a position where only one of them can play at a time, the value of each player is significantly reduced by having to split playing time. For instance, going into the 2015 offseason, the Royals already had catcher Salvador Perez on their roster, who accounted for 3.1 WAR in 2014 and is generally considered a well above average catcher (Fangraphs, c). However, the model still suggested that the Royals sign free agent catcher Russell Martin, who was worth 5 WAR in 2014 (Fangraphs, c). While it would be possible to use one of the players as a designated hitter when not catching, only one of them would be able to play catcher at a time, reducing the playing time, and therefore value, of each. In its current form, the model does not account for this at all. The model is also unable to account for future payroll obligations or back loaded contracts, which leads to the risk of recommending players who are being paid large sums of money while their skills are declining. This means that, though the players selected may return the most value of any potential roster in the upcoming season, their selection may negatively impact the team in future seasons. In addition to this, the model is also unable to account for trades or include players with no major league experience, which can limit the ability of the model to find good solutions. The Royals', for example, made several trades prior to the trade deadline in 2015 in an attempt to improve their roster on the way to winning the World Series. As a result, their World Series roster actually had a 2015 WAR of 43.2, which is higher than any of our optimized rosters for the team (Fangraphs, c; Pekarsky, 2015). While the Royals were required to give up several of their own players in order to make these trades, this still demonstrates that there are other ways to create value for a team outside of the free agent market. The final, and potentially most important, weakness is that the model is reliant on the assumption that every free agent would be willing to sign a contract with the team being modeled at the price listed. This assumption is not extremely realistic, as all players have different preferences on where they would like to play and whether or not they would take a discount to play in certain cities.

One technique that may help address some of these potential issues would be to run the model incrementally as players are signed or become unavailable. In order to demonstrate this concept, the Boston Red Sox roster is considered along with the most recently discussed optimization technique that utilizes separate regression equations for each position group. One of the players that the model recommends for the Red Sox in this instance is starting pitcher Jon Lester. It is very possible that, as the Red Sox are attempting to reach a contract agreement with Jon Lester, one or more of the other recommended players would sign a contract with a different team. If we assume that the Red Sox were able to sign Jon Lester on December 15, 2014, (the date Lester actually did sign with the Chicago Cubs), several other players would have already signed elsewhere. One such player is Russell Martin, who was recommended by the model, but agreed to a contract with the Toronto Blue Jays on November 18 (MLB.com, b). With Martin unavailable, the optimal roster the Red Sox should pursue has changed. Therefore, it is necessary to adjust the model and run it again after Jon Lester is signed by the Red Sox. In order to do this, Lester's 2015 salary is subtracted from the budget and he is added to the set of players already available on the Red Sox payroll. All players who have signed contracts with other teams are removed from the model. The model is then run again, using the same constraints and objective as before. The roster recommended by this optimization is shown in Appendix S. It is interesting to note that two players recommended by the initial optimization are no longer recommended by the model despite the fact that they are still available. Though it may be unrealistic to assume the Red Sox would just allow one of their free agent targets to join another team without any effort to sign him, this example is still useful in demonstrating how incremental usage of the model can be leveraged to drive the highest amount of value possible in a dynamic free agent market that is constantly changing.

Conclusion

Overall, the model was able to successfully demonstrate how a structured approach can be applied to the construction of a Major League Baseball team. Using portfolio optimization techniques and related concepts, three different optimization methods were created that were able to successfully assemble rosters that returned a large amount of value while meeting several necessary constraints. Though not without weaknesses, these methods were able to illustrate the power and usefulness of portfolio optimization methodology, and how they can be successfully applied to scenarios where this type of problem solving is not traditionally implemented.

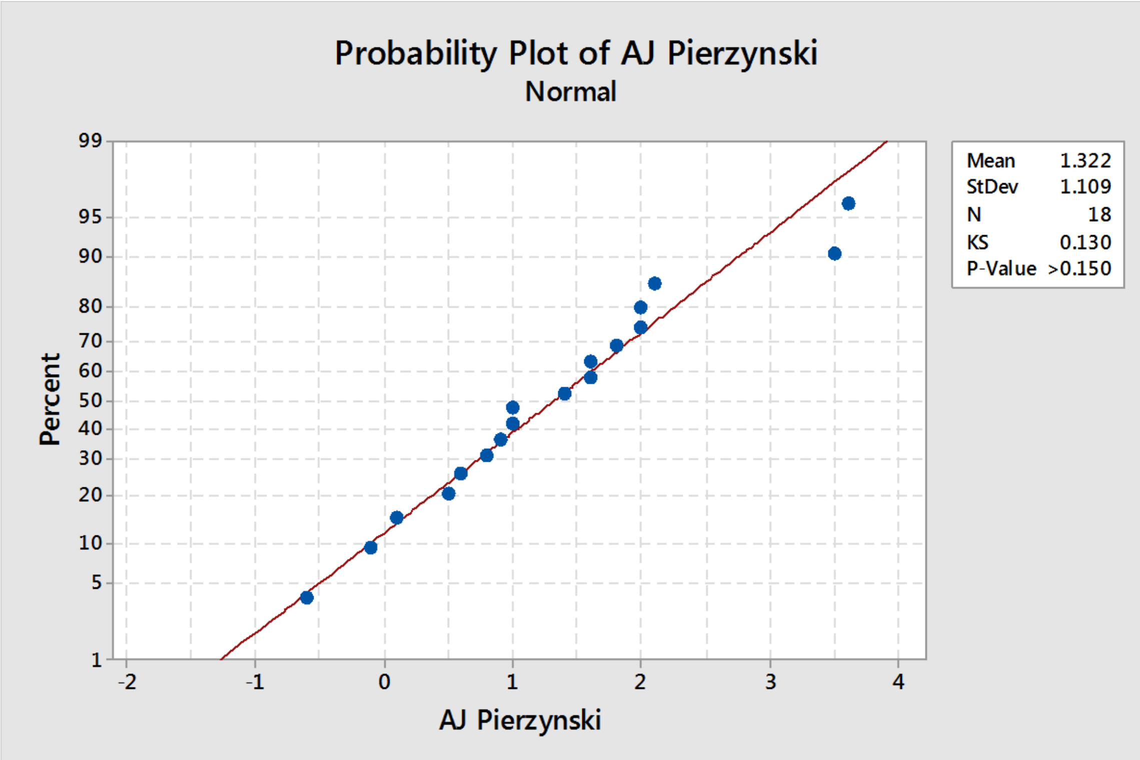
Appendix A: Portfolio Optimization Article Summary

Article Title	Author	Year
Portfolio Selection	Harry Markowitz	1952
Portfolio Optimization: A Primer	Speidell, Miller, and Ullman	1989
Monte Carlo Simulation in Risk Analysis	Hercules E. Haralambides	1991
Global Portfolio Optimization	Fischer Black and Robert Litterman	1992
Portfolio Optimization in Practice	Philippe Jorion	1992
Using Monte Carlo Simulation to Improve Long-Term Investment Decisions	Charles F. Kelliher and Lois S. Mahoney	2000
A Systematic Approach Integrating Risk and Strategy Management to Optimize Portfolios of Industrial Assets	David A. Wood	2001
The Problems with Monte Carlo Simulation	David Nawrocki	2001
Simulation Modeling to Optimize Stochastic Manufacturing Processes and Resources by a Dynamic Monte Carlo Method	Roberto F. Lu and Guixo Qiao	2003
A Stochastic Programming Approach to Power Portfolio Optimization	Sen, Yu, and Genc	2006
Exploring Monte Carlo Simulation Applications for Project Management	Young Hoon Kwak and Lisa Ingall	2007
Portfolio Optimization Using Stochastic Programming	Erhan Deniz and James T. Luxhoj	2008
Accelerated Ensemble Monte Carlo Simulation	Kevin Thompson and Alistair McLeod	2009
Multiobjective Optimization Using Differential Evolution for Real-World Portfolio Optimization	Thiemo Krink and Sandra Paterlini	2011
A Recommended Financial Model for the Selection of Safest Portfolio by Using Simulation and Optimization Techniques	Kirti Arekar and Sanjeevani Kumar	2011
The Graceful Aging of Mean-Variance Optimization	Mark Kritzman	2011
Portfolio Selection Under Model Uncertainty: A Penalized Moment-Based Optimization Approach	Jonathan Y. Li and Roy H. Kwon	2012
A Simulation Model to Analyze the Impact of Golf Skills and a Scenario-based Approach to Options Portfolio Optimization	Soonmin Ko	2012
Markowitz's Portfolio Selection Model and Related Problems	Abhijit Ravipati	2012

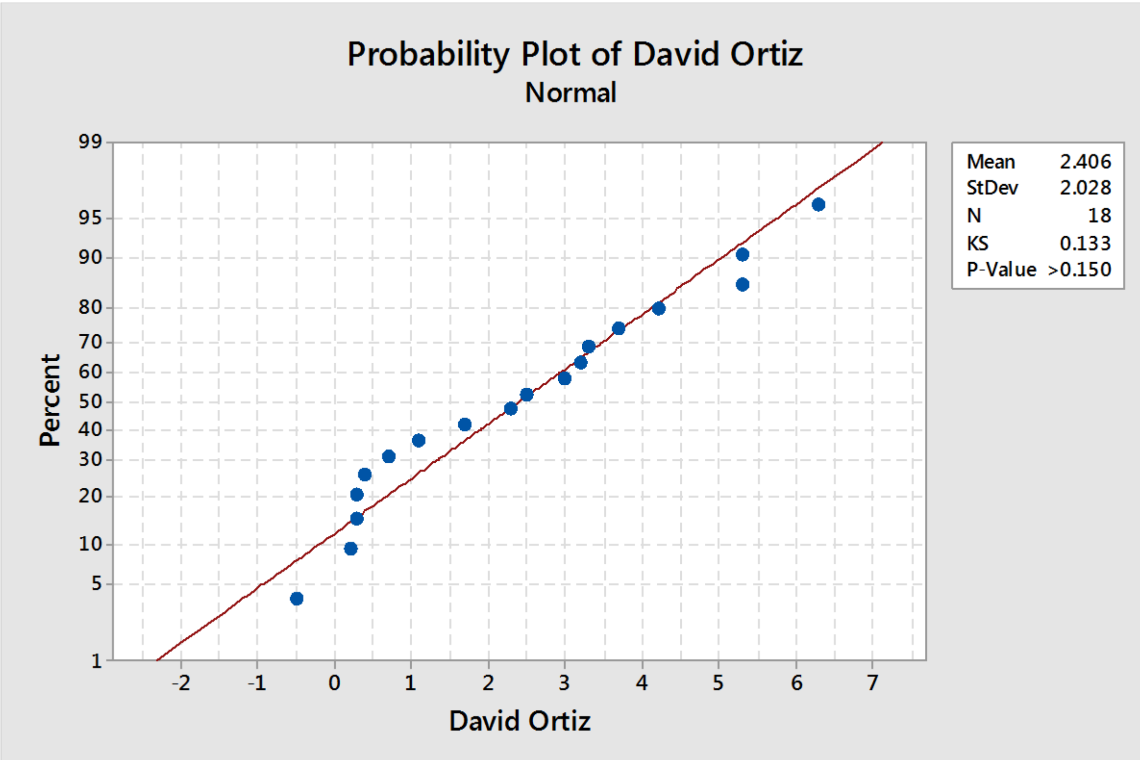
Appendix B: Baseball Performance Metric Article Summary

Article Title	Author	Location	Year
What is WAR?	Fangraphs	Fangraphs	n.d.
The Book: Playing Percentages in Baseball	Tango, Lichtman, and Dolphin	Potomac Books, Inc.	2007
WAR for Position Players	Fangraphs	Fangraphs	n.d.
WAR for Pitchers	Fangraphs	Fangraphs	n.d.
The Relationship between WAR and Team Wins	Sports Reference	Sports Reference	2012
The Beginner's Guide to Replacement Level	Neil Weinberg	Fangraphs	2015
Ultimate Base Running Primer	Mitchel Lichtman	Fangraphs	2011
wSB	Fangraphs	Fangraphs	n.d.
wRAA	Fangraphs	Fangraphs	n.d.
FIP	Fangraphs	Fangraphs	n.d.
Pitching and Defense: How Much Control Do Hurlers Have?	Voros McCracken	Baseball Prospectus	2001
BsR	Fangraphs	Fangraphs	n.d.
UZR	Fangraphs	Fangraphs	n.d.
Def	Fangraphs	Fangraphs	n.d.
Regression to the Mean and Beta Distributions	Kincaid	3-D Baseball	2011
Randomness, Stabilization, & Regression	Steve Staude	Fangraphs	2013
Converting Runs to Wins	Graham MacAree	Fangraphs	n.d.
The Fangraphs UZR Primer	Mitchel Lichtman	Fangraphs	2010
Catcher Defense	Fangraphs	Fangraphs	n.d.
The Average Number of Pitches Thrown Per Game is Rising	Andy	Baseball-Reference	2010
wOBA	Fangraphs	Fangraphs	n.d.

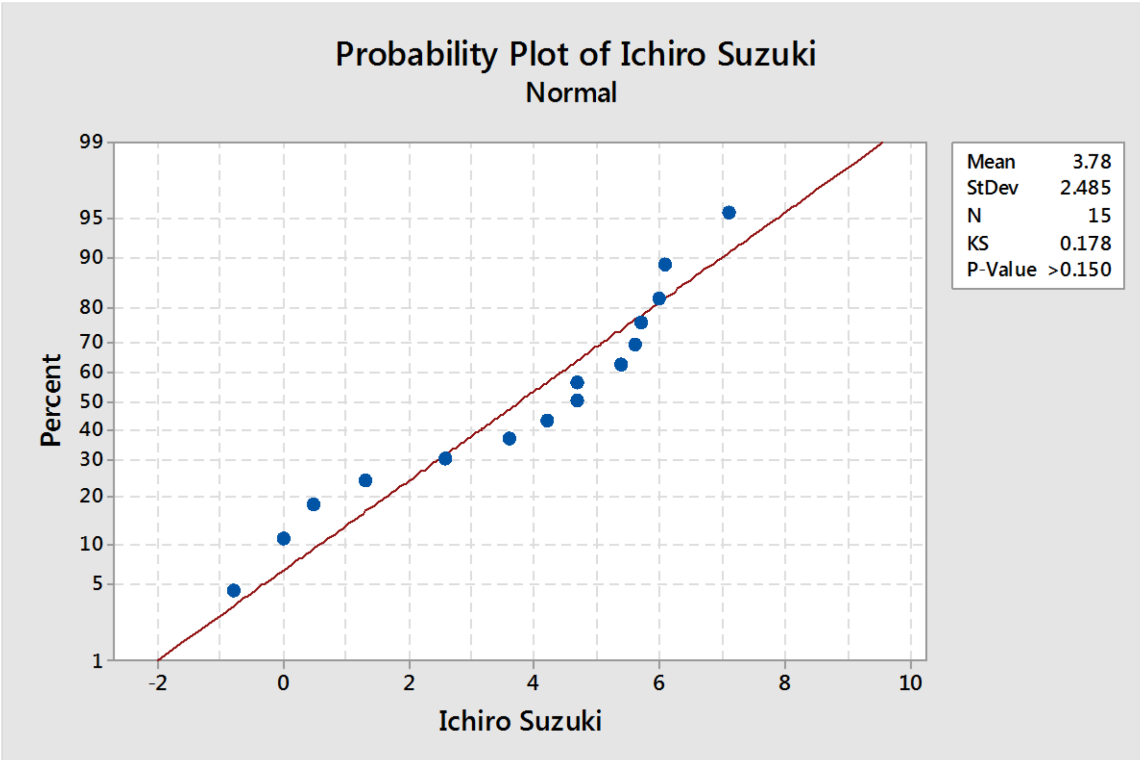
Appendix C: Kolmogorov-Smirnov Test for A.J. Pierzynski Career WAR Values



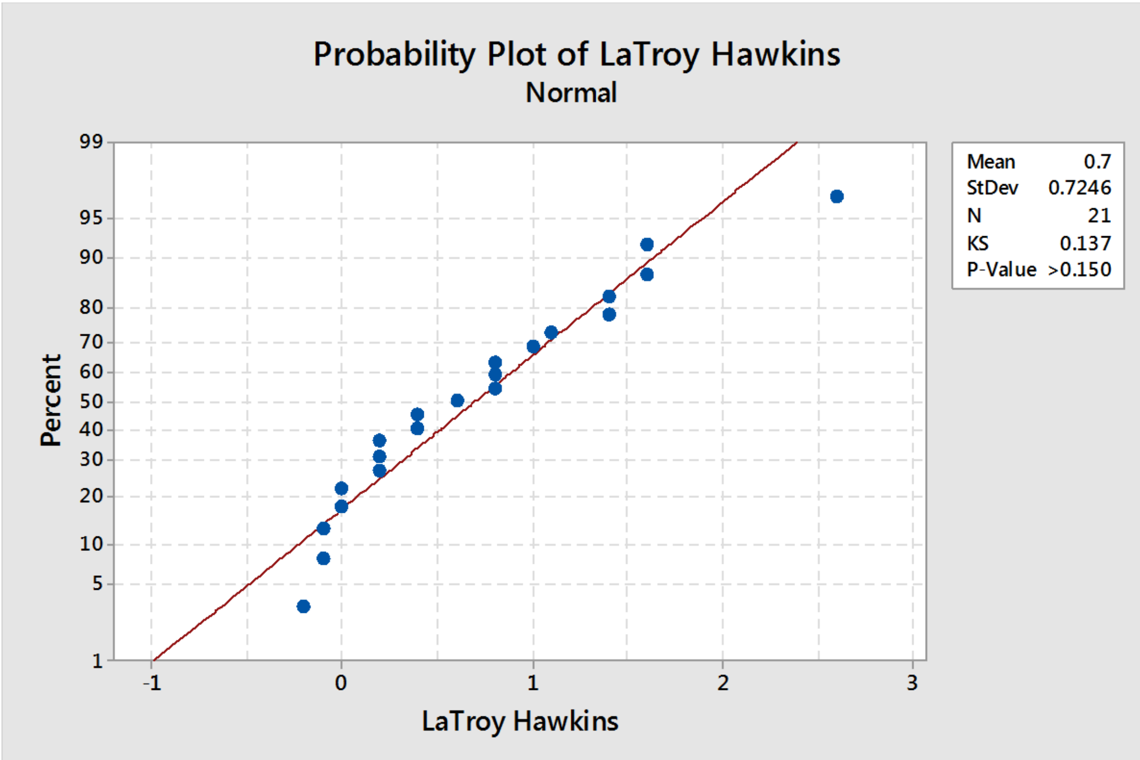
Appendix D: Kolmogorov-Smirnov Test for David Ortiz Career WAR Values



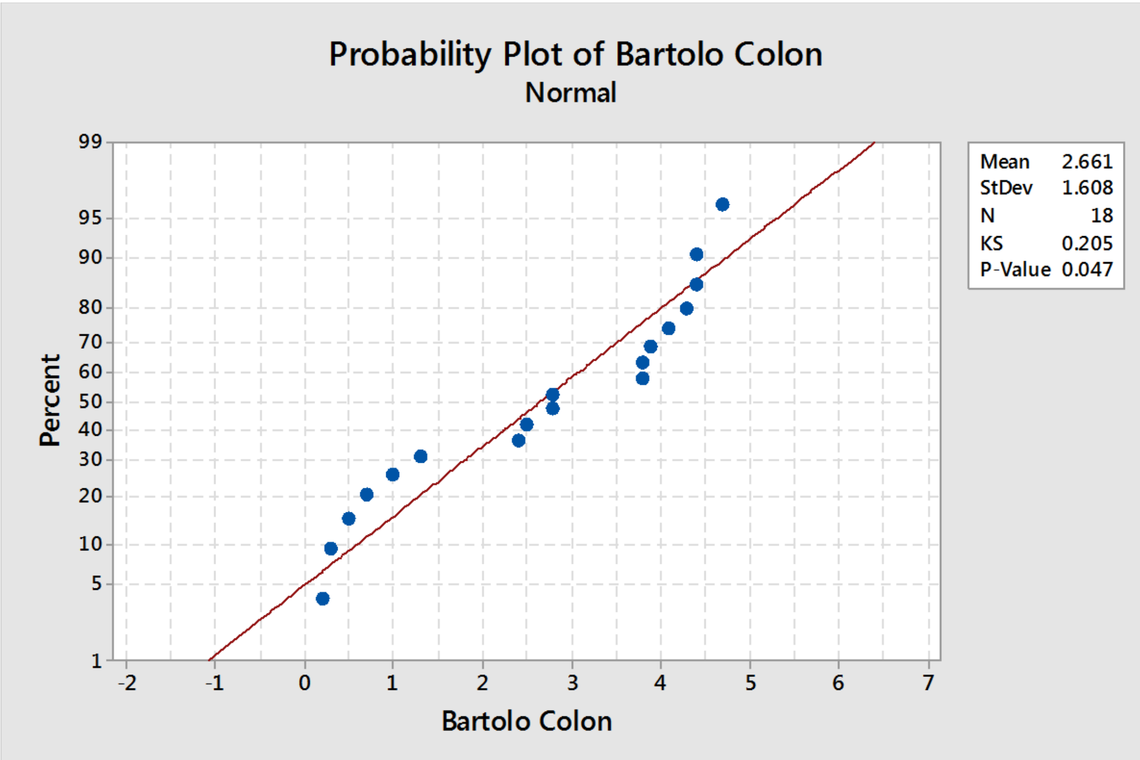
Appendix E: Kolmogorov-Smirnov Test for Ichiro Suzuki Career WAR Values



Appendix F: Kolmogorov-Smirnov Test for LaTroy Hawkins Career WAR Values



Appendix G: Kolmogorov-Smirnov Test for Bartolo Colon Career WAR Values



Appendix H: Starting Pitcher Regression Results

	Coefficients	Standard Error	t Statistic	P-value
Intercept	2.478172537	1.258812438	1.96865908	0.053695692
Expected Wins	0.696277941	0.134120046	5.19145321	2.70267E-06
Age	-0.08262571	0.04180108	-1.9766405	0.052762403

Appendix I: Relief Pitcher Regression Results

	Coefficients	Standard Error	t Statistic	P-value
Intercept	0.410099811	0.746440679	0.549407103	0.584358322
Expected Wins	0.458400488	0.141486212	3.239895114	0.001783638
Age	-0.012975072	0.023541425	-0.551159173	0.58316283

Appendix J: Catcher Regression Results

	Coefficients	Standard Error	t Statistic	P-value
Intercept	3.3204115	3.287530752	1.010001655	0.330924681
Expected Wins	0.922375145	0.314559188	2.932278508	0.01166222
Age	-0.115752855	0.100933346	-1.146824709	0.272123978

Appendix K: Corner Infielder Regression Results

	Coefficients	Standard Error	t Statistic	P-value
Intercept	-3.216330186	2.213254287	-1.453213129	0.155608701
Expected Wins	0.428979417	0.187940648	2.282526011	0.029034621
Age	0.108704266	0.075281107	1.443978036	0.158172829

Appendix L: Middle Infielder Regression Results

	Coefficients	Standard Error	t Statistic	P-value
Intercept	2.366594021	2.393957368	0.988569827	0.334127977
Expected Wins	0.474226324	0.194105272	2.443139838	0.023482305
Age	-0.074844167	0.083361899	-0.897822242	0.379458972

Appendix M: Outfielder Regression Results

	Coefficients	Standard Error	t Statistic	P-value
Intercept	3.147476434	2.139700677	1.470989128	0.149117767
Expected Wins	0.625094347	0.180353921	3.465931557	0.001276141
Age	-0.105724441	0.070418783	-1.501367069	0.141114108

Appendix N: Boston Red Sox Actual 2015 Opening Day Roster

Name	Position 1	Position 2	Position 3	2015 Salary (in millions of dollars)	2015 WAR
Mookie Betts	CF	2B	RF	0.5145	4.8
Xander Bogaerts	SS	3B		0.543	4.3
Craig Breslow	RP			2	-0.6
Clay Buchholz	SP			12.25	3.2
Allen Craig	1B	RF	LF	5.5	-0.9
Ryan Hanigan	C			3.5	0.6
Brock Holt	2B	3B	RF	0.5305	2.4
Tommy Layne	RP			0.557	0.1
Sandy Leon	C			0.5104	-0.5
Justin Masterson	SP			9.5	-0.2
Wade Miley	SP			3.666667	2.6
Edward Mujica	RP			4.75	-0.5
Mike Napoli	1B			16	0.7
Daniel Nava	RF	LF	1B	1.85	-0.2
Alexi Ogando	SP	RP		1.5	-0.9
David Ortiz	DH	1B		16	2.8
Dustin Pedroia	2B			12.625	2.5
Rick Porcello	SP			12.5	1.6
Hanley Ramirez	SS	3B	LF	19.75	-1.8
Robbie Ross	RP			0.5665	0.1
Pablo Sandoval	3B			17.6	-2
Junichi Tazawa	RP			2.25	1.1
Anthony Varvaro	RP			0.5765	0.1
Shane Victorino	RF	CF	LF	13	0
Steven Wright	RP	SP		0.5105	-0.2

Appendix O: Boston Red Sox Roster via Stochastic Optimization

Name	Position 1	Position 2	Position 3	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Nori Aoki*	RF			4.7	1.999	1.5
Mookie Betts	CF	2B	RF	0.5145	1.800	4.8
Yoenis Cespedes	LF			10.5	2.955	6.7
Brock Holt	2B	3B	RF	0.5305	1.964	2.4
Jed Lowrie*	2B	SS		8	2.590	1
Russell Martin*	C			7	4.178	3.5
Edward Mujica	RP			4.75	0.183	-0.5
Mike Napoli	1B			16	2.910	0.7
Daniel Nava	RF	LF	1B	1.85	2.037	-0.2
David Ortiz	DH	1B		16	2.733	2.8
Dustin Pedroia	2B			12.625	4.573	2.5
Joel Peralta*	RP			2.5	0.667	-0.3
Yohan Pino*	SP			0.5075	0.700	0.2
David Robertson*	RP			10	1.616	1.9
David Ross*	C			2.5	0.652	0.1
Sergio Santos*	RP			0.5075	0.062	-0.1
Max Scherzer*	SP			17.142857	5.384	6.4
James Shields*	SP			10	3.635	1.1
Ichiro Suzuki*	RF			2	1.279	-0.8
Junichi Tazawa	RP			2.25	0.936	1.1
Jose Veras*	RP			0.5075	0.283	0
Shane Victorino	RF	CF	LF	13	3.887	0
Ryan Vogelsong*	SP			4	1.147	0
Randy Wolf*	SP			0.5075	0.207	0.1
Jamey Wright*	RP			0.5075	0.433	0

* Indicates a player that was a free agent prior to optimization

Appendix P: Boston Red Sox Roster via Linear Optimization

Name	Position 1	Position 2	Position 3	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Mookie Betts	CF	2B	RF	0.5145	1.800	4.8
Clay Buchholz	SP			12.25	1.663	3.2
John Buck*	C			0.5075	1.009	0
Chris Capuano*	SP	RP		5	1.488	-0.1
Yoenis Cespedes	LF			10.5	2.955	6.7
Bruce Chen*	SP	RP		0.5075	1.023	-0.2
Dane De La Rosa*	RP			0.5075	0.916	0
Chris Denorfia*	LF	RF		2.6	1.928	0.8
Chase Headley*	3B			13	4.654	1.5
Brock Holt	2B	3B	RF	0.5305	1.964	2.4
Jed Lowrie*	2B	SS		8	2.590	1
Russell Martin*	C			7	4.178	3.5
Kris Medlen*	SP	RP		2	1.551	0.5
Brandon Morrow*	SP	RP		2.5	0.955	0.5
Mike Napoli	1B			16	2.910	0.7
Daniel Nava	RF	LF	1B	1.85	2.037	-0.2
David Ortiz	DH	1B		16	2.733	2.8
Dustin Pedroia	2B			12.625	4.573	2.5
Wandy Rodriguez*	SP			0.5075	1.187	1
Max Scherzer*	SP			17.142857	5.384	6.4
James Shields*	SP			10	3.635	1.1
Eric Stults*	SP			0.5075	1.436	-0.3
Junichi Tazawa	RP			2.25	0.936	1.1
Shane Victorino	RF	CF	LF	13	3.887	0
Carlos Villanueva*	SP	RP		0.5075	0.994	0.2

* Indicates a player that was a free agent prior to optimization

Appendix Q: Boston Red Sox Roster via Linear Optimization with Linear Regression Projections

Name	Position 1	Position 2	Position 3	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Mookie Betts	CF	2B	RF	0.5145	1.447	4.8
Clay Buchholz	SP			12.25	0.999	3.2
Yoenis Cespedes	LF			10.5	1.823	6.7
Dane de la Rosa*	RP			0.5075	0.458	0
Chase Headley*	3B			13	2.755	1.5
Brock Holt	2B	3B	RF	0.5305	1.317	2.4
Joe Kelly	SP			0.603	0.495	1.2
Tommy Layne	RP			0.557	0.127	0.1
Jed Lowrie*	2B	SS		8	1.511	1
Russell Martin*	C			7	2.422	3.5
Kris Medlen*	SP	RP		2	0.977	0.5
Brandon Morrow*	SP	RP		2.5	0.573	0.5
Mike Napoli	1B			16	1.613	0.7
Daniel Nava	RF	LF	1B	1.85	1.133	-0.2
Alexi Ogando*	SP	RP		1.5	0.379	-0.9
David Ortiz	DH	1B		16	1.232	2.8
Dustin Pedroia	2B			12.625	2.706	2.5
Colby Rasmus*	CF	LF	RF	8	1.499	2.8
Max Scherzer*	SP			17.142857	3.241	6.4
James Shields*	SP			10	2.049	1.1
Eric Stults*	SP			0.5075	0.634	-0.3
Junichi Tazawa	RP			2.25	0.607	1.1
Christian Vazquez	C			0.5125	0.633	0
Shane Victorino	RF	CF	LF	13	2.156	0
Brandon Workman	SP	RP		0.5395	0.390	0

* Indicates a player that was a free agent prior to optimization

Appendix R: Boston Red Sox Roster via Linear Optimization with Separate Linear Regression Projections by Position

Name	Position 1	Position 2	Position 3	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Mookie Betts	CF	2B	RF	0.5145	1.947	4.8
Xander Bogaerts	SS	3B		0.543	0.857	4.3
Clay Buchholz	SP			12.25	1.157	3.2
Rusney Castillo	RF	LF	CF	11.271429	0.793	0.4
Yoenis Cespedes	LF			10.5	1.929	6.7
Dane De La Rosa*	RP			0.5075	0.415	0
Brock Holt	2B	3B	RF	0.5305	1.277	2.4
Joe Kelly	SP			0.603	0.664	1.2
Jon Lester*	SP			20	2.954	5
Russell Martin*	C			7	3.470	3.5
Kris Medlen*	SP	RP		2	1.162	0.5
Brandon Morrow*	SP	RP		2.5	0.664	0.5
Mike Napoli	1B			16	1.619	0.7
Daniel Nava	RF	LF	1B	1.85	1.037	-0.2
Alexi Ogando*	SP	RP		1.5	0.411	-0.9
David Ortiz	DH	1B		16	2.196	2.8
Dustin Pedroia	2B			12.625	2.215	2.5
Colby Rasmus*	CF	LF	RF	8	1.651	2.8
Max Scherzer*	SP			17.142857	3.749	6.4
James Shields*	SP			10	2.282	1.1
Junichi Tazawa	RP			2.25	0.463	1.1
Christian Vazquez	C			0.5125	1.096	0
Shane Victorino	RF	CF	LF	13	1.983	0
Carlos Villanueva*	SP	RP		0.5075	0.609	0.2
Brandon Workman	SP	RP		0.5395	0.572	0

* Indicates a player that was a free agent prior to optimization

Appendix S: Boston Red Sox Roster via Incremental Optimization Example

Name	Position 1	Position 2	Position 3	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Mookie Betts	CF	2B	RF	0.5145	1.947	4.8
Xander Bogaerts	SS	3B		0.543	0.857	4.3
Clay Buchholz	SP			12.25	1.157	3.2
Rusney Castillo	RF	LF	CF	11.271429	0.793	0.4
Yoenis Cespedes	LF			10.5	1.929	6.7
Dane De La Rosa*	RP			0.5075	0.415	0
Brock Holt	2B	3B	RF	0.5305	1.277	2.4
Nick Hundley*	C			3.1	0.594	2.3
Kyle Kendrick*	SP			5.5	0.915	-1
Jon Lester*	SP			20	2.954	5
Kris Medlen*	SP	RP		2	1.162	0.5
Brandon Morrow*	SP	RP		2.5	0.664	0.5
Mike Napoli	1B			16	1.619	0.7
Daniel Nava	RF	LF	1B	1.85	1.037	-0.2
David Ortiz	DH	1B		16	2.196	2.8
Dustin Pedroia	2B			12.625	2.215	2.5
Colby Rasmus*	CF	LF	RF	8	1.651	2.8
Max Scherzer*	SP			17.142857	3.749	6.4
James Shields*	SP			10	2.282	1.1
Anthony Swarzak*	SP	RP		0.5075	0.360	0.1
Junichi Tazawa	RP			2.25	0.463	1.1
Christian Vazquez	C			0.5125	1.096	0
Shane Victorino	RF	CF	LF	13	1.983	0
Carlos Villanueva*	SP	RP		0.5075	0.609	0.2
Brandon Workman	SP	RP		0.5395	0.572	0

* Indicates a player that was a free agent prior to optimization

Appendix T: Kansas City Royals' Actual 2015 Opening Day Roster

Name	Position 1	Position 2	Position 3	2015 Salary (in millions of dollars)	2015 WAR
Lorenzo Cain	CF	RF		2.725	6.6
Christian Colon	2B	SS	3B	0.509525	0.2
Wade Davis	RP			7	2
Danny Duffy	SP			2.425	1.2
Jarrod Dyson	CF			1.225	1.8
Alcides Escobar	SS			3	1.5
Jason Frasor	RP			1.8	0
Alex Gordon	LF			14	2.8
Jeremy Guthrie	SP			9	-0.8
Kelvin Herrera	RP			1.6	0.6
Greg Holland	RP			8.25	0.8
Eric Hosmer	1B			5.65	3.5
Omar Infante	2B			7.5	-0.9
Erik Kratz	C			0.5325	-0.2
Ryan Madson	RP			0.85	0.9
Kendrys Morales	1B	DH		6.5	2.1
Franklin Morales	SP	RP		0.5075	0.4
Mike Moustakas	3B			2.64	3.8
Paulo Orlando	RF	LF		0.5075	1
Salvador Perez	C			1.75	1.6
Alex Rios	RF			11	0.2
Jason Vargas	SP			8.5	0.4
Yordano Ventura	SP			0.95	2.7
Edinson Volquez	SP			7.5	2.6
Chris Young	SP	RP		0.675	0.9

Appendix U: Kansas City Royals' Roster via Stochastic Optimization

Name	Position 1	Position 2	Position 3	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Lorenzo Cain	CF	RF		2.725	3.865	6.6
Mike Carp*	1B	LF		0.5075	0.307	0
Christian Colon	2B	SS	3B	0.509525	0.700	0.2
Wade Davis	RP			7	1.983	2
Dane de la Rosa*	RP			0.5075	0.916	0
Andy Dirks*	LF			0.5075	0.960	0
Danny Duffy	SP			2.425	1.696	1.2
Jarrod Dyson	CF			1.225	2.595	1.8
Alcides Escobar	SS			3	2.350	1.5
Brandon Finnegan	RP			0.5075	0.300	-0.1
Alex Gordon	LF			14	5.402	2.8
Jeremy Guthrie	SP			9	0.766	-0.8
Kelvin Herrera	RP			1.6	1.124	0.6
Greg Holland	RP			8.25	2.459	0.8
Eric Hosmer	1B			5.65	0.972	3.5
Omar Infante	2B			7.5	1.690	-0.9
Russell Martin*	C			7	4.178	3.5
Mike Moustakas	3B			2.64	1.391	3.8
Salvador Perez	C			1.75	3.222	1.6
Max Scherzer*	SP			17.142857	5.384	6.4
Tim Stauffer*	RP			2.2	0.369	-0.5
Joe Thatcher*	RP			0.5075	0.271	0.3
Jason Vargas	SP			8.5	1.756	0.4
Yordano Ventura	SP			0.95	2.268	2.7
Eric Young*	LF			0.5075	1.078	-0.5

* Indicates a player that was a free agent prior to optimization

Appendix V: Kansas City Royals' Roster via Linear Optimization

Name	Position 1	Position 2	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Lorenzo Cain	CF	RF	2.725	3.865	6.6
Bruce Chen*	SP	RP	0.5075	1.023	-0.2
Wade Davis	RP		7	1.983	2
Dane De La Rosa*	RP		0.5075	0.916	0
Chris Denorfia*	LF	RF	2.6	1.928	0.8
Danny Duffy	SP		2.425	1.696	1.2
Jarrod Dyson	CF		1.225	2.595	1.8
Alcides Escobar	SS		3	2.350	1.5
Alex Gordon	LF		14	5.402	2.8
Chase Headley*	3B		13	4.654	1.5
Kelvin Herrera	RP		1.6	1.124	0.6
Greg Holland	RP		8.25	2.459	0.8
Eric Hosmer	1B		5.65	0.972	3.5
Omar Infante	2B		7.5	1.690	-0.9
Russell Martin*	C		7	4.178	3.5
Kris Medlen*	SP	RP	2	1.551	0.5
Mike Moustakas	3B		2.64	1.391	3.8
Salvador Perez	C		1.75	3.222	1.6
Wandy Rodriguez*	SP		0.5075	1.187	1
Eric Stults*	SP		0.5075	1.436	-0.3
Ichiro Suzuki*	RF		2	1.279	-0.8
Dan Uggla*	2B		0.5075	1.058	-0.1
Jason Vargas	SP		8.5	1.756	0.4
Yordano Ventura	SP		0.95	2.268	2.7
Carlos Villanueva*	SP	RP	0.5075	0.994	0.2

* Indicates a player that was a free agent prior to optimization

Appendix W: Kansas City Royals' Roster via Linear Optimization with Linear Regression Projections

Name	Position 1	Position 2	Position 3	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Everth Cabrera*	SS	2B		2.4	0.769	-0.5
Lorenzo Cain	CF	RF		2.725	2.371	6.6
Tim Collins	RP			1.475	0.478	0
Christian Colon	2B	SS	3B	0.509525	0.602	0.2
Wade Davis	RP			7	1.237	2
Andy Dirks*	LF			0.5075	0.621	0
Danny Duffy	SP			2.425	1.202	1.2
Jarrod Dyson	CF			1.225	1.560	1.8
Alcides Escobar	SS			3	1.504	1.5
Brandon Finnegan	RP			0.5075	0.544	-0.1
Alex Gordon	LF			14	3.205	2.8
Kelvin Herrera	RP			1.6	0.903	0.6
Greg Holland	RP			8.25	1.524	0.8
Eric Hosmer	1B			5.65	0.812	3.5
Omar Infante	2B			7.5	0.878	-0.9
Russell Martin*	C			7	2.422	3.5
Kris Medlen*	SP	RP		2	0.977	0.5
Mike Moustakas	3B			2.64	1.018	3.8
Salvador Perez	C			1.75	2.167	1.6
Max Scherzer*	SP			17.142857	3.241	6.4
Eric Stults*	SP			0.5075	0.634	-0.3
Jason Vargas	SP			8.5	0.964	0.4
Yordano Ventura	SP			0.95	1.638	2.7
Carlos Villanueva*	SP	RP		0.5075	0.551	0.2
Eric Young*	LF			0.5075	0.647	-0.5

* Indicates a player that was a free agent prior to optimization

Appendix X: Kansas City Royals' Roster via Linear Optimization with Separate Linear Regression Projections by Position

Name	Position 1	Position 2	Position 3	Position 4	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Lorenzo Cain	CF	RF			2.725	2.498	6.6
Tim Collins	RP				1.475	0.278	0
Christian Colon	2B	SS	3B		0.509525	0.753	0.2
Wade Davis	RP				7	0.943	2
Dane De La Rosa*	RP				0.5075	0.415	0
Andy Dirks*	LF				0.5075	0.681	0
Danny Duffy	SP				2.425	1.511	1.2
Jarrod Dyson	CF				1.225	1.598	1.8
Alcides Escobar	SS				3	1.385	1.5
Alex Gordon	LF				14	3.247	2.8
Jack Hannahan*	1B	3B			1	0.512	0
Kelvin Herrera	RP				1.6	0.601	0.6
Greg Holland	RP				8.25	1.161	0.8
Omar Infante	2B				7.5	0.698	-0.9
Kelly Johnson*	1B	2B	3B	LF	0.5075	0.681	0.3
Kevin Kouzmanoff*	3B				0.5075	0.393	0
Russell Martin*	C				7	3.470	3.5
Kris Medlen*	SP	RP			2	1.162	0.5
Salvador Perez	C				1.75	3.398	1.6
Max Scherzer*	SP				17.142857	3.749	6.4
Eric Stults*	SP				0.5075	0.586	-0.3
Jason Vargas	SP				8.5	1.057	0.4
Yordano Ventura	SP				0.95	2.074	2.7
Carlos Villanueva*	SP	RP			0.5075	0.609	0.2
Eric Young*	LF				0.5075	0.650	-0.5

* Indicates a player that was a free agent prior to optimization

Appendix Y: San Diego Padres' Actual 2015 Opening Day Roster

Name	Position 1	Position 2	Position 3	2015 Salary (in millions of dollars)	2015 WAR
Yonder Alonso	1B			1.65	1.1
Alexi Amarista	SS	2B	CF	1.15	-0.8
Clint Barmes	SS	2B		1.5	0
Joaquin Benoit	RP			8	0.3
Andrew Cashner	SP			4.05	2.3
Odrisamer Despaigne	SP	RP		0.5173	0
Frank Garces	RP			0.5085	-0.9
Jedd Gyorko	2B	SS		2	0.7
Shawn Kelley	RP			2.835	1
Matt Kemp	RF	CF	LF	21.25	0.4
Ian Kennedy	SP			9.85	0.8
Craig Kimbrel	RP			9.25	1.5
Will Middlebrooks	3B			0.5405	-0.4
Brandon Morrow	SP	RP		2.5	0.5
Wil Myers	RF	CF	1B	0.5198	0.6
Wil Nieves	C			0.5075	-0.1
Derek Norris	C			0.545	2.4
Tyson Ross	SP			5.25	4.4
James Shields	SP			10	1.1
Yangervis Solarte	3B	2B		0.5164	1.6
Cory Spangenberg	2B	3B		0.5085	2.1
Dale Thayer	RP			1.375	-0.3
Justin Upton	LF	RF		14.708333	3.6
Will Venable	CF	RF		4.25	1.2
Nick Vincent	RP			0.5253	0.3

Appendix Z: San Diego Padres' Roster via Stochastic Optimization

Name	Position 1	Position 2	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Yonder Alonso	1B		1.65	0.777	1.1
Joaquin Benoit	RP		8	1.105	0.3
Emilio Bonifacio*	2B	CF	4	1.404	-0.7
Blaine Boyer	RP		0.75	0.165	0.2
John Buck*	C		0.5075	1.009	0
Andrew Cashner	SP		4.05	2.368	2.3
Dane de la Rosa*	RP		0.5075	0.916	0
Odrisamer Despaigne	SP	RP	0.5173	1.100	0
Jedd Gyorko	2B	SS	2	1.059	0.7
Ian Kennedy	SP		9.85	2.427	0.8
Jed Lowrie*	2B	SS	8	2.590	1
Russell Martin*	C		7	4.178	3.5
Cameron Maybin	CF		7.1	0.884	1
Kevin Quackenbush	RP		0.5147	0.900	0.2
Carlos Quentin	LF	RF	8	1.043	0
Rene Rivera	C		1.2	2.174	-0.9
Tyson Ross	SP		5.25	2.601	4.4
Max Scherzer*	SP		17.142857	5.384	6.4
Seth Smith	LF	RF	6	1.893	2.2
Yangervis Solarte	3B	2B	0.5164	1.600	1.6
Eric Stults*	SP		0.5075	1.436	-0.3
Joe Thatcher*	RP		0.5075	0.271	0.3
Dan Uggla*	2B		0.5075	1.058	-0.1
Will Venable	CF	RF	4.25	1.861	1.2
Nick Vincent	RP		0.5253	0.961	0.3

* Indicates a player that was a free agent prior to optimization

Appendix AA: San Diego Padres' Roster via Linear Optimization

Name	Position 1	Position 2	Position 3	Position 4	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Yonder Alonso	1B				1.65	0.777	1.1
Joaquin Benoit	RP				8	1.105	0.3
Andrew Cashner	SP				4.05	2.368	2.3
Bruce Chen*	SP	RP			0.5075	1.023	-0.2
Dane De La Rosa*	RP				0.5075	0.916	0
Chris Denorfia*	LF	RF			2.6	1.928	0.8
Odrisamer Despaigne	SP	RP			0.5173	1.100	0
Jedd Gyorko	2B	SS			2	1.059	0.7
Chase Headley*	3B				13	4.654	1.5
Kelly Johnson*	1B	2B	3B	LF	0.5075	0.724	0.3
Ian Kennedy	SP				9.85	2.427	0.8
Russell Martin*	C				7	4.178	3.5
Kris Medlen*	SP	RP			2	1.551	0.5
Carlos Quentin	LF	RF			8	1.043	0
Rene Rivera	C				1.2	2.174	-0.9
Tyson Ross	SP				5.25	2.601	4.4
James Shields*	SP				10	3.635	1.1
Seth Smith	LF	RF			6	1.893	2.2
Yangervis Solarte	3B	2B			0.5164	1.600	1.6
Eric Stults*	SP				0.5075	1.436	-0.3
Dan Uggla*	2B				0.5075	1.058	-0.1
Will Venable	CF	RF			4.25	1.861	1.2
Carlos Villanueva*	SP	RP			0.5075	0.994	0.2
Nick Vincent	RP				0.5253	0.961	0.3
Eric Young*	LF				0.5075	1.078	-0.5

* Indicates a player that was a free agent prior to optimization

Appendix AB: San Diego Padres' Roster via Linear Optimization with Linear Regression Projections

Name	Position 1	Position 2	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Yonder Alonso	1B		1.65	0.557	1.1
Joaquin Benoit	RP		8	0.343	0.3
Everth Cabrera*	SS	2B	2.4	0.769	-0.5
Andrew Cashner	SP		4.05	1.515	2.3
Dane de la Rosa*	RP		0.5075	0.458	0
Chris Denorfia*	LF	RF	2.6	0.976	0.8
Odrisamer Despaigne	SP	RP	0.5173	0.751	0
Jedd Gyorko	2B	SS	2	0.818	0.7
Chase Headley*	3B		13	2.755	1.5
Ian Kennedy	SP		9.85	1.459	0.8
Russell Martin*	C		7	2.422	3.5
Cameron Maybin	CF		7.1	0.621	1
Kris Medlen*	SP	RP	2	0.977	0.5
Kevin Quackenbush	RP		0.5147	0.722	0.2
Carlos Quentin	LF	RF	8	0.534	0
Rene Rivera	C		1.2	1.261	-0.9
Tyson Ross	SP		5.25	1.656	4.4
James Shields*	SP		10	2.049	1.1
Seth Smith	LF	RF	6	1.046	2.2
Yangervis Solarte	3B	2B	0.5164	1.098	1.6
Cory Spangenberg	2B	3B	0.5085	0.272	2.1
Eric Stults*	SP		0.5075	0.634	-0.3
Will Venable	CF	RF	4.25	1.027	1.2
Carlos Villanueva*	SP	RP	0.5075	0.551	0.2
Nick Vincent	RP		0.5253	0.668	0.3

* Indicates a player that was a free agent prior to optimization

Appendix AC: San Diego Padres' Roster via Linear Optimization with Separate Linear Regression Projections by Position

Name	Position 1	Position 2	Position 3	Position 4	2015 Salary (in millions of dollars)	Expected WAR	Actual WAR
Yonder Alonso	1B				1.65	0.161	1.1
Alexi Amarista	SS	2B	CF		1.15	0.604	-0.8
Joaquin Benoit	RP				8	0.437	0.3
Andrew Cashner	SP				4.05	1.813	2.3
Dane De La Rosa*	RP				0.5075	0.415	0
Odrisamer Despaigne	SP	RP			0.5173	0.931	0
Andy Dirks*	LF				0.5075	0.681	0
Jedd Gyorko	2B	SS			2	0.923	0.7
Kelly Johnson*	1B	2B	3B	LF	0.5075	0.681	0.3
Ian Kennedy	SP				9.85	1.689	0.8
Kevin Kouzmanoff*	3B				0.5075	0.393	0
Russell Martin*	C				7	3.470	3.5
Cameron Maybin	CF				7.1	0.740	1
Kris Medlen*	SP	RP			2	1.162	0.5
Kevin Quackenbush	RP				0.5147	0.485	0.2
Rene Rivera	C				1.2	1.738	-0.9
Tyson Ross	SP				5.25	1.976	4.4
Max Scherzer*	SP				17.142857	3.749	6.4
James Shields*	SP				10	2.282	1.1
Seth Smith	LF	RF			6	0.947	2.2
Yangervis Solarte	3B	2B			0.5164	0.405	1.6
Cory Spangenberg	2B	3B			0.5085	0.570	2.1
Will Venable	CF	RF			4.25	0.928	1.2
Carlos Villanueva*	SP	RP			0.5075	0.609	0.2
Nick Vincent	RP				0.5253	0.487	0.3

* Indicates a player that was a free agent prior to optimization

References

Adams, S. (2014). Offseason Outlook: Kansas City Royals. *MLB Trade Rumors*. Retrieved from http://www.mlbtraderumors.com/2014/11/offseason-outlook-kansas-city-royals-5.html

Baseball Prospectus. (n.d.). Cot's Baseball Contracts. *Baseball Prospectus*. Retrieved from http://www.baseballprospectus.com/compensation/cots/

Baseball Reference. (n.d.). *Baseball-Reference.com*. Retrieved from www.baseball-reference.com/

CBS Sports. (n.d.). *MLB Salaries.* Retrieved from http://www.cbssports.com/mlb/salaries/avgsalaries

Dataplot Reference Manual. (1996). *Weighted Standard Deviation*. Retrieved from http://www.itl.nist.gov/div898/software/dataplot/refman2/ch2/weightsd.pdf

Fangraphs. (n.d.). a. *BsR*. Retrieved from http://www.fangraphs.com/library/offense/bsr/

Fangraphs. (n.d.). b. *Catcher Defense.* Retrieved from http://www.fangraphs.com/library/defense/catcher-defense/

Fangraphs. (n.d.). c. *Fangraphs.com*. Retrieved from http://www.fangraphs.com/

Fangraphs. (n.d.). d. *FIP*. Retrieved from http://www.fangraphs.com/library/pitching/fip/

Fangraphs. (n.d.). e. *UZR*. Retrieved from http://www.fangraphs.com/library/defense/uzr/

Fangraphs. (n.d.). f. *WAR for Pitchers*. Retrieved from http://www.fangraphs.com/library/war/calculating-war-pitchers/

Fangraphs. (n.d.). g. *WAR for Position Players*. Retrieved from http://www.fangraphs.com/library/war/war-position-players/

Fangraphs. (n.d.). h. *What is WAR?* Retrieved from http://www.fangraphs.com/library/misc/war/

Fangraphs. (n.d). i. *wOBA*. Retrieved from http://www.fangraphs.com/library/offense/woba/

Fangraphs. (n.d.). j. *wRAA*. Retrieved from http://www.fangraphs.com/library/offense/wraa/

Frontline Solvers. (n.d.). *Genetic Algorithms and Evolutionary Algorithms – Introduction*. Retrieved from http://www.solver.com/genetic-evolutionary-introduction

Glaser, K. (2015). Padres: Opening Day roster set, including Kimbrel and Upton Jr. The Press-Enterprise. Retrieved from http://www.pe.com/articles/roster-764002-day-kimbrel.html

Google Scholar. (n.d.). *Google Scholar*. Retrieved from https://0-scholar.google.com.library.uark.edu/schhp?hl=en&as_sdt=0,4

Jorion, P. (1992). Portfolio Optimization in Practice. *Financial Analysts Journal, 48(1),* 68-74. Retrieved from http://0-www.jstor.org.library.uark.edu/stable/4479507

Kelliher, C. F., & Mahoney, L. S. (2000). Using Monte Carlo Simulation to Improve Long-Term Investment Decisions. *The Appraisal Journal*, *68(1)*,44-56. Retrieved from http://0-search.proquest.com.library.uark.edu/docview/199961549?accountid=8361

Kritzman, M. (2011). The Graceful Aging of Mean-Variance Optimization. *Journal of Portfolio Management, 37(2)*, 3-5. Retrieved from http://0-search.proquest.com.library.uark.edu/docview/851512017?accountid=8361

Kwak, Y. H., & Ingall, L. (2007). Exploring Monte Carlo Simulation Applications for Project Management. *Risk Management, 9(1),* 44-57. Retrieved from http://0-www.jstor.org.library.uark.edu/stable/4143844

Lampe, N. (2015). Player Contracts and Investment Diversification. Retrieved from http://www.beyondtheboxscore.com/2015/5/6/8485845/player-contracts-and-investment-diversification

Links, Z. (2014). Offseason Outlook: San Diego Padres. MLB Trade Rumors. Retrieved from http://www.mlbtraderumors.com/2014/10/offseason-outlook-san-diego-padres-6.html

Major League Baseball Players Association. (n.d.). MLBPA Info. Retrieved from http://mlb.mlb.com/pa/info/faq.jsp#minimum

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance, 7(1),* 77-91. doi: 10.2307/2975974

Mathematics Stack Exchange. (n.d.). *Standard Deviation of the Weighted Mean*. Retrieved from http://math.stackexchange.com/questions/320441/standard-deviation-of-the-weighted-mean

McAdam, S. (2015). Red Sox announce Opening Day roster. Comcast SportsNet New England. Retrieved from http://www.csnne.com/blog/red-sox-talk/red-sox-announce-opening-day-roster

MLB Trade Rumors. (n.d.). *2014-2015 MLB Free Agent Tracker*. Retrieved from http://www.mlbtraderumors.com/2015-mlb-free-agent-tracker

MLB.com. (2015). a. *Standings*. Retrieved from http://mlb.mlb.com/mlb/standings/#20151004

MLB.com. (n.d.). b. *Transactions*. Retrieved from http://mlb.mlb.com/mlb/transactions/#month=12&year=2014

Pekarsky, M. (2015). KC Royals release World Series roster. *Fox4kc.com*. Retrieved from http://fox4kc.com/2015/10/27/kc-royals-release-world-series-roster/

Polishuk, M. (2014). Offseason Outlook: Boston Red Sox. MLB Trade Rumors. Retrieved from http://www.mlbtraderumors.com/2014/10/offseason-outlook-boston-red-sox-5.html

Rieper, M. (2015). Royals finalize 25-man opening day roster. *Royals Review*. Retrieved from http://www.royalsreview.com/2015/4/5/8347067/royals-finalize-25-man-opening-day-roster

Speidell, L., Miller, D., Ullman, J. (1989). Portfolio Optimization: A Primer. *Financial Analysts Journal, 45(1)*, 22-30. Retrieved from http://0-www.jstor.org.library.uark.edu/stable/4479184

Thompson, K., & McLeod, A. (2009). Accelerated Ensemble Monte Carlo Simulation. *Risk, 22(4)*, 86-91. Retrieved from http://0-search.proquest.com.library.uark.edu/docview/201373649?accountid=8361