5-2015

Evaluating Return on Investment in the MLB Rule IV Draft

Christopher Campione
Clemson University

EVALUATING RETURN ON INVESTMENT IN THE MLB RULE IV DRAFT

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Masters of Arts
Economics

---

by
Christopher Campione

May 2015

Accepted by:
Dr. F. Andrew Hanssen, Committee Chair
Dr. Raymond D. Sauer
Dr. Scott L. Baier

ABSTRACT

Every June, the 30 clubs of Major League Baseball gather on a conference call to select amateur players from high school and college into their organization. These players may be position players or pitchers, and when selected, the players are awarded a signing bonus to entice them to join the organization. Using regression techniques and statistical tests, position players will be compared to pitchers as well as high school players to college players in terms of return on investment. First multiple regression techniques are used to develop a model to determine what a player's value is to his team based off of his marginal revenue product (MRP). This value is then compared to yearly compensation including initial signing bonus for the player's pre-arbitration seasons. The difference in total MRP and total compensation formulate a return on investment. These distributions of returns on investment are compared by group using a Kolmogorov-Smirnov procedure. When performed, it is found that there is no statistical difference in return by position outside of a rare exception in 2005 where position players significantly outperformed pitchers in the first round. There is statistical evidence suggesting college players return higher returns on investment on average than high school players in all rounds, but when broken down by position, the significance only holds for infielders and outfielders.

DEDICATION

I dedicate this Thesis to my loving and supporting family who I work every day to make proud.

ACKNOWLEDGMENTS

I want to thank my advisor, Dr. Hanssen, as well as my other committee members, Drs. Sauer and Baier, for their valuable time and insights as I worked through the various ideas and drafts of this Thesis. I would also like to thank Gary Cohen of The Baseball Cube and Lewie Pollis for providing me with some of the data necessary to perform this analysis.

TABLE OF CONTENTS

Page

TITLE PAGE i

ABSTRACT ii

DEDICATION iii

ACKNOWLEDGMENTS iv

LIST OF FIGURES AND TABLES vi

CHAPTER

I. Introduction 1

II. Determining a Player's Value 5

III. Model Methodology and Data Gathering 9

IV. Player Value Model Results 14

V. Previous Draft Analysis and Theoretical Hypothesis 17

VI. Performance Metrics and Methodology 21

VII. Discussion and Results 38

VIII. Conclusions 48

IX. Limitations and Future Research 53

APPENDICES 55

A: Figures 58
B: Tables 73

REFERENCES 89

LIST OF FIGURES AND TABLES

Page

Figure

A-1 Run Differential and Win Percentage……………………………….56

A-2 MLB Revenue By Year……………………………………………..57

A-3 Total Revenue and Fan Cost Index………………………………… 58

A-4 Total Revenue and Run Differential……………………………… .58

A-5 Position Player and Pitcher Mean Return by Year………………… …59

A-6 Position Player Mean Return by Position and Year………………..59

A-7 High School and College Mean Return by Year…………………… ...60

A-8 Return on Investment Distribution…………………………….……61

A-9 Position Player and Pitcher Return Distribution……………………..62

A-10 First Round Position Player and Pitcher Return Distribution………..…63

A-11 Other Round Position Player and Pitcher Return Distribution……….64

A-12 Offensive Return Distribution by Position…………………………... .65

A-13 High School and College Player Return Distribution………………....66

A-14 First Round High School and College Player Return Distribution…...67

A-15 Other Round High School and College Player Return Distribution.…68

A-16 High School and College Catcher Return Distribution……………..….69

A-17 High School and College Infielder Return Distribution…………….. ..70

A-18 High School and College Outfielder Return Distribution…………..….71

A-19 High School and College Pitcher Return Distribution……………..…72

List of Figures and Tables (Continued)

Table Page

B-1 Model Variable Description 73

B-2 Summary Statistics 73

B-3 MRP Model Results 74

B-4 Fielding Positional Adjustments 75

B-5 MRP Model v One-Year Free Agent Contracts Using Replacement 76

B-6 MRP Model v One-Year Free Agent Contracts Raw 76

B-7 MRP Model v One-Year Free Agent Contracts Using Average 77

B-8 Highest Annual MRPs 78

B-9 Top 50 MLB Salaries for 2015 79

B-10 Highest Return on Investment 81

B-11 Position Players v Pitchers 82

B-12 First Round Position Players v Pitchers 83

B-13 First Round Position Players v Pitchers Excluding 2005 83

B-14 Other Round Position Players v Pitchers 84

B-15 Offensive Position Comparison 85

B-16 High School v College Players 85

B-17 First Round High School v College 86

B-18 Other Round High School v College 86

B-19 High School v College Catchers 87

B-20 High School v College Infielders 87

List of Figures and Tables (Continued)

Page

B-21 High School v College Outfielders 88

B-22 High School v College Pitchers 88

CHAPTER ONE

Introduction

Major League Baseball is America's oldest professional sports league, dating back to 1869. While the game itself has gone through some changes over the years such as moving fences in on playing fields to increase homeruns, and decreasing the amount of innings starting pitchers are allowed to throw in a season, the sport of baseball as a whole is not much different today than when greats like Babe Ruth donned Yankee pinstripes in the 1920s. That said, where baseball has changed drastically in recent times is in how the game is thought of and analyzed from a front office perspective. The days of simply looking at the back of a baseball player's trading card and evaluating his batting average or earned run average to successfully value a player's worth are over, as Bill James' creation of sabermetrics in the mid 1990s has led to a revolution of in depth statistical analysis of Major League Baseball. In this paper, some of the concepts developed by these sabermatricians will be utilized in conjunction with economic theory to investigate the Major League Baseball Rule IV Draft. The first part of this paper will analyze various methods of valuing a player, eventually settling on a model that will produce a player's worth to his team based off of quantifiable statistics. After that, metrics will be selected as the basis for which to be inserted into the model, for means of comparison between what a player was worth and what the player was actually paid for that season. Finally, drafted players can be compared for differences in return on investment between hitters and pitchers as well as between high school and college players to see if there are any groups of athletes that have performed significantly better than others, leading to a possible optimal drafting strategy for teams.

The MLB Rule IV Draft is a conference call that takes place every June between the 30 MLB clubs. During this call, the teams take turns in a specified order based off of the previous year's standings, selecting amateur players eligible to be drafted. According to the MLB's official website, these eligible players include, "high school players who have graduated and have not yet attended college, college players from four-year colleges who have completed either their junior or senior years or are at least 21 years old, and junior college players regardless of how many years of school completed" (MLB). Additionally, these players must be residents of the United States, Canada, or a United States territory such as Puerto Rico to be eligible. Once players are selected, a club retains this player's rights until August 15th, where they must try and come to terms on a signing bonus and ensuing contract to entice the player to join their organization. These signing bonuses vary, and are typically higher based off of what round the player was selected in, but this is not always the case. There are times when teams look to save money by drafting a player they know will sign for less, and times when high spending teams roll the dice on athletes later in the draft who have fallen because of teams' concerns over the player's "signability", or actual probability of agreeing to forgo other athletic opportunities such as college baseball or football in favor of joining a minor league system.

Unlike other sports such as the National Basketball Association, and the National Football League, most players selected in the MLB Rule IV Draft are not immediately ready to produce on a big league club. Many of these players are drafted for their raw skills, and are expected to need some time developing in minor league systems, sometimes for time periods close to a decade. For this reason, MLB draftees can be looked at as investments made by their teams. The teams usually sign the player to a signing bonus that reaches levels of a major league salary, however, outside of a few rare exceptions, they pay them very small annual salaries until they are eligible for salary arbitration. In other words, MLB teams pay a relatively high fee up front to acquire the player, but then they have a window of time pre arbitration to recoup their investment, as well as produce additional value. Because of the nature of the draft, however, there is volatility on these returns, as some players pan out in a big way, while others with high expectations never even make it to the major leagues for various reasons. While previous research has concluded that on average, drafted players provide a significant return to their clubs, there remains the question on whether or not there is a group of players that maximize this return relative to another. Using economic theory as a guide, if the MLB Draft is operating efficiently, players' compensation in the form of signing bonuses and salary should be a representation of their expected contribution to their club, and thus, there should be no exploitable group of players that teams can pick from that would provide a better return on investment than another. However, this paper will investigate whether that theory truly holds in this platform.

In the draft, MLB teams can select from different groups of players. The first group is classified by position. MLB teams, at its simplest form can select either pitchers or position players. Additionally, within the position player group, there are subgroups by actual position, such as infielders, outfielders, and catchers. Along with this, there are two distinct age groups to select from. The first is high school players. These players are usually 18 years old, and are typically thought of as high upside prospects in need of some professional coaching to maximize potential. The second group is college players. These players vary in age from 19 to 24, and while they are thought to be more pro ready than their high school counterparts on average, they come at a price of added age, and potentially smaller upside. Given these various different groups of players to pick from, MLB front offices are tasked with the job of trying to determine if there is an optimal strategy for what type of player they should select in certain draft positions on the basis of return on investment, and it is the goal of this paper to dive into that question in more detail. If economic theory holds, and players are compensated at a consistent rate between groups for their future performance in the form of signing bonuses and salaries, my hypothesis that there will be no exploitable drafting strategy should be proven true in this study.

CHAPTER TWO

Determining a Player's Value

One of the core hurdles this analysis must overcome is determining what a player's value is to a major league team. If a player's value can be calculated, this total can then be compared to the player's compensation to formulate a return. This is fundamental in the ability to evaluate draft picks in the second part of the paper. Fortunately, this idea has been studied previously by many sports economists. The classic paper on this topic, and the one that is most heavily referenced and talked about is Gerald Scully's 1974 paper titled, "Pay and Performance in Major League Baseball." The basis of Scully's research, is that if the baseball labor market was perfectly competitive, "the player salaries would be equated with player marginal revenue products (MRP)" (Scully 1974). This is rooted in a basic labor economic theory that salaries should reflect the additional revenue added by the worker. Scully goes on to explain that teams' winning percentage is a function of "a vector of player skills," and, "a vector of other non player inputs such as managers, coaches, capital, etc." (Scully 1974). Additionally, he states that, "teams derive revenue essentially from two main sources: gate receipts and the sale of radio and television rights" (Scully 1974). What Scully concludes, then, is that gate receipts and broadcasting rights increase as a function of win percentage, and seeing as win percentage increases as a function of player skills, these player performances have an impact on total revenue realized by the team.

While Scully concedes determining a player's MRP is not an easy task, one can determine a player's impact on winning, and then determine the impact of winning on revenue, resulting in a crude estimate on what the player's MRP actually is, and thus what his salary should be in a perfectly competitive market. Scully then goes on to estimate two different linear regression models, the first regressing winning percentage as a dependent variable as a function of the independent variables of team slugging average, team strikeout-to-walk ratio, league, contender status, and status as a habitual winner or loser. In this instance, Scully uses slugging average as a proxy for total offensive contribution and strikeout-to-walk ratio as a proxy for pitching. The second model Scully estimates is a linear model regressing team revenue on winning percentage, market size, team differences in attendance over time, league, stadium age, and percentage of minority players. When both of these models have been estimated, Scully takes the coefficient on slugging average from the first model and multiplies that by the coefficient of winning percentage on the second model to determine the MRP of an offensive player per point of slugging average. He does the same for pitchers, however, he uses strikeout-to-walk ratio instead of slugging average. In other words, in the Scully model, one can find the MRP for any player by simply having that players' slugging average or strikeout-to-walk ratio and multiplying it by the value calculated from estimating the two previous regression models.

While Scully's paper was innovative in how people thought about professional player salaries, as Scully points out himself in future work, "one needs to be cautious about the results, since the estimates of player marginal products are crude" (Scully 1989). In other words, the Scully model needs some updating and retooling to be an effective model to measure MLB players' value. One economist, Anthony Krautmann discards Scully's method altogether in his paper, "What's Wrong With Scully-Estimates of a Player's Marginal Revenue Product." The basis of Krautmann's argument is that at the time of Scully's work, players were not permitted to become free agents. However, in the present day, a free agent market exists where players can be bid on by various teams, and thus, should receive compensation that aligns them with what their value is to teams. In Krautmann's free market returns approach, he regresses free agents' wages on their performance using an ex ante measure of a free agent's performance, time trends, and a fixed-effects parameter for teams to control for "team-invariant factors, including managerial quality" (Krautmann 1999). By calculating this model, one can determine a player's MRP without having to know a team's revenue function, but rather by using past free agent information to essentially estimate what a player would be worth on the free market, and thus, what his value truly is.

The Krautmann free market returns approach was directly challenged in a subsequent article written by John Charles Bradbury of Kennesaw State University in a paper titled, "What's Right with Scully-Estimates of a Player's Marginal Revenue Product." Bradbury does concede flaws in the Scully method as explained by Krautmann, including the fact that Scully includes revenue that is shared amongst teams, and the revenues of MLB teams are never actually officially reported, and thus, the model is grounded in an estimated revenue value, which could lead to inaccurate results. However, Bradbury finds further flaws in Krautmann's method, explaining that "the market prices on which the Krautmann model depends may not properly reflect players' true marginal revenue contributions" (Bradbury 2013). He gives various examples of why this may be the case, including, “the availability of cheap substitutes, players’ willingness to accept discounts for non-pecuniary wages and reduce risk, past evidence of inefficiency in professional sports labor markets, limited competition for player services, and the non-linear relationship between performance and revenue” (Bradbury 2013) Because of this, Bradbury concludes that while the Krautmann method is a useful method to value players, a less flawed approach would be to create an updated version of Scully’s model to better reflect information and statistics available to today’s baseball researcher.

Given the background information on this topic, it is clear that two different methods are available to use as player valuation techniques. The first, is described by Scully and supported by Bradbury, and this is a model that is based on team revenues. The second is the model described by Krautmann, which is grounded in a free agent market approach. While I see the value in Krautmann’s approach, I feel there are too many potential flaws in evaluating free agent salaries as a basis for MRP. The idea of free agent salaries determining a player’s value requires that players would accept the highest bid on them in free agency. This is simply not the case very often, as players take less money for longer deals, and for situations that they feel better fit themselves as players and individuals. While free agent salaries can help determine a player’s value, there are too many potential causes for error in the human element of offering and agreeing to compensation terms that leave me with reservations in using this approach. For that reason, the valuation approach used in this paper will be an updated version of Scully’s method, using a model created by Bradbury as a strong reference.

CHAPTER THREE

Model Methodology and Data Gathering

In order to determine a player's MRP, a model had to be created to measure team revenue as a function of various different explanatory variables. As a reference for this type of model, I used J.C. Bradbury's book *Hot Stove Economics. Understanding Baseball's Second Season*. In this book, Bradbury attempts to create a model that is similar in theory to Scully's MRP model; however, he tries to make it more precise and a better reflection of a player's true value by updating the explanatory variables. As Bradbury explains, "a performance metric should be judged according to three criteria: (1) how it correlates with winning, (2) the degree to which it separates true ability from random chance, and (3) whether or not the information it conveys regarding performance matches reasonable intuition about what constitutes good performance" (Bradbury 2010). The performance metric that Bradbury finds correlates the strongest with winning is runs scored less runs allowed, also known as run differential. This is intuitive, as baseball is a game of scoring more runs than the other team. Because of this, if a team scores a lot more runs than it allows, the chances are high that it will win a lot more games than a team who does the opposite. When carrying this concept out further, Bradbury looks at the correlation of various offensive metrics such as batting average, on base percentage, and slugging average, and finds they correlate strongly with runs scored. Additionally, common pitching metrics such as earned run average, strikeout rate, and walk rate correlate strongly with runs allowed. For this reason, run differential served as a useful metric, as it could represent winning percentage while still remaining a stat that would be quantifiable for individual performance.

Once Bradbury discovers that run differential serves as a good metric for a team's ability to win, he then begins to construct a regression model to explain team revenue as a function of run differential and other tested factors. Bradbury collects team revenue from Forbes annual reports, adjusting by a constant amount to account for national revenue that is shared amongst each team, as including this value would artificially inflate MRP estimates. He tests various different factors, and finds that the only significant ones to include in his model are population of the metropolitan area of the team, and whether or not the team is playing in a relatively new ballpark. For the latter variable, Bradbury uses previous research to conclude additional revenue can be expected for a team playing in a ballpark that is eight years or younger, so he creates a dummy variable called honeymoon that receives a value of one for any team playing in a ballpark that fits this criteria. Finally, when investigating the relationship between run differential and team revenue, Bradbury finds the relationship is not linear. In fact, he finds that run differential has a greater positive effect on team revenue as run differential increases. This is intuitive, as it says winning games generates more revenue for teams that approach a certain threshold, most likely a playoff spot. If a team is on the verge of making the playoffs or winning a championship, each increased win would likely generate more revenue than a team who is out of contention, as fans are more likely to spend money to attend games and purchase merchandise for teams that are good. Bradbury combines all of this analysis to form a pooled panel regression model that estimates team revenue as a function of run differential, the square of run differential, the cube of run differential, MSA population, and honeymoon status of the team's ballpark.

To create my model, I used very similar concepts to those explained by Bradbury, with a few minor alterations. The first alteration was the time period to be utilized. While Bradbury used the years 2003 to 2007, I planned on using a broader data set. For this reason the years 2000 to 2013 were utilized. Much like Bradbury, I gathered run differentials and winning percentages for teams over the 14-year window from ESPN's website. As *Figure A-1* shows, the relationship between both run differential and winning was strongly linear, and had a correlation of 93%. Given this information, it was clear that much like Bradbury's model, run differential would serve as a sound proxy for win percentage.

The team revenue data was gathered from various different Forbes Business of Baseball Reports over the 14-year period. These values were raw revenue and had to be adjusted for inflation. To do this, all revenue values were converted to 2013 US dollars by using CPI indices for sake of comparison. One main difference between the data I utilized and the data Bradbury utilized was that he adjusted the revenue estimates by subtracting out a constant value for national television revenue that is given to each team in the league. The reasoning for this is sound, as if this value was included in the revenue estimates, players' MRP calculations would wind up being artificially high, as this national revenue is awarded to every team regardless of their performance, and thus individual player contributions have no effect on this revenue stream. However, there is a problem with how Bradbury accounts for this value. To find a national revenue estimate

Bradbury finds a 2007 value for national revenue and applies that to every team over the life of his data set. As *Figure A-2* demonstrates, however, this is not the best way to estimate national revenues, because MLB revenues have been increasing significantly over time, and a lot of that growth is reflected in growing national television contracts. In other words, the national revenue a team received in 2005 would not necessarily be consistent with national revenue a team received in 2002.

There were a few ways to account for this difference in national revenue by year. The first idea I had was to adjust national revenue by year-over-year growth rates of MLB league revenues so that the estimates reacted more consistently with how it truly moved. While this method would have been more accurate than the Bradbury method, it was still a crude estimate, as other factors besides national television contracts could be responsible for MLB league revenue growth. For this reason, I decided to simply add dummy variables for every year in the study. By adding these time effects, several different factors across years could be captured, whether that be varying disbursements of TV contracts or other nonobvious changes that caused revenue to fluctuate from year to year. Because national television revenue is simply a lump sum value allotted to every team, adding dummy variables for year should allow unadjusted Forbes revenue estimates to be utilized without being affected.

Much like in Bradbury's model, it was decided that population and honeymoon values would be gathered to test in the model. The population values were gathered from US Census data for the metropolitan area of the team. For years where Census data was not reported, the values were interpolated between the two nearest end points. Ballpark age data was gathered from Baseball Reference, with eight years once again being the cutoff for the honeymoon dummy variable. Finally, I decided to gather data on Fan Cost Index as reported by Team Marketing Report. This index is a value in United States dollars which represents the cost of two adult average priced tickets, two children priced average tickets, four small soft drinks, two small beers, four hot dogs, two programs, parking, and two adult-sized caps. While obviously not every fan goes to a ballpark and purchases all of this, the FCI is supposed to represent the average cost for a family to attend a game at a given ballpark. It was intuitive that this value would correlate strongly to total revenue, and as *Figure A-3* shows, when FCI was plotted with total revenue it showed a strong positive relationship with a correlation of 75%. With this being the case, it was useful to include FCI.

With the addition of FCI, the data required to build a revenue projection model was compiled, and a regression analysis that was very similar to the one explained in J.C. Bradbury's book could be performed.

CHAPTER FOUR

Player Value Model Results

With the data compiled, the next step was to fit a model that could serve as a reliable predictor of player MRP based off of individual results. *Table B-1* shows a description of the variables discussed in the previous section that would be used and *Table B-2* displays the summary statistics. The first models I attempted were to run the same exact regressions as Bradbury does in his book, with and without FCI and yearly dummy variables included in the model. Much like when Bradbury created his model, the regression analysis performed here would be estimates of panel data, or, "data of a cross-section of units over multiple observations" (Bradbury 2010). Additionally, much like when Bradbury ran his model, the unadjusted pooled regression would face the problems of heteroskedasticity and first order serial correlation. The heteroskedasticity meant that there was unequal variability of the residual values across all predicted values and the first order serial correlation meant that current estimates were affected by the previous estimate, resulting in error terms that were correlated over time. While this did not affect the bias of the estimators in the model, it did affect the efficiency of them, resulting in estimates of confidence intervals and t-statistics that were unreliable. To correct for this, a Newey-West correction of the standard errors was used, much like how Bradbury corrects for this issue in his book.

As models 1 and 2 in *Table B-3* demonstrates when the regression model that Bradbury settled on both with and without the inclusion of the FCI and yearly dummy variables were run, the cube of run differential variable was statistically insignificant both times. To investigate why this was occurring, I decided to plot the relationship of total revenue and run differential in *Figure A-4* similar to what Bradbury does in his book. As this image shows, the relationship between total revenue and run differential clearly is not linear, much like what Bradbury found. That said, the area of this graph in which the relevant data lies does not provide significant evidence that the relationship must be cubic. While a cubic graph and a quadratic graph have very distinguishing features when fully drawn out, the area of the curve where the data points in this study lie only show a convex relationship, without any clear inflection points, which is a characteristic of both graphs. For that reason, run differential and total revenue was treated as a quadratic to see if that relationship better fit the data.

Model 3 and model 4 in *Table B-3* shows the same models as 1 and 2, with the relationship of run differential and total revenue being a quadratic rather than a cubic. In both models, the square of run differential was significant, and there was a stark difference in adjusted R^2 when FCI and yearly dummy variables were added to the equation meaning model 4 explained more of the variability in total revenue as explained by the independent variables than did model 3. Also, as shown in the previous chapter, FCI had a strong linear relationship with total revenue, and it made sense that it should be included in the model, while the discussion on national television revenue explains the usefulness of adding yearly dummy variables. Given these reasons, model 4 appeared to fit the data the best over the alternatives. While other models could have been tested including interaction variables or fixed- and random-effect estimators for teams, Bradbury sheds light on the decision to not move forward with these models in his book.

As he explains, his interaction models were all highly insignificant. Along with this, his reasoning to not move forward with random- and fixed-effects models was that “the known characteristics of teams includes their propensity to win” (Bradbury 2010). In other words, random- and fixed-effect methods, “attribute a large part of a team’s unique characteristics to revenue generation, which leads to a near-equal apportionment of revenue to all players” (Bradbury 2010). This obviously is not the goal of this model, as based off of real MLB salaries we can confirm that is not how the pay structure in the league actually works. For this reason, I decided to omit these types of procedures, and settled on the model 4 as my MRP projection model for this study.

CHAPTER FIVE

Previous Draft Analysis and Theoretical Hypothesis

With a model set in place for valuing players' MRP, the focus could be shifted to answering the main question of this study, as to whether or not there is an optimal strategy to pursue in the MLB Rule IV Draft. Some economists and baseball analysts have done some research into this field, and have offered fairly consistent results. The first study I referenced was a 2006 article on Baseball Prospectus' website written by Rany Jazayerli. In Jazayerli's study, he looks at wins above replacement player values for MLB draft picks. Wins above replacement player is a metric created by sabermatricians that accounts for every statistic a player attributes to his team and assigns that player an amount of wins he contributed to his team over a replacement player, or a player who could be acquired for the league minimum. Jazayerli looks at discounted values of 15-year wins above replacement level player for draft picks, and separates groups by high school, college, and junior college players. He then finds an expected value of discounted wins based on pick number that a player should have hypothetically generated over their career. He then subtracts the expected value a player would have generated based off of his pick number and the actual discounted value he produced to calculate a difference as well as a margin of return based off of actual wins above replacement over expected wins above replacement. While he does not separate by position, he concludes that on average in the time period of 1992-1999, "college draft picks yielded approximately 25% more value than high school players" (Jazayerli 2006).

A follow up study was performed in 2009 by Victor Wang, now a member of the Cleveland Indians front office. In his study, he finds players' wins above replacement per year for each player's cost controlled six year window, and breaks the groups up by high school and college as well as hitters and pitchers. He then looks at different rounds the players were selected in, breaking them up by first round, supplemental and second round picks, and third round picks. The results that he finds, are that in the first round, college players have a slight edge over high school players, and hitters have an advantage over pitchers. He finds, however, that college players lose their advantage after the first round as do hitters over pitchers in the years 1990-1997. He concludes with an optimal strategy of, "hitters first, pitchers next" (Wang 2009). This strategy is fairly consistent with Bill James' 1984 findings that college players produced more value than high school draftees, and that position players provided more value than pitchers. Finally, General Manager of the Oakland Athletics, Billy Beane has been noted to be of the opinion that high school players are overvalued compared to college players, and thus college players provide more return.

In summation, most previous research on this topic leads to consistent viewpoints that there is an optimal drafting strategy, and that is to place preference on college hitters in the draft, especially in the first round. While it is possible that this strategy may hold true, there are some important factors some of this previous research has left out, that may lead to a different result. The first, is that these studies have not taken into account the cost of acquisition of a player. In the MLB Draft, all signing bonuses are not created equal. Some players command high bonuses, and some require being offered MLB contracts before they agree to join an organization. For example, in the 2001 draft, Mark Teixeira was selected 5th overall out of Georgia Tech. In order to get him to sign with the Texas Rangers, the club had to offer him a four year contract worth $9.5 million in addition to his $2.5 million signing bonus before he ever stepped foot on a field. This cost of acquisition is incredibly high over a player who could be obtained at the end of the round such as Mike Fontenot who signed on board with the Cubs for a $1.4 million signing bonus at pick number 19 in the same draft. So, while Teixeira may have produced better numbers, and thus a higher wins above replacement over the early years of his career, did he produce significantly more than Fontenot to justify the much higher cost of acquisition? In addition, these studies do not take into account the lost expenses on players who do not make the major leagues, accounting for their loss from signing bonus as well as their minor league salaries while they tried to make it to the big leagues. In other words, these studies do not account for the variability between players of different age and positions. To see that a certain group produces higher average production is useful, but it does not tell the whole story, as it does not necessarily confirm that that group actually provides more return on investment.

If players are being drafted efficiently, and signing bonuses are a fair proportional representation of future production on average, economic theory would suggest that there should be no optimal drafting strategy or exploitation of the market, but rather a necessity for good scouting and decision making by pick. Theoretically, there should be no difference in return on pitchers or hitters, or college or high school players, as what they get paid to sign and in ensuing salary should be a reflection on what they will produce in the future. In other words, if on average, hitters do actually perform better than pitchers, they should require a higher cost of acquisition to align them with this production. Thus, their return on investment, measured by their MRP subtracted by their compensation, should fall in line with their pitcher counterparts who would see a lower cost of acquisition for lower expected production. While MLB rules would make it reasonable to assume the top players on average provide a positive return over their MRP's until they reach a service time of salary adjustment, the returns should be consistent amongst positions and age. These papers have touched on this topic, however none of them have truly assigned monetary value to production versus acquisition, and it is my hypothesis that when this procedure is done, it will shed light on the fact that there truly is no exploitable strategy in the Rule IV Draft.

CHAPTER SIX

Performance Metrics and Methodology

With the expectations of results set, it was important to decide on performance metrics to plug into the model built in Chapter 4 to test this paper's hypothesis. The first place I turned to for this was Bradbury's book. While the model built for this paper has some slight differences to Bradbury's, the general methodology was consistent, so it was useful to revisit Bradbury's analysis and determine how he utilized his model to evaluate a player's marginal revenue product. Bradbury decides to use "park-adjusted linear weights and plus/minus estimates of run contributions for each player" (Bradbury 2010). In other words, he finds estimates of the amount of runs players generated on offense and defense. He then compares this value to an average player in Major League Baseball. He multiplies this runs above average number to the weights found in his model for run differential, so that that number multiplies by the weight for run differential, the total runs above average squared multiplies by the weight for the square of run differential, and cubed runs above average multiplies by the weight for run differential cubes. This value provides Bradbury with a sum of the value above an average player that can be factored into a player's MRP calculation. Once he has this result, he must determine what an average player's value would be so that he can add this to the number he just found to calculate a player's total value. Bradbury uses the intercept of his model as the total revenue an average team would expect to generate, and thus he develops a formula to divide this intercept up for both offensive and defensive contributions. By using the ratio of plate appearances a player makes of the entire team's plate appearances, and positional adjustments for defense, Bradbury calculates the value of an average player with the same playing time as the player being studied. He then adds the player's value above average calculation to the value of an average player and finds the total MRP of the player.

For pitchers, Bradbury uses a slightly different approach. He creates a model using players' defensive independent pitching statistics and estimates the expected runs they would allow based on strikeouts, walks, and home runs. He adjusts this value to account for home park influence. He does this at the team level, so to calculate a player's MRP, he simply multiplies the percentage of batters faced by a pitcher in relation to the total batters faced by the team to the value he finds in his total runs prevented model. He once again uses a fraction of the intercept to account for the value of an average pitcher who pitched the same amount of innings as the pitcher being studied. When this average value is added to the value above averaged, he arrives at an unadjusted MRP for pitchers. The final adjustment Bradbury makes for pitchers is to account for the amount of innings pitched from the 7th as he finds runs allowed in these innings influence winning probability more. When this adjustment is made, he has his final estimate for pitcher MRP.

I found Bradbury's method to be interesting, and understandable, however, I was not convinced that using runs above average was necessarily the best way to move forward with my model. Bradbury does not simply plug a runs created value into his model to estimate a player's MRP, because given the context of baseball, if a player did not play, someone would take his spot both on the 25 man roster as well as his playing time. This substitute player can be thought of as the 26th man on the team, or the first replacement not currently on the MLB roster. Thus, in order to calculate a player's MRP, the added runs that player produces over what his substitute would produce in the same positions and playing time is that player's true run contribution rather than just the sum of his total runs. Of course, it is possible should a player not be on a team, a player currently on the bench would take his playing time, but for the sake of analysis assume all bench players are of about equal quality as this 26th man, so the substitute level of play would be the same whether it is someone who would be a bench player or off the team entirely. The Bradbury method suggests that this roster spot and playing time would be filled by an average Major League Baseball player who would otherwise be on the bench or in the minor leagues. The problem with this assumption is that an average player in Major League Baseball is usually not a readily available substitute. If average players were easily accessible, it would be the assumption that every MLB team should finish with as many wins as loses or better. This is obviously not the case, and thus, leads to the more realistic thought that non-starters in the major leagues are typically below average players.

As explained earlier in the previous chapter, a replacement player in sabermetric research is considered a player who can be acquired for virtually no cost, or in this instance, the league minimum salary. This can include below average veteran players, or minor league players. The definition of a replacement player in sabermetrics states that a team full of replacement players would have a winning percentage of .294, down from the expected winning percentage of .500 that a team full of average players would be able to achieve. While it is likely that some teams may have substitutes available who are better players than theoretical replacement level, it is more reasonable to assume that teams' bench and high minor league level players are closer to replacement quality than the average level assumed by Bradbury. Because of this, a similar approach could be used in my model as Bradbury's, however, a runs above replacement metric would be a more accurate representation of a player's MRP than runs above average, and thus, that is what I decided to use for this study.

Much like in Bradbury's method, I had to find a quantifiable value for the total runs produced by a player. The metric I referenced for this value was Bill James' total runs created metric. Total runs created is a value reported by Bill James ever year that produces a value that a player contributes to his team summing up offensive and defensive runs. This is a counting statistic and is not scaled to zero for league average or replacement players. For offensive runs, James uses a metric called runs created. James' runs created is a statistic that measures a hitter's value to his team based off of his ability to score runs. In its simplest form, runs created is measured by hits plus walks, multiplied by total bases. This total is divided by at bats plus walks. Since the metric was invented in the 1970's, however, it has gone through many technical changes, and is now a lot more advanced than the simple formula it started as. Rather than using James' runs created formula, I decided I would use Fangraphs weighted runs created (wRC), which "is an improved version of Bill James' Runs Created (RC) statistic which is based off Weighted On-Base Average (wOBA)," (Fangraphs 2015) developed by Tom Tango. The formula for weighted runs created is:

wRC = ((wOBA-League wOBA/wOBA Scale) + (League Runs/Plate Appearances))* Plate Appearances

To understand wRC, it is essential to understand wOBA. Weighted On-Base Average is grounded on the concept that "not all hits are created equal" (Fangraphs 2014). The classic metric of on-base percentage weights every trip on base the same, whether that be a single or a home run. Slugging percentage adjust this slightly, adding weights to hits, but it is not accurate as research has shown a double is not truly worth twice as much as a single. This new metric of wOBA "combines all different aspects of hitting into one metric, weighting each of them in proportion to their actual run value" (Fangraphs 2014). In other words, this metric more accurately measures what each trip on base does for a player's total production to offense using linear weights. There is a strong relationship between wRC and Bill James' runs created, so either of these metrics could have been utilized and would have produced very similar results, however, the technical adjustments from wOBA that produce the wRC led me to believe that wRC would be a slightly more accurate representation of a player's offensive runs created.

With wRC settled on as an offensive metric, I then had to decide on how to value a player's contribution on defense and on the base paths. For this, I once again turned to Bill James. In James' total runs created calculation, he uses John Dewan's defensive runs saved (DRS) from his *Fielding Bible* books. Defensive runs saved utilizes film study and computer comparisons to calculate how players make plays in the field relative to the league average. As Joe Posnanski of Sports Illustrated explains, "if a shortstop makes a play that only 24% of shortstops make, he will get .76 of a point. If a shortstop blows a play that 82% of shortstops make, then you subtract .82 of a point" (Fangraphs 2014). The total value of runs saved takes into account stolen bases saved by pitchers and catchers, double plays turned by middle infielders, bunts saved by corner infielders, and outfielders' ability to prevent runners from advancing extra bases. When everything is added together, a counting metric is produced to represent runs saved above an average fielder. The total runs created value, however, was not supposed to be a comparison to average players, but rather a counting metric from zero. To adjust for this, Dewan also adds in a positional adjustment, which accounts for the difference in difficulty and importance of defense in certain positions such as shortstop and catcher. *Table B-4* reports the positional adjustments used, and they are reflections of playing a full season, or nine innings in every one of the 162 games of a MLB season at a position. To measure a player's positional adjustment, I simply found the ratio of innings they played at a position relative to a full season of innings and multiplied it by the respective adjustment. All of the positions were added up to get a total positional adjustment. This adjustment serves two purposes. First, it allows players across positions to be compared. Average defense at shortstop is worth more to a team's chances of winning than average play at first base, as shortstop is the more difficult defensive position. Additionally, it levels the playing field slightly for players who produce less offensively, but play great defense at those important positions over players who just go out and hit homeruns at positions like designated hitter. Along with this, adding a positional adjustment builds in the value produced by an average defender. If a player has zero runs saved, they are still credited for providing that average value by receiving their adjustment based off their innings played, which is a similar concept to what Bradbury uses in his analysis to calculate average defensive value.

With hitting and defense squared away, the next metric added into a total runs calculation is base running. For this value, I used the what Bill James uses in his total runs created calculation, which is to add a quantifiable value for base running plays such as steals, advancing to extra bases on base hits, and tagging up on fly balls to name a few. The metric utilized for this paper was reported on Fangraphs as base running (BsR), and this value was consistent with that utilized by James.

When values for offensive runs created, defensive runs saved, base running runs, and a positional adjustment were added together, a total runs calculation could be made for position players. This value is the sum of all of the runs a player added to his team's total during his season of play. Before any analysis could be carried out, however, it was necessary to create a metric for pitchers that would utilize the same scale as hitters. For this, I referenced a 2006 article by David Gassko titled, "Pitching Runs Created." This article devises a method in which a pitcher's production in run prevention can be converted to what a team would have to produce offensively to match that production. To devise this method, Gassko calls on what is called the Pythagorean Theorem of Baseball, also an invention of Bill James which defines expected winning percentage to be a function of the square of runs scored divided by the square of runs scored plus the square of runs allowed. Gassko uses this formula to explain that a pitcher "who allows one run a game is exactly as valuable as a lineup that scores 15.4" (Gassko 2006). The procedure to convert these runs allowed to runs scored involves finding a player's runs allowed average. This value is essentially the same as the classic earned run average, except it also takes into account unearned runs. Next, the league average for runs per game is found, whether it be the AL or the NL, and added to the runs allowed average of the pitcher to create a run environment. This run environment is raised to the .287 power, to create a custom exponent to use in the Pythagorean formula. The average runs per game of the league and the runs allowed average of the pitcher are plugged into the Pythagorean formula to produce a pitcher's custom winning percentage. The average runs per game is then switched from runs scored to runs allowed in the formula, to find how many runs a team would need to score to match that winning percentage in the run environment defined. This value is then multiplied by the innings a pitcher pitched in the season and divided by nine. Finally, Gassko uses an adjustment for defense. He finds that pitchers who have higher strikeout rates should be given a bigger share of the defensive credit for a team, and thus he multiplies his pitching runs created value by a percentage of credit the pitcher should receive based off of their strikeout rate, that he has calculated through empirical data. The value calculated after this is a pitching runs created value that is on a consistent scale as offensive runs created.

I decided to replicate Gassko's procedure, with a few alterations. The first, was that I felt a more accurate metric could be used than runs allowed average. Additionally, I felt if a more accurate metric could be utilized, the adjustment made for strikeout rate would be unnecessary. The metric I ultimately decided to use was skill interactive earned run average (SIERA), developed by Matt Swartz and Eric Seidman of Baseball Prospectus in 2010. This metric is an ERA estimator, and is thus on the same scale as that classic statistic. SIERA is used to evaluate what the overall skill level of a pitcher is, giving higher credit to strikeouts, which is consistent with Gassko's findings, and makes adjustments for a pitcher's ground ball and fly ball rates, home ballparks, and run environments. Much like other metrics such as fielding independent pitching (FIP), SIERA represents a value of what a pitcher did individually, and does not account for the defense behind him. Thus, SIERA could be used in replacement of runs allowed average, as it is a more accurate picture of a pitcher's skill level, and there is no need to adjust the initial pitching runs created value by anything, as SIERA already accounts for the added skills Gassko talks about. This method of calculating pitching runs created is similar to Bill James' in his total runs created report, however, he uses component ERA rather than SIERA. While this is essentially a choice of personal preference, SIERA has performed well as an ERA predictor since its inception, and I felt it would be the most accurate representation of a pitcher's talent.

With pitching runs created now added, to find the total runs created for a pitcher, the only steps remaining were to add in offensive runs created, base running, defense, and positional adjustment, much like with their position player counterparts. Once this was completed, a consistent input variable was created for both pitchers and hitters that would make them easily comparable for the draft analysis.

With total runs created calculated for both hitters and pitchers, it was essential to find a way to estimate a replacement player's total runs created for each player to compare to much like how Bradbury devises a method to find an average player's value. To do this, I once again utilized the Pythagorean Theorem of Baseball. As stated previously, sabermetric research has defined a team full of replacement players to be a team that would have a .294 win percentage over a full season. Given that definition, I found the value of a replacement's total offensive runs created for a particular position player by plugging the .294 win percentage into the Pythagorean formula to generate a value for the total runs a team full of replacement hitters would score based off the amount of runs that player's team allowed over the course of the season. I then found the total number of plate appearances that player's team had over the course of the year, and multiplied the total runs scored for the replacement team by the ratio of plate appearances the player took as a reflection of his team's total plate appearances. This represented the offensive runs a replacement player would generate on the same team as the player in question given the same playing time. To find the defensive runs a replacement player would generate, I assumed they would save no defensive runs, and thus would be an average fielder. I added the same positional adjustment as the player in question to the replacement player's offensive runs, and this provided me with a total runs created value for the replacement player playing the same amount of time as the player being evaluated. Given this value, I could subtract a hitter's total runs created value by their replacement player's total runs created value to find their runs produced above the replacement.

For pitchers, a much easier method could be used to find a replacement player's total runs created. All that had to be done in this instance was to follow the same procedure used to calculate pitcher's runs created, however, rather than using the calculated win percentage of the pitcher from their SIERA, the win percentage used in the Pythagorean formula was the replacement team win percentage of .294. This formula then produced a replacement runs per game, and multiplying this value by the pitcher in question's innings pitched and dividing by nine produced a replacement pitcher's pitching runs created given the same amount of playing time on the same team. Finally, the positional adjustment was added to this value to calculate the replacement pitchers total runs created. Once again, the total runs created of the pitcher being studied could be subtracted by the replacement pitcher's total runs created to find the pitcher's runs created above replacement.

With every player now having a value for runs above replacement as well as a total runs created for their personal replacement player, the MRP model could be utilized to evaluate a player's value for that season. To do this, the player's runs above replacement was plugged into the equation to find the player's value above replacement added to the team due to their performance rather than the replacement player who would take their playing time if they did not play. To find the player's total value, then, the replacement runs created was plugged into the model to find the value that the replacement player would be worth given the same playing time. The value above replacement could be added to the value of a replacement player to calculate the total MRP of a player. This is a nearly identical procedure Bradbury takes to calculate MRP, with the utilization of a replacement player rather than an average player.

I did consider adjusting MRP calculations based off of the run differential for team in which the player was a member of, I ultimately decided to not do this. The theory behind this adjustment follows with the idea that wins produce more revenue the better the team is. This is seen by the nonlinear relationship of run differential and total revenue. Thus, a player who produces equal amounts of runs on a good team as someone on a bad team, would generate more revenue than the player on a bad team. This adjustment would be very useful in determining what a particular team would be willing to pay to acquire a player in free agency, and could provide more accurate measures of what a player's value truly was in a given year. However, in the context of MLB Draft analysis making this adjustment could alter the true results of the question at hand. If traditionally good teams follow a consistent drafting strategy that differs from a bad team, a clear difference amongst position or age would be evident even if the players were providing equal production. It could be argued that the drafting strategies are the reason these teams are better, however, that would be ignoring several important factors such as quality free agent additions and trades that are added to drafted players to fill out a roster. Thus, for the sake of this analysis, the best bet was to consider all of these drafted players playing on otherwise average teams.

The question still remained, however, as to whether or not the method explained in this chapter would provide estimates consistent with MRP, much like Bradbury's model had. To test this, I decided to use my model to create MRP estimates and compare them with MLB player salaries. To do this, however, specific types of players had to be chosen. Initially, my idea was to compare the MRP model with a random collection of players and see how it correlated with salary. This, however, is not necessarily the best way to test the model. The biggest reason for this is multi-year contracts in Major League Baseball. Player production usually sees some fluctuation from year to year. Some seasons, players have career years and breakout performances, and thus, their production exceeds their salary, while other years, players can get hurt or underachieve and not live up to expectations. A proposed solution to this, then would be to evaluate free agent contracts in regards to previous performance. Once again, however, this is not a great way to test the model. As discussed in earlier parts of this paper, free agency is not necessarily a great indicator of MRP for various reasons. One large one, again, is due to multi-year deals. When a player signs a contract over a number of years, a MLB team does not simply take their marginal revenue product, and multiply that over the life of the contract to arrive at a total value. They must project out a player's performance over the life of the deal, and often times, that results in deals that are smaller than if a player had signed a one-year deal every season over the same time period. Why then do players sign multi-year contracts? The answer to this is for security. MLB contracts are fully guaranteed, and when a player signs a long-term deal, they are guaranteed payment no matter what happens to their health or production level. Teams know this, and this also will cause them to offer less than a player's current MRP multiplied over the length of the contract. For this reason, free agent contracts are a potentially unstable measure of a player's true worth.

The best bet to compare my model to a player's real MRP value, then, was to look at player's who had signed one-year contracts. A one-year deal is the closest we can get to a player receiving what his actual MRP is in baseball. In theory, when a team agrees to terms with a player on a one-year basis, they are paying him exactly what their expectation is based off of past performance as to what he will produce on the field. As with all free agent deals, this can be slightly skewed, as players at the end of their careers could be willing to accept less than their worth to chase a championship, or player's can take a home town discount. That said, evaluating one-year deals to MRP projections should provide some insight to the power of the model. As *Table B-5* shows, I evaluated the five highest hitter and pitcher deals agreed upon in the 2015 offseason. For the inputs to calculate a total runs created value, I used Fangraphs Steamer projections for 2015 numbers, as well as weighted averages of the last three seasons for any values not present in their projection system. To find replacement values, I used the total at bats and runs allowed of the team the player was joining from the previous season. As the results show, the model was about $1.7 million off from the real contract on average. Additionally, in *Table B-6* I calculated salary projections using the same model with just raw total runs created as the input rather than the adjustment made for replacement. Finally, in *Table B-7* I mirrored the procedure I used to calculate MRP with replacements, however, I set the values to now account for average players. This is not the exact procedure that Bradbury takes, but it is the same in theory. As the tables show, the MRP calculation when replacement players were used performed the best across the board, which is what I had anticipated. While there was some error in the projections, this is certainly to be expected, as projection systems are merely educated guesses, and there could be other factors at play effecting agreed upon deals. That said, I felt that an error under $2 million was sufficient to conclude that total runs created above replacement would serve as an effective performance metric to use in the model.

Finally, with the model created and the performance metrics devised, it was time to develop the methodology to evaluate the MLB Rule IV Draft. The first decision I had to make was what players to use and from what time period. I decided to use the picks that were selected in the top 100 picks of the draft that had agreed to sign contracts over the 10-year time period of 2000-2009. Any years after 2009 would likely have not been that relevant, as many players would not have reached the big leagues yet. I accounted for the player's drafted position, whether they were drafted out of high school or college, and their inflation adjusted signing bonus. For positions, I grouped them into infielders, outfielders, catchers, or pitchers. I did not separate by infield or outfield position as many players were drafted without specific positions.

The next big factor to decide was how many years of a player's career I wanted to utilize. In Major League Baseball, a player has no negotiating power in his annual salary until he reaches three years of service time. At this point, a player is eligible to have a hearing in front of a salary arbitrator where both the player and the team submit salary offers, and the arbitrator selects the player's salary based off the cases made by both teams. Some players who have significant playing time in their first two years, known as "Super Two" players reach this status early. Arbitration is the first time, salaries are aligned with performance in a player's career, although they are still less than what the player would find in the open market. After six years of service, a player can become a free agent, and thus, can sign with any team on the free agent market. Some draft analysis papers have looked at a player's pre-arbitration years while others have looked at the full six seasons. In this study, I decided to only look at the years of pre-arbitration. One reason for this, is simply that not every player ever gets to arbitration. In fact, most players are signed to extensions that buy out the player's arbitration years for the team. To evaluate a player's return over their salary during an arbitration extension, would be adding into account how well teams execute strategies, their foresight for player progression, and negotiating skills. The goal of this paper is to evaluate the draft, and the arbitration process begins to add outside noise to the analysis. Additionally, some smaller market teams find when players get to arbitration they become too expensive and they feel the need to trade or release them. On the other hand, pre-arbitration players are almost never moved due to salary, and thus, in theory, a team who drafts a player is essentially guaranteed a player's pre-arbitration efforts if they want them. For these reasons, pre-arbitration years made the most sense for this analysis.

While the first season of arbitration served as a cutoff for players who made the majors, I could carry out the analysis of a player's total return on investment. To do this, I found a player's salary value for each year up until their first year of arbitration. For salary data that was unavailable, I used the league minimum based off the season in question reported by Statista. I then calculated the player's total runs created above replacement for each season and plugged that into the MRP estimation model. I added this value to the value of the player's replacement to find the player's total MRP. I then subtracted their salary from their MRP to calculate a return. I summed up their return over each season to arrive at a total return on investment. For the first season a player was drafted, I added their signing bonus to their salary, and utilizing a 2014 ESPN article that revealed that minor leaguers typically make anywhere from $3,000-$11,000 a year, I decided then, for every season a player played in the minors, I would assign them a salary value of $5,000. This study included both players who made it to the majors as well as players who never did. With the methodology and metrics in place, I could evaluate the draft picks to look for any potential trends in the data.

CHAPTER SEVEN

Discussion and Results

Once the database of players was built, the distribution of top performances by position was analyzed. *Table B-8* lists the top performances in terms of dollar worth of all the players included in the study in addition to their runs above replacement and total replacement runs. The distribution of hitters to pitchers was 35-15. By fielder position, infielders held 16 of the top 50 scores each, outfielders had 13 and catcher had six. This distribution was compared to the distribution of the top 50 MLB salaries for the 2015 season as reported in *Table B-9* from Spotrac. This shows that 30 of the top 50 highest paid players in baseball are hitters, with 20 being pitchers. By position, pitchers lead the pack with 20, infielders with 17, outfielders with 11, and catchers with two. The two distributions are very similar, and further sheds light on the fact that the runs above replacement method is predicting reasonably. While hitters appear to dominate the distribution of top seasons, they also dominate the distribution of top salaries.

Table B-9 shows the top 50 returns on investment in the study. The distribution, as expected, comes much closer to an even split, with hitters accounting for 27 of the top 50 and pitchers accounting for 23. *Figure A-5* shows the trend of mean returns between hitters and pitchers by Draft year over time, and *Figure A-6* shows the same with hitter positions broken down specifically. Finally, *Figure A-7* shows how the mean of returns has varied over time broken up by high school and college players. While these visuals depict some interesting information, they do not reveal anything conclusive.

While the mean of return on investment provides some useful information, it is not the only metric that can be analyzed to reveal the true relationship between return on investment between position and age. The reason for this is demonstrated by *Figure A-8,* which is a depiction of the distribution of returns of all players in the study. As this image illustrates, the majority of returns fall right around zero. This is intuitive, as one would expect signing bonuses and salaries to be a reflection of future performance as explained previously in this paper. Therefore, if this compensation is efficiently given out, one would expect most players to produce exactly at the level they are paid. Therefore, the majority of players will provide a return on investment right around zero, with quite a large amount of players who never make the MLB providing negative values, and some who become legitimate contributors to provide positive values. If this was not the case, players would likely lobby for higher signing bonuses and salaries if returns were typically much higher than zero, and the MLB team owners would argue for lower compensation if the reverse were true. That said, as illustrated by this image, there is still a portion of this distribution that moves to the far right of zero, with returns reaching as high as $60,000,000. It is the goal of all teams to draft as many players who fall in this area of the curve while drafting as few as possible who fall to the left of zero. Given the shape of the distribution, it is clear that while the mean has value in reaching conclusions, it is not the only metric that can be looked at, as this value is clearly being pulled to the right by the extreme values displayed. For that reason to compare returns between pitchers and position players as well as high school and college players, it must be the entire distribution that is compared rather than just an average. By comparing the entire distribution, it can be statistically concluded whether or not, on average, the probability of selecting a player who provides positive or negative return is more likely given a certain group of players, and this will reveal the conclusions desired in this study.

To test for difference in distributions, a two-sample Kolmogorov-Smirnov test was used. This test is a nonparametric test that evaluates against the null hypothesis that two independent sample distributions come from the same population. In other words, this test will take into the account the characteristics of the distributions and will provide a conclusion as to whether or not there is significant evidence of a difference in distributions between two samples, whether that difference be from central tendency, spread, or shape. This test, along with analyzing some summary data and visuals between to distributions should allow some significant conclusions to be drawn.

The first group analyzed was position players and pitchers. *Figure A-9* shows a plot of the approximate distributions of return on investment of these two groups. Visually analyzing this graph, it appears that pitchers have a higher probability of falling within a central point right around zero. It appears as though position players provide a slightly higher probability of falling to the right of the curve at returns around $20,000,000, with this phenomenon beginning to stop at the higher extremes. However, they also appear to have a higher probability of providing returns that are below zero than pitchers. To build off of this eye test, some numbers were brought into the analysis. *Table B-11* displays the 25th percentile of the distribution, median return, 75th percentile of the distribution, interquartile range, maximum value, minimum value, and mean return on investment for both position players and pitchers. These numbers are consistent with the analysis of the distribution curves. Both groups have similar slightly negative medians that lie right around zero. The position player group has a higher mean value, in addition to a higher value at the 75th percentile than pitchers. An interesting observation to note is that position players provided a higher maximum return, but also a lower minimum return, which is consistent with the observation from the distribution plots. When the Kolmogorov-Smirnov test was run against the null hypothesis, there was insignificant evidence to conclude that returns of position players were different than pitchers. Taking all of the relevant information into account, it appears that position players provide a higher risk and higher reward than pitchers, but there is no clear exploitable strategy between the two.

The next part of the analysis was to look at the difference between position players and pitchers selected in the first round. *Figure A-10* displays the distribution of returns on investment between these groups of players. This distribution plot looks quite a bit different than the previous. Once again, pitchers appear to have a higher probability of falling within a return around zero than position players, however, in this graph, the difference in positive returns seems to be of much greater significance in the favor of position players. When analyzing the values provided in *Table B-12*, it can be seen that position players appeared to provide a higher mean and median return on investment than pitchers, with a higher value for the 75th percentile as well. For this group, the Kolmogorov-Smirnov test provided significant evidence to conclude a difference in returns on investment between hitters and pitchers in the first round exists, in favor of the position players. This was an interesting result, and fell in line with some previous research referenced in this paper, however, it did not follow economic theory. To investigate this further, I looked at disparities in median and mean return on investments in the first round by year. The most notable difference was demonstrated in 2005. When I looked further into the players selected in this year, I found that position players selected included Justin Upton, Alex Gordon, Ryan Zimmerman, Ryan Braun, Troy Tulowitzki, Andrew McCutchen, Jay Bruce, and Jacoby Ellsbury, all of who became very successful players. It was possible then, that 2005 was the outlier, and was creating a relationship that did not truly exist. To test this idea, I reran the summary statistics with the exclusion of the year 2005. As seen in *Table B-13,* the results appeared to be a lot more consistent between the two groups in this instance. When the Kolmogorov-Smirnov test was run without 2005 included, the resulting test statistic was insignificant, leading one to question whether position players truly provide higher returns on investment in the first round.

After analyzing the first round, the same procedure was done for rounds other than the first. *Figure A-11* displays the results of these distributions, and by analyzing them with the naked eye, they look very similar. *Table B-14* confirms the similarity in these distributions, as many of the values between the two groups are not very far apart. The Kolmogorov-Smirnov test failed to provide significant evidence of a difference in return on investment between hitters and pitchers in rounds outside of the first, which is consistent with the expected result.

Finally, the last positional tests that were of interest were by actual positions of the hitters. *Figure A-6* shows the average return by year of these positions, and given the large fluctuations by year, it is unlikely that any of these positions provide a significant advantage over the others consistently, however, the test was still run. *Figure A-12* displays the distributions of return on investment for catchers, infielders, and outfielders. It is clear from this visual that infielders and outfielders have very similar distributions, while the distribution of catchers appears to be more centrally located around zero than the other two. *Table B-15* displays that each position recorded very similar median and mean returns. Catchers provided a lower maximum return as well as a higher minimum return than infielders and outfielders, also demonstrating a lower spread displayed by the IQR value. This could once again suggest catchers are less of a risk-reward pick than the other two positional groups. A Kolmogorov-Smirnov test was run for each of the three positional groups, and failed to show a significant difference between any two of them. This is logical, and falls in line with the expectations brought about by the visual and summary metrics.

With the analysis of position providing very little conclusive evidence of a difference in return on investment, the focus could be shifted to the high school and college player groups. The overall distribution as seen in *Figure A-13* shows a much higher percentage of returns falling right around zero for high school players, while college players appear to have a larger percentage of returns immediately to the right of zero than high schoolers, with the behavior at the right extremes appearing to be consistent. *Table B-16* shows that while both groups the college players have a slightly higher mean than the high school players, as the 75^{th} percentile is much higher for this group. This remains consistent with the visual, as it appears that college players provide more consistent positive returns than do high school players. The Kolmogorov-Smirnov test confirms this, as it shows significant evidence of a difference between returns of college and high school players. Given this result and the other information mentioned, it can be concluded that college players are the preferable selection over high school players overall.

The result of the high school and college analysis found concluding evidence against my hypothesis, so this prompted me to look further. I decided to evaluate whether this relationship was consistent amongst all rounds, or if it could be explained by just the first. *Figure A-14* shows the distribution of returns for first round players. This image is very telling, as it appears high school players not only have a higher probability of landing around zero, but also have a higher probability of providing a negative return on investment. The right side of the image shows that college players are have a much higher proportion of positive returns relative to high school with the behavior at the extremes looking fairly similar. The statistics in *Table B-17* show that college players have a much higher median return on investment, while the mean shows an advantage in the favor of college players as well. The maximum return for high school players is higher than that of college, but the minimum return is also lower, demonstrating the risk involved with selected such a player. The Kolmogorov-Smirnov test shows results similar to the overall high school and college results, providing statistical evidence to conclude that there is a difference in returns between the two groups. It is clear, that college players are the more preferable selection overall in the first round.

This same analysis was carried out for rounds outside of the first to see if the results were consistent with the previous two. *Figure A-15* displays the very peculiar graph of distributions between high school and college players in rounds outside of the first. It is hard to draw many conclusions from this graph, however, it appears that high school players have a larger proportion of returns around zero, and a lower proportion to the direct right of zero. *Table B-18* shows that the medians and the means are fairly consistent, as are the minimum and maximum values. The major difference between these two groups appears to lie in the 75th percentile, where college players see a much higher value than their high school counterparts. The Kolmogorov-Smirnov test statistic was significant against the null hypothesis, and thus, it could be concluded that there is a difference in returns on investment in favor of college players across the entire MLB Draft.

It looked clear that college players were providing more return on investment on average than high school players, however, I decided to try and evaluate this relationship by position to see if there was any position that was driving this relationship, or if college players seemed to be the best pick on average in the draft no matter what. The first position I evaluated was catcher. The distribution of returns can be seen in *Figure A-16,* and this image appears to show that high school catchers have a larger probability of falling around a return of zero with a lower probability of falling just to the right of that and just to the left of that than college catchers. The relationship as the returns on investment approach about $10,000,000 seem to show inconsistency between the two groups, and thus, it was hard to draw anything definitive from the image alone. The values provided in *Table B-19* show that the medians and means appear to be about the same, with the 75th percentile being higher for college players. The up and down movements of the distributions however make it hard to draw any conclusions without the use of a statistical test, and when the Kolmogorov-Smirnov test was run, there was no significant evidence to conclude a difference between return on investment between college and high school catchers.

Figure A-17 illustrates the distributions of college and high school infielders' returns on investment. This graph appears to demonstrate a preference to the college player, as it provides higher probability of values above zero by what appears to be quite a bit. *Table B-20* confirms this hunch, showing that college infielders have a much higher median return, mean return, maximum return, minimum return, and 75th percentile value. Finally, when the Kolmogorov-Smirnov test was run, there was significant evidence to conclude that the distribution of college infielder returns on investment and high school infielder returns on investment was different, and it was clear that college infielders showed a significant advantage over the high school players.

The next position evaluated was the outfielder position. The distribution of high school and college outfielders as seen in *Figure A-18* appeared to show a very similar relationship to that of the infielders. Once again, it looked as though outfielders have a much higher proportion of players in the positive returns, with the only noticeable time the high schoolers taking an advantage being at the extreme points. When the summary statistics were analyzed in *Table B-21*, similar median and mean returns were reported between the two groups. However, the 75th percentile for college players was much higher, which appears consistent with the image of the distributions. The Kolmogorov-Smirnov test provided significant evidence to conclude there is a difference between high school and college outfielders in terms of their return on investment, and thus, the conclusion could be drawn that the college outfielder was the superior choice over the high school player on average.

The final position to evaluate in the high school and college tests was pitchers. The image of the distributions of these two groups, seen in *Figure A-19,* appears to show similar curves. The high school pitchers appear to have a higher probability around zero, with a slightly lower probability to the returns around $10,000,000, however, other than that, the graphs look very similar. The mean and median values of return on investment reported in *Table B-22* show very similar results, and there is no value that is glaringly obvious to suggest a difference in the distributions. When the Kolmogorov-Smirnov test was run, the test statistic was insignificant, and thus, it could not be concluded that there is a difference in returns between high school and college pitchers.

Overall, the data showed no significant difference between pitchers and position players as a whole, with the only significant difference coming in the first round. However, when one particular draft was accounted for, this result also became insignificant. On the other hand, college players showed overall significance of higher returns on investment over high school players in all rounds of the draft. When broken down by position, this result was mirrored by infielders and outfielders, with catchers and pitchers failing to provide statistical evidence of a difference in return between high school and college players.

CHAPTER EIGHT

Conclusions

After performing the analysis on hitters against pitchers as well as high school players versus college players, there were a few interesting results to comment on. First off, I can conclude through this study, that there does not appear to be any exploitable strategy by position in the draft, whether that be hitter versus pitchers or by hitter position. While the first round results did show a significant difference in preference of hitters over pitchers, these results were explained strongly by one draft in particular, leading one to question whether that difference truly exists. This all falls in line with my expected outcome, and with economic theory, that suggests that signing bonuses and compensation should reflected expected performance, and thus, return on investment should not be different between hitters and pitchers if the draft operates efficiently.

Based off of the distributions of return on investment between hitters and pitchers, it appears that pitchers provide a little more certainty in return, while hitters are more of a boom or bust group. There could be many reasons for this, but I believe the most likely explanation to the increased volatility in hitters' return on investment as opposed to pitchers is that hitters' skillsets are harder to project long-term. For example, if a pitching prospect throws a 94-mile per hour fastball, and has plus off speed pitches with good command, his skillset is very clear. While he may need further seasoning and development, one can project this pitcher's ability to succeed in the long run based off of what he throws and how similar pitchers have faired in the MLB. Hitters on the other hand are harder to project. Metrics such as bat speed and strikeout rate, and physical size are components that can help shape a hitter's expectations, however, there is no guaranteed way to project how a hitter's skills and stats will carry over to facing better pitching in the pro game. In baseball, the pitcher controls the game, as the ball starts in his hands. For that reason, you know what a pitcher can give you consistently based off of his skills, but a hitter's performance depends on his ability to square up the ball that is thrown to him, which could vary as that pitching he faces improves. While the data shows when a team successfully drafts a hitter, it rewards them with a bigger return, that ability to draft one can be difficult, which would explain the higher probability of negative values over pitchers.

The analysis of high school players against college players showed that college players provided higher returns during pre-arbitration years on average in the top 100 picks, regardless of round selected. That said, when broken down by position only infielders and outfielders provided a significant difference in return between the two groups of players, while pitchers and catchers did not. While this does not follow economic theory, there are a few possible explanations for this. One explanation is that high school infielders and outfielders generate more off field revenue than college infielders and outfielders do. This is possible, as many times organizations try to sell a young prospect as the face of their future. This can draw more fans to minor league games to watch this young player and can raise merchandise sales when he finally does arrive to the MLB team. College players typically spend less time in the minor leagues, and thus they cannot build as much hype as their high school counterparts before joining the MLB club. If MLB teams feel they can sell a prospect as "the next Derek Jeter," it is possible that they are willing to pay more to acquire an inferior talent, as other sources of revenue will cancel out the difference on the field, and in reality, high school infielders and outfielders do generate equal return to college infielders and outfielders.

While this could be an explanation, if this was the case, it would be likely that the relationship between high school and college returns would be consistent amongst all the positions, not just the two mentioned. For that reason, it is also entirely possible that teams simply draft high school and college infielders and outfielders inefficiently. It is easy for teams to fall in love with a young prospect and believe that he will be the next Mike Trout, however, it is reasonable to assume that teams have a lot more information available with players who played college baseball than those coming out of high school. As mentioned before, offensive production can be hard to project, and this can become increasingly difficult when trying to measure a high school player's skillset against local high school and summer league showcase pitching. A lot of high school hitters can appear as a big fish even if they are in a small pond, and that can increase teams' viewpoints on this player. A college player usually has fully developed physically, and has played years against higher level talent, thus, more information is available as to how good that player actually is. With more available information, teams can select college players more efficiently than they can with high school players. For that reason, high school players provide more risk, even if they provide more reward for hitting the jackpot on a prospect. This risk should be reflected in the spot high school players get drafted as well as how much money they receive in compensation, however, it appears that this is not the case. With teams willing to pay equal or greater signing bonuses to high school players, they make it clear that they are pursuing the extreme to the far right of the distribution without necessarily understanding the probability that they are making a bad investment.

Additionally, given the rules of the MLB Draft, teams may be willing to assume risk on a high school player knowing that they will not get a chance to select that player for three years if they decide to attend a college program. With this being the case, teams know they may never have another chance to select this particular player, and may be willing to assume a loss with the potential of large upside that they may not be able to obtain at any other future point.

It would make sense also, that infielders and outfielders are the only two positions where this difference is seen. A lot of high school players are selected at these positions to be difference making offensive threats. As discussed before, it is very hard to project how a player will develop as a hitter, and the younger that player is, the harder the projection becomes for a MLB career. Thus, the information for these positions can be much hazier for high school players than college players. Catchers, on the other hand, are known as a defensive position. They are often drafted for their arms, blocking ability, and ability to understand and control the pitching game. Offense at this position is usually just a bonus. Pitchers, as stated before, are selected for the pitches they throw, and their ability to harness their pitches to get batters out. Both of these skillsets are a little easier to project long-term, and thus, make the knowledge gap of high school and college prospects smaller, and allow teams to compensate them accordingly, as demonstrated by this study.

Overall, this paper confirms that there is not a lot of supporting evidence to conclude that the MLB Rule IV Draft operates inefficiently, by position, but there is evidence to believe college players outperform high school players on average at the infield and outfield positions. This paper suggests that teams should evaluate their organization, and determine where they can find the best fit to plug holes. If an organization is in a fortunate spot where they have a sound and deep farm system, their best bet is to draft the best player on their board. The team should not be concerned with what position they are selecting, however, if they are choosing an infielder or an outfielder, they should give preference to the college player unless they have strong inclination to believe the high school player will be one of the extremes found on the right side of the distribution. This conclusion also provides evidence that teams should allocate more resources to scouting the best college players than high school. Finally, for teams to have successful drafts, they must evaluate the board, depth of talent, and team needs. While this paper shows that there is not a statistical difference amongst positions on average, years like 2005 where the draft is hitter-heavy, teams may be best suited picking one of them early, and waiting on pitching, as it is likely offense will be going first if the talent is skewed in that direction. If teams evaluate the flow of the draft, they can potentially optimize their return in that given year.

CHAPTER NINE

Limitations and Future Research

While it is my belief that this paper does provide some significant insight into the MLB Rule IV Draft, there are some limitations in how much the results can tell us. First, the MRP model built in this paper is formulated on MLB team revenue estimates, which are almost certainly not totally correct. MLB teams are not required to release revenue information, and while Forbes estimates are the best bet for possible numbers, there is certainly some residual effect between actual revenues and Forbes estimates. For that reason, the MRP model, while theoretically correct, may be produce incorrect results depending on how far off the revenue estimates are from reality. Additionally, there are some players included in this study who have not yet reached their potential, and are still developing in the minors. Some of these players may wind up having very successful MLB careers, and that could alter the results found in this paper. Finally, because this study limits itself to pre-arbitration, for reasons listed earlier, it excludes any return from players who develop later into their six-year pre-free agency period, and it is possible that some positions do not realize true success until after they are into arbitration. It is also entirely possible that none of these factors would change the results any, but they are worth considering when evaluating what the analysis has found.

In the future, there are some opportunities to expand upon the research in this paper. Some interesting tests I think would be worth investigating would be the difference between left-handed and right-handed pitchers. Also, it would be interesting to note the success rates of players based off of their geographic regions, as some research has attempted to do in the past. Finally, an analysis on MLB Draft picks as opposed to international free agents would be very interesting to study, as the means and cost of acquisition of these two different groups of players are very different. I believe that this study has cast light on where the MLB Draft operates efficiently and where it does not, and by combining this paper with future research, much more can be learned about the MLB Rule IV Draft and baseball prospects as a whole.

APPENDICES

APPENDIX A

Figure A-1

Run Differential and Win Percentage

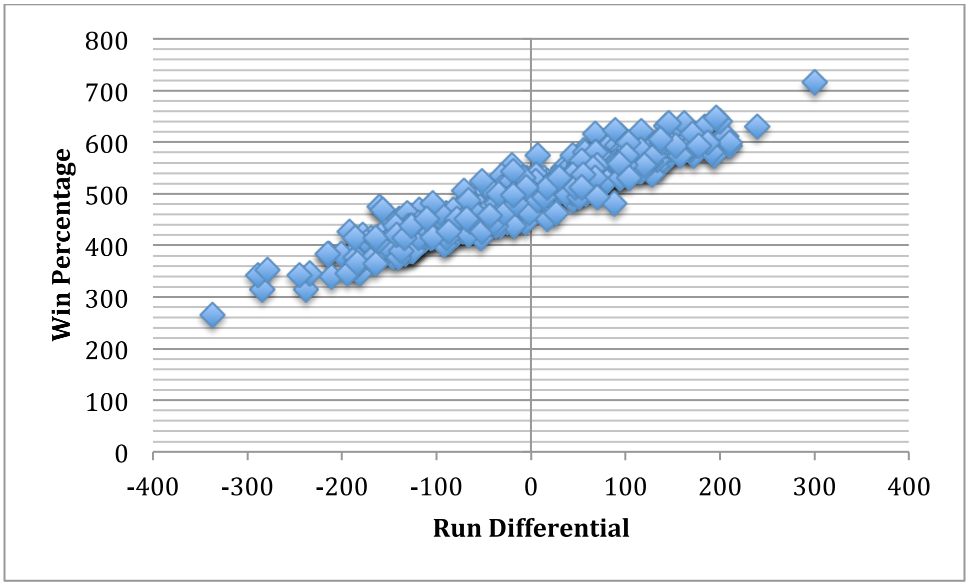
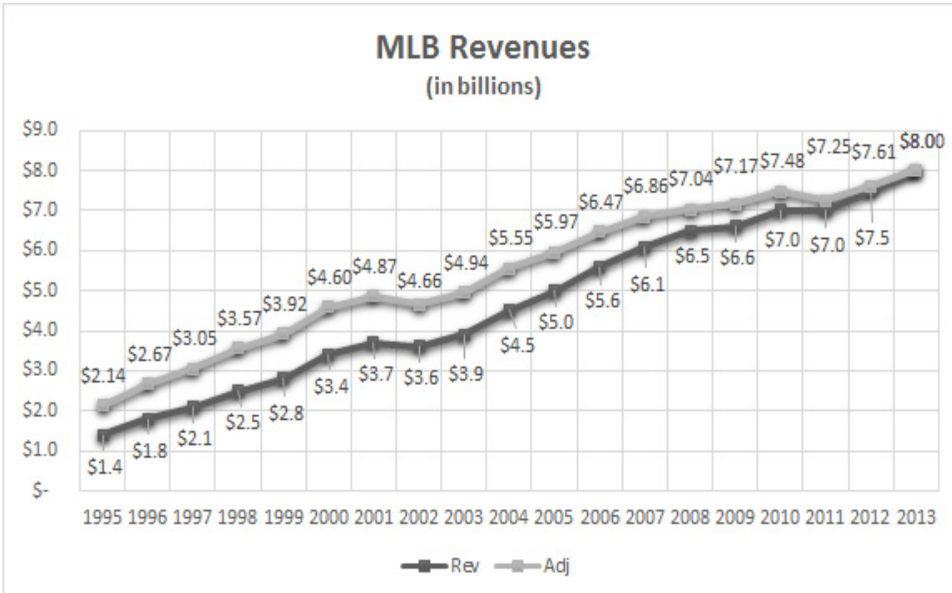


Figure A-2

<u>MLB Revenue By Year</u>



Source: Forbes

Figure A-3

Total Revenue and Fan Cost Index

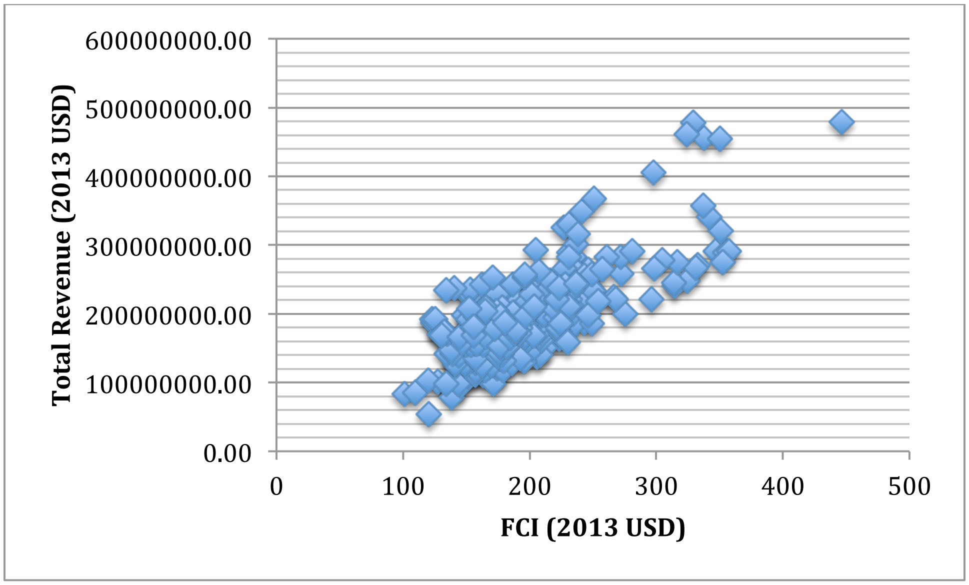


Figure A-4

Total Revenue and Run Differential

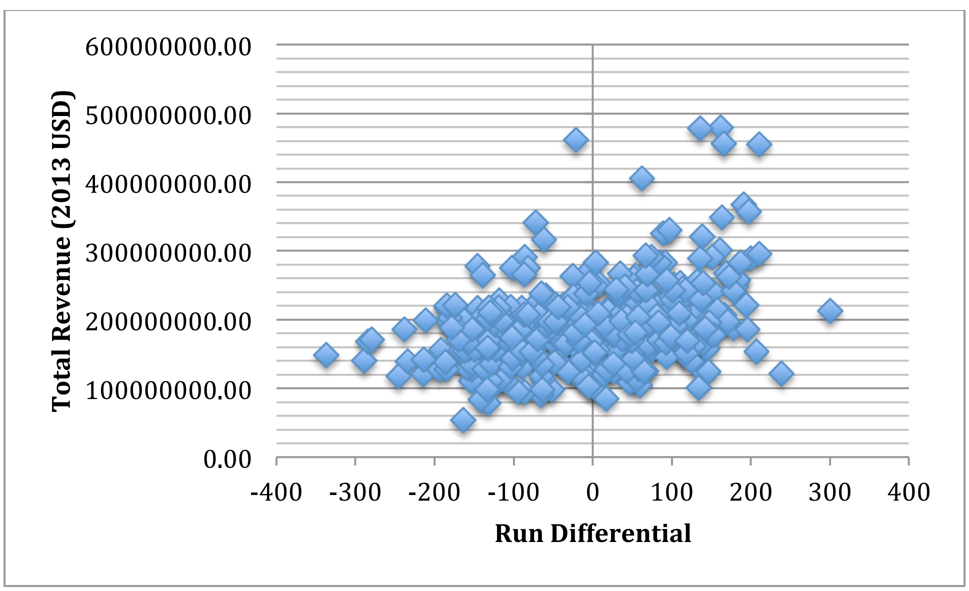
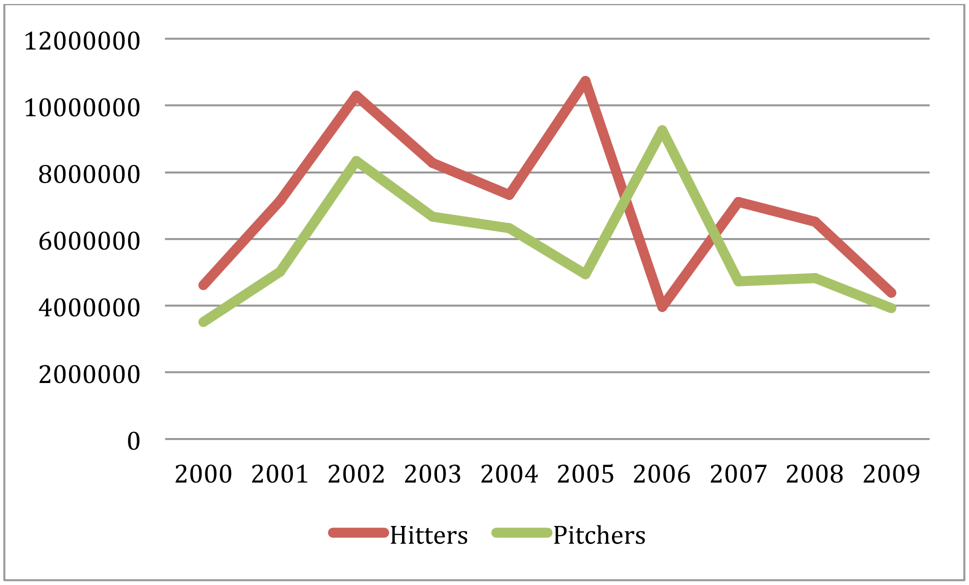


Figure A-5

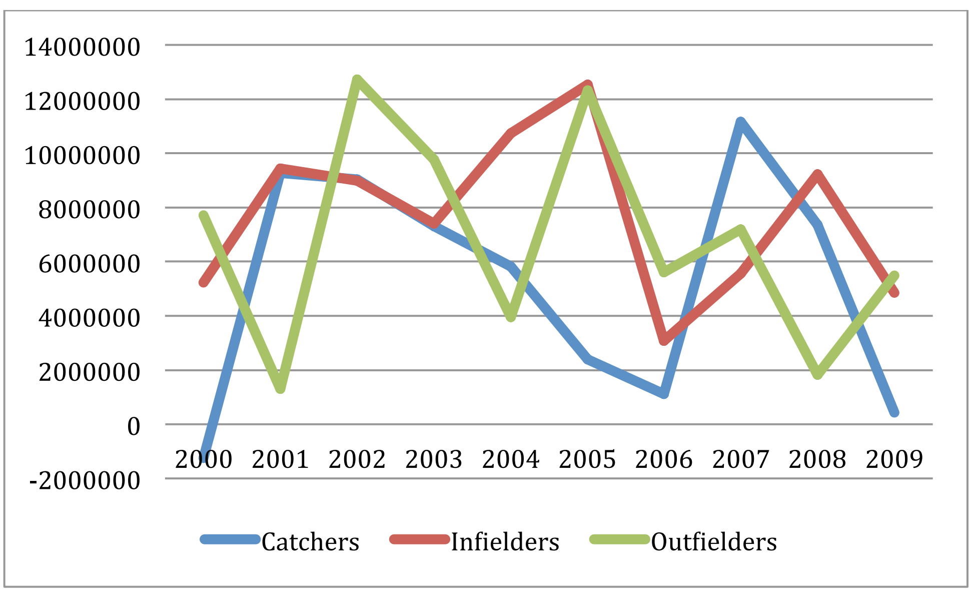
Position Player and Pitcher Mean Return by Year



Returns measured in 2013 USD

Figure A-6

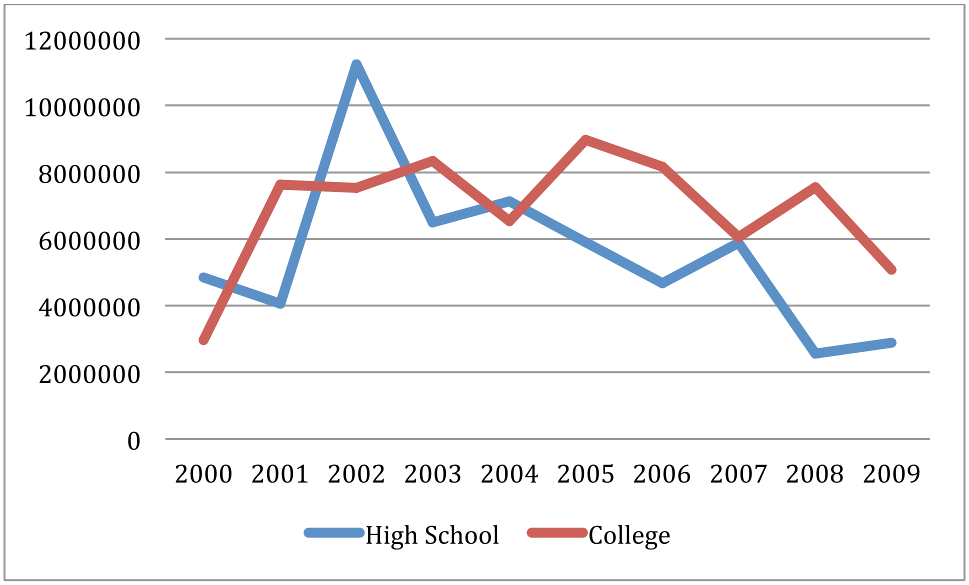
Position Player Mean Return by Position and Year



Returns measured in 2013 USD

Figure A-7

<u>High School and College Mean Return by Year</u>



Returns measured in 2013 USD

Figure A-8

Return on Investment Distribution

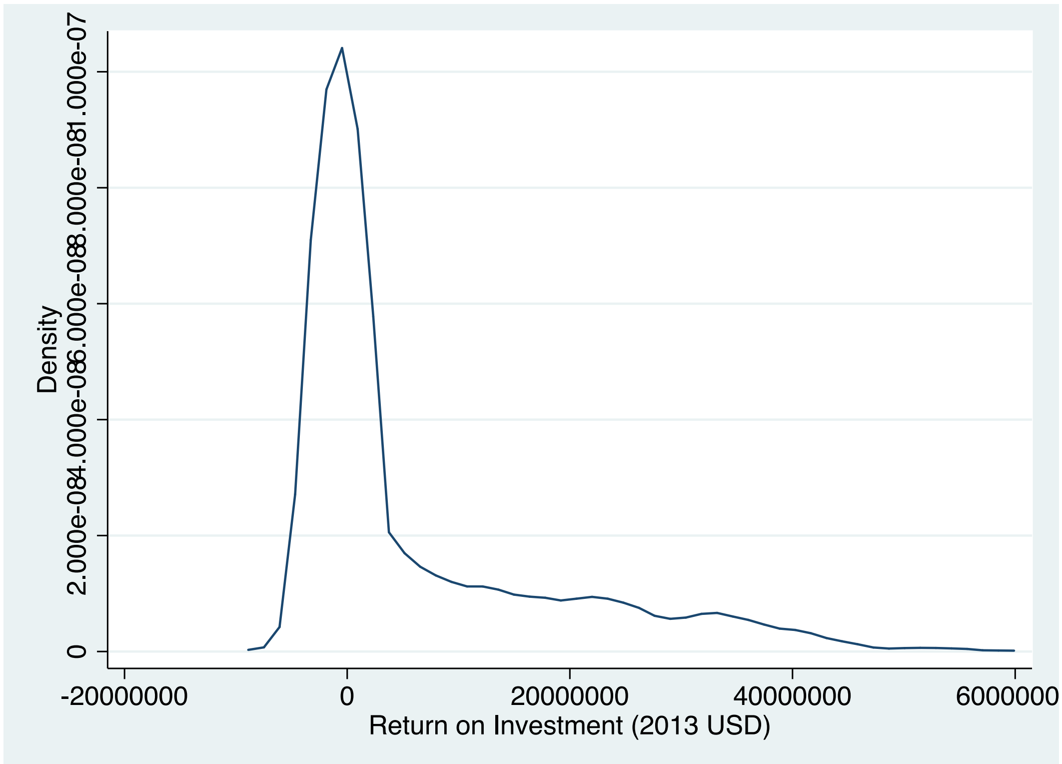


Figure A-9

<u>Position Player and Pitcher Return Distribution</u>

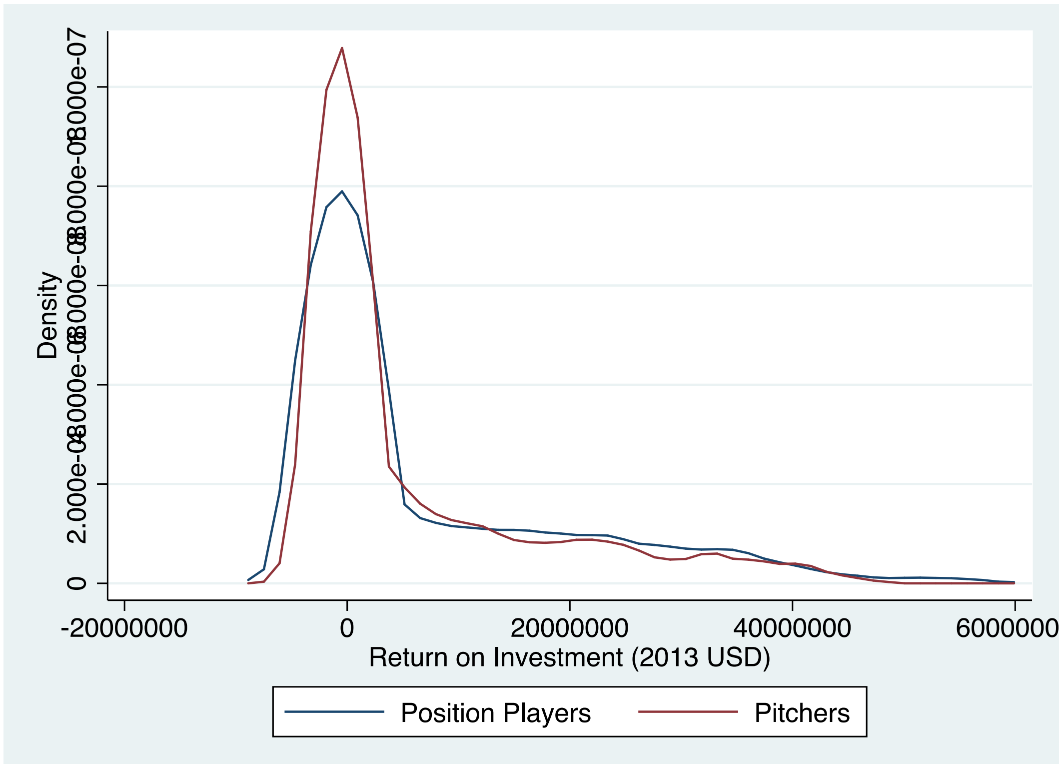


Figure A-10

First Round Position Player and Pitcher Return Distribution

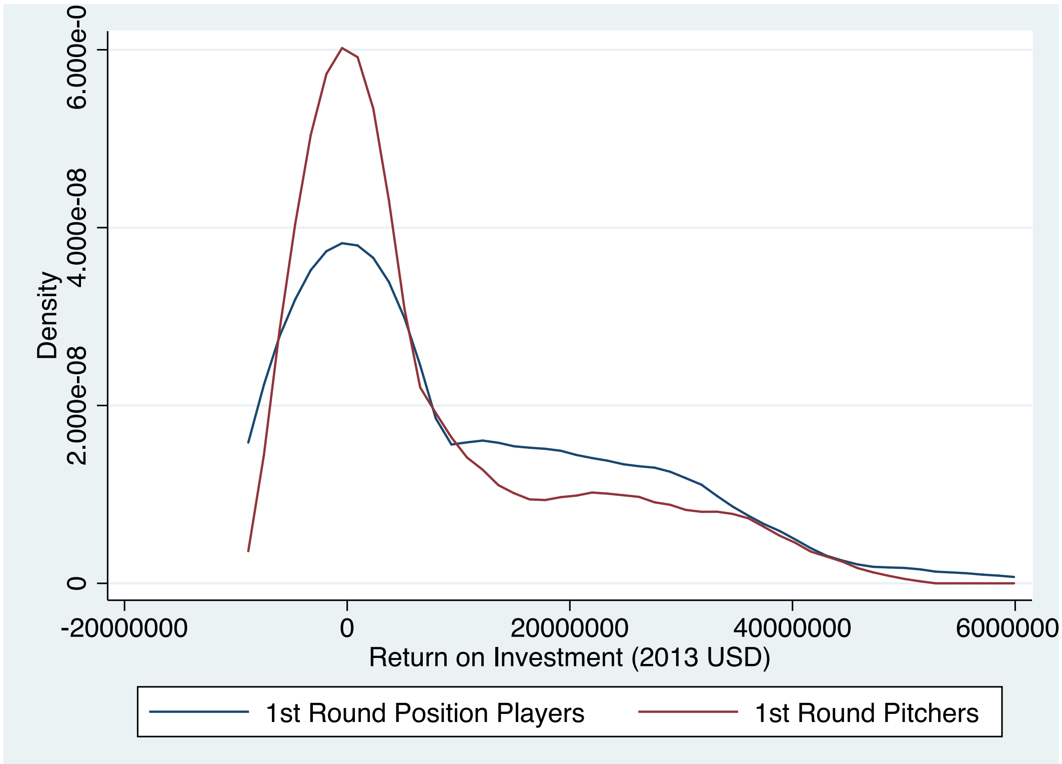


Figure A-11

Other Round Position Player and Pitcher Return Distribution

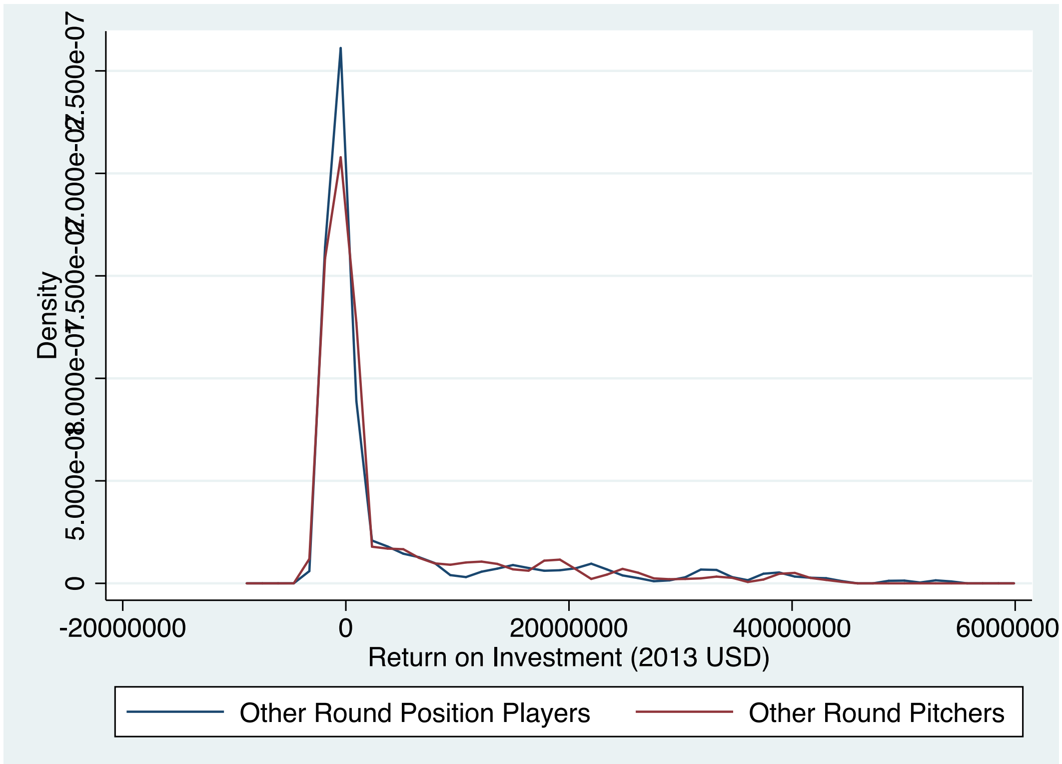


Figure A-12

Offensive Return Distribution by Position

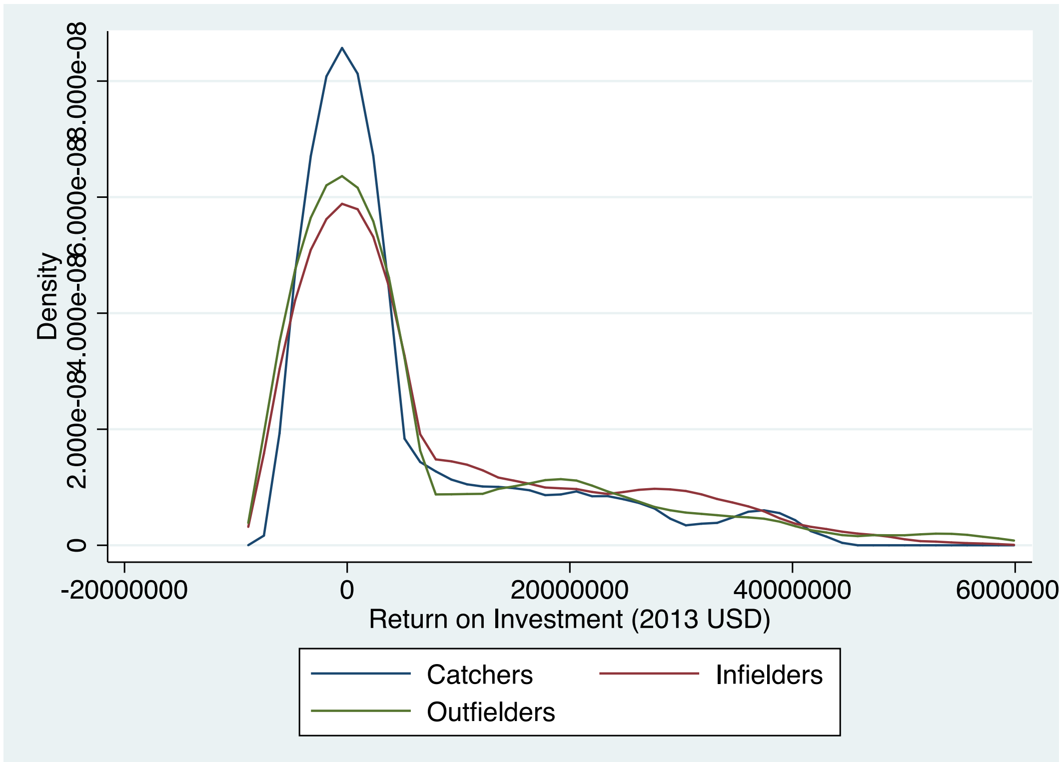


Figure A-13

<u>High School and College Player Return Distribution</u>

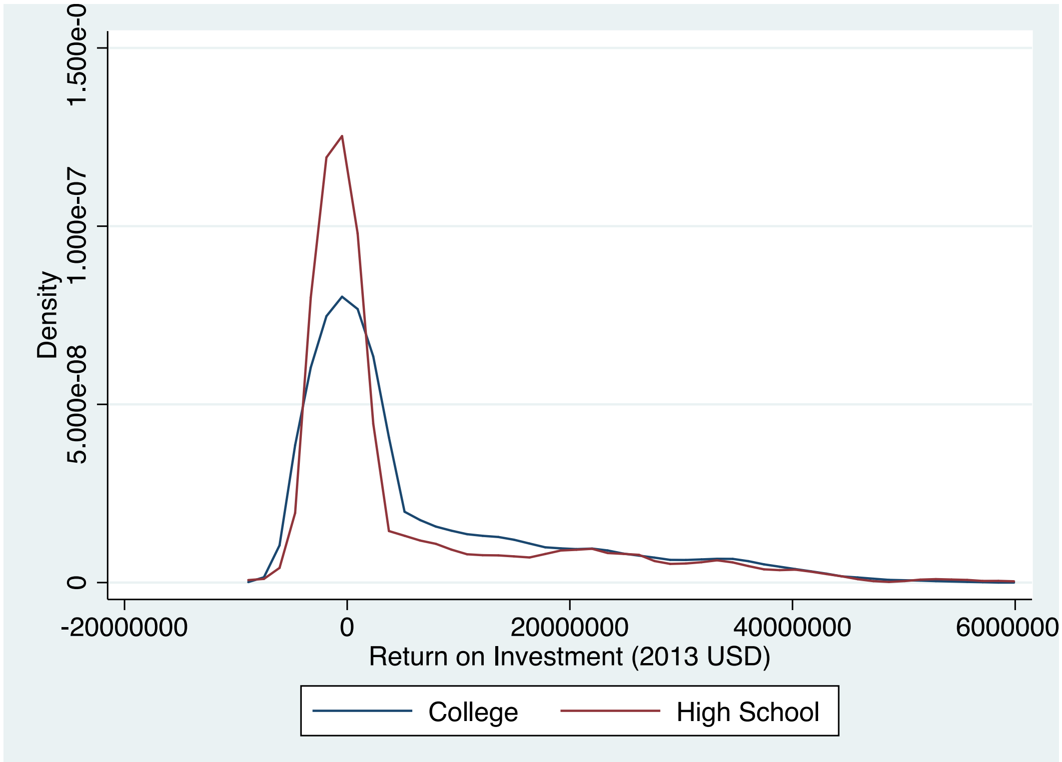


Figure A-14

First Round High School and College Player Return Distribution

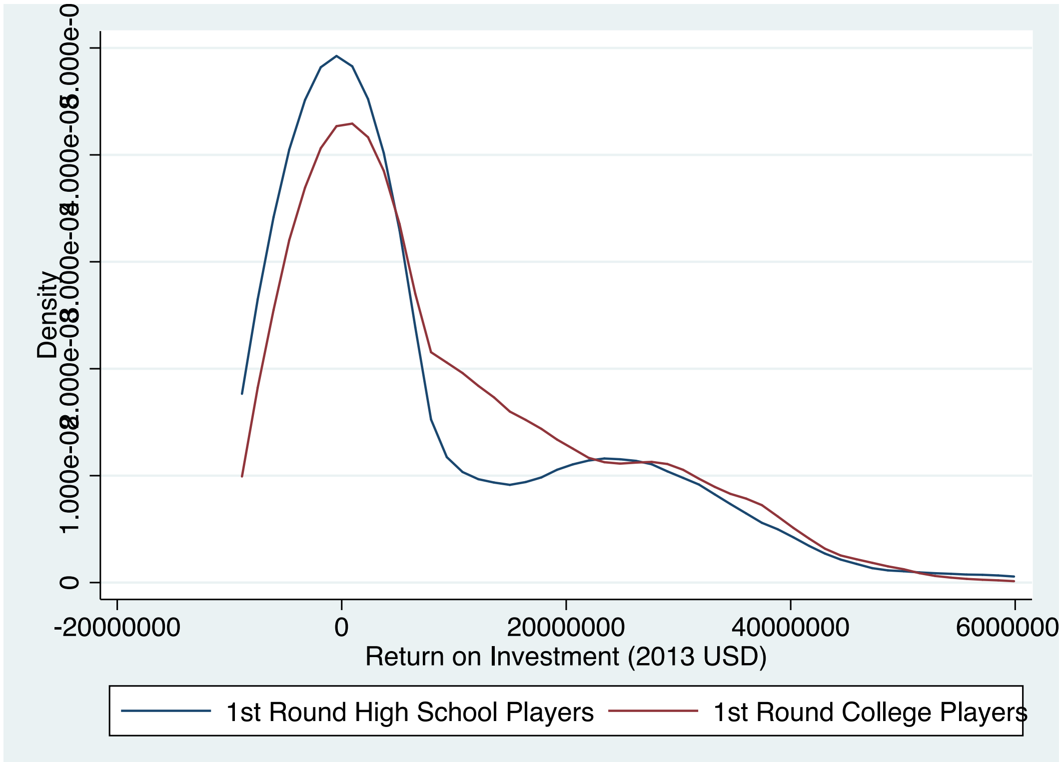


Figure A-15

Other Round High School and College Player Return Distribution

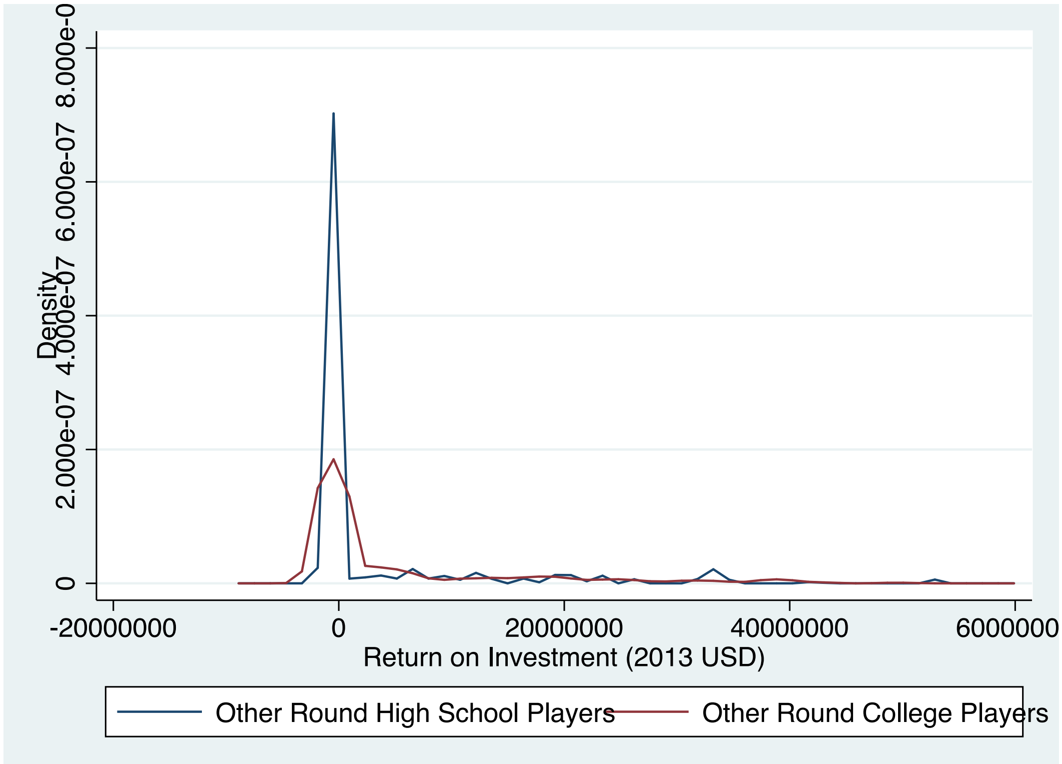


Figure A-16

High School and College Catcher Return Distribution

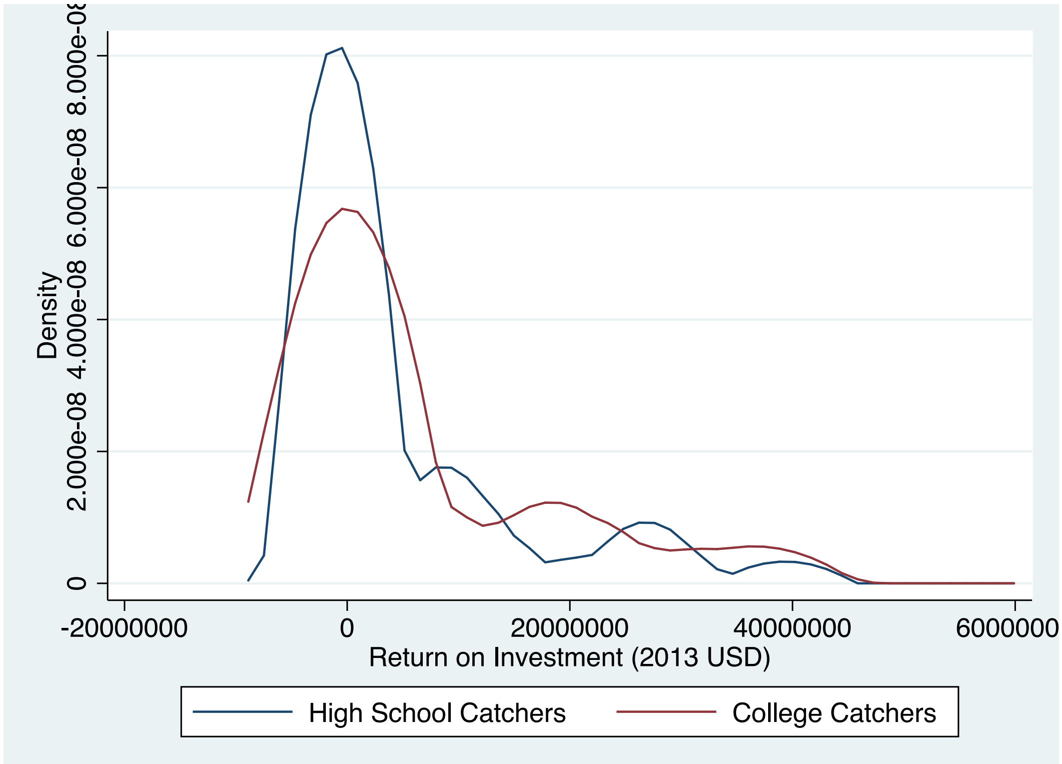


Figure A-17

High School and College Infielder Return Distribution

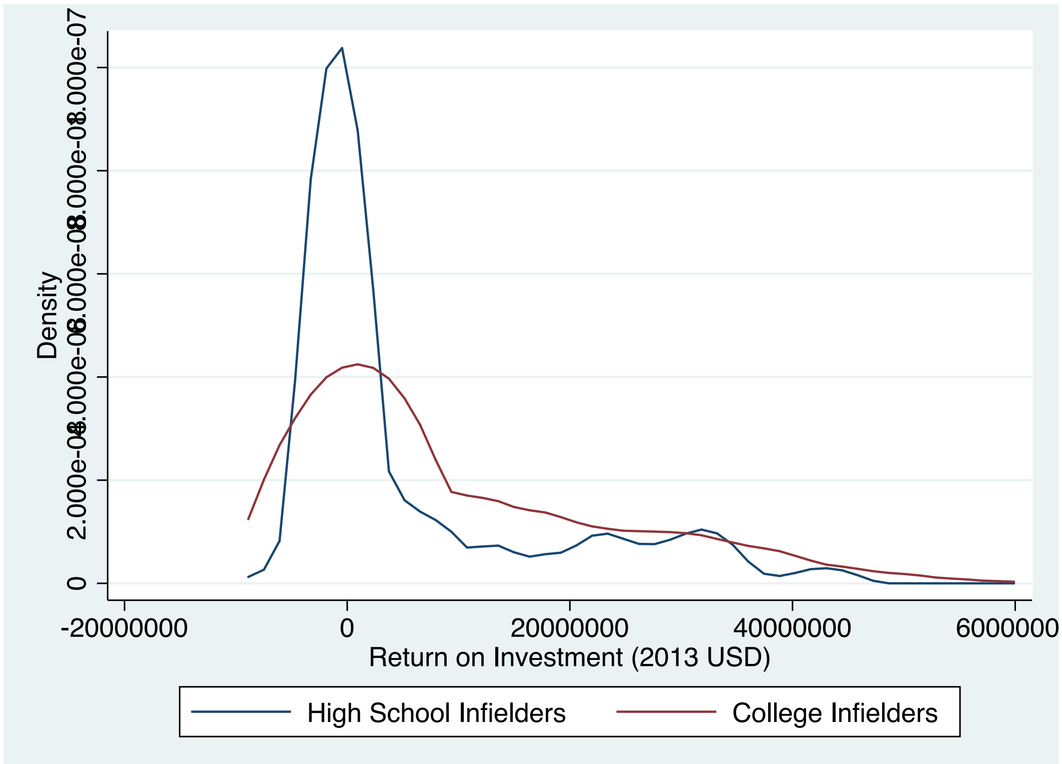


Figure A-18

<u>High School and College Outfielder Return Distribution</u>

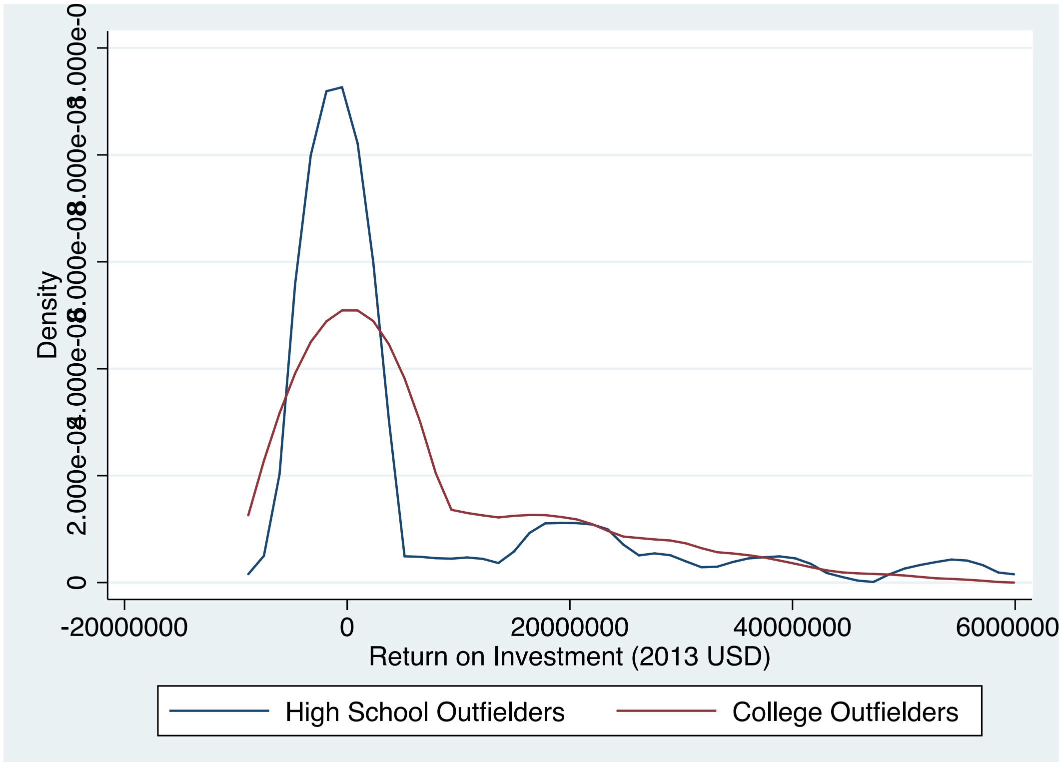
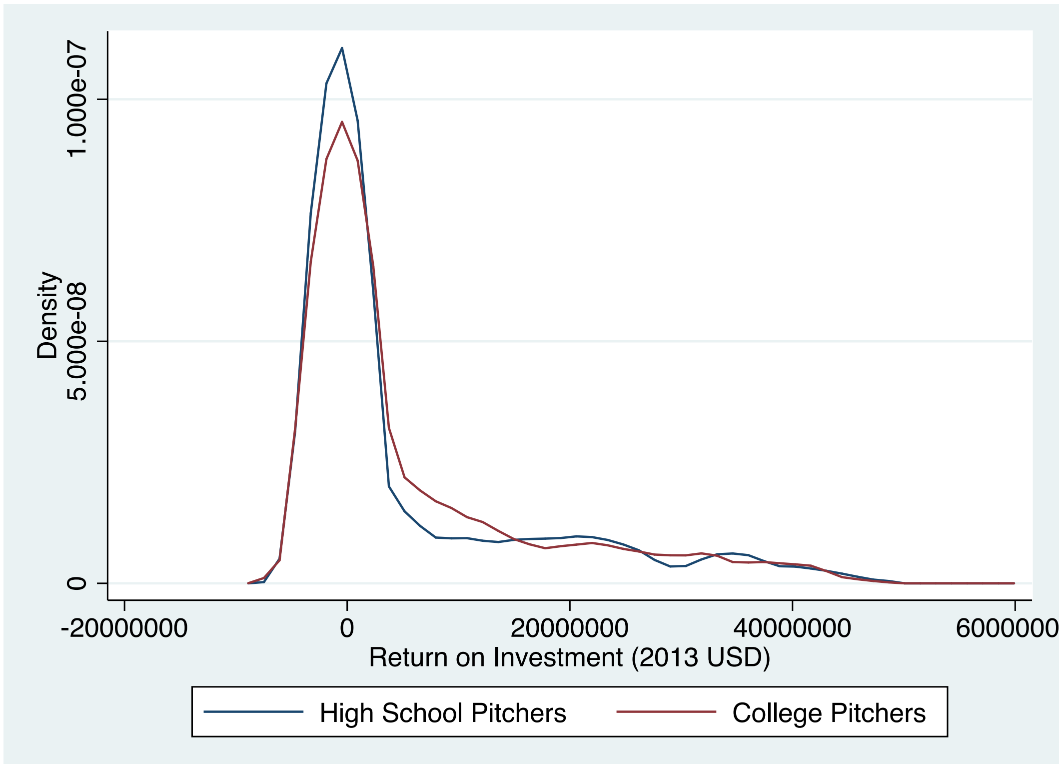


Figure A-19

High School and College Pitcher Return Distribution



APPENDIX B

Table B-1

Model Variable Description

Variable	Description	Measure
totrevmil	Total Team Revenue	Millions of 2013 USD
rd	Team Runs Scored- Team Runs Allowed	Runs
pop	Metropolitan Area Population	People
honeymoon	Team Plays In a New Ballpark	1 if 8 years or younger, 0 if not
fci	Fan Cost Index	2013 USD
year	Revenue Estimate for Year i	1 if from Year i, 0 if not

Table B-2

Summary Statistics

Variable	Obs	Mean	Std. Dev	Min	Max
totrevmil	420	194.0138	57.4325	-337	300
rd	420	0.2238	107.8387	1	113,569.0000
rdsquare	420	11601.54	14,302.18	-33,800,000	27,000,000
rdcube	420	-258,939.1	407,643	1,502,305	19,900,000
pop	420	5,864,319	4,542,058	1,502,305	19,900,000
honeymoon	420	0.319	0.4667	0	1
fci	420	197.863	47.1929	101.04	446.16
year2000	420	0.0714	0.2578	0	1
year2001	420	0.0714	0.2578	0	1
year2002	420	0.0714	0.2578	0	1
year2003	420	0.0714	0.2578	0	1
year2004	420	0.0714	0.2578	0	1
year2005	420	0.0714	0.2578	0	1
year2006	420	0.0714	0.2578	0	1
year2007	420	0.0714	0.2578	0	1
year2008	420	0.0714	0.2578	0	1
year2009	420	0.0714	0.2578	0	1
year2010	420	0.0714	0.2578	0	1
year2011	420	0.0714	0.2578	0	1
year2012	420	0.0714	0.2578	0	1
year2013	420	0.0714	0.2578	0	1

Table B-3

MRP Model Results

Variable	Model 1	Model 2	Model 3	Model 4
Run Difference	0.0918 [0.0341]***	0.0855 [0.0198]***	0.1301 [0.0272]***	0.09376 [0.0143]***
Run Difference^2	0.0005 [0.0002]**	0.0003 [0.0001]***	0.0004 [0.0002]***	0.0003 [0.0001]***
Run Difference^3	1.36E-06 [1.15E-06]	4.29E-07 [5.98E-07]		
MSA Population	6.94E-06 [09.74E-07]***	4.35E-06 [6.66e-07]***	6.92E-06 [9.75E-07]***	4.33E-06 [6.61e-07]***
Honeymoon	19.6288 [6.14674]***	13.7019 [3.9212]***	19.7280 [6.1965]***	13.6779 [3.9101]***
FCI		0.5561 [0.0411]***		0.5592 [0.0404]***
Year 2000		-74.0140 [7.5077]***		-73.9911 [-7.4829]***
Year 2001		-65.6641 [8.0461]***		-65.2966 [7.9831]***
Year 2002		-67.7572 [7.7949]***		-67.7493 [7.7989]***
Year 2003		-60.1615 [7.5843]***		-60.5336 [7.6411]***
Year 2004		-50.6138 [8.0337]***		-50.8140 [8.0300]***
Year 2005		-34.1317 [7.8918]***		-34.2467 [7.9141]***
Year 2006		-31.6930 [7.9536]***		-31.8829 [7.9679]***
Year 2007		-23.5719 [7.8136]***		-23.5092 [7.8057]***
Year 2008		-24.5042 [8.5359]***		-24.5925 [8.5382]***
Year 2009		-23.9492 [7.8894]***		-24.0367 [7.8968]***

Year 2010		-16.5202 [8.5798]**		-16.8388 [8.5703]**
Year 2011		-13.5385 [8.3834]		-13.4693 [8.3831]
Year 2012		-5.9114 [6.8350]		-6.1311 [6.8113]
Constant	141.4984 [7.4146]***	85.3087 [9.7715]***	142.1545 [7.4013]***	85.0972 [9.7655]***
Observations	420	420	420	420
R^2	0.41	0.82	0.42	0.82

***,**,*: Significant at 1%, 5%, and 10%, respectively

Table B-4

Fielding Positional Adjustments

Position	Adjustment
Catcher	42
First Base	13
Second Base	32
Shortstop	36
Third Base	25
Left Field	19
Center Field	29
Right Field	20
Designated Hitter	-7
Pitcher	2

As reported in the *Fielding Bible* by John Dewan

Table B-5

MRP Model v One-Year Free Agent Contracts Using Replacement

Player	2015 Salary	MRP Estimate	Absolute Value Difference
Alex Rios	10,824,407.78	8,149,137.55	2,675,270.23
Torii Hunter	10,332,389.24	9,908,996.38	423,392.86
Colby Rasmus	7,872,296.57	9,884,799.36	2,012,502.79
Asdrubal Cabrera	7,380,278.03	9,474,259.70	2,093,981.67
Stephen Drew	4,920,185.35	6,332,068.90	1,411,883.55
Hitter Average Difference	1,723,406.22	Hitter St. Dev Difference	853,526.39
Brett Anderson	9,840,370.71	4,796,053.51	5,044,317.20
Justin Masterson	9,348,352.17	10,934,389.75	1,586,037.58
AJ Burnett	8,364,315.10	9,875,143.75	1,510,828.65
Kyle Kendrick	5,412,203.89	5,685,630.32	273,426.43
Chris Capuano	4,920,185.35	4,586,954.00	333,231.35
Pitcher Average Difference	1,749,568.24	Pitcher St. Dev Difference	1,944,486.72
Total Average Difference	1,736,487.23	Total St. Dev Difference	1,415,778.34

All monetary values are adjusted to 2013 USD

Table B-6

MRP Model v One-Year Free Agent Contracts Raw

Player	2015 Salary	MRP Estimate	Absolute Value Difference
Alex Rios	10,824,407.78	8,793,755.05	2,030,652.73
Torii Hunter	10,332,389.24	10,803,027.59	470,638.35
Colby Rasmus	7,872,296.57	11,095,533.87	3,223,237.30
Asdrubal Cabrera	7,380,278.03	10,094,634.10	2,714,356.07
Stephen Drew	4,920,185.35	6,551,922.06	1,631,736.71
Hitter Average Difference	2,014,124.23	Hitter St. Dev Difference	1,058,432.94
Brett Anderson	9,840,370.71	5,099,163.40	4,741,207.31
Justin Masterson	9,348,352.17	12,336,229.77	2,987,877.60
AJ Burnett	8,364,315.10	10,998,779.07	2,634,463.97
Kyle Kendrick	5,412,203.89	6,131,849.96	719,646.07
Chris Capuano	4,920,185.35	4,883,902.82	36,282.53
Pitcher Average Difference	2,223,895.50	Pitcher St. Dev Difference	1,879,994.06
Total Average Difference	2,119,009.86	Total St. Dev Difference	1,442,553.36

All monetary values are adjusted to 2013 USD

Table B-7

<u>MRP Model v One-Year Free Agent Contracts Using Average</u>

Player	2015 Salary	MRP Estimate	Absolute Value Difference
Alex Rios	10,824,407.78	8,901,139.45	1,923,268.33
Torii Hunter	10,332,389.24	11,132,950.86	800,561.62
Colby Rasmus	7,872,296.57	10,269,934.07	2,397,637.50
Asdrubal Cabrera	7,380,278.03	10,497,587.91	3,117,309.88
Stephen Drew	4,920,185.35	6,961,192.90	2,041,007.55
Hitter Average Difference	2,055,956.97	Hitter St. Dev Difference	842,179.14
Brett Anderson	9,840,370.71	5,169,760.33	4,670,610.38
Justin Masterson	9,348,352.17	12,260,674.78	2,912,322.61
AJ Burnett	8,364,315.10	11,235,226.09	2,870,910.99
Kyle Kendrick	5,412,203.89	6,011,177.85	598,973.96
Chris Capuano	4,920,185.35	4,825,138.01	95,047.34
Pitcher Average Difference	2,229,573.05	Pitcher St. Dev Difference	1,874,245.05
Total Average Difference	2,142,765.01	Total St. Dev Difference	1,372,896.25

All monetary values are adjusted to 2013 USD

Table B-8

Highest Annual MRPs

Draft Year	Pick #	Season	Player	Pos	HS/C	Rep TRC	RAR	$ Worth
2000	15	2006	Chase Utley	IF	C	89.62	94.91	23,077,421.58
2001	2	2003	Mark Prior	P	C	68.28	112.64	22,819,284.91
2008	96	2012	Craig Kimbrel	P	C	17.53	144.41	22,114,544.67
2009	25	2012	Mike Trout	OF	HS	69.15	106.91	22,004,252.13
2005	7	2007	Troy Tulowitzki	IF	C	85.29	83.46	20,701,762.64
2006	7	2011	Clayton Kershaw	P	HS	67.01	99.70	20,560,463.97
2002	57	2009	Jon Lester	P	HS	69.25	96.42	20,357,469.72
2004	65	2008	Dustin Pedroia	IF	C	80.99	84.11	20,165,486.14
2006	10	2009	Tim Lincecum	P	C	69.95	93.33	19,978,691.54
2002	80	2007	Curtis Granderson	OF	C	80.26	83.57	19,977,496.69
2000	75	2006	Grady Sizemore	OF	HS	87.52	75.27	19,847,203.41
2009	25	2013	Mike Trout	OF	HS	77.60	84.70	19,760,443.34
2003	7	2008	Nick Markakis	OF	C	81.68	80.37	19,716,905.55
2000	15	2005	Chase Utley	IF	C	73.09	88.43	19,674,654.65
2001	5	2005	Mark Teixeira	IF	C	75.91	85.55	19,644,667.70
2006	10	2008	Tim Lincecum	P	C	72.47	86.57	19,307,835.42
2001	72	2006	Dan Haren	P	C	79.21	79.08	19,170,257.50
2005	7	2009	Troy Tulowitzki	IF	C	78.38	76.57	18,688,106.52
2006	3	2010	Evan Longoria	IF	C	66.97	86.95	18,598,999.10
2006	3	2009	Evan Longoria	IF	C	74.67	77.14	18,237,879.69
2007	1	2011	David Price	P	C	70.60	80.67	18,175,512.76
2000	75	2007	Grady Sizemore	OF	HS	81.35	69.23	18,083,275.73
2002	15	2007	Scott Kazmir	P	HS	72.15	78.53	18,081,997.28
2002	44	2010	Joey Votto	IF	HS	57.02	91.72	17,980,688.06
2005	11	2012	Andrew McCutchen	OF	HS	75.80	73.53	17,884,671.69
2009	25	2014	Mike Trout	OF	HS	71.40	76.19	17,640,403.46
2002	17	2008	Cole Hamels	P	HS	73.13	73.19	17,458,393.40
2007	48	2013	Josh Donaldson	C	C	66.88	78.71	17,375,210.80
2000	75	2005	Grady Sizemore	OF	HS	74.01	70.84	17,251,444.32
2002	24	2007	Joe Blanton	P	C	80.98	63.24	17,208,245.97
2005	5	2009	Ryan Braun	IF	C	77.13	67.24	17,197,820.89
2005	4	2007	Ryan Zimmerman	IF	C	83.34	57.37	16,770,148.07
2001	72	2005	Dan Haren	P	C	73.68	67.42	16,729,585.03
2005	5	2010	Ryan Braun	IF	C	73.61	67.42	16,720,928.93
2001	1	2006	Joe Mauer	C	HS	72.81	67.97	16,682,888.58
2003	13	2007	Aaron Hill	IF	C	78.77	61.38	16,636,343.66

2007	14	2012	Jason Heyward	OF	HS	59.85	79.75	16,575,326.07
2002	55	2006	David Bush	P	C	71.28	67.81	16,445,543.03
2008	5	2012	Buster Posey	C	C	71.03	66.69	16,257,118.12
2007	48	2014	Josh Donaldson	C	C	63.34	72.12	15,952,089.74
2005	5	2008	Ryan Braun	IF	C	64.18	70.15	15,790,819.55
2003	7	2007	Nick Markakis	OF	C	82.69	50.01	15,721,427.12
2004	2	2007	Justin Verlander	P	C	70.81	62.80	15,695,459.96
2002	64	2008	Brian McCann	C	C	78.11	54.82	15,674,727.60
2005	11	2011	Andrew McCutchen	OF	HS	78.25	54.43	15,643,920.34
2007	14	2010	Jason Heyward	OF	HS	56.86	75.45	15,559,624.17
2002	9	2007	Jeff Francis	P	C	72.76	59.39	15,512,479.29
2001	38	2005	David Wright	IF	HS	63.20	69.08	15,508,699.22
2001	38	2006	David Wright	IF	HS	67.01	64.86	15,447,783.47
2003	28	2010	Daric Barton	C	HS	57.20	74.30	15,439,555.43

Table B-9

Top 50 MLB Salaries for 2015

Player	Position	2015 Salary
Clayton Kershaw	P	32,571,428.00
Justin Verlander	P	28,000,000.00
Josh Hamilton	OF	25,400,000.00
Cliff Lee	P	25,000,000.00
Ryan Howard	IF	25,000,000.00
Zack Greinke	P	25,000,000.00
Felix Hernandez	P	24,857,142.00
Albert Pujols	IF	24,000,000.00
Robinson Cano	IF	24,000,000.00
Prince Fielder	IF	24,000,000.00
Cole Hamels	P	23,500,000.00
Mark Teixeira	IF	23,125,000.00
C.C. Sbathia	P	23,000,000.00
Joe Mauer	IF	23,000,000.00
Jose Reyes	IF	22,000,000.00
Masahiro Tanaka	P	22,000,000.00
Alex Rodriguez	IF	22,000,000.00
Miguel Cabrera	IF	22,000,000.00
Adrian Gonzalez	IF	21,857,142.00
Jayson Werth	OF	21,571,428.00
Carl Crawford	OF	21,357,142.00
Matt Kemp	OF	21,250,000.00

Jacoby Ellsbury	OF	21,142,857.00
Matt Cain	P	20,833,333.00
Jon Lester	P	20,000,000.00
David Wright	IF	20,000,000.00
Troy Tulowitzki	IF	20,000,000.00
Mark Buehrle	P	20,000,000.00
David Price	P	19,750,000.00
Hanley Ramirez	OF	19,750,000.00
Adam Wainwright	P	19,500,000.00
C.J. Wilson	P	18,500,000.00
Hunter Pence	OF	18,500,000.00
Jered Weaver	P	18,200,000.00
Andre Ethier	OF	18,000,000.00
Time Lincecum	P	18,000,000.00
Pablo Sandoval	IF	17,600,000.00
Buster Posey	C	17,277,777.00
Max Scherzer	P	17,142,857.00
Brian McCann	C	17,000,000.00
Matt Holliday	OF	17,000,000.00
Anibal Sanchez	P	16,800,000.00
Jordan Zimmermann	P	16,500,000.00
Carlos Gonzalez	OF	16,428,571.00
Adrian Beltre	IF	16,000,000.00
Ian Kinsler	IF	16,000,000.00
David Ortiz	IF	16,000,000.00
Mike Napoli	IF	16,000,000.00
Curtis Granderson	OF	16,000,000.00
John Danks	P	15,750,000.00

As reported by Spotrac, all values are 2015 USD

Table B-10

Highest Return on Investment

Draft Year	Round Picked	Overall	Player	Position	Return (2013 USD)
2009	1	25	Mike Trout	OF	57,956,578.34
2000	3	75	Grady Sizemore	OF	53,097,381.04
2005	1	5	Ryan Braun	IF	52,870,562.36
2005	1	11	Andrew McCutchen	OF	52,693,630.54
2002	3	80	Curtis Granderson	OF	49,455,925.26
2000	1	15	Chase Utley	IF	47,252,485.49
2006	1	7	Clayton Kershaw	P	45,662,531.76
2006	1	10	Tim Lincecum	P	44,930,467.91
2006	1	3	Evan Longoria	IF	43,524,400.29
2005	2	75	Yunel Escobar	IF	43,495,801.83
2005	1	7	Troy Tulowitzki	IF	42,985,996.86
2003	2	58	Scott Baker	P	42,935,771.99
2003	1	7	Nick Markakis	OF	42,891,420.80
2007	1	10	Madison Bumgarner	P	42,818,868.84
2002	2	44	Joey Votto	IF	42,056,990.97
2002	1	15	Scott Kazmir	P	41,243,987.48
2002	2	55	David Bush	P	41,152,597.31
2007	2	76	Giancarlo Stanton	OF	40,877,115.24
2008	3	96	Craig Kimbrel	P	39,659,322.08
2004	1	12	Jered Weaver	P	39,489,863.05
2002	2	57	Jon Lester	P	39,405,667.81
2002	2	64	Brian McCann	C	39,355,783.09
2006	2	71	Justin Masterson	P	39,186,372.18
2001	2	72	Dan Haren	P	39,125,436.14
2001	1	2	Mark Prior	P	39,044,672.00
2002	1	17	Cole Hamels	P	38,990,213.46
2002	1	24	Joe Blanton	P	38,931,137.12
2001	3	78	Ryan Theriot	IF	38,852,255.12
2004	2	67	Kurt Suzuki	C	38,116,440.36
2009	3	82	Kyle Seager	IF	37,927,600.13
2002	1	23	Jeff Francoeur	OF	37,828,635.51
2007	1	14	Jason Heyward	OF	37,564,059.57
2003	1	13	Aaron Hill	IF	37,028,146.54
2003	1	24	Chad Billingsley	P	36,861,507.03
2004	2	64	Hunter Pence	OF	36,817,154.33
2005	1	4	Ryan Zimmerman	IF	36,349,813.22
2007	1	48	Josh Donaldson	C	35,744,887.89
2007	1	5	Matt Wieters	C	35,617,764.39

2004	1	2	Justin Verlander	P	35,562,080.02
2001	1	38	David Wright	IF	35,410,507.98
2002	1	25	Matt Cain	P	35,117,780.66
2001	1	26	Jeremy Bonderman	P	34,596,173.07
2006	1	11	Max Scherzer	P	34,573,025.19
2002	1	9	Jeff Francis	P	34,505,378.92
2006	2	66	Trevor Cahill	P	34,390,550.26
2002	1	16	Nick Swisher	OF	34,374,406.62
2009	2	63	Jason Kipnis	OF	34,336,554.17
2006	1	21	Ian Kennedy	P	34,128,497.88
2001	1	5	Mark Teixeira	IF	34,101,737.16
2003	1	9	John Danks	P	34,037,693.81

Return is measured in 2013 USD

Table B-11

<u>Position Players v Pitchers</u>

Group	Position Players	Pitchers
Obs	469	492
25%	-882,737	-892,322
Median	-517,462	-477,155
75%	12,500,000	8,880,495
IQR	13,382,737	9,772,817
Max	58,000,000	45,700,000
Min	-6,947,075	-4,271,194
Mean	7,087,449	5,761,469
D Stat	0.0622	

* Represents significance of D statistic at 5% level

Table B-12

First Round Position Players v Pitchers

Group	Position Players	Pitchers
Obs	204	235
25%	-1,455,200	-1,246,073
Median	5,506,365	1,319,972
75%	20,400,000	13,000,000
IQR	21,855,200	14,246,073
Max	58,000,000	45,700,000
Min	-6,947,075	-4,271,194
Mean	10,400,000	7,680,994
D Stat	0.1331*	

* Represents significance of D statistic at 5% level

Table B-13

First Round Position Players v Pitchers Excluding 2005

Group	Position Players	Pitchers
Obs	182	210
25%	-1,497,896	-1,253,807
Median	2,630,536	1,396,153
75%	19,300,000	13,200,000
IQR	20,797,896	14,453,807
Max	58,000,000	45,700,000
Min	-6,947,075	-4,271,194
Mean	9,449,917	7,890,775
D Stat	0.0985	

* Represents significance of D statistic at 5% level

Table B-14

<u>Other Round Position Players v Pitchers</u>

Group	Position Players	Pitchers
Obs	265	257
25%	-760,374	-771,233
Median	-553,251	-543,250
75%	3,119,521	4,256,810
IQR	3,879,895	5,028,043
Max	53,100,000	42,900,000
Min	-2,791,855	-1,606,428
Mean	4,524,859	4,006,261
D Stat	0.0394	

* Represents significance of D statistic at 5% level

Table B-15

<u>Offensive Position Comparison</u>

Group	Catchers	Infielders	Outfielders
Obs	78	230	161
25%	-939,607	-899,630	-881,482
Median	-572,405	-425,335	-544,712
75%	8,534,248	14,400,000	13,800,000
IQR	9,473,855	15,299,630	14,681,482
Max	39,400,000	52,900,000	58,000,000
Min	-3,004,278	-6,947,075	-6,812,643
Mean	5,377,773	7,771,003	6,939,234
C v IF D Stat	0.1179		
C v OF D Stat	0.0814		
IF v OF D Stat	0.0988		

* Represents significance of D statistic at 5% level

Table B-16

<u>High School v College Players</u>

Group	High School	College
Obs	414	547
25%	-1,019,922	-820,230.90
Median	-601,177.40	-3,188,468
75%	7,407,729	12,300,000
IQR	8,427,651	13,120,231
Max	58,000,000	52,900,000
Min	-6,947,075	-4,271,194
Mean	5,641,068	6,989,496
D Stat	0.1571*	

* Represents significance of D statistic at 5% level

Table B-17

First Round High School v College

Group	High School	College
Obs	193	246
25%	-1,559,146	-1,178,751
Median	-357,217	5,276,402
75%	16,800,000	17,600,000
IQR	18,359,146	18,778,751
Max	58,000,000	52,900,000
Min	-6,947,075	-4,271,194
Mean	7,855,764	9,812,184
D Stat	0.1483*	

* Represents significance of D statistic at 5% level

Table B-18

Other Round High School v College

Group	High School	College
Obs	221	301
25%	-819,124	-716,167
Median	-602,500	-486,318
75%	116,251	4,827,423
IQR	935,375	5,543,590
Max	53,100,000	49,500,000
Min	-2,791,855	-2,661,463
Mean	3,706,966	4,682,582
D Stat	0.1808*	

* Represents significance of D statistic at 5% level

Table B-19

<u>High School v College Catchers</u>

Group	High School	College
Obs	36	42
25%	-1,080,142	-821,337
Median	-606,409	-387,126
75%	7,574,768	12,500,000
IQR	8,654,910	13,321,367
Max	39,400,000	38,100,000
Min	-3,004,278	-1,982,863
Mean	4,536,815	6,098,595
D Stat	0.1667	

* Represents significance of D statistic at 5% level

Table B-20

<u>High School v College Infielders</u>

Group	High School	College
Obs	109	121
25%	-1,101,734	-745,303
Median	-602,500	2,457,344
75%	6,733,935	16,600,000
IQR	7,835,669	17,345,303
Max	43,500,000	52,900,000
Min	-6,947,075	-2,927,754
Mean	5,628,829	9,700,731
D Stat	0.2116*	

* Represents significance of D statistic at 5% level

Table B-21

High School v College Outfielders

Group	High School	College
Obs	76	85
25%	-1,096,771	-729,662
Median	-673,170	-319,232
75%	7,844,363	14,900,000
IQR	8,941,134	15,629,662
Max	58,000,000	49,500,000
Min	-6,812,643	-3,128,590
Mean	6,501,663	7,330,473
D Stat	0.2406*	

* Represents significance of D statistic at 5% level

Table B-22

High School v College Pitchers

Group	High School	College
Obs	193	299
25%	-968,242	-865,345
Median	-587,415	-362,556
75%	7,407,729	9,028,367
IQR	8,375,971	9,893,712
Max	45,700,000	44,900,000
Min	-3,437,153	-4,271,194
Mean	5,515,067	5,920,518
D Stat	0.1198	

* Represents significance of D statistic at 5% level

REFERENCES

Badanhausen, Kurt, Michael Ozanian, and Christina Settimi. "Baseball's Most Valuable Teams." *Forbes*. 1 Mar. 2014. Web.

Bradbury, John Charles (2010). *Hot Stove Economics: Understanding Baseball's Second Season.* New York: Copernicus. (2010).

Bradbury, John Charles (2013). "What's Right with Scully-Estimates of a Player's Marginal Revenue Product." *Kennesaw State University*. 1 Jan. 2013. Web.

Dewan, John. *The Fielding Bible*. Skokie, IL: ACTA Sports, 2009. Print.

"Fan Cost Index." *Team Marketing Report*. N.p., n.d. Web. Feb. 2014.

Fangraphs Sabermetrics Library. 1 Jan. 2014. Web.

"First Year Player Draft Official Rules." *MLB.com*. MLB Advanced Media, 1 Jan. 2014. Web.

Gassko, David. "Pitching Runs Created." *Hardball Times*. 1 Jan. 2006. Web.

James, Bill. The Bill James Baseball Abstract 1988. New York: Ballantine Books, 1988. Print.

Jazayerli, Rany. "Doctoring the Numbers: Part 12." Baseball Prospectus, 05 Jun 2006. Web..

Kimes, Mina. "Minor Leaguers Deserve Fair Pay." *ESPN.com*. 1 Jan. 2014. Web.

Krautmann, Anthony C. "What's Wrong With Scully-Estimates Of A Player's Marginal Revenue Product." *Economic Inquiry* 37.2 (1999): 369-81. *Western Economic Association International*. Web.

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

"Minimum Players Salaries in Major League Baseball From 2003-2014." *Statista*. 1 Jan. 2014. Web.

"MLB Free Agent Tracker." *ESPN.com*. 1 Jan. 2015. Web.

"MLB Salary Rankings." *Spotrac*. 1 Jan. 2015. Web.

"Runs Created." *Baseball Reference*. 1 Jan. 2014. Web.

Schwartz, Peter, Max Raskin, Scott Soshnick, and Pamela Roux. "Major League Baseball Franchise Valuations." *Bloomberg.com*. 1 Jan. 2013. Web.

Scully, Gerald (1974). "Pay and Performance in Major League Baseball." *The American Economic Review* 64.6 (1974): 915-30. *Vanderbilt Economics*. Web.

Scully, Gerald (1989). "The Business of Major League Baseball, Chicago; University of Chicago Press, 1989, 151-170

Wang, Victor. "Valuing the Draft (Part 2)." The Hardball Times, 02 Feb 2009. Web

"wRC and wRC+." *Fangraphs*. 1 Jan. 2015. Web.

"2001 Team-By-Team Revenues and Expense Forecast." *USA Today*. Gannett Co, INC, 1 Jan. 2001. Web.