2016

Defensive Ability and Salary Determination in Major League Baseball

Christopher B. Shorey

Defensive Ability and Salary Determination in Major League Baseball

Chris Shorey

Readers: Professor Dave Findlay and Professor Tim Hubbard

Abstract

The process of salary determination in Major League Baseball (MLB) includes multiple levels of bargaining power and performance determinants. Previous studies of MLB salary determination have used a variety of measures of player performance. This paper examines the effect defensive ability has on salary determination for arbitration eligible players and for free agent players. Specifically, it will analyze player salary/contract data negotiated during the 2012-2015 period along with performance data from past seasons to examine the extent to which fielding percentage, errors, and the more recently developed measures of defensive ability affect player salary. Particular attention is paid to matching the negotiated contract/salary data to previous seasons' performance data in order to replicate the informational conditions known to both the team and the player at the time of negotiation. I also included offensive performance, player race and player ethnicity in all models. Results will examine how much emphasis is placed on defensive ability when determining a player's value.

I. Introduction

There has been much discussion of the topic about player salary determination in Major League Baseball since the restructuring of the labor market in 1970s. This restructuring caused a movement in bargaining power from the team to the player. As a result, different tiers of player bargaining power have emerged. In this paper, I will examine the effects of defensive ability on the salary determination model within Major League Baseball (MLB) in each of those tiers. This paper will focus on two key issues surrounding that topic. First, I will examine the manner in which salary data are correctly matched with performance data across the different tiers of player bargaining power. Second, I will examine how the different statistical measures of defensive ability alter the salary determination model. This research will examine the effect defensive metrics have on the salary determination model, examining statistical measures such as fielding percentage, defensive wins above average, defensive runs saved, errors, and defensive wins above replacement. The primary purpose of this research is to examine how defensive statistics affect the salary determination model.

In order to examine these two issues, I constructed a number of models that estimate player salary based upon player bargaining power. The first model will estimate the salary of players that are eligible for free agency, who, after obtaining six years of MLB experience, have the ability to negotiate their salary on the open market. The second model will also examine free agent players, however, players in this model are only free agents because they were released by their previous team. The third model will examine all arbitration eligible players who signed contracts that began during the 2012-2015 seasons. The fourth model will estimate the salaries for arbitration eligible players that did not file for arbitration, despite having enough experience to do so. The fifth model will examine arbitration eligible players who filed for arbitration and the salary value received at the end of the process, regardless of the outcome.[1] The sixth model will estimate the salary figure put forth by arbitration eligible players after they file for arbitration. The seventh and final model will examine team salary offers for players that filed for arbitration.

My research is organized is the following manner: the second section of this paper will discuss the salary determination process in the MLB, outlining how players are separated into different levels of bargaining power. The third section will provide an inventory of defensive statistics that exist, including how they are collected and calculated. Included in this section will be a discussion of each of the five defensive statistics that are used in all of my model, and the period of time for which they are available. Section four of my paper will be my review of relevant literature and any findings that came from that review. Next, in section five, I will discuss my three samples and the data collection process. In this section, I will also outline the specification of my core salary determination equation. Section six of this paper will discuss the results of my regression analysis of players in the MLB free agent market. This section will include separate results for the two types of free agents in my sample: free agents with six or more years of MLB experience, and free agents who have less than six years of experience and only entered the market after being released by their previous teams. The seventh section with discuss the results of regression analysis preformed on salary observations of players with arbitration eligibility. This section with include five sets of regression results, including: salaries received by all players in the sample, salaries received by players in the sample who did not file for arbitration, salaries received by players who filed for arbitration, the salary offers put forth by players during the arbitration process, and the salary offers put forth by teams during the arbitration process. The eighth section will include a discussion of the results from the defensive

[1] Salaries in the model are either negotiated between the player and the team or decided by an arbitration panel based upon offers made by both the player and the team.

statistics within my regressions. The ninth section will outline my overall conclusions from my regression analysis, with special attention paid to the area of interest of my research: the effect defensive statistics have on player salary.

II. Salary Determination in the MLB

For much of baseball's history, player bargaining power was virtually nonexistent. The reserve clause was implemented in the 1800s, restricting a player's ability to negotiate his own contract. Until 1975, the player's original team retained the exclusive rights to a player's services for the duration of the player's career. This meant that players were not allowed to negotiate contracts with other teams, and the player's original team controlled movement from team to team. This system was altered following the creation of the Major League Baseball Players Association (MLBPA), an entity that has worked to increase the amount of bargaining power players have since its creation. In 1973, following a labor strike, players gained the ability to file for arbitration and have an independent third party determine the player's salary for the following year. In 1975, Peter Seitz, a federal judge at the time, ruled that the reserve clause guaranteed a team exclusive rights over a player's services only for the first six years of a player's career. This ruling legalized free agency, a player's ability to let teams bid for his services on the open market.

The 1975 ruling led to the current system of player contract determination. The MLB currently has a three-tier system of bargaining, with bargaining power dependent on a player's MLB service. MLB service is the measure of how long a player has spent on a MLB roster, and it is measured in days and years, with 172 days equaling one year. For the first three years of a player's Major League career, he is considered arbitration ineligible and subject to a minimum salary depending upon their minor or major league status. During this time, teams control a player's rights exclusively. After logging three years of Major League service, a player is considered arbitration eligible. This means that in the case when the player feels his team is undervaluing his services, that player can take his case to an independent arbitrator, who will help the team and the player decide the salary player will receive in exchange for his services. Some players that have less than three years of MLB experience are eligible for arbitration. These players are known as "Super Two" players, and they are defined as players that have more than two, but less than three years of experience. In order to qualify as a "Super Two" player, a player must have at least 86 days of experience in the previous season and rank in the top 22% of all "Super Two" eligible players in terms of service time. The cutoff for "Super Two" status in 2014-2015 offseason was 2 years and 133 days of Major League service. This means that if a player spent 2 years and 133 days on a Major League roster by the end of the 2014 season, then that player would be considered arbitration eligible.

The arbitration process in Major League Baseball is meant to give players a chance to have some control over their compensation. The process begins with the player's original team tendering him an offer for his services in the next season. If the player deems this offer to be insufficient, he can decide to file for arbitration. The deadline for this to happen is typically during the second week in January. After the player files for arbitration, both the player's team and the player will submit a value for the player's services in the following season. These two values are sent on to the arbitration panel. The player and the team can continue negotiations throughout this process. If the two parties are unable to reach a deal by early February, then there is an arbitration hearing. During this hearing, the arbitration panel with decide between the player's value and the value put forth by the player's team. The value that the panel picks will be the player's compensation for the next season. The arbitration panel considers factors such as player performance, player leadership, public appeal, and comparable major league salaries when making their decision between the two values (MLB Collective Bargaining Agreement). A player will go through this process for determining their salary until he reaches six years of MLB service, at which time he is eligible to become a free agent.

The third-tier of player bargaining power is free agency. After six years of Major League Baseball service, a player is allowed to enter the open market and have teams bid competitively for his services. The player's former team can offer a qualifying offer, of an amount equal to the average of the top-125 player salaries in the league. If the former team offers a qualifying offer and the player signs with another team, the former team receives a draft pick as compensation. The process is as follows: following the conclusion of the World Series all players with expired contracts and more than six years of MLB service become free agents. During the first five days following the conclusion of the World Series, a player can either negotiate a new contract with his former team or discuss the possibility of playing elsewhere with other teams. It is during this time that the player's former club is able to make a qualifying offer. If a qualifying offer is not made, then following the fifth day the player will enter the open market, able to negotiate a new contract with any team. If a qualifying offer is made, then that player has seven days to accept that offer. If he does not, then movement to a new team could prove costly, as a new team would have to send a draft pick as compensation to the player's original team. A player does not necessarily have to have six years of MLB service to become a free agent. If a player is released by his original team prior to his sixth year of service, he is granted free agency as well.

Free agency has had a great impact on player salaries since its implementation in the 1970s. Since 1975, the highest salary in the league has increase by nearly 121 times, while the average salary has increased by 93 times (Haupert 2014). Player salary continues to increase along with player contract length, which now includes multi-year deals for guaranteed money. Players are now paid far more than ever before, and salaries continue to trend in an upward direction. Competitively bidding for the rights to a player through free agency completely changed the landscape of how a player's salary is determined and the factors considered when the salaries are being negotiated.

III. Defensive Statistics

This section will include a description of different measures of defensive ability, including how they are calculated and how long they have existed. Defensive statistics have evolved over time, and there is a lot more data available now than there was twenty years ago.

A. *Official MLB Statistics- Errors, Assists and Fielding Percentage*

There are multiple ways to measure defensive ability in Major League Baseball. Major League Baseball collects a few statistics officially, including errors, putouts, assists, and fielding percentage. An error is defined as, "a statistic charged against a fielder whose action has assisted the team on offense"(MLB Official Rulebook). The calculation of an error is left to the opinion of an official scorer, who gives the ruling on all plays in the field. This means that the standard for what is and what is not an error is subjective. Putouts and assists are the number of observations when a player makes a play in the field. A putout is, "is a statistic credited to a fielder whose action causes the out of a batter-runner or runner," while an assist is, "is a statistic credited to a fielder whose action contributes to a batter-runner or runner being put out" (MLB Official Rulebook). These three statistics combine to form a players fielding percentage. A player's fielding percentage is equal the number of putouts and assists divided by putouts and assists plus errors. A fielding percentage close to one typically corresponds with a high-level of defensive ability. Fielding percentage, along with values for putouts, assists, and errors, are available back through the 1871 season.

B. *Range Factor-*

Range factor is a statistic calculated using the same components as fielding percentage. Range factor is equal to putouts plus assists per defensive inning played or per games played. It calculates the number of players a player makes per inning. This statistic is also available through the 1871 season.

C. Ultimate Zone Rating-

Ultimate zone rating is a defensive statistic that puts a run value to defensive ability. It is an aggregation of the following four stats: outfield arm runs, double play runs, range runs, and error runs, each of which I have in the disaggregated form. To calculate each of those stats, player performance is compared to league average. A fielder of average ability would possess a UZR of 0. UZR is available back to the 2002 season. This stat does not include a positional adjustment, therefore, positions like 2B and SS are more likely to have double play runs than OF, among other possible forms of bias. This is a value, not a percentage, thus the career totals are aggregate from season to season. A fielder of average ability would possess a UZR of 0, as this would indicate that the player makes almost all routine plays, or plays that a player of average ability would make, within his "zone." A player of above average fielding ability would possess a UZR of +5. UZR data is available back to the 2002 season, however, it is not available for the position of catchers.

D. Defensive Runs Saved (DRS)-

Similar to UZR, Defensive Runs Saved measures the number of runs a player saves with his defensive ability compared to a player of average ability at the position. It is available back to 2003, and its algorithms are preserved but not shared by the organization Fielding Bible. Fielding Bible is a sabermetric organization that calculates unique statistics to evaluate a player's defensive ability. DRS is the primary statistic they calculate. This statistic does not include a positional adjustment. This means that more difficult positions, such as shortstop, are more likely to have more defensive runs saved than an easier position, such as first base. It is an aggregation of the following: rSB (stolen bases runs saved), rBU (bunts run saved), rGDP (double play runs saved), rARM (outfield arm runs saved), rHR (robbed home runs runs saved), and rPM (plus minus runs saved).

E. *Defensive Wins Above Replacement (DWAR)-*

This statistic converts defensive ability into win shares, and compares the ability of an average player to a player's defensive ability. It does this by determining the number of runs a player saves with his defensive ability relative to a player of average defensive ability. From there, it is calculated by dividing the number of runs a player saves by the number of runs necessary to generate a win. This statistic includes a positional adjustment, which means it weigh different components of calculation relative to the position a player plays. DWAR's computation is dependent on play-by-play data. For this reason, it is only available for all players through the 1974 season. Incomplete data is available for the 1938-1973 seasons.

F. *Defensive Runs Above Average (DEF)-*

This statistic measures a player's defensive value relative to the league average. It calculates the number of runs a player's defensive ability saves compared to a player of average defensive ability, and it includes positional adjustment. This statistic is available through the 2003 season, and its calculation is based upon play-by-play data.

IV. **Literature Review**

In reviewing relevant literature, I looked for two key indicators within each model. These indicators pertained to how economists dealt with two different statistical components of their salary determination model. The first involves defensive ability, an ability with very few metrics prior to the revolution of "sabermetrics" in the 1980s. The second is the manner in which authors work to match salary negotiation with player performance measures. When a salary is negotiated, it is negotiated based upon a player's performance prior to when a player signed the contract. In some studies, for example Stone and Pantuosco (2008), the model does not match performance statistics to when the salary was signed, but rather look to explain how previous statistics match up with the salary received in a given year. In the model mentioned above, the economists run a regression for all salaries in the 2003 season, but some of the salaries received in that year were the result of contracts signed prior to the 2003 offseason. For this reason, 2002 performance data would be insufficient to explain salaries received in 2003, as the contracts were signed, in some cases, prior to the 2002 season. This is important because in actual terms, contracts are negotiated in response to past performance, which is why statistics from the previous year are not necessarily going to give the best estimate of a salary negotiated three-years ago as part of a multi-year contract.

Salary determination models have evolved significantly since Gerald Scully (1974) introduced the first modern salary determination model in 1974, a few years prior to the transition to the modern Major League Baseball (MLB) labor market structure. Scully's original model examined team revenue, player performance, and team performance. In order to evaluate player performance for position players he included the statistical measures of slugging percentage and strikeout to walk ratios. He did not include a defensive metric, and since free agency as it stands today was nonexistent at the time, he did not match the statistical measures to when the player salaries were negotiated, because in some cases he analyzed multi-year contracts and used performance data from the 1969 season to determine salaries received in 1970. This is insufficient because in those cases, 1969 performance data was not available to the player or the team when those multi-year contracts were signed. The salary determination model has evolved significantly since this original model, but those two components appear inconsistently in the post-Scully models. In the 22 papers that I have examined while researching for this work, eight some included measures of defensive ability, and 11 worked to attempt to match player performance measures with the time when player salary was determined.

Link and Yosifov (2012) sought to examine if players are willing to trade-off salary amount for contract length in Major League Baseball. They did this by examining the 1025 free agent contracts signed between 1984-1994 and 2003-2006. Their model for salary determination included measures of player performance, contract length, player experience, team revenue and player race. The variables they examined for player performance were a three-year average of slugging percentage prior to free agent negotiations, and a three-year average of at-bats per year prior to free agent negotiations. They also built a second model that uses win shares as the measure of player performance instead of at-bats and slugging percentage. They found that there was no significant difference between the models and in both models there was a statistically significant trade-off between salary amount and contract length. This meant that free agent players were typically willing to accept lower salaries in return for longer guaranteed contracts.

Link and Yosifov did not include defensive metrics in either of their models. Their model did work to align performance metrics with the time of contract negotiation. They did this by focusing on free agent contract negotiations and examining the three-year averages of performance measures in the period leading up to the negotiations. While this allowed proper contract negotiation alignment with performance measures, it did limit the size of Link and Yosifov's sample size, causing them to have to expand the range of time they examined to a period of 14 years. Their research sought to only examine free agency, and in doing so, left out the other tiers of negotiation in the Major League labor market.

Link and Yosifov's research brings forth a third important component. Player and team behavior can change over time. Their results indicated that players were willing to trade-off contract length for salary amount in the 1980's but not as much during the 1990's and 2000's.

This brings forth a concept worth examining. How did player and team negotiating behavior change over time, and did the statistical measures used as explanatory variables vary in level of significance over time. Link and Yosifov never attempt to answer this question. However, their work will be relevant to future research that examines the effect certain types of performance variables, such as defensive performance variables, have on salary determination over time.

Rockerbie (2009) examined whether a greater supply of free-agent baseball players at a certain position has an effect on negotiated free-agent salaries. In order to do this, he examined 303 free agent salaries, by position, negotiated between the 1997 and 2002 seasons. He then built a salary determination model that included measures of player performance, team wins in the previous year, team payroll in the previous year, and the number of free agents available between seasons. The performance metrics he used in his model were slugging percentage, fielding percentage, and games played, all in the previous year. His results suggest that free-agent player ability affected player salary for the positions of shortstops and catchers, but not for other positions. He hypothesized that Major League teams tend to carry more players of the other positions that can easily cover the loss of one to free agency, but do not for the positions of shortstop and catcher.

Rockerbie included the defensive metric of fielding percentage in his model. His reasoning for doing so was because it is a relatively uniform metric across positions, and his model examined each position independently. The use of fielding percentage as a statistical measure of defensive ability is a method that has met high levels of critique in recent years, primarily because it is a subjective statistic (fangraphs.com, 2015). Fielding percentage is left to the opinion of the official scorer, as errors are judged by this impartial entity that decides whether a defensive play should be made on a consistent basis. Fielding percentage is also an imperfect statistical measure of performance because it measures the number of times an attempted defensive play is made. It fails to factor in superior defensive ability that allows more defensive plays to be made through a player's range. For example, a player like Derek Jeter, who had limited range throughout his career, posted a high fielding percentage due to his ability to make plays that he was able to reach (fangraphs.com, 2015). Fielding percentage merely measures conversion rates on attempted defensive plays. Rockerbie's decision to include fielding percentage in his model was an interesting choice on that basis. That being said, fielding percentage is also one of the few official defensive statistics measures Major League Baseball collects.

Rockerbie also attempted to match player performance metrics with the time when salaries were negotiated by focusing on only free-agent contracts negotiated between 1997 and 2002. By examining only new contracts, he was able to ensure the performance metrics were relevant at the time of negotiations and thus match them to the salaries received by the player. His use of only free agent salaries did, however, work to limit his sample size as the number of free agent contracts signed from 1997-2002 was far less than the number of total new contracts signed between 1997 and 2002. It also worked to limit his analysis to only one-tier of bargaining power.

Miller (2000) examined the difference between player salaries negotiated in final offer arbitration and salaries negotiated in free agency. He did this by building a model for salary determination and comparing the effects on the two-levels of bargaining power. This model included variables for player performance, player durability, and team winning percentage in the previous year. He analyzed cases of free agency and arbitration eligible players between the years of 1991 and 1993. The measures he used for player performance were runs created and defensive runs saved. Miller found that contracts negotiated under final offer arbitration, where an arbitration panel decides between a player's offer and the team's offer, were different than contracts negotiated under free agency. Miller noted that the results follow his hypothesis, as players have more bargaining power under free agency than they do under the final offer arbitration system. The study also found that final offer arbitration salaries are dependent upon free agency salaries, as arbitration panels factor free agent salaries of other players into their arbitration decisions.

This model incorporates a defensive measure of performance in the form of defensive runs saved. This metric examines how many runs better or worse a player is compared to an average player at that position. It is considered by many sabermetricians to be a defensive metric that captures a player's total defensive ability (fangraphs.com, 2015). Miller's efforts to match performance measures with the time when the contract is negotiated are also of note. He does this matching by examining performance data from the year leading up to when the salary is determined, and by only examining the 303 new free-agent contracts and new cases of final offer arbitration for the 1991-1994 seasons. His model includes a variable for what tier of bargaining power a player falls into, whether it be free agency eligible or arbitration eligible.

Marburger (1994) examined the relationship between bargaining power and the structure of Major League Baseball contracts. He did this by examining the three-tiers of bargaining power that exist in Major League Baseball, and the effect each has on player salary amounts. These three-tiers of bargaining power include: non-arbitration eligible players, players eligible for arbitration, and players eligible for free agency. In order to do this, he built a model that examined player performance, player salary in the current season, and player experience. In order to account for player performance, he included the measures of home runs in the previous season, runs batted in (RBI) in the previous season, runs scored in the previous season, player fielding percentage divided by the average fielding percentage at the player's position, career runs scored, career RBIs, and career home runs. He ran three separate regressions based on a player's bargaining power: arbitration ineligible, arbitration eligible, and free agency eligible. His results indicate that the arbitration process does not usually result in a player being paid the amount he would receive if he were a free agent. This shows that player bargaining power plays a significant role in the salary determination.

The Marburger model does use a measure of defensive ability in the form of fielding percentage compared to the league average at the player's position. He did not explain his use of this metric, and while the statistic of fielding percentage does have its pros and cons, Marburger's efforts to compare the metric to a league average is worth noting. This alteration does work to examine how performance in that metric alone affects salary determination. The significance of the measure in the model differs depending on bargaining power. In the arbitration eligible model, the coefficient for fielding percentage compared to league average is significant at the five percent level. In the other two models, fielding percentage compared to league average is not significant at the five percent level. Marburger's model does not match the performance metrics with the time of salary negotiation. In his work, he examines all of the player salaries of the 1992 season, regardless of when they were negotiated. In the cases of players that were arbitration eligible this would in some cases present a match with the performance measures. However, in the other two tiers of negotiating power, free agents and non-arbitration eligible players, such a match does not exist. While the model does sort players by bargaining power eligibility, it merely examines their salaries over that time instead only examining new contracts signed in the time period studied.

Brown (2008) examined what factors led to the determination of final offer arbitration salaries from a third party arbitrator in Major League Baseball. In order to do this, he examined all 2681 instances of contract negotiations that led to a third party arbitrator between the years of 1984 and 2007. Brown found the key component that differed from players that went on to final offer arbitration and those that did not was the size of the difference between offers between players and the teams with which they were negotiating. He also examined the reasoning behind why players filed for arbitration, and the number of times within their five-year window prior to becoming free agent eligible that they did so. His model included many different components including player experience, player performance, team performance, and a dummy variable for player position. For player performance, he examined the measures of on-base percentage plus slugging for the two-year period prior to negotiations, at-bats per year for the two-year period prior to negotiations, number of All-Star game appearances, number of Silver Slugger awards, and number of Gold Glove awards. His results showed that the highest variable correlated with final offer arbitration of the variables he examined was at-bats per year. This would indicate that player durability, the ability to play on a regular basis, is a significant factor in the salary determination model for final offer arbitration.

Brown attempted to match his performance variables with the time of contract negotiation by examining only cases of final offer arbitration. This allowed him to examine how a player was preforming at the point of negotiation, and how the player was compensated for that performance. However, by narrowing his scope of examination to one level of bargaining power, he also had to widen the range of dates of the data he observed. He examined final offer arbitration data that ranged from 1984 to 2007. For this reason, he ran four separate regressions for five-year increments of time. He did this to factor in changes in the overall salary determination landscape over time. A five-year range is still a period of time in which performance measure's effect on salary determination can change, which made his choice, which appears arbitrary, to calculate a model for a range of years noteworthy.

In terms of defensive performance measures used, Brown chose to incorporate only a measure for number of Gold Gloves won at the time of negotiations. The incorporation of this measure proved to be an interesting choice, as the mean number of Gold Gloves won in the 2,693 cases of final offer arbitration he observed was 0.11. This indicates that not many players win Gold Glove awards in their first five years in the Major Leagues. However, it is of note that the use of Gold Glove awards won at time of negotiation is statistically significant at the one-percent level in two of Brown's four models.

Holmes (2011) examines how players are compensated based upon their race. He uses a salary determination model that incorporates a variety of different variables including: player performance, team revenue, population of the city the player signs his contract in, and race. In order to measure player performance, Holmes uses the measures of defensive zone rating, stolen base runs, on base percentage, slugging percentage, and the number of Gold Glove awards a player has won for their defensive ability. It is noteworthy that, similar to Brown (2010), Holmes chose not to include any official defensive statistics in his analysis. However, he did include an aggregated statistic known as zone rating. Zone rating is calculated by examining the number of balls hit in a player's zone, and comparing that figure to the number of plays a player makes. The measure is relatively new compared to other statistics and data only exists for players since 1987. Holmes does not explain his choice of zone rating over other defensive measures, however, he does conclude that the reason other salary determination models typically find evidence of salary discrimination among race is their failure to include measures of defensive and running ability.

A distinguishing feature of this study is the manner in which the author attempted to match performance measures with the date of signing a new contract. In constructing his model, Holmes only examined the 511 players who signed new contracts as free agents between the years of 1998 and 2006. Holmes discussed how free agent contracts tend to be influenced more by recent performance than their total body of work. Holmes examined the statistical measures of player slugging percentage and on base percentage of the three years prior to signing new free agent contracts. He did this instead of only looking at previous year or career to date figures.

Holmes believes that teams are more likely to evaluate performance over the past three years when determining a free agent salary offer. This method of evaluation is noteworthy because it is unique when compared to other salary determination models, as only four of the 24 models examined followed this method.

An examination of previous literature shows evidence that salary determination models have attempted, albeit on an inconsistent basis, to incorporate defensive metrics of player performance into the model. Scholars are far from a consensus on both the mere presence of a metric and, in the case of a defensive metric's presence, which metric to use. The second problem is not unique to defensive metrics, as there is variation across the models on how to best incorporate offensive ability. Previous models offer differing suggestions to the solution of this problem, including the incorporation of fielding percentage, defensive runs saved, zone rating, and the number of Gold Glove awards won, but no definitive trend has emerged. The problem surrounding salary matching to player performance is similar in nature. Economists have dealt with the issue of data matching in unique ways, including building models that focus solely on new free agent contracts or models that examine arbitration settlements, but again, no consensus or definitive trend has emerged. My research focus on these two areas, including finding the important data nuances to properly match salary data with player performance data in salary determination models, and examining different defensive statistical measures of player performance, how each one influences the salary determination model, and how that influence has changed over time.

V. Data

A. Sample

The data I collected consist of all arbitration eligible and free-agent eligible players that signed contracts that would begin during the 2012-2015 seasons. This window was chosen due to the implementation of a new collective bargaining agreement following the end of the 2011 season. However, there are nine players in my sample that signed contracts prior to the implementation of the new collective bargaining agreement. These salary observations represented situations in which players signed a contract extension with their original teams that took effect once their previous contracts expired. For example, Todd Helton signed a contract extension prior to the beginning of the 2010 season, despite being under contract through the end of the 2011 season. This means that his contract extension began in the 2012 season, and for that reason is included in the sample. With the exception of the nine players that meet this criteria, all of the observations in my sample followed the contract negotiation guidelines set forth by the 2011 MLB Collective Bargaining Agreement. This means that the process of salary determination is consistent for all of the other players in the sample, as these processes, which are discussed in detail in section 3 of this paper, remained the same for all contracts negotiated following the conclusion of the 2011 season, until the conclusion of the 2016 season. My sample includes all new contracts signed by players that were either eligible for arbitration or free agency during this time period, along with any contract extensions that were signed prior to the end of the 2011 season, but began in the 2012-2015 seasons.

Following this criteria of selection, the sample is sorted by player bargaining power. This means that the sample is broken into sub-samples based on the amount of MLB service each player has at the time each contract is signed. There are 260 observations of free agent contracts signed by position players for the 2012-2015 seasons.[2] Over the same period of time there are 320 observations of position players who were eligible for arbitration for the 2012-2015 seasons. Of these 320 observations, 235 are players who elected not to file for arbitration and signed contracts with their original team through independent negotiations. The remaining 85 players filed for arbitration, exchanging figures with their original team. This means that for these 85 players, I obtained three separate salary figures: the player offers, the team offers, and the settled salary amounts. Of those 85 observations, nine players had their salary determined by an arbitration panel. Of the nine players who had their salaries determined by an arbitration panel, four had the panel side with the player and five had the panel side with the team. The remaining 57 observations in the sample are players that were arbitration eligible, having less than six years of MLB service, but were released by their original team and thus entered the free agent market.

B. *Chow Test Results*

1. Free Agent and Arbitration Eligible Player Samples

The results of the chow test conducted between the free agent and arbitration eligible player samples suggest that there is a structural break in the data between the two samples. This means that the variables that comprise the specification of the core equation have a different impact on free agent salaries and arbitration eligible player salaries. The f-test indicates that the difference in impact is statistically significant at the 5% level of significance. This means that the salary determination equation for free agents and arbitration eligible players must be examined separately, as the performance variables affect salary differently based upon which process the player uses to determine his salary. This finding is in line with intuition, as bargaining power

---

[2] I exclude pitchers from the sample due to the fact that the determinants of their salary varies greatly with the determinants of a positon player's salary. This is because the skill set of a pitcher is far different from that of a position player, thus the specification of a salary determination equation would be very different.

varies greatly between these two groups. Free agent players have the ability to negotiate in the open market, whereas, arbitration eligible players are restricted to negotiating with their original team.

As opposed to analyzing the model as a whole, when analyzing the variables on an individual level, the results indicate that there is a significantly significant difference of the impact of each variable on salary, with the exception of OBP. According to the results of the Chow test, OBP does not have a different impact on salary across the models that is statistically significant. The same is true of the constant estimates from each sample. These estimates fail to show a statistically significant structural break in the data. This suggests that, all else fixed, the starting point of salary negotiation does not vary significantly between the two samples. This finding also indicates that the error term is representing similar sources of variance in salary across the two samples. This makes sense given that the specification of the two models does not differ across samples.

2. Released Free Agent and Free Agent Player Samples

The results of the Chow test run on estimates from the released free agent and free agent with six or more years of experience samples indicate that there is a statistically significant break in the salary data between the two samples. This finding is in line with intuition given the differences in player quality across the two groups. Players in the released free agent sample were released by their former teams, which indicates that they held little value to those teams, and as such their salary determination equation should differ to the equation for players in the other sample, which included free agent players with more than six years of MLB experience. The players in the non-released free agent sample were not released by their former teams, instead choosing to enter the open market to negotiate a value for their services. Thus, demand for players in the two samples varies greatly.

3. Released Free Agent and Arbitration Eligible Player Samples

The results of the Chow test run between the release free agent and arbitration eligible player samples suggests that, at the 5% level of significance, there is a statistically significant structural break in the data between the two samples. This result is in line with intuition, given the difference in the process of salary determination between the two samples and the different levels of bargaining power between players in the two samples. Released free agents, while not necessarily in high demand, are able to negotiate a contract with any team, while arbitration eligible players are only able to negotiate with their original team. On an individual level, the results for most of the variables in the indicate that the difference in impact across models is not significant, with the primary exception being the constant estimates. These estimates represent where the structural break occurs between the two samples.

C. Contract Data Collection and Matching

One of the focuses of this research is to determine how to properly match player contract data with player performance data for the purposes of building a salary determination model. As outlined in my review of preexisting literature, a number of studies have examined player salary as an independent figure, determined on a year-to-year basis, without taking into account that salaries can be negotiated as part of a multi-year agreement between the team and the player. For this reason, it is insufficient to merely examine a player's salary when building a salary determination model. Player salary is not determined on a year-to-year basis in the MLB. Players and teams can engage in multi-year contracts, which is broadly defined as an agreed upon dollar amount in exchange for multiple years of a player's service. Due to this fact, when I collected my data, I found that it would be insufficient to merely pull salary data as the measure of player value I would be examining. Instead, I pulled data of player contracts and from this contract data, calculated the average annual value over the life of contract to use as the dependent variable in my analysis. For example, if a player signed a two-year contract with a total value of $10 million, the average annual value would be $5 million.

Multi-year contracts create problems with matching performance statistics, because the salary a player receives in years after the first year of the contract are determined based upon the player's performance when the contract was initially signed. For example, if a player signed a two-year contract before the 2012 season, the player's performance during the 2012 season did not play a role in determining the salary that the player would receive in 2013. In this example the player's 2013 salary was determined prior to the 2012 season. It is for this reason that when matching player performance to the salary the player would receive, it is insufficient to analyze only the salary a player receives from year-to-year. These salaries are determined based upon player contracts, which in turn are determined based upon the information available to both the team and the player at the time of negotiation.

In order to correctly match performance data with player salary, I began by collecting data on all of the contracts negotiated to begin during the 2012-2015 period. Important to note, at this point in my data collection, the contracts were sorted based upon when they began. This means that I had not yet obtained the data on when the contracts were actually negotiated, this step would take place later in the process. Recall that in the previous section I noted that some contract extensions were negotiated prior to the 2012-2015 period, despite beginning in this period. In short, I obtained the contract data based on when the contract took effect, not based upon when the contract was signed. All of the contract data were obtained from Cot's Contracts, a section of Baseballprospectus.com. This was done by extracting the contract data for all players on MLB Opening Day rosters for all 30 MLB teams. I did this for all four seasons from 2012-2015 and I separated the data by year. This means that I then obtained a list of contracts that were active during the 2012 season, 2013 season, 2014 season, and 2015 season respectively, regardless of when the contracts began. From these lists, I eliminated all contracts that began prior to the year in question. For example, from the list of 2012 contracts, I eliminated all contracts that began prior to the 2012 season, such as a multi-year contract signed in 2011 or before. This was an important step because it eliminated possible duplicate observations from the sample. Without taking this step, a multi-year contract signed in 2012 would have appeared again in the sample under contracts signed for 2013. I also eliminated all contracts signed for pitchers, as my research focuses only on position players.

Following the elimination of the extraneous data, I combined the four lists into one, and sorted the players by MLB service time. The MLB service time figures in the dataset represented the amount of time the player had spent on a MLB roster at the beginning of the season in question. Thus for contracts signed for the 2012 season, the MLB service time figure I had for the player corresponded to the amount of service that player had at the end of the 2011 season. This is important because that is the figure that determined the process the player was eligible to go through for salary determination: whether it be arbitration or free agency.

Once the data set of player contracts that began during the 2012-2015 period had been compiled, I next split the players into groups based on player bargaining power. Players with six years or more of MLB service were placed into the free agent group. Players with less than six years of experience but more than the "Super Two" cutoff amounts were placed into the arbitration eligible group. The "Super Two" cutoffs used were 2.146 years for arbitration eligible player contracts signed for the 2012 season, 2.140 for contracts signed for the 2013 season, 2.122

for contracts signed for the 2014 season, and 2.133 for contracts signed for the 2015 season.[3] The remaining contracts, those for players with less service time than the "Super Two" cutoff, were eliminated from the data set, as their salary is set exclusively by their team subject to the league minimum.

1. Free Agent Contract and Performance Data Matching

The free agent group data had a final step of preparation before performance statistics could be matched. I went through the 260 observations and collected the date that each contract was signed. This was an important step because the contract data I pulled were of contracts negotiated to begin during the 2012-2015 seasons, without indication of the actual date those contracts were negotiated and signed. It is possible that a contract that begins in a given season could have been negotiated and signed years in the past. As previously mentioned, I identified nine observations in my sample that met this criterion. The signing date of each contract was essential to allow me to identify these contract observations. It was important for me to identify these contracts within my dataset because performance data would have to be adjusted to reflect a player's statistics at the time when the contract was signed. This means that the player's performance data in data set had to be matched to reflect his career through the end of the nearest full season prior to when the contract was signed. This adjustment had to be made for 39 observations in the sample.[4]

[3] MLB service times are not explicitly in years, but rather in years and days. For example, a service time of 2.122 would be equal to 2 years and 122 days. A full MLB season is considered to be 172 days

[4] When adjusting the performance data of players who signed extensions prior to the end of the year before the contract would take effect, partial season data was not used. This means that if a player signed a contract extension during the 2011 season that would begin in the 2012 season, data from the 2011 season were not used.

Once the dates that the contracts were signed were determined, data were matched to reflect a player's career performance through the last season prior to signing the contract. This means that for all players that signed contracts during the off-season between the 2011 and 2012 seasons, the performance data were chosen to reflect the player's career up to the end of the 2011 season. This was done for all players in the sample. The statistics matched included basic batting statistics, basic fielding statistics, advanced fielding metrics, advanced player value metrics, and race[5] and ethnicity data.

2. Arbitration Eligible Contract and Performance Data Matching

The process for matching arbitration eligible player's contracts and performance data was largely done in a similar manner as the observations in the free agent group. The first step was to analyze the "Super Two" players in the dataset and make sure that the cutoff dates were reflective of whether or not the player was arbitration eligible with less than three years of service. Next, each observation was analyzed to determine whether the player accepted his team's tender, negotiated a contract extension with his team, filed for arbitration, or was released. There were 57 players in the dataset who were released by their original team, and thus entered the free agent market to sign a contract with a new team. These players were removed from the arbitration eligible dataset and placed in their own group, because their salary determination process was not the same as the rest of the players in the arbitration eligible group. For players that signed extensions with their original teams, the date that the extension was negotiated and signed was collected so that adjustments could be made to the performance stats, if necessary. This was process was completed the same way it was for the free agent players.

[5] Race was determined by multiple individuals analyzing a photo of a player and determining whether or not the player was perceived to be a player of color

For players who filed for arbitration, values from the arbitration process were recorded. These values included the amount put forth by the player, the amount put forth by the team, and the outcome of the process. Once this was done, performance statistics were matched in the same way as they were matched for free agent contracts. Performance statistics represented each player's career prior to when the contract was signed. For players who went through the arbitration process to determine their salary for the 2012 season, performance statistics represent the player's career through the end of the 2011 season.

D. Performance Data

Most of the performance data were collected from the Sean Lehman database, which is an aggregation of all data collected by Major League Baseball. Performance data was also collected from BaseballReference and FanGraphs. As previously noted, the performance data reflect a player's career prior to when the contract was signed and negotiated. This was done because that data represents the information available to both the team and the player regarding the player's value at the time the contract was negotiated. From the Sean Lehman database, the following offensive player data were acquired: games played, at bats, runs scored, hits, doubles, triples, home runs, runs batted in, stolen bases, number of times a player is caught stealing, walks, strikeouts, intentional walks, number of times a player was hit by a pitch, sacrifice hits (bunts and ground balls), sacrifice flies, and the number of times a player grounded into a double play. The following defensive player data were collected from the Sean Lehman database: defensive innings played, putouts, assists, errors, and double plays.[6] From Baseball Reference the following statistics were collected: wins above replacement, offensive wins above replacement, and defensive wins above replacement. From FanGraphs, the following statistics were collected:

[6] The catcher position specific statistics of passed balls, runners caught stealing, runners allowed to steal, and wild pitches were also collected.

defensive runs saved (DRS), ultimate zone rating (UZR), and defensive runs above average (DEF).

The FanGraphs data had to be adjusted, as the advanced defensive metrics were only available starting with the 2003 season.[7] A number of players in the free agent dataset began playing prior to the 2003 season. Thus, when adjusting these statistics on a per season basis, a separate statistic of adjusted player experience had to be calculated. The first adjusted player experience measure is equal to the number of seasons a player played from 2002 on, and it was used to calculate UZR on a per year basis.[8] DRS and DEF per season were calculated using another adjusted player experience statistic, which represented the number of seasons the player played from 2003 on. I will discuss my reason for adjusting these statistics in the next section.

Race and ethnicity data were also collected. The race dummy variable in my data is based on the evaluation of multiple people to determine whether or not the player was a person of color based upon a photo of the player.[9] The ethnicity dummy variable in my data denotes whether a player was born in a Latin American country or not. Home countries of players in the sample that indicated a Hispanic ethnicity included: Venezuela, Curacao, Dominican Republic, Puerto Rico, Panama, and Cuba. The evaluation of whether or not a player was a person of color was only made for players not born in a Latin American country. Thus, a non-white player would not show up in the data as both Black and Hispanic, but instead as one or the other.

[7] UZR was only available back to the 2002 season
[8] UZR data are not available for catchers. UZR data were collected to examine models without players that play the catcher position. This would serve as a robustness check.
[9] The evaluation of whether or not a player was a person of color was only made for players not born in a Latin American country. Thus, a non-white player would not show up in the data as both Black and Hispanic, but instead as one or the other.

E. Model

The core salary determination model used was as follows:

$$\text{Salary} = \beta_0 + \beta_1 (\text{Slugging Percentage}) + \beta_2 (\text{On-base percentage}) + \beta_3 (\text{At bats per year})$$
$$\beta_4 (\text{Stolen Bases per year}) + \beta_5 (\text{Middle Infielder Position dummy}) +$$
$$\beta_6 (\text{Latin American dummy}) + \beta_7 (\text{Black dummy}) + \beta_8 (\text{Total Seasons}) +$$
$$\beta_9 (\text{Defensive metric})$$

1. Non-defensive variables in the Model

   a. *Slugging percentage-*

Slugging percentage (SLG) measures a player's ability to hit the ball for power. It is calculated by dividing a player's total bases by at bats. At bats is the number of times a player either creates an unproductive out or gets a hit. [10] Total bases is the number of bases a player collects from a hit, and is calculated by adding the number of singles a player hits, the number of doubles a player hits times two, the number of triples a player hits times three and the number of home runs a player hits times four. Values for slugging percentage were calculated using the statistics from the Sean Lehman dataset. SLG is a measure of efficiency, which means it represents, on average, how often a player will generate offense each at bat. This means that players with a higher SLG are considered more productive players than those with lower SLG. For this reason, one would expect a player with a high SLG to receive a high salary.

[10] Sacrifice flies and hits, which are instances where a player creates an out but advances a runner on base to the next base, are not included in the measure of at bats.

b. On base percentage

On base percentage(OBP) represents the rate at which a player reaches base. It is calculated by dividing walks, number of times hit by a pitch and hits by the number of plate appearances a player has. Plate appearances is a value that represents the number of times a player attempts to get a hit against the pitcher. With the Lehman data, this value was acquired by adding at bats, walks, number of times hit by a pitch, sacrifice hits, and sacrifice flies. OBP is an efficiency measure, as it indicates how often a player reaches base per attempt. As such, a higher OBP usually indicates that a player has a greater amount of offensive ability. For this reason, one would expect a player with a high OBP to receive a high salary.

c. *At bats per year*

At bats per year (AB per year) indicates the average number of at bats a player has during an average year of his career.[11] As previously mentioned, at bats is a measure of how many times a player gets a hit or makes an unproductive out. This means that walks, intentional walks, sacrifice hits, and sacrifice flies are not included in the calculation of at bats. AB per year is included in the model to supplement both OBP and SLG. Recall that both OBP and SLG are efficiency measures, representing the likelihood of success per attempt. In contrast, AB per year indicates how long a player is able to sustain his performance offensively. For example, for players with equal values for SLG, the player with more AB per year is likely to be more valuable, because he sustained that level of efficiency over a greater number of attempts. Also, players with a higher number AB per year are players that are unlikely to miss games, due to injury or other reason, during a typical year. It is for these reasons that at bats is included in my model on a per year basis, it indicates how often the player performs at a certain level during an

[11] Per year was calculated by using a measure of seasons (full or partial) played. This means that if a player only played in 30 games one season of his career, that is still counted as a full season.

average season. One would expect a player with a higher number of AB per year to receive a higher salary, because they are able to sustain their performance on a more consistent basis than a player with less AB per year.

d. Middle infield binary variable

The middle infield dummy variable indicates the position that a player primarily plays. If a player primarily plays catcher, second base, or shortstop, this variable is equal to 1. If a player primarily players a different position, it is equal to 0. Since salaries vary highly by position, and position is not indicative of a player's ability, this variable's effect on salary is unknown. Of note, corner infield variable was dropped from the model's specification because it was not significant in any regression results.

e. Total seasons

This variable is a measure of the number of seasons a player has played. This is an integer value, so all partial seasons that a player plays are counted as 1. Since the data are segmented by bargaining power, the hypothesized effect on salary will differ. For free agents, this variable will likely have a negative effect on salary, because all players in that group have a minimum of six years of experience and will likely receive lower salaries as they age. For arbitration eligible players, this variable will likely have a positive effect on salary as those players are in the earlier stages of their career, all having less than six years of experience.

2. Defensive variables in the model

a. Fielding Percentage

Fielding percentage equals the number of putouts and assists a player completes divided by putouts plus assists and errors. Putouts are defined as, "a statistic credited to a fielder whose action causes the out of a batter-runner or runner" (MLB Official Rulebook). Assists are, "a statistic credited to a fielder whose action contributes to a batter-runner or runner being put out"

(MLB Official Rule Book). Errors are, “a statistic charged against a fielder whose action has assisted the team on offense” (MLB Official Rulebook). Fielding percentage is an efficiency measure, with a higher value indicating that a player is a better fielder than one with a lower number. For this reason, one would expect fielding percentage to have a positive effect on salary.

b. *Defensive runs above average*

Defensive runs above average (DEF) measures a player’s defensive value relative to the league average. It represents the number of runs a player saves with his defensive ability compared to the defensive ability of an average player. This statistic includes a positional adjustment, which means that components of its calculation are weighted differently based upon the position the player plays. This statistic is available beginning in the 2003 season, and its calculation is based upon play-by-play data. Due to the fact that DEF is not an efficiency measure, but rather a value that can be aggregated from year-to-yea, it appears in my model as a per year figure. This means that its value indicates the number of runs a player saves per year. This variable was calculated by dividing career DEF by an adjusted experience statistic, because statistics are only available back to the 2003 season. DEF is an indicator of defensive ability, which means that players with higher DEF statistics are better fielders than players with lower ones. For this reason, one would expect DEF to have a positive effect on salary.

c. *Defensive runs saved*

Defensive runs saved (DRS) is a measure of runs saved with defensive ability compared to a player of average ability. This statistic does not include a positional adjustment. This means that more difficult positions, such as shortstop, are more likely to have more defensive runs saved than an easier position, such as first base. It is an aggregation of the following: rSB (stolen bases runs saved), rBU (bunts run saved), rGDP (double play runs saved), rARM (outfield arm runs saved), rHR (robbed home runs runs saved), and rPM (plus minus runs saved). Each of these measures is calculated based on play by play data and compared to a player of average ability. DRS is included on a per year basis because it is not an efficiency measure, thus it can be aggregated from year to year, so players who have play longer will have higher DRS values than those that have not played as long. This variable was calculated by dividing career DRS by an adjusted experience statistic, because statistics are only available back to the 2003 season. DRS is an indicator of defensive ability, which means that players who are better at defense have higher DRS statistics than players that are not as skilled defensively. For this reason, one would expect DRS to have a positive effect on salary.

d. Errors per year

Errors per year is calculated by dividing the total number of errors committed by number of seasons played. Errors represent the number of times a player makes a mistake on defense, this means that players with a higher number of errors, in theory, possess less defensive ability than players that commit less errors. That being said, players with more defensive chances, such as infielders, also likely commit more errors due to higher volume. However, all else fixed, one would expect errors per year to have a negative effect on salary. This means that players with more errors per year would be expected to have a lower salary than players with less errors per year, all else fixed. This variable is included on a per year basis because it can be aggregated from season to season. As such, including a variable for total errors committed in a player's career would not be a fair measure of a player's defensive ability, as the longer a player plays, the more likely he is to commit errors. Instead, the including errors on a per season basis allows me to cancel out the effects of player experience in the variable.

e. Defensive wins above replacement

Defensive wins above replacement (DWAR) is a calculation of wins generated from a player's defensive ability compared to a player of average ability. This statistic includes a positional adjustment, which means that the weights of its components are adjusted based on which position a player typically plays. It is calculated by dividing the number of runs saved by a player's defensive ability compared to an average player by the number of runs saved necessary to account for a win. Due to the fact that it is not an efficiency measure, DWAR is included in the model on a per year basis. DWAR can be aggregated from year to year, so players with more time in the league have higher career DWAR values than players with less time. To eliminate this bias, it was calculated on a per year basis by dividing career DWAR by total seasons played. DWAR is available back to 1934, so an adjusted experience statistic was not used to adjust it on a per season basis. This statistic is an indicator of defensive ability, which means that players with greater defensive ability have a higher DWAR statistic. For this reason, one would expect DWAR to have a positive effect on salary.

VI. Free Agent Results

A. Non-released Free Agent Player Results

1. Description of Free Agent Sample

The summary statistics of this sample indicate that the sample is largely representative of the demographics of the MLB. 60.4% of all players in the sample are White, while 29.6% are Hispanic, and the remaining 10% are Black. These values are all within 5% of actual demographic estimates of the entire MLB. In this sample, Hispanic players receive the highest mean salaries, while White players receive the lowest mean salary among these three demographic groups.

2. Results

a. Equation 1.1

The results of equation 1.1 indicate that SLG, AB, SB, MIF, and Hispanic all have a statistically significant positive effect on salary, while total seasons played has a statistically significant negative effect. All else fixed, a 1% increase in slugging percentage would cause a 10.56% increase in salary. This result was found to be significant at all reasonable levels of significance. The results for OBP would indicate that at the 5% level of significance, OBP does not have a statistically significant impact on salary. The results for AB indicate that, for example, an extra 10 AB per year would cause a 2.01% increase in salary, all else fixed. This result was found to be significant at all reasonable levels of significance. The results for SB per year indicate that an increase in stolen bases per year by one causes a 2.48% increase in salary. This result was found to be significant at all reasonable levels of significance.

Table 1
Summary Statistics of Free Agent Players

	White	**Hispanic**	**Black**
	Mean (sd, max, min)	Mean (sd, max, min)	Mean (sd, max, min)
Salary (millions)	5.249 (5.316,25,0.586)	6.692 (6.326,24,0.718)	6.391 (6.207,25,0.700)
SLG	0.433 (0.0496,0.567,0.301)	0.419 (0.0630,0.617,0.314)	0.417 (0.0490,0.586,0.349)
OBP	0.335 (0.0276,0.427,0.250)	0.323 (0.0306,0.420,0.265)	0.329 (0.0192,0.376,0.261)
AB	360.0 (99.57,673.8,146.1)	374.4 (110.9,592.9,143.9)	437.5 (91.96,571.5,149.3)
SB	5.665 (6.519,37.67,0)	6.595 (8.013,41.11,0.143)	17.72 (14.11,46.17,1.429)
MIF	0.363 (0.482,1,0)	0.545 (0.501,1,0)	0.154 (0.368,1,0)
Total Seasons	9.510 (2.805,20,6)	10.44 (3.447,23,6)	9.538 (2.470,16,7)
Fielding Percentage	0.985 (0.00866,0.998,0.954)	0.982 (0.0112,0.995,0.940)	0.986 (0.00763,0.994,0.969)
DEF	0.496 (6.702,14.26,-12.87)	2.708 (6.633,15.82,-12.48)	-0.429 (7.277,12.26,-14.15)
DRS	0.897 (4.932,14.43,-12.25)	0.724 (4.147,12.10,-10.25)	0.578 (4.899,9.857,-11.40)
Errors	5.422 (3.560,20.14,0.636)	7.576 (4.811,19.58,1.235)	4.930 (3.207,14.10,1)
DWAR	0.0554 (0.743,2.057,-1.450)	0.250 (0.671,1.538,-1.086)	0.0121 (0.677,1.229,-1.225)
N	157	77	26

Table 2
Free Agents Results

	1.1	1.2	1.3	1.4	1.5	1.6
SLG	10.56***	10.60***	11.00***	11.06***	10.68***	11.12***
	(1.582)	(1.583)	(1.541)	(1.558)	(1.570)	(1.547)
OBP	4.599	4.279	5.720*	4.577	4.182	5.925*
	(2.757)	(2.781)	(2.693)	(2.702)	(2.742)	(2.706)
AB	0.00201***	0.00204***	0.00214***	0.00188**	0.00260***	0.00188**
	(0.000600)	(0.000601)	(0.000584)	(0.000589)	(0.000651)	(0.000585)
SB	0.0248***	0.0252***	0.0176*	0.0230***	0.0240***	0.0190**
	(0.00680)	(0.00682)	(0.00684)	(0.00668)	(0.00676)	(0.00679)
MIF	0.605***	0.607***	0.358**	0.554***	0.714***	0.376**
	(0.119)	(0.119)	(0.131)	(0.118)	(0.128)	(0.130)
Hispanic	0.357**	0.370**	0.345**	0.379***	0.401***	0.361***
	(0.111)	(0.112)	(0.108)	(0.109)	(0.112)	(0.108)
Total Seasons	-0.116***	-0.117***	-0.127***	-0.119***	-0.119***	-0.122***
	(0.0169)	(0.0169)	(0.0166)	(0.0166)	(0.0168)	(0.0165)
Fielding Percentage		4.639				
		(5.229)				
DEF			0.0366***			
			(0.00916)			
DRS				0.0350***		
				(0.0105)		
Errors					-0.0319*	
					(0.0143)	
DWAR						0.330***
						(0.0856)
Constant	8.892***	4.407	8.505***	8.764***	8.920***	8.415***
	(0.649)	(5.097)	(0.638)	(0.638)	(0.644)	(0.644)
Adjusted R^2	0.465	0.464	0.495	0.485	0.473	0.492
N	260	260	260	260	260	260

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results for MIF indicate that, all else fixed, middle infielders are paid 67.4% more than players at other positions.[12] This result was significant at all reasonable levels of significance. This finding suggests that when offensive performance statistics are held constant, teams value a middle infielder more than either a corner infielder or outfielder. This is likely because middle infielders that possess average levels of offensive ability are at a premium in the MLB. This is reflected within my sample, as the mean OBP and SLG for middle infielders are 7.2% and 14.8% below the mean value of OBP and SLG for all other players. For this reason, it makes sense that given a middle infielder and non-middle infielder of equal offensive ability, a team would pay more for the middle infielder. The results for the Hispanic variable indicate that, all else fixed, Hispanic players are paid 52.6% more than non-Hispanic players.[13] This result was found to be significant at all reasonable levels of significance. There could be a variety of reasons behind this finding. First, this could be attributed to a level of bias, as Hispanic players receive the highest mean salary of all race/ethnicities within the sample. Correlation coefficients do not suggest that there is a high level of correlation between the Hispanic variable and other data collected. The results for Total Seasons indicates that, all else fixed, each additional number of seasons played causes a 11.6% drop in salary for free agent players. This result was significant at all reasonable levels of significance. The negative coefficient estimate of this variable likely represents the fact that all players in this sample have at least six years of MLB experience. This means that this sample has primarily older players, and as players play more seasons, their performance declines, resulting in a lower salary.

[12] This coefficient estimate was adjusted using the method suggested by Giles (2011), as will the coefficient estimates for all binary variables in this paper.
[13] Coefficient estimates for the Black binary variable were not significant

b. Equation 1.2

In equation 1.2, the fielding percentage variable is added to the core model. The results from this equation indicate that career fielding percentage is not significant at the 5% level of significance. For that reason, I would fail to reject the null hypothesis that fielding percentage has no impact on salary. This result is noteworthy for a number of reasons. Fielding percentage is a statistic that has been the standard for measuring a player's defensive ability for a majority of the history of the MLB. It is calculated using statistics that are officially collected by the MLB, indicating the rate of efficiency that a player successfully completes a defensive play that an MLB official scorer decides he should be able to complete. The results from this equation indicate that this statistic, despite its historical prevalence, is not used in the salary determination process for free agent players. It is possible that with the advent of new statistics in recent years, including some of the statistics used in subsequent equations, both players and teams are relying on different measures of ability. An interesting area for further research would be to examine the significance of fielding percentage in the salary determination model over time. However, my sample only includes observations dating back to 2012.

c. Equation 1.3

In equation 1.3, defensive runs above average per year is added to the core equation. Recall that DEF is a measure of runs saved by defensive ability by a player compared to a player of average ability. This value can be either positive or negative, with a value of zero indicating average defensive ability. The results for DEF indicate that, all else fixed, an increase in DEF by 1 causes a 3.66% increase in salary. This result was found to be significant at all reasonable levels of significance. Given the results for equation 1.2, the results for DEF suggest that players and teams are taking advantage of defensive metrics other than fielding percentage to make their salary determination decision. DEF is a newer statistic, created in the 2000s, and of note is its inclusion of a positional adjustment. The calculation of DEF varies based on the position a player plays. This adjustment cancels out differences in volume based on position. That is to say, statistics such as fielding percentage assume all positions are equal. This means that there is no adjustment for the fact that some positions, such as shortstop, receive more opportunities to make a defensive play than other positions. The qualities that make a player excel defensively vary by position, and DEF takes this variance into account by weighting components differently based on position. In doing so, the DEF statistic gives an output that is comparable across positions to signify a player's overall defensive ability. This fact is likely the reason why it was found to be significant in the salary determination model, whereas fielding percentage was not.

d. Equation 1.4

In equation 1.4, the DRS variable was added to the core equation. Recall that similar to DEF, DRS is a measure of runs saved by defensive ability compared to a player of average ability. The primary difference between the two statistics is that DRS does not include a positional adjustment. This value can be positive or negative, with a value of zero indicating average defensive ability. The results for DRS indicate that, all else fixed, an increase in DRS per year by one causes a 3.5% increase in salary. This result was found to be significant at all reasonable levels of significance. For that reason, I would reject the null hypothesis that DRS has no impact on salary.

e. Equation 1.5

The results for the Errors variable indicate that, all else fixed, each additional error committed per year causes a 3.19% decrease in salary. This result was found to be significant at all reasonable levels of significance. For that reason, I would reject the null hypothesis that Errors has no impact on salary. This result is in line with intuition, as errors are associated with poor defensive ability. This result is noteworthy because errors are a component of fielding percentage calculation, which the results from equation 1.2 show is not statistically significant. The Errors variable is correlated with the position a player plays, and that holds true in this sample. The average infielder in this sample commits 7.23 errors per year, while the average outfielder commits 3.23 errors per year. With errors serving as the measure of defensive ability, this statistical results suggests that the average outfielder possesses more than twice the defensive ability than the average infielder. This finding is not in line with the results of any of the other defensive statistics in this sample. In fact, for each DEF, DRS, and DWAR, the mean value in this sample is lower for outfielders than non-outfielders, suggesting a lower level of defensive ability. Considering those facts, the results from equation 1.5 likely speak more to the positive correlation between outfielders and salary and the low mean values of errors per year for outfielders relative to the rest of the sample. This finding once again shows the flaws in officially collected MLB defensive statistics, as errors has been a defensive statistic collected throughout the history of the MLB.

f. Equation 1.6

In equation 1.6, the DWAR variable is added to the core equation. Recall that DWAR is the measure of wins generated by a player's defensive ability compared to a player of average ability. This value can be either positive or negative, with average ability equaling a value of zero. This statistic includes a positional adjustment, similar to DEF. The results for DWAR indicate that, all else fixed, an increase in DWAR per year by one causes a 33.0% increase in salary. This result was found to be significant at all reasonable levels of significance.

g. Robustness Tests

I also conducted a series of tests to determine the robustness of the results and to determine whether the significance of the results is due to a possible omitted variable. While not technically a test for robustness, I will first discuss the implications of the inclusion of both the

Black and corner infield (CIF) binary variables. Both of these variables were a part of the original specification of the core model. However, the estimated coefficients for both variables are never statistically significant. Furthermore, the inclusion of the Black binary variable did little to alter the significance or the coefficient estimates of other variables in the model. The same findings held for the CIF variable. These results indicate that both of these variables do not appear to be relevant in the salary determination process.

The next set of tests sought to determine if the interaction of defensive statistics with the Hispanic and MIF binary variables yielded significant results or altered the coefficient estimates of other variables. The results of interacting the MIF variable with each defensive statistic where inconsistent, altering the significance of some variables, leaving the significance of others unchanged. The inclusion of the interaction variable with fielding percentage and DEF caused MIF to lose significance at the 5% level. It also caused the Errors variable to lose significance at the 5% level. In all cases, the interaction variable was not significant, and it did little to alter the coefficient estimates for the remaining variables.

The results were the same when the MIF variable was disaggregated and the positions of catcher, shortstop, and second base were included separately, each with its own interaction variables. These results show that middle infielders do not have a statistically significant increase in salary due to defensive ability relative to other positions. The results from interacting the Hispanic binary variable with defensive statistics showed that the inclusion of the interaction variable did little to alter significance levels or coefficient estimates of other variables, and in all cases the interaction variable is not significant at the 5% level. These results suggest that Hispanic players do not receive higher salaries for their defensive ability relative to other players.[14]

[14] The same tests were run for the Black and CIF binary variables, yielding the same statistically insignificant results.

The final set of tests I conducted sought to hold external variables surrounding the player's contract negotiation fixed. These external variables can be sorted into two major categories: team fixed effects and year fixed effects. In order to examine the team fixed effects, I introduced a binary variable for 29 of the 30 MLB teams, and examined the impact the inclusion of these variables had on the significance and coefficient estimates of other variables.[15] The binary variable is equal to one if the player signed with that team and equal to 0 if the player signed with a different team. These team-specific binary variables work to hold constant variables such as team payroll, market size, and team success. The inclusion of these variables did not alter the coefficient estimates or significance of other variables in the model, suggesting that the team with which a player signs does not have a significant effect on salary. In order to hold the effects of a year fixed, I introduced a binary variable for the first year the contract took effect into the model, and examined how the inclusion of these variables affected the coefficient estimates and significance of other variables. If the player's contract began in 2012, then the 2012 binary is equal to one, and it began during a different season, then the 2012 binary is equal to zero. I introduced such a binary for the 2012, 2013 and 2014 seasons.[16] The inclusion of these binary variables did not significantly alter the coefficient estimates or significance levels of other variables in the model. The results of this robustness test shows that the year the player's contract began does not have a significant impact on salary.

h. Summary of Results

The results from these regressions suggest that defensive ability has a significant impact on

[15] The 30th team acted as the reference group. Of the coefficient estimates of these binary variables, only one, Philadelphia Phillies, was statistically significant. It had a negative coefficient estimate.
[16] 2015 was the reference group. Of the coefficient estimates of these binary variables, the 2012 and 2013 variables were both statistically significant and negative

player salary for free agent players. The movement of the adjustment r-squared value shows that the defensive variables that cause the largest increase are the positionally adjusted statistics of DWAR and DEF. DEF caused the adjusted r-squared value to increase from .465 to .495 relative to the core equation, while DWAR caused an increase to .492. While not positionally adjusted, DRS also caused a sizable increase in the r-squared value, while the Errors variable caused the smallest increase of all statistically significant variables. This suggests that the Errors variable had the least amount of explanatory power of all statistically significant variables used. This finding furthers the hypothesis that officially collected MLB statistics such as errors and fielding percentage are less commonly used for assessing a player's value at least within the free agent market.

B. Released Free Agents

1. Model

This samples includes players that were arbitration eligible, but were released by their former team and effectively became free agents as a result. This group includes 57 players, all of which went through a different salary determination process than players that had similar levels of experience, as all of these players have less than six years of experience, and are only free agents due to their release. The core equation used to examine this salary determination was the same as the core equation used in previous sections. This equation includes SLG, OBP, AB, SB, MIF, Latin American, and Total Seasons as the explanatory variables. In each subsequent regression the defensive variables of fielding percentage, DEF, DRS, Errors, and DWAR were added individually.

The demographics of this sample suggest that it is relatively representative of the demographics of the entire MLB. 64.9% of the sample is White, 21.1% is Hispanic, and 14.0% is

Table 3
Summary Statistics of Arbitration Eligible Free Agent Players

	White	**Hispanic**	**Black**
	Mean (sd, max, min)	Mean (sd, max, min)	Mean (sd, max, min)
Salary (millions)	1.496	1.310	1.085
	(1.407,6,0.503)	(0.702,2.750,0.750)	(0.405,1.850,0.505)
SLG	0.383	0.346	0.374
	(0.0461,0.494,0.308)	(0.0400,0.437,0.296)	(0.0445,0.429,0.318)
OBP	0.312	0.301	0.315
	(0.0237,0.359,0.259)	(0.0206,0.335,0.270)	(0.0109,0.333,0.300)
AB	286.9	232.4	264.8
	(125.0,536,98.25)	(72.61,346.6,127.6)	(67.64,382.4,154.7)
SB	4.668	5.161	14.42
	(4.795,22.33,0)	(8.159,27.20,0.200)	(10.32,27.60,2.800)
MIF	0.405	0.833	0
	(0.498,1,0)	(0.389,1,0)	(0,0,0)
Total Seasons	4.730	5.333	4.625
	(0.962,6,3)	(0.651,6,4)	(0.916,6,3)
Fielding Percentage	0.978	0.984	0.985
	(0.0121,0.997,0.956)	(0.00905,0.994,0.968)	(0.00551,0.992,0.975)
DEF	-1.732	2.880	0.506
	(4.636,5.500,-11.48)	(2.784,7.333,-1.820)	(5.746,10.76,-4.900)
DRS	-0.513	1.053	0.423
	(4.318,7.400,-10.83)	(1.823,3.833,-2.333)	(4.392,9.400,-4)
Errors	5.698	4.893	2.312
	(4.111,21.67,0.250)	(3.237,13,1.200)	(0.925,3.750,1)
DWAR	-0.160	0.343	-0.116
	(0.586,0.900,-1.400)	(0.181,0.667,-0.0600)	(0.588,1.040,-0.720)
N	37	12	8

Black. This suggests that White and Black players are slightly overrepresented in this sample, while Hispanic players are underrepresented. White players have the highest mean salary in this sample, while Black players have the lowest. This finding is interesting given that the none of the demographic groups consistently has the highest average performance statistics.

2. Results

   a. *Equation 2.1*

The results of equation 2.1 indicate that the SLG, AB, and SB variables all have a statistically significant positive effect on arbitration eligible free agent salary. All else fixed, a 1% increase in slugging percentage would cause a 6.20% increase in team salary offer. This result was found to be significant at all reasonable levels of significance. The results for OBP are marginally significant, but not significant at the 5% level. The results for AB would indicate that an extra 10 AB per year would cause a 2.49% increase in salary, all else fixed. This result was found to be significant at all reasonable levels of significance. The results for SB per year indicate that each additional stolen bases per year corresponds to a 2.04% increase in salary. This result is significant at the 4.7% level of significance. The results for MIF also indicate that the positional variable does not have a significant impact on salary. The results for the Hispanic variable are not significant at the 5% level, indicating that ethnicity does not have a significant effect on an arbitration eligible free agent's salary.[17] The results for the Total Seasons variable are not significant at the 5% level of significance, indicating this variable does not have a significant role in determining arbitration eligible free agent salary.[18] The adjusted r-squared value of this equation was .494.

---

[17] The Hispanic variable was marginally statistically significant, with a p-value of .11
[18] Total Seasons was also marginally statistically significant, with a p-value of .11 as well

Table 4
Arbitration Eligible Free Agents

	2.1	2.2	2.3	2.4	2.5	2.6
SLG	6.205**	6.252**	6.518**	7.259***	6.124**	7.141**
	(1.916)	(1.934)	(2.110)	(2.050)	(1.933)	(2.149)
OBP	-6.143	-6.486	-6.505	-5.940	-5.351	-6.489
	(3.984)	(4.080)	(4.136)	(3.952)	(4.212)	(4.002)
AB	0.00249**	0.00251**	0.00258**	0.00269**	0.00218*	0.00265**
	(0.000808)	(0.000815)	(0.000845)	(0.000813)	(0.000959)	(0.000824)
SB	0.0204*	0.0209*	0.0200	0.0200*	0.0202*	0.0200
	(0.00997)	(0.0101)	(0.0101)	(0.00989)	(0.0100)	(0.00999)
MIF	0.0296	0.0441	0.0330	0.120	-0.00644	0.0358
	(0.146)	(0.150)	(0.148)	(0.159)	(0.158)	(0.146)
Hispanic	0.256	0.232	0.246	0.236	0.274	0.234
	(0.157)	(0.166)	(0.161)	(0.157)	(0.161)	(0.159)
Total Seasons	0.112	0.117	0.102	0.0955	0.102	0.0916
	(0.0689)	(0.0702)	(0.0743)	(0.0694)	(0.0715)	(0.0722)
Fielding Percentage		2.648				
		(5.576)				
DEF			0.00584			
			(0.0158)			
DRS				0.0245		
				(0.0179)		
Errors					0.0124	
					(0.0202)	
DWAR						0.136
						(0.141)
Constant	12.11***	9.575	12.14***	11.65***	11.98***	11.94***
	(1.109)	(5.458)	(1.121)	(1.150)	(1.136)	(1.124)
Adjusted R^2	0.494	0.486	0.485	0.503	0.487	0.493
N	57	57	57	57	57	57

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

b. Equation 2.2

The addition of the fielding percentage variable to the core equation did not yield significant results. The results of equation 2.2 indicate that fielding percentage does not have a statistically significant effect on the salaries of released free agents. Adding this variable to the equation also resulted in a decrease in the adjusted r-squared value relative to the core equation from .494 to .486. The variable also causes the coefficient estimate of the constant variable to lose statistical significance at the 5% level. That being said, it did little to alter the coefficient estimates of other significant variables in the equation.

c. Equation 2.3

Adding DEF to the equation also does not yield statistically significant results, indicating that changes in the DEF variable do not have a statistically significant effect on released free agent salary. The inclusion of this variable also causes a decrease in the adjusted r-squared value, from .494 to .485. This variable's inclusion causes SB to become insignificant at the 5% level. DEF also causes an increase in the coefficient estimate of SLG, however it does not alter its significance. That being said, the inclusion of this variable does little to alter the coefficient estimates of the other statistically significant variables in the equation.

d. Equation 2.4

The results indicate that changes in the DRS variable does not have a statistically significant on salaries for released free agent player. That being said, the inclusion of this variable does cause the adjusted r-squared value to increase relative to the core equation, from .494 to .503. Adding this variable to the equation causes a large increase in the coefficient estimate of SLG, from 6.20 to 7.26. However, it did little to alter the coefficient estimates or significance of the other variables.

e. *Equation 2.5*

The results indicate that changes in the Errors variable does not have a statistically significant effect on arbitration eligible free agent salaries. This variable is found to be significant at the 54.2% level, thus, at the 5% level of significance, I fail to reject the null hypothesis that Errors does not impact salary. The inclusion of this variable also caused the adjusted r-squared value to decrease from .494 to .487. However, the inclusion of this variable did little to alter the coefficient estimates of other significant variables.

f. *Equation 2.6*

In equation 2.6, DWAR is added to the salary determination model. The results indicate that the inclusion of this variable does not have a statistically significant impact on salaries for arbitration eligible free agent players. The inclusion of this variable caused SB to become insignificant at the 5% level. DWAR also caused the coefficient estimate of SLG to increase from 6.20 to 7.14. However, it had a negligible impact on the coefficient estimates of other significant variables relative to the core equation.

g. *Robustness Tests*

I conducted a series of tests to check the robustness of my salary estimation results and determine whether the significance of any of the variables in the model was the result of omitted variable bias. First, I will mention that the Black and CIF variable were included in the original specification of the model, but both variables were found to never be significant. The inclusion of each of these variables did not alter the significance of other variables. Second, I included an interaction variable between defensive statistics and both the Hispanic and MIF variables. The coefficient estimates for all of the interaction variables were not found to be significant, and the inclusion of these variables did not alter the significance of other variables.

The final tests I conducted aimed to examine how the coefficient estimates of variables in the core model behaved when the effects of both team and year were held constant. I introduced a team binary variable to hold team effects constant. These binary variables represented the team with which each player signed. The inclusion of these team binary variables caused the SB variable to become insignificant. Beyond changing the significance of this variable, the inclusion of team binary variables did not alter the coefficient estimates or significance of other variables in the model. These results suggest that the team a player signs with does not significantly impact the player's salary. In order to hold year effects constant, I introduced a binary variable that indicated the year each player's contract began. The inclusion of these variables also made the SB variable insignificant at the 5% level. However, beyond that change, these variables did not alter the coefficient estimates of other variables. This indicates that the results from my regression analysis, with the exception of the predicted value of the SB variable, the year each player's contract began did not alter the salary determination process.

h. *Summary of Results*

The observations in this sample poses an interesting puzzle. These players do not have enough experience to enter the free agent market conventionally, however, they are granted free agency after being released by their original teams. As such, their bargaining power increases as they are able to have all 30 MLB teams compete for their services. Ceretis paribus, this change in bargaining power should be associated with a higher salary than players with similar levels of experience, who are limited to salary determination through the free agent process. However, the data indicated that these released players are unable to enjoy the positive salary benefits associated with extra bargaining power. The mean salary for an arbitration eligible free agent in my dataset is less than half the mean salary for an arbitration eligible player who is limited to negotiated with his current team. This suggests that these released players were likely released for a reason, and that reason is poor performance. The release of a player is a sign that that player is not as skilled as a player that a team chooses not to release. For this reason, it is difficult to find a salary determination equation for these players, as demand for their services is low. Teams are able to acquire their services for low salaries because of low demand. As such, performance variables do not have a large impact on salary for these arbitration eligible free agents. It is likely for this reason that so few of the variables in my equations are significant. It is difficult to predict the salary of low quality players.

The results from these equations show that SLG, AB and SB are significant in determining salary. However, when DEF and DWAR are added, SB becomes insignificant at the 5% level, whereas the addition of fielding percentage, DRS and Errors did not cause this to happen. Since both DEF and DWAR include positional adjustments, while the other three defensive variables do not, it is likely that when significant, SB is explaining some of the positional variance in salary. None of the defensive variables in these equations are found to be significant. The variable that caused the largest increase in the adjusted r-squared value was DRS, but that increase was less than .01. From a theory standpoint, these results do not line up with intuition. In many cases, low quality players are signed by teams to specialize in a certain area, whether it be offensively or defensively related. The mean values of the four offensive variables in this sample are well below the mean values for offensive variables in the other two samples I examined, while the defensive statistics are comparable. This would suggest that if an arbitration eligible free agent player were to be signed by a team in order to specialize, it would likely be for their defensive ability. However, the regression results to not follow this theory.

There are a variety of reasons for why the results from this sample indicate weaker relationships between salary and player performance. First, the sample size is small, so there could be some bias in the sample. Second, players in this sample have received less playing time throughout their career than players in the other samples, so offensive efficiency measures, such as SLG and OBP, hold greater explanatory power than statistics that can be aggregated from season to season such as DEF, DRS, Errors and DWAR. Third and lastly, these salaries are on average close to the league minimum, thus performance metrics have little impact on general salary determination for these players. The third explanation is the most likely. Standard deviations for all of the variables in this sample are far smaller than in the other two samples, indicating that most players in the sample possess similar levels of ability. This means that there is not much to differentiate one from another in terms of ability. Therefore, salary is going to be less dependent on player performance, and more dependent on the market price for this type of player, which in this case is more in line with the league minimum salary, which is less than one standard deviation away from the mean salary for players in this sample.

VII. Arbitration Eligible Player Results

A. All Arbitration Eligible Players

1. Model

There are some key differences between this dataset and the free agent dataset. First, all of the players in this dataset had less than six years of MLB experience beginning in 2012. For this reason, there was no need to adjust the advanced defensive metrics with an adjusted experience variable, as all of the players in this sample entered the league after the 2003 season. The logarithmic transformation was once again used on average annual value of a player's contract to account for the average value of the player to his team when the contract was signed. The core

Table 5
Summary Statistics of Arbitration Eligible Players

	White	Hispanic	Black
	Mean (sd, max, min)	Mean (sd, max, min)	Mean (sd, max, min)
Salary (millions)	3.519	3.444	4.764
	(3.317,24.08,0.560)	(3.364,25,0.505)	(3.868,20,0.988)
SLG	0.403	0.389	0.407
	(0.0498,0.544,0.262)	(0.0545,0.540,0.308)	(0.0381,0.496,0.324)
OBP	0.322	0.318	0.330
	(0.0267,0.404,0.226)	(0.0219,0.364,0.266)	(0.0160,0.362,0.304)
AB	318.5	307.5	402.5
	(110.2,578.8,87.50)	(106.8,583,102.5)	(106.8,584,141)
SB	5.625	7.250	16.60
	(6.527,35,0)	(7.718,34,0)	(10.81,43.33,1)
MIF	0.417	0.613	0.0556
	(0.494,1,0)	(0.490,1,0)	(0.232,1,0)
Total Seasons	4.456	4.537	4.389
	(0.895,6,3)	(0.993,6,3)	(1.022,6,3)
Fielding Percentage	0.983	0.980	0.986
	(0.0121,0.998,0.935)	(0.0124,0.995,0.934)	(0.00654,0.997,0.964)
DEF	0.437	1.560	0.400
	(6.258,16.40,-21.10)	(4.180,13.50,-9.433)	(6.635,12.43,-12.73)
DRS	0.618	1.009	2.489
	(5.021,19.25,-13.67)	(4.096,15.20,-10.50)	(7.489,16.75,-8.333)
Errors	5.667	6.656	4.094
	(3.902,19.60,0.400)	(4.716,22,0.600)	(2.664,15.67,1)
DWAR	0.0607	0.241	0.159
	(0.691,2.183,-2.100)	(0.457,1.300,-1.033)	(0.864,2,-1.733)
N	204	80	36

equation used to examine this salary determination was the same as the core equation used in previous sections. This equation includes SLG, OBP, AB, SB, MIF, Latin American, and Total Seasons as the explanatory variables. In each subsequent regression the defensive variables of fielding percentage, DEF, DRS, Errors, and DWAR were added individually.

A key difference from the spread of summary statistics in this sample was the simple fact that this sample was not as representative of the demographic dispersion of the total Major Leagues. 25% of this sample identified as Hispanic, while estimates indicate that 29.4% of the MLB is Hispanic. 12% of this sample was of players that are Black, while 8.3% players in the MLB identify as Black. 63.75% of the players in this sample are considered to be White, while in 2015 58.8% of players in the MLB were White. This is a slight deviation from the other sample in terms of being representative of the MLB, but it is still within 5% for each race and ethnicity. In most cases, White players have lower mean values for the advanced defensive metrics, with the exception of DEF. Black players have a much larger mean value for stolen bases, which likely led to the African-American variable being insignificant in all regression results. In this sample, Black players have the highest mean salary, while Latin American players have the lowest. The average experience level for all of the players is comparable across race and ethnicity, with the largest difference in means being less than .2 seasons. This indicates that the spread of player experience in the sample is likely uniform across race and ethnicity.

2. Results

a. Equation 3.1

The results of Equation 3.1 indicate that SLG, OBP, AB, SB, MIF, and Total Seasons all have a statistically significant positive effect on salary. All else fixed, a 1% increase in slugging percentage would cause a 6.37% increase in salary. This result was found to be significant at all

Table 6 Arbitration Eligible Results						
	3.1	**3.2**	**3.3**	**3.4**	**3.5**	**3.6**
SLG	6.367*** (0.561)	6.389*** (0.562)	6.690*** (0.554)	6.408*** (0.566)	6.349*** (0.561)	6.658*** (0.563)
OBP	2.348* (1.017)	2.255* (1.029)	2.406* (0.995)	2.338* (1.018)	2.486* (1.026)	2.328* (1.005)
AB	0.00372*** (0.000219)	0.00374*** (0.000224)	0.00380*** (0.000216)	0.00372*** (0.000220)	0.00358*** (0.000259)	0.00369*** (0.000217)
SB	0.00788**	0.00787**	0.00500	0.00770**	0.00809**	0.00639*
	(0.00292)	(0.00292)	(0.00295)	(0.00294)	(0.00293)	(0.00293)
MIF	0.163***	0.164***	0.118*	0.166***	0.144**	0.127**
	(0.0465)	(0.0465)	(0.0469)	(0.0468)	(0.0501)	(0.0475)
Hispanic	0.0373	0.0418	0.0397	0.0370	0.0315	0.0345
	(0.0464)	(0.0469)	(0.0454)	(0.0464)	(0.0467)	(0.0458)
Total Seasons	0.278***	0.277***	0.271***	0.278***	0.281***	0.275***
	(0.0214)	(0.0215)	(0.0210)	(0.0214)	(0.0215)	(0.0212)
Fielding Percentage		1.110				
		(1.730)				
DEF			0.0146***			
			(0.00372)			
DRS				0.00217		
				(0.00391)		
Errors					0.00610	
					(0.00606)	
DWAR						0.0942**
						(0.0325)
Constant	8.894***	7.819***	8.779***	8.881***	8.863***	8.821***
	(0.285)	(1.699)	(0.280)	(0.287)	(0.287)	(0.283)
Adjusted R^2	0.808	0.808	0.816	0.808	0.808	0.812
N	320	320	320	320	320	320
Standard errors in parentheses * p<0.05, ** p < 0.01, *** p < 0.001						

reasonable levels of significance. The results for OBP would indicate that a 1% increase in OBP would cause a 2.35% increase in salary. This result was found to be significant at the 2.2% level of significance. The results for AB indicate that each additional 10 AB per year would cause a 3.72% increase in salary, all else fixed. This result was found to be significant at all reasonable levels of significance. The results for SB per year indicate that each additional stolen base per year causes a 0.79% increase in salary. This result was found to be significant at the 1% level of significance. The results for MIF indicate that, all else fixed, middle infielders are paid 43.3% more than players at other positions.[19] This result was significant at all reasonable levels of significance. The results for the Hispanic variable indicate that being of Hispanic ethnicity has no significant impact on salary. The results for Total Seasons indicates that, all else fixed, an additional season played causes a 27.8% increase in salary for arbitration eligible players. This result was significant at all reasonable levels of significance. This result is of the opposite sign as the results from the free agent sample. This is likely because players in this dataset have played less seasons than players in the free agent dataset, indicating that they are in earlier stages of their career. As such, their performance is not expected to be in decline.

As noted above, the drop in significance for the Hispanic variable was likely a result of the composition of the sample, as the mean salary for Hispanic players was the lowest among race and ethnicity groups, despite being comparable to the mean salary for White players. However, in the free agent sample, where the Hispanic variable is positive and significant, Hispanic players have the highest mean salary.

[19] Adjusted following the method put forth by Giles (2011)

b. Equation 3.2

In equation 3.2, the fielding percentage variable is added to the core equation. The results show that the fielding percentage variable is not significant at the 5% level of significance and as such does not have a statistically significant effect on arbitration eligible player salaries. The addition of fielding percentage to the core model did not alter largely alter any of the coefficients of significant variables.

c. Equation 3.3

The addition of the DEF variable caused some movements of the coefficients of significant variables from the core model. Most notably, including DEF caused stolen bases to become insignificant at the 5% level. The inclusion of this variable also caused a -27.6% change in the coefficient estimate for the MIF variable. However, this estimate is still within the 95% confidence interval of the estimate from the core equation. The results show that the DEF variable is significant at any reasonable level of significance, and indicating that, all else fixed, an increase in DEF by one would cause a 1.46% increase in player salary.

d. Equation 3.4

Adding DRS to the core model caused little movement in the coefficients of significant variables from the core equation. The stolen bases per year variable is once again significant at the 5% level. However, the results for DRS show that the variable was not significant at the 5% level of significance, indicating that changes in the DRS variable have no statistically significant effect on arbitration eligible player salaries.

e. Equation 3.5

The inclusion of the Errors variable did not prove to be a significant addition to the core model. The results show that the Errors variable is not significant at the 5% level, indicating that changes in errors per year does not have a statistically significant effect on salary. That being said, the inclusion of the Errors variable did not considerably alter the significance levels or coefficient estimates of any of the variables from the core model.

f. Equation 3.6

In equation 3.6, I included the defensive wins above replacement variable. Its inclusion increased the adjusted r-squared value relative to the core model from .808 to .813. DWAR was also found to be significant at the 0.4% level of significance, indicating that an increase in defensive wins above replacement per year by 1 is associated with a 9.42% increase in salary. Therefore, at the 5% level of significance, I would reject the null hypothesis that DWAR has no effect on salary. The inclusion of DWAR did not significantly alter any of the coefficient estimates of other variables relative to the core equation, nor did it alter the level of significance of other variables.

g. Summary of Results

The results of the arbitration eligible equations showed indicate that both the DEF and DWAR variable are significant at the 5% level. Both of these defensive variables are positionally adjusted, while the defensive variables that were not statistically significant were not significant at the 5% level in any of these equations. It is important to note that these results were for the entire arbitration eligible dataset. The salaries of players in this dataset were determined by a number of different processes, including: player and team negotiations prior to arbitration and the arbitration process. For players that signed contracts with their team prior to filing for arbitration, the process of negotiation between the team and the player was far shorter than for players that filed for arbitration. For this reason, in the following four sets of models, I segment the arbitration eligible dataset by salary determination process. Players that are arbitration eligible either avoid arbitration, file for arbitration and settle their salary prior to the arbitration hearing, or they file for arbitration and have their salary determined by an arbitration panel. Each outcome represents a different process, and while be explored in subsequent sections.

B. *Arbitration Eligible Players, Players Who Avoided Arbitration*

1. Model

As mentioned, arbitration eligible players have a variety of options for the manner in which their salaries can be determined. Players can choose to file for arbitration, or they can choose to re-sign with their previous team after negotiating a tender value or a new multi-year contracts.[20] Of the 320 players in the arbitration eligible dataset, 235 of them choose to not file for arbitration. Thus, their salaries were the product of informal negotiations with their team. From these negotiations, a value was settled upon before the arbitration file date, indicating a different salary determination process than the other 85 players in the dataset. The core model to examine salary determination for these players was the same as the one used for free agent players and all arbitration eligible players in previous section. This model included the variables SLG, OBP, AB, SB, MIF, Hispanic, and Total Seasons.

The summary statistics indicate that this sub-sample is fairly representative of the demographics of the Major Leagues. 63.4% of the players in this sample are White, 25.5% are Hispanic, and 11.1% are Black. Each is within 5% of the actual values of players within the Major Leagues. Black players have the highest mean salary in this sub-sample, and they also have the highest mean value for most performance statistics. Mean salary values are once again comparable for White and Hispanic players.

[20] Player tenders are associated with single year contracts

Table 7 Summary Statistics of Arbitration Eligible Players Who Avoided Arbitration			
	White	**Hispanic**	**Black**
	Mean (sd, max, min)	Mean (sd, max, min)	Mean (sd, max, min)
Salary (millions)	3.180	3.307	4.786
	(3.372 24.08 0.560)	(3.610 25 0.505)	(4.269 20 0.988)
SLG	0.396	0.386	0.410
	(0.0511 0.544 0.262)	(0.0540 0.540 0.308)	(0.0375 0.496 0.324)
OBP	0.319	0.317	0.328
	(0.0276 0.404 0.226)	(0.0230 0.364 0.266)	(0.0148 0.352 0.304)
AB	300.6	297.9	407.2
	(110.6 578.8 87.50)	(105.3 583 102.5)	(107.8 584 141)
SB	5.652	6.865	16.94
	(6.723 35 0)	(7.658 29.50 0)	(11.66 43.33 1)
MIF	0.450	0.600	0.0769
	(0.499 1 0)	(0.494 1 0)	(0.272 1 0)
Total Seasons	4.483	4.517	4.423
	(0.875 6 3)	(1.033 6 3)	(1.102 6 3)
Fielding Percentage	0.984	0.981	0.986
	(0.0116 0.998 0.935)	(0.0121 0.995 0.935)	(0.00707 0.997 0.964)
DEF	0.875	1.510	0.0614
	(5.659 16.40 -18.30)	(4.307 9.980 -9.433)	(6.404 12.43 -12.73)
DRS	0.666	0.853	2.315
	(4.889 15.83 -13.33)	(3.816 12.50 -10.50)	(7.318 16.67 -8.333)
Errors	5.181	6.396	4.224
	(3.775 19.60 0.400)	(4.426 21.25 1)	(3.013 15.67 1)
DWAR	0.0931	0.221	0.0893
	(0.654 2.183 -1.460)	(0.443 1.125 -1.033)	(0.896 2 -1.733)
N	149	60	26

2. Results

   a. *Equation 4.1*

The results of Model 1 indicate that SLG, AB, SB, MIF, and Total Seasons all have a positive effect on salary. All else fixed, a 1% increase in slugging percentage would cause a 6.74% increase in salary. This result was found to be significant at all reasonable levels of significance. The results for OBP were not significant at the 5% level, therefore, I would fail to reject the null hypothesis that OBP has no effect on salary. The results for AB would indicate that an extra ten AB per year would cause a 3.72% increase in salary, all else fixed. This result was found to be significant at all reasonable levels of significance. The results for SB per year indicate that a stolen base during an average causes a 0.94% increase in salary. This result was found to be significant at the 1% level of significance, making SB statistically significant at the 5% level. The results for MIF indicate that, all else fixed, middle infielders are paid 44.9% more than players at other positions.[21] This result was significant at all reasonable levels of significance. The results for the Hispanic variable were not found to be significant, indicating that being of Hispanic ethnicity has no significant impact on salary for players that avoided arbitration. The results for Total Seasons indicates that, all else fixed, an extra year of experience causes a 27.2% increase in salary for arbitration eligible players. This result was significant at all reasonable levels of significance.

[21] This value was adjusted using the Giles (2011) method.

Table 8 Arbitration Eligible Players Who Avoided Arbitration						
	4.1	**4.2**	**4.3**	**4.4**	**4.5**	**4.6**
SLG	6.742***	6.793***	7.269***	6.772***	6.711***	7.119***
	(0.668)	(0.672)	(0.658)	(0.681)	(0.670)	(0.677)
OBP	1.335	1.171	1.201	1.321	1.478	1.266
	(1.191)	(1.209)	(1.151)	(1.194)	(1.205)	(1.177)
AB	0.00372***	0.00377***	0.00384***	0.00372***	0.00359***	0.00370***
	(0.000267)	(0.000274)	(0.000260)	(0.000268)	(0.000317)	(0.000264)
SB	0.00943**	0.00951**	0.00625	0.00938**	0.00958**	0.00820*
	(0.00345)	(0.00346)	(0.00342)	(0.00347)	(0.00346)	(0.00345)
MIF	0.199***	0.204***	0.142*	0.201***	0.179**	0.166**
	(0.0563)	(0.0567)	(0.0561)	(0.0571)	(0.0620)	(0.0571)
Hispanic	0.0666	0.0724	0.0703	0.0666	0.0601	0.0637
	(0.0551)	(0.0556)	(0.0532)	(0.0552)	(0.0557)	(0.0544)
Total Seasons	0.272***	0.269***	0.261***	0.271***	0.274***	0.267***
	(0.0256)	(0.0258)	(0.0249)	(0.0258)	(0.0258)	(0.0254)
Fielding Percentage		1.748				
		(2.204)				
DEF			0.0200***			
			(0.00483)			
DRS				0.00121		
				(0.00495)		
Errors					0.00607	
					(0.00775)	
DWAR						0.104*
						(0.0407)
Constant	9.037***	7.342***	8.913***	9.030***	9.010***	8.950***
	(0.336)	(2.163)	(0.326)	(0.338)	(0.338)	(0.334)
Adjusted R^2	0.802	0.801	0.815	0.801	0.801	0.806
N	235	235	235	235	235	235

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The core model for this set of players found that both the Hispanic and OBP variables are not even marginally significant. This evidence supports the idea that the variables affecting salary determination for players differ based on the process by which the salary is determined.

b. Equation 4.2

In this equation, the fielding percentage variable is added to the salary determination equation. The results indicate that the fielding percentage variable is not statistically significant at the 5% level.[22] Therefore, I would fail to reject the null hypothesis that fielding percentage has no effect on player salary. That being said, despite it being an irrelevant variable, the inclusion of fielding percentage did not significantly alter the coefficient estimates of significant variables compared to the core model.

c. Equation 4.3

In this equation, the DEF variable is added to the salary determination equation. The results indicate that the DEF variable has a statistically significant positive effect on player salary. Recall that DEF is a measure of runs saved by defensive ability compared to a player of average ability. This statistic includes a positional adjustment, and its values in this sample range from positive to negative, with a value of zero indicating average defensive ability. The DEF variable is significant at all level of significance, indicating that an increase of DEF by one corresponds to a 2.00% increase in player salary. Therefore, I would reject the null hypothesis at the 5% level of significance that DEF has no impact on player salary. As expected, the inclusion of a relevant defensive variable decreased the magnitude of impact of the MIF variable. This DEF variable also made the previously statistically significant SB variable insignificant at 5% level. DEF did not significantly alter the coefficient estimates or the significance of any of the

[22] It was significant at the 42.8% level

other variables from the core model. These results indicate that DEF is a relevant variable in salary determination for arbitration eligible players that did not file for arbitration.

d. Equation 4.4

Adding DRS to the model caused a decline in the adjusted r-squared model relative to the core from, and the results indicate that the variable is not statistically significant. DRS does not appear to be a relevant variable in this salary determination equation. That being said, the inclusion of DRS did not alter the coefficient estimates or the significance of variables from the core model.

e. Equation 4.5

The errors variable was not found to be significant in this model at the 5% level. This is further evidenced by the fact that the inclusion of the errors variable in the model decreased the adjusted r-square value relative to the core model, suggesting that this variable does not appear to be relevant in this salary determination equation.

f. Equation 4.6

The inclusion of the DWAR variable was found to have a significant impact on the arbitration eligible salary determination model for players that avoided arbitration. This is shown by the fact that the adjusted r-squared value increased relative to the core model, while only slightly, this result is of note given that three of the other four defensive variables had a negative impact on the adjusted r-squared value. On top of that, DWAR was found to be significant at the 1.1% level of significance. Recall that DWAR is a measure of wins produced by a player's defensive ability compared to a player of average defensive ability. This statistic includes a positional adjustment. The results indicate that an increase in DWAR by one corresponds with a 10.4% increase in salary. From these two findings, I would conclude that DWAR is a relevant variable in the salary determination model for arbitration eligible players that avoided arbitration.

g. *Summary of Results*

The results of the salary determination equation of arbitration eligible players who did not file for arbitration indicate that the SLG, AB, SB and MIF variables have a positive statistically significant impact on salary. These findings held true across all the results of all six regressions run for this sample, which the exception of one equation for the SB variable. When the DEF variable is added to the core equation, the SB variable becomes insignificant at the 5% level. In terms of defensive statistics, the trend continues, with the results for DEF and DWAR indicating that both variables have a positive statistically significant on player salary for arbitration eligible players who did not file for arbitration. The other three defensive statistics examined did not yield significant results.

C. *Arbitration Eligible Players- Players Who Filed for Arbitration*

1. Model

As mentioned above, the arbitration eligible dataset can be further segmented between players that filed for arbitration and those that did not. There are 85 players in the dataset that filed for arbitration. This set of equations will examine their final salaries, the salaries that were agreed upon between the team and the player following the arbitration process, or, in the case of nine players in the sample, the salary chosen by the arbitration panel. The core equation used to examine salary determination was the same as the core equation used in previous sections. This equation includes SLG, OBP, AB, SB, MIF, Latin American, and Total Seasons as the explanatory variables in the core model. In each subsequent regression the defensive variables of fielding percentage, DEF, DRS, Errors, and DWAR are added.

Table 9
Summary Statistics of Arbitration Eligible Players Who Filed for Arbitration

	White	Hispanic	Black
	Mean (sd, max, min)	Mean (sd, max, min)	Mean (sd, max, min)
Salary (millions)	4.437	3.854	4.706
	(3.004,16.88,1.075)	(2.521,10,0.900)	(2.751,9.500,1.225)
SLG	0.420	0.398	0.399
	(0.0415,0.514,0.310)	(0.0564,0.504,0.324)	(0.0405,0.443,0.330)
OBP	0.330	0.322	0.335
	(0.0225,0.381,0.280)	(0.0183,0.356,0.286)	(0.0187,0.362,0.316)
AB	367.0	336.2	390.4
	(93.84,571.2,180.2)	(108.8,551.7,144)	(108.8,492,192)
SB	5.552	8.406	15.71
	(6.022,33.75,0)	(7.982,34,0.333)	(8.700,32,6.500)
MIF	0.327	0.650	0
	(0.474,1,0)	(0.489,1,0)	(0,0,0)
Total Seasons	4.382	4.600	4.300
	(0.952,6,3)	(0.883,6,3)	(0.823,6,3)
Fielding Percentage	0.981	0.977	0.986
	(0.0134,0.996,0.948)	(0.0128,0.994,0.934)	(0.00520,0.994,0.979)
DEF	-0.751	1.710	1.281
	(7.583,16.20,-21.10)	(3.875,13.50,-4.340)	(7.489,11.68,-8.700)
DRS	0.487	1.476	2.942
	(5.408,19.25,-13.67)	(4.924,15.20,-5.600)	(8.309,16.75,-7.333)
Errors	6.983	7.438	3.757
	(3.972,18,1.800)	(5.547,22,0.600)	(1.487,5.600,1.750)
DWAR	-0.0271	0.300	0.341
	(0.781,1.825,-2.100)	(0.504,1.300,-0.440)	(0.788,1.900,-0.550)
N	55	20	10

The summary statistics of this sample indicate that the sample was roughly representative of the demographic spread of players in the MLB. 64.7% of players in the sample were identified to be White, while 23.5% are Hispanic, and 11.7% is Black. Hispanic players are slightly underrepresented in this sample relative to the demographics of the MLB. In turn, White and Black players are slightly overrepresented relative to the same standard. Black players in this sample have the highest average salary, while also having the best performance statistics on average for a majority of the categories in Table 7.

2. Results

   a. *Equation 5.1*

The results of equation 5.1 indicate that SLG, OBP, AB, and Total Seasons all have a statistically significant positive effect on salary. All else fixed, a 1% increase in slugging percentage would cause a 5.13% increase in salary. This result was found to be significant at all reasonable levels of significance. The results for OBP would indicate that a 1% increase in OBP would cause a 4.75% increase in salary. This result was found to be significant at the 4.6% level of significance. For this reason, I would reject the null hypothesis that OBP has no effect on salary at the 5% level of significance. The results for AB would indicate that each additional 10 AB per year would cause a 3.37% increase in salary, all else fixed. This result was found to be significant at all reasonable levels of significance. The results for SB per year indicate this variable does not have a statistically significant impact on salary determination for the agreed upon salary of players that file for arbitration. The results for MIF indicate that the positional variable does not have a significant impact on player salary. The results for the Hispanic variable indicate that this variable does not have a statistically significant effect on salary. The results for Total Seasons indicates that, all else fixed, each additional year of experience causes a 30.3%

Table 10 Arbitration Eligible Players Who Filed for Arbitration						
	5.1	**5.2**	**5.3**	**5.4**	**5.5**	**5.6**
SLG	5.132***	5.085***	5.158***	5.083***	5.137***	5.212***
	(1.025)	(1.030)	(1.018)	(1.022)	(1.032)	(1.012)
OBP	4.752*	4.634*	5.134*	4.902*	4.784*	4.921*
	(1.976)	(1.990)	(1.981)	(1.973)	(2.004)	(1.951)
AB	0.00337***	0.00339***	0.00341***	0.00335***	0.00334***	0.00334***
	(0.000379)	(0.000382)	(0.000378)	(0.000378)	(0.000435)	(0.000374)
SB	0.00454	0.00408	0.00181	0.00269	0.00460	0.00167
	(0.00559)	(0.00565)	(0.00588)	(0.00578)	(0.00565)	(0.00575)
MIF	0.0843	0.0778	0.0549	0.0862	0.0815	0.0353
	(0.0783)	(0.0791)	(0.0805)	(0.0781)	(0.0818)	(0.0820)
Hispanic	-0.0409	-0.0291	-0.0370	-0.0449	-0.0417	-0.0411
	(0.0845)	(0.0865)	(0.0840)	(0.0843)	(0.0853)	(0.0833)
Total Seasons	0.303***	0.302***	0.299***	0.306***	0.303***	0.302***
	(0.0384)	(0.0385)	(0.0382)	(0.0384)	(0.0387)	(0.0379)
Fielding Percentage		1.870				
		(2.691)				
DEF			0.00773			
			(0.00542)			
DRS				0.00747		
				(0.00614)		
Errors					0.00118	
					(0.00916)	
DWAR						0.0901
						(0.0508)
Constant	8.773***	6.994**	8.670***	8.740***	8.761***	8.727***
	(0.554)	(2.620)	(0.555)	(0.553)	(0.565)	(0.547)
Adjusted R^2	0.804	0.803	0.807	0.805	0.801	0.809
N	85	85	85	85	85	85

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

increase in salary for arbitration eligible players. This result was significant at all reasonable levels of significance, as such I conclude that this variable is a significant determinant of salary.

b. Equation 5.2

The addition of the fielding percentage variable in equation 5.2 found that fielding percentage was not statistically significant when added to the regression equation. The adjusted r-squared decreased relative to the core model with the addition of this variable. Based on these results, I conclude that fielding percentage is not a significant determinant of salary for arbitration eligible players that file for arbitration. That being said, the inclusion of this variable did not alter the coefficient estimates of other significant variables in the model.

c. Equation 5.3

When added to the salary determination model, DEF was not found to be significant at the 5% level.[23] However, the inclusion of the DEF variable did cause the adjusted r-squared value to increase relative to the core equation. The inclusion of the DEF variable also caused the coefficient estimate of OBP to increase relative to the core equation estimate. This suggests that when defensive ability is controlled for with the DEF variable, the effects of the OBP variable increase. There was no large change in the coefficient estimations of other variables in the equation with the addition of the DEF variable. However, the fact that DEF is not significant indicates that it does not appear to be relevant explanatory variable in the equation for determining the settled salary of players that filed for arbitration.

d. Equation 5.4

The DRS variable was not found to be statistically significant at the 5% level. That being said, its inclusion did increase the adjusted r-squared value relative to the core equation from

[23] Found to be significant at the 15.7% level

.804 to .805. The inclusion of DRS increases in the coefficient estimate of OBP relative to the core model, however, this increase is not as large as the increase seen when DEF is included, as seen in equation 5.3. DRS did not cause a large change in the coefficient estimates of any other significant variables. Based on these results, I conclude that DRS is not a significant determinant of the settled salary for players that file for arbitration.

e. Equation 5.5

The errors variable is not significant at the 5% level when added to the core equation. This indicates that it does not appear to be a relevant variable for the salary determination of arbitration eligible players that file for arbitration. This is further evidenced by the fact that the adjusted r-squared value decreased relative to the core equation with the inclusion of the errors variable. That being said, the inclusion of the errors variable in the model did not cause a large change in any of the coefficient estimates of significant variables relative to the base model.

f. Equation 5.6

The inclusion of the DWAR variable did not yield significant results. DWAR was not found to be significant at the 5% level of significance. That being said, it was marginally significant, being found significant at the 8.0% level of significance. Also, it's inclusion did cause an increase in the adjusted r-squared value relative to the core model. DWAR causes the coefficient estimates for OBP to increase relative to the core equation estimates. The DWAR variable did not cause a large change in any of the other coefficient estimates for significant variables.

g. Summary of Results

The fact that none of the defensive variables in these equations were found to be statistically significant would appear to suggest that defensive ability is not factored into the salary decision for arbitration eligible players. However, this conclusion is misleading at face value, and other factors need to be considered. It would be incorrect to definitively conclude that defensive ability is not considered when determining the salary of players that file for arbitration. This is because the process of salary determination for players that file for arbitration and negotiate a final salary prior to their arbitration hearing represents a departure from what I will call "conventional formulaic salary determination."[24]

For many players in the sample, the settled salary amount was the result of a process of negotiation within a given range. The arbitration process dictates that a player and his team put forth a value for the player's services in the following year. The next step in the process is a hearing, where a panel choose one of the values as the player's salary. Of the 85 players in this sample, all of them filed for arbitration, which means that the player and the team put forward values for the player's services. However, it is not that case that one of those values was the player's eventual salary for the next season. Only, nine of the players in this sample had their salaries decided by an arbitration panel. This means that for those nine players, either the team value or the player value was chosen for the player's salary. Throughout this sample, there were

[24] Conventional formulaic salary determination occurs when the salaries are determined by a process of negotiation, where each party, the player and the team, is acting with limited knowledge of the valuation made by the other party. An example of this would be the process of free agency, where the anchor for negotiations is set by the team, which makes an offer to the players from which subsequent negotiations are based. The process of determination for players that do not file for arbitration follows the same process, with the exception that players have the ability to negotiate with all 32 teams in free agency, compared to only their original team when they are arbitration eligible. Salary negotiation after arbitration is not considered to fall under this category because two anchors are set, one by the team and an additional anchor by the player. While it is likely that other negotiations follow this process with counteroffers being made by the player. Arbitration filings have the added element that if the valuation made by one party is considered more suitable than the valuation made by the other, then the player will receive a salary equal to the valuation that an arbitration panel considers more suitable. In conventional formulaic salary determination there is no such penalty for an incorrect valuation, thus each party has less of an incentive to depart from their respective valuations.

76 players who filed for arbitration, but had their salaries determined through negotiations with their original team, as they signed a deal before the arbitration hearing. These salaries followed a very different pattern of determination.

There are 76 observations in this dataset that had salary determined by a process of negotiation following the arbitration filing. The final negotiated salary value, which acted as the dependent variable in these regressions, followed a three-step process of determination. Each step likely considered each of the variables in the model differently. This suggests that the final salary for these 76 observations were the result of three different salary determination methods. These methods are how a player determines his own value, how a team determines a player's value, and the method for determining the final negotiated value. This model examined the third process, and shows the factors that are considered for determining the negotiated value between the range of the initial team offer and player offer. 18 of the 76 players in the sample negotiated a salary that was equal to the midpoint between the player and team value. 36 of these 76 players received salaries above the midpoint value, and the remaining 31 received salaries below the midpoint. This spread of results for negotiated salaries after arbitration filings suggests that in all cases negotiations were within the range between the player offer and team offer from the arbitration filing. The result is not random, but in these cases the final negotiated amount is likely a departure from how players and teams make their initial salary determination. The factors that influence the outcome of these negotiations are shown in the results of the equations in this section. These factors are the basis for both parties, the player and team, to examine the likelihood that they will win in the arbitration hearing, and based on this evaluation how they should adjust their initial valuations. If the player is considered likely to win, he will receive a salary that is above the midpoint of the values from the filing. In contrast, if the team is likely to win, then the player will accept a salary that is below the midpoint. In all cases for these 76 observations the negotiated salary is equal to neither the team offer or the player offer. This finding suggests that players and teams are willing to depart from their initial evaluations based on the consideration of different factors.

With that being said, the conclusions from these results must be compounded in a different light. The results of these regressions suggest that the negotiate salaries are determined by the significant variables of SLG, OBP, AB, and total seasons. These variables represent the factors given consideration to depart from initial valuations by each party. They also represent factors that both the team and player consider to be key in determining the likelihood of success against an arbitration panel. These results suggest that defensive ability is not a significant factor of consideration during this negotiation process. However, as I previously stated, that does not indicate that defensive ability is not a significant determinant of player salary for arbitration eligible players. In order to make that determination, the anchor values for negotiation, the player offer and team offer, must be examined. This is done in section 5e and 5f of this paper.

D. Arbitration Eligible Players- Player Offer

1. Model

As mentioned above, the arbitration eligible dataset can be further segmented between players that filed for arbitration and those that did not. There were 85 players in the dataset that did not file for arbitration. This set of equations will examine player offers, the values put forth by the player at the exchange date prior to arbitration hearings. This means that the dependent variable does not represent a salary earned by the player, but rather a player's self-evaluation of his worth. The core equation used to examine this salary determination was the same as the core equation used in previous sections. This equation includes SLG, OBP, AB, SB, MIF, Latin

American, and Total Seasons as the explanatory variables in the core model. In each subsequent

Table 11
Summary Statistics of Arbitration Eligible Players Who Filed for Arbitration Salary Amounts

	White	Hispanic	Black
	Mean (sd, max, min)	Mean (sd, max, min)	Mean (sd, max, min)
Settled Salary (millions)	4.437	3.854	4.706
	(3.004,16.88,1.075)	(2.521,10,0.900)	(2.751,9.500,1.225)
Player Offer (millions)	4.672	3.837	4.698
	(2.597,11.80,1.425)	(2.126,7.500,1)	(2.813,10.80,1.600)
Team Offer (millions)	3.371	3.051	3.512
	(2.105,9,0.900)	(1.849,6.650,0.750)	(2.381,8.500,0.900)
N	55	20	10

regression the defensive variables of fielding percentage, DEF, DRS, Errors, and DWAR were added.

Table 9 includes only the summary statistics for the salary figures of players who filed for arbitration. This table includes the mean values of settled salary, which is the salary the player received as a result of the arbitration process, player salary offers, and team salary offers. The table does not include a summary of player performance statistics. The reason for the is because the summary statistics for the rest of the variables in the model are the same as the statistics in Table 7. This is because the observations in this sample are the same as the sample used in the previous section. The lone difference is that in this section the dependent variable will be the player's salary offer from the arbitration process. The trends of this variable indicate that on average, Black players put forth the highest self-evaluated value, while Hispanic players put forth the lowest.

Table 12
Arbitration Eligible Players Player Salary Offer

	6.1	6.2	6.3	6.4	6.5	6.6
SLG	4.525***	4.503***	4.560***	4.465***	4.537***	4.621***
	(0.767)	(0.772)	(0.736)	(0.753)	(0.771)	(0.732)
OBP	3.337*	3.281*	3.868**	3.518*	3.422*	3.541*
	(1.478)	(1.492)	(1.432)	(1.453)	(1.498)	(1.411)
AB	0.00280***	0.00281***	0.00285***	0.00278***	0.00273***	0.00277***
	(0.000284)	(0.000286)	(0.000273)	(0.000279)	(0.000325)	(0.000271)
SB	0.00385	0.00364	0.0000597	0.00162	0.00403	0.000403
	(0.00419)	(0.00424)	(0.00425)	(0.00426)	(0.00422)	(0.00416)
MIF	0.0941	0.0911	0.0532	0.0964	0.0866	0.0351
	(0.0586)	(0.0593)	(0.0582)	(0.0575)	(0.0612)	(0.0593)
Hispanic	-0.129*	-0.124	-0.124*	-0.134*	-0.131*	-0.129*
	(0.0632)	(0.0648)	(0.0607)	(0.0621)	(0.0637)	(0.0603)
Total Seasons	0.314***	0.314***	0.309***	0.319***	0.315***	0.313***
	(0.0287)	(0.0289)	(0.0277)	(0.0283)	(0.0290)	(0.0274)
Fielding Percentage		0.887				
		(2.018)				
DEF			0.0108**			
			(0.00392)			
DRS				0.00906*		
				(0.00452)		
Errors					0.00312	
					(0.00684)	
DWAR						0.108**
						(0.0367)
Constant	9.751***	8.907***	9.607***	9.711***	9.720***	9.696***
	(0.415)	(1.965)	(0.402)	(0.407)	(0.422)	(0.396)
Adjusted R^2	0.855	0.853	0.866	0.860	0.853	0.868
N	85	85	85	85	85	85

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2. Results

   a. *Equation 6.1*

The results of equation 6.1 indicate that the SLG, OBP, AB, Total Seasons, DEF, DRS, and DWAR variables all have a statistically significant positive effect on player salary offer. The results also show that the Hispanic variable has a statistically significant negative effect on player salary offer. All else fixed, a 1% increase in slugging percentage would cause a 4.52% increase in salary. This result was found to be significant at all reasonable levels of significance. The results for OBP would indicate that a 1% increase in OBP would cause a 3.34% increase in player salary offer. This result was found to be significant at the 2.7% level of significance, thus, at the 5% level of significance I reject the null hypothesis that OBP has no effect on player salary offer. The results for AB would indicate that an extra 10 AB per year would cause a 2.80% increase in player salary offer, all else fixed. This result was found to be significant at all reasonable levels of significance. The results for SB per year indicate this variable does not have a statistically significant impact on a player's salary offer for arbitration filings. The results for MIF indicate that the positional variable does not have a significant impact on a player's salary offer. The results for the Hispanic variable indicate that, all else fixed, a Hispanic player offers a salary that is 41.8% below other ethnicities. This result was significant at the 4.4% level of significance, thus at the 5% level, I reject the null hypothesis that ethnicity has no impact on a player's salary offer. The results for Total Seasons indicates that, all else fixed, each addition season a player plays causes a 31.4% increase in player salary offer for arbitration filings. This result was significant at all reasonable levels of significance, as such I conclude that this variable is a significant determinant of a player's evaluation of his own value. The adjusted r-squared value of this equation was .855.

b. Equation 6.2

Fielding percentage was not found to be a significant variable in this equation. The inclusion of this variable caused the adjusted r-squared value to decline relative to the core equation. These results indicate that fielding percentage is not a relevant variable in determining a player's salary offer for arbitration filings. That being said, the inclusion of this variable did not alter the coefficient estimates of significant variables relative to the core equation.

c. Equation 6.3

The addition of DEF to the core equation yielded statistically significant results. Recall that DEF is a measure of fielding runs saved, which includes a positional adjustment. The results showed that an increase in DEF by one corresponds with a 1.08% increase in player salary offer. This result was significant at the 0.8% level of significance. This means that at the 5% level, I reject the null hypothesis that DEF has no impact on salary. The inclusion of this variable also caused the adjusted r-squared value to increase relative to the core equation, from .855 to .866. The inclusion of this variable caused a relatively large increase in the estimated coefficient of OBP, from 3.34 to 3.87. As explained earlier, this is likely caused by a negative correlation between the two variables. Apart from OBP, DEF did little to alter the coefficient estimates of the other significant variables relative to the core equation.

d. Equation 6.4

DRS is statistically significant when added to this model. Recall that DRS is a measure of fielding runs saved, without a positional adjustment. The results show that an increase in DRS by corresponds with a 0.91% increase in player salary offer. This result was significant at the 4.9% level of significance. Therefore, at the 5% level of significance, I reject the null hypothesis that DRS has no impact on salary. The inclusion of this variable caused the r-squared value to increase relative to the core equation, from .855 to .860. The inclusion of this variable did little to alter the coefficient estimates of significant variables relative to the core equation. I conclude that DRS is a relevant variable in a player's determination of his salary offer for arbitration filings.

e. *Equation 6.5*

The Errors variable is not statistically significant in this equation. Its inclusion caused the adjusted r-squared value to decline relative to the core equation. As such, I conclude that the Errors variable is not relevant in a player's determination of his salary offer for arbitration filings. However, the inclusion of this variable did little to alter the coefficient estimates of significant variables relative to the core equation.

f. *Equation 6.6*

The inclusion of the DWAR variable yields significant results. Recall that DWAR is the number of wins a player generates with his defensive ability compared to a player of average ability, and the variable includes a positional adjustment. The results show that an increase in DWAR corresponds with a 10.8% increase in salary. This result is significant at the 0.8% level of significance, meaning that at the 5% level, I reject the null hypothesis that DWAR has no impact on salary. The inclusion of this variable did little to alter the coefficient estimates of significant variables relative to the core equation. It also caused the adjusted r-squared value to increase from .855 to .868. I conclude that the DWAR variable is relevant to a player's determination of his salary offer for arbitration filings.

g. *Summary of Results*

These equations represent one of the three methods of final salary determination for player's that file for arbitration. The results reflect how a player evaluates his own value and determines what salary offer to put forward in his arbitration filing. The dependent variable in these equations is unique, because the salary value is not the product of negotiations, but rather the product of a what a player finds to be a fair representation of his worth. That note comes with one caveat, as in making these evaluations, the player is required to be a rational actor. The reason for this is because if he salary offer is well above what is considered reasonable by an independent third party, then he is subject to receive a salary determined unilaterally by his team.

Across all equations, the SLG, OBP, AB, Hispanic, and Total Seasons variables were significant. The defensive variables of DEF, DRS and DWAR were also found to be significant when individually added to the core equation. The inclusion of DWAR yielded the largest increase in the adjusted r-squared value relative to the core equation, with DEF not far behind it. DRS caused an increase in the adjusted r-squared value as well, however, this increase was not as large in magnitude as the other variables. This is an interesting finding, because the DWAR and DEF variables include a positional adjustment, while DRS does not. This means the values of DWAR and DEF are adjusted based upon which position a player plays. In these equations, this positional adjustment increased the explanatory power of the defensive variable. This result suggests that the impact defensive ability has on a player's evaluation of his own value varies by position.

E. Arbitration Eligible Players-Team Offer

1. Model

As mentioned above, the arbitration eligible dataset can be further segmented between players that filed for arbitration and those that did not. The sample of players that filed for arbitration can includes three different dependent variables that must be examined: the final salary amount settled as a result of the process, the offer made by the player, and the offer made by the player. There were 85 players in the dataset that filed for arbitration. This set of equations will examine team offers, the values put forth by the team at the exchange date prior to arbitration hearings. This means that the dependent variable does not represent a salary earned by the player, but rather a team's evaluation of the player's worth. The core equation used to examine this salary determination was the same as the core equation used in previous sections. This equation includes SLG, OBP, AB, SB, MIF, Latin American, and Total Seasons as the explanatory variables. In each subsequent regression the defensive variables of fielding percentage, DEF, DRS, Errors, and DWAR were added individually.

Table 9 includes the summary statistics for salaries in this model. Table 7 includes the summary of performance statistics of players in the model. A new table was not necessary here, because the observations in this sample are the same as the sample used in the previous section, with the exception of the dependent variable of team salary offer, which is included in Table 9. This section's dependent variable is the team's salary offer from the arbitration process. The trends of this variable indicate that on average, Black players put receive the highest offer value from their teams, while Hispanic players receive the lowest.

2. Results

a. Equation 7.1

The results of equation 7.1 indicate that the SLG, OBP, AB, Total Seasons, DEF, and DWAR variables all have a statistically significant positive effect on team salary offer. All else fixed, a 1% increase in slugging percentage would cause a 4.79% increase in team salary offer. This result was found to be significant at all reasonable levels of significance. The results for OBP would indicate that a 1% increase in OBP would cause a 4.14% increase in team salary offer. This result was found to be significant at the 0.8% level of significance, thus, at the 5%

Table 13
Arbitration Eligible Players Team Salary Offer

	7.1	7.2	7.3	7.4	7.5	7.6
SLG	4.787^{***}	4.768^{***}	4.817^{***}	4.747^{***}	4.789^{***}	4.861^{***}
	(0.791)	(0.797)	(0.772)	(0.788)	(0.797)	(0.774)
OBP	4.135^{**}	4.087^{**}	4.583^{**}	4.257^{**}	4.148^{**}	4.291^{**}
	(1.525)	(1.540)	(1.502)	(1.522)	(1.548)	(1.493)
AB	0.00346^{***}	0.00347^{***}	0.00350^{***}	0.00344^{***}	0.00345^{***}	0.00343^{***}
	(0.000293)	(0.000295)	(0.000286)	(0.000292)	(0.000336)	(0.000286)
SB	0.00391	0.00373	0.000712	0.00240	0.00394	0.00126
	(0.00432)	(0.00437)	(0.00446)	(0.00446)	(0.00437)	(0.00440)
MIF	0.106	0.103	0.0712	0.107	0.105	0.0602
	(0.0605)	(0.0612)	(0.0610)	(0.0602)	(0.0632)	(0.0628)
Hispanic	-0.0104	-0.00560	-0.00581	-0.0136	-0.0107	-0.0105
	(0.0652)	(0.0669)	(0.0637)	(0.0650)	(0.0658)	(0.0638)
Total Seasons	0.342^{***}	0.342^{***}	0.338^{***}	0.346^{***}	0.343^{***}	0.342^{***}
	(0.0296)	(0.0298)	(0.0290)	(0.0296)	(0.0299)	(0.0290)
Fielding Percentage		0.753				
		(2.083)				
DEF			0.00907^{*}			
			(0.00411)			
DRS				0.00612		
				(0.00474)		
Errors					0.000472	
					(0.00707)	
DWAR						0.0835^{*}
						(0.0389)
Constant	8.644^{***}	7.928^{***}	8.524^{***}	8.618^{***}	8.640^{***}	8.602^{***}
	(0.428)	(2.028)	(0.421)	(0.427)	(0.436)	(0.419)
Adjusted R^2	0.876	0.875	0.882	0.877	0.874	0.882
N	85	85	85	85	85	85

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

level of significance I reject the null hypothesis that OBP has no effect on player salary offer. The results for AB would indicate that each additional 10 AB per year would cause a 3.46% increase in team salary offer, all else fixed. This result was found to be significant at all reasonable levels of significance. The results for SB per year indicate this variable does not have a statistically significant impact on a team's salary offer for arbitration filings. The results for MIF indicate that the positional variable does not have a significant impact on a team's salary offer. The results for the Hispanic variable are not significant at the 5% level, indicating that ethnicity does not have a significant effect on a team's salary offer for arbitration filings. The results for Total Seasons indicates that, all else fixed, each additional season played causes a 34.2% increase in a team's salary offer for arbitration filings. This result was significant at all reasonable levels of significance, as such I conclude that this variable is a significant determinant of a player's evaluation of his own value. The adjusted r-squared value of this equation was .876.

b. Equation 7.2

The fielding percentage variable is not significant in this equation, registering a result that was significant at the 71.9% level when added to the core equation. At the 5% level of significance, I fail to reject the null hypothesis that fielding percentage has no impact on team salary offer. The inclusion of this variable caused the adjusted r-squared value to decrease relative to the core equation. These findings suggest that fielding percentage does not appear to be a relevant variable that teams consider when making their salary offer for arbitration filings. The addition of this variable did cause a large decrease in the coefficient estimate of the Constant relative to the core equation. That being said, including fielding percentage did little to alter the coefficient estimates of other significant variables in the equation.

c. Equation 7.3

The inclusion of the DEF variable yielded significant results. Recall that DEF is a measure of runs saved by a player's defensive ability compared to an average player, including a positional adjustment. The results show that an in DEF by 1 causes an increase in salary by

0.91%. DEF was found to be significant at the 3.0% level of significance, thus, I reject the null hypothesis that DEF does not have an impact on a team's salary offer for arbitration filings. The inclusion of this variable also caused the adjusted r-squared value to increase relative to the core equation from .876 to .882. This finding indicates that the DEF variable is a relevant determinant of a team's salary offer. Adding DEF to the equation causes the coefficient estimate of OBP to increase by .448. That being said, the inclusion of this variable did little to alter the coefficient estimates of other significant variables.

d. Equation 7.4

Including DRS in the salary determination equation did not yield significant results. DRS is significant at the 20.0% level of significance. Therefore, at the 5% level, I fail to reject the null hypothesis that DRS has no impact on team salary offer. However, the inclusion of this variable did cause a small increase in the adjusted r-squared value relative to the core equation. DRS also did little to alter the coefficient estimates of statistically significant variables relative to the core equation. I conclude that DRS is not a relevant variable in a team's salary offer determination for arbitration filings.

e. Equation 7.5

The errors variable is not statistically significant in determining the team salary offer, as it was found to be significant at the 94.7% level of significance. Thus, at the 5% level, I would fail to reject that the errors variable has an impact on this salary determination. The inclusion of this variable also caused the adjusted r-squared value to decrease, indicating the addition of an irrelevant variable. That being said, the errors variable did little to alter the coefficient estimates of statistically significant variables relative to the core equation.

f. Equation 7.6

In this equation DWAR is added to the core equation. The addition of this variable yielded significant results, as DWAR is significant at the 3.5% level of significance. Thus, at the 5% level, I reject the null hypothesis that DWAR has no impact on team salary offer. Recall that DWAR is measure of how many wins a player generates with his defensive ability compared to a player of average ability. This statistic includes a positional adjustment. The results show that an increase in DWAR by one causes an 8.35% increase in team salary offer. The inclusion of this variable did little to alter the coefficient estimates of statistically significant variables relative to the core equation. It also caused the adjusted r-squared value to increase from .876 to .882. Therefore, I conclude that DWAR is a relevant variable used by teams to determine their salary offer for the arbitration filing process.

g. Summary of Results

These equations examine the factors teams consider when making their decision on how much to offer a player at the time of arbitration filings. As with the player salary offer variable, the dependent variable in these equations is unique. It is a salary value that is not the result of a negotiation process. Teams decide how much to offer players that file for arbitration unilaterally, and the value they put forward serves as an anchor for future negotiations. That being said, teams are forced to be rational actors when deciding how much to offer. The reason for this is because if a team does not offer a value that is perceived to be a reasonable value for the player's services by a third party, then the player's salary will be determined solely by the player. The arbitration process dictates that if a team does not give a player a fair offer, then the player will receive a salary equal to his offer, and the team will be unable to impact how much it must pay. The team must make an offer that is competitive to avoid this scenario.

The results of these equations show that when making this decision, teams consider the variables of SLG, OBP, AB, and Total Seasons. Beyond that, they also consider the defensive variables of DEF and DWAR when making their salary offer determination. The consistent difference between the statistically significant defensive variables and the ones that are not significant is once again the positional adjustment. Both DEF and DWAR includes a positional adjustment, altering their calculation based upon what position a player plays. Both of these variables caused an equal increase in the adjusted r-squared value when added to the equation.

F. Robustness Tests

I conducted a variety of tests to check for the robustness of the results in this sample. The first was to include an interaction variable which interacted both the positional and defensive variables. The results of the inclusion of this variable were not significant and did little to alter the significance of other coefficient estimates in the model. In short, I conclude that players within this level of bargaining power are not compensated differently for their defensive ability based on the position they play. The same result held true when introducing an interaction variable between race/ethnicity and the defensive metrics. This finding indicates that arbitration eligible players are not compensated differently for their defensive ability based upon their race or ethnicity.

The next set of robustness tests centered around holding both year and team effects fixed. In order to hold team effects fixed, I introduced a binary variable for 29 of the 30 MLB teams into the model.[25] The results from the inclusion of these variables found that none of the coefficient estimates for these variables were significant, and they did little to alter the coefficient estimates of other variables in the model. These results indicate that teams do not

[25] The 30th team was the reference group

compensate arbitration eligible players in a statistically significant manner. These results held across the five different models used to analyze salaries for arbitration eligible players.

The final set of robustness tests I ran centered around fixing the effects of the different years that players signed their contracts. In order to do this, I introduced a binary variable that indicated the year a player's contract began. For example, if the players contract began in 2012, then the Year 2012 variable was equal to one. If the contract began in a different year, then this variable was equal to zero.[26] I included a year binary variable for the 2012, 2013, and 2014 seasons in order to hold fixed different conditions that could change on a yearly basis, such as MLB popularity and revenues. The results of the inclusion of these variables found that these variables did little to alter the coefficient estimates of other significant variables in the core model. The being said, these binary year variables were found to be significant at the 5% level of significance. This finding indicates that players are compensated differently by year. This finding led me to conduct an additional Chow test in order to see if there was a significant break in the data from year to year. The results of this chow test found that within this level of bargaining power, there is not a significant difference in the salary determination process. This means that the significance of these yearly binary variables are likely due to the omission of a significant variable, likely either MLB revenues, or GDP. The important finding from this test was that when holding the effects of year fixed, the coefficient estimates of the core model did not differ significantly.

VIII. Defensive Statistics Discussion

The results across the seven sets of equations indicate that defensive ability's importance in the salary determination model differs across different markets within Major League Baseball.

[26] 2015 was the reference group

In the free agent market, the DEF, DRS, errors, and DWAR variables are all statistically significant, indicating that each statistic effects salary. The DEF, DRS, and DWAR variables have a positive statistically significant impact on salary, while the errors variable has a statistically significant negative impact. The results for the career fielding percentage variable indicate that it does not have a statistically significant effect of free agent salary. The primary difference in the results from the free agent market and the arbitration eligible market is the significance of the Errors variable, which is not significant is any of the arbitration eligible player equations. The results indicate that the DRS variable is marginally significant in arbitration eligible player equations, including being significant in the model for estimating the player salary offer of players that file for arbitration. However, both the DWAR and DEF variables are statistically significant in all arbitration player equations, except for the salary determination model for the settled salary of players that file for arbitration. In that model, none of the five defensive statistics I examine are statistically significant. The same is true of the salary determination model for released free agents. From these results, a number of conclusions concerning defensive ability in the salary determination model emerge.

First, positionally-adjusted defensive statistics are significant in all markets where negotiations between the team and the player are not constrained. I will explain this conclusion in two steps, beginning with the conditions under which salary negotiations are constrained. I mentioned that there were two sets of equations where DWAR and DEF were not statistically significant. These are in the released free agent market and in the model for determining the settled salary for players that file for arbitration. Each of these sets of equations estimate a salary determined under a constrained negotiation process. By constrained negotiation process, I am referring to a process where the explanatory power of player performance variables is reduced because the negotiations are anchored around a value that determined by a separate process. In the case of released free agents, the anchor value is the league minimum salary for veteran players. I previously mentioned that the mean salary for released free agent players is less than one standard deviation away from the league minimum. This suggests that players within this market receive salaries that do not largely vary from this league minimum salary. The league minimum salary is determined by the league's collective bargaining agreement, and is not dependent on player performance. Since many players in this sample are receiving salaries that are at or near this league minimum, the explanatory power of their performance statistics is reduced. This is seen by the fact that only two of the seven explanatory variables in the core model are statistically significant across all equations. By contrast, in all other markets at least four variables are significant across all equations. This finding is paired with the fact that the adjusted r-squared values for equations in this market are the lowest of all markets examined. This suggests that either there is a problem with the model's specification, or there is little correlation between performance statistics and player salaries in this sample.

The other set of equations where defensive statistics were not statistically significant was the salary determination equations for the settled salary amount of players that filed for arbitration. Again, this salary value is determined by a constrained negotiation process. All of the players in this sample filed for arbitration and exchanged salary figures with their team. These exchanged salary figures represent the upper and lower bounds of the negotiation for the settled salary amount. For all salaries in this sample that were not determined by an arbitration panel, the settled salary amount was greater than the team offer and less than the player offer. This suggests that these values were used in every case as the starting values for the negotiation process. Defensive statistics are significant in determining each of these starting values, as shown by the results of equations 6.1-6.7 and 7.1-7.7. However, the results of the equations that predict the results of these constrained negotiations show that defensive ability does not have a significant impact on the outcome of these salary negotiations. As I previously mentioned, this does not indicate that defensive ability is not determinant of salary for these players that filed for arbitration. Instead, these results indicate that defensive ability is not a significant determinant of salary for the final step of the negotiation process. Both the player and the team consider defensive variables when making their initial offers, but not when negotiating away from these offers. The results of equations 5.1-5.7 indicate that offensive ability is the primary determinant of those negotiation outcomes. This process represents a constrained negotiation process.

In all other markets, positionally-adjusted defensive variables are statistically significant in determining player salary. This indicates that the defensive statistics that serve as the best predictors of player salaries are the statistics that have their calculations adjusted based upon the position a player plays. This means that a position, such as shortstop, that has a high volume of opportunities to make defensive plays relative to other positions, will not have a higher statistical value due to these extra opportunities. DWAR and DEF adjust based on a player's defensive ability relative to other players at his position. This is important because it means these values can be compared across positions. The DWAR statistic of a second baseman is comparable to the DWAR statistic of a left fielder because each statistic represents the two player's defensive ability relative to the rest of the league. DWAR and DEF place an all else fixed value on each player's ability to play defense. This fact is the reason that it makes sense that these two variables are the defensive measures that are statistically significant across both the free agent and arbitration markets. From this finding, I conclude that defensive ability is a significant determinant of player salary in both markets.

Another conclusion that can be made from the regression results regards the use of both the errors and fielding percentage variables in the salary determination model. Across all of the markets examined, fielding percentage is not significantly significant in the salary determination model. Errors, on the other hand, is only statistically significant in one of seven equations. This finding is of note considering the importance that is placed on these variables by the MLB. Major League Baseball keeps a variety of records and statistics during a league baseball game, among those are: errors, assists, and putouts. All three of these statistics are defined and explained in the official MLB rulebook. These three statistics also serve as the inputs for the calculation of fielding percentage. As such, fielding percentage values have been calculated throughout the history of the MLB, dating back to 1871. The historical presence of this statistic, along with the errors statistic, likely accounts for its emphasis. Fielding percentage is the only defensive efficiency statistic that can be calculated from official MLB data. As a result, fielding percentage is assumed to be telling of a player's defensive ability. Until the advent of advanced defensive statistics in the 1980s, fielding percentage was the primary measure of defensive ability, representing the frequency at which a player successfully completes a defensive play per attempt. Errors also acted as an important statistic until advanced measures were introduced, as they indicated the number of unsuccessful attempted defensive plays a player made. However, the significance of these variables when included in my salary determination equation suggest that when determining player salary, neither statistic is given careful consideration. This finding suggests that either fielding percentage and errors make a poor proxy for defensive ability, or they represent a component of defensive ability that is not considered valuable to teams. In my opinion, both are correct.

Given the significance of the DEF and DWAR variables in the salary determination equations across all levels of bargaining power, it is difficult to look at my results and suggest that defensive ability does factor into the determination of a player's salary. On a consistent basis, these variables had a statistically significant positive effect on salary. DRS was not as consistent, but it lacked a positional adjustment, as I have previously discussed. It is important to note that each defensive statistic examined in my equations were serving as a proxy for defensive ability. Each variable was included to explain variance in salary that was not explained by variables in the core model. Since the core model did not include any defensive statistics, the variance explained by the inclusion of the five defensive statistics I examined was hypothesized to be defensive ability. However, including the fielding percentage and errors variables, in most cases, did not increase the explanatory power of the equation. With the exception of errors in the free agent salary determination equation, including fielding percentage and errors caused the adjusted r-squared value for each equation to either decrease or remain the same relative to the core equation. This suggests that neither of these two variables were explaining variance in salary, despite the fact that the results of other defensive variables indicated that some of this variance is caused by defensive ability. This suggests that both the fielding percentage and errors variables were a poor proxy for defensive ability. This could be for a variety of reasons, but in my opinion, it is related to bias created from the calculation method of both variables.

Fielding percentage is calculated by dividing putouts and assists by putouts, assists, and errors. This suggests that the variable is calculated by dividing number of successfully completed plays by number of successfully completed plays plus errors. This can cause bias on two fronts. First, number of successfully completed plays is a function of how many opportunities a player has to complete a play. The number of opportunities a player has to complete a play is a function of player position. Players playing in the infield have more opportunities to make defensive plays than players playing in the outfield. Thus, the calculation of this statistic is not equal. An outfielder who commits one error is likely to have a lower fielding percentage than an infielder who commits one error, because that infielder will have more opportunities obtain putouts and assists due to his position. This positional bias aside, the determination of the error statistic also leaves room for bias. In the MLB official rulebook, a player is charged with an error when:

> [His] action has assisted the team on offense, as set forth in this Rule 10.12.The official scorer shall charge an error against any fielder: (1) whose misplay (fumble, muff or wild throw) prolongs the time at bat of a batter, prolongs the presence on the bases of a runner or permits a runner to advance one or more bases, unless, in the judgment of the official scorer…If a ground ball goes through a fielder's legs or a fly ball falls untouched and, in the scorer's judgment, the fielder could have handled the ball with ordinary effort, the official scorer shall charge such fielder with an error… For example, the official scorer shall charge an infielder with an error when a ground ball passes to either side of such infielder if, in the official scorer's judgment, a fielder at that position making ordinary effort would have fielded such ground ball and retired a runner. The official scorer shall charge an outfielder with an error if such outfielder allows a fly ball to drop to the ground if, in the official scorer's judgment, an outfielder at that position making ordinary effort would have caught such fly ball. If a throw is low, wide or high, or strikes the ground, and a runner reaches base who otherwise would have been put out by such throw, the official scorer shall charge the player making the throw with an error (MLB Official Rulebook: Rule 10.12).

The above section is a sample of Rule 10.12 in the MLB Official Rulebook. This rule outlines the criteria an official MLB scorer uses when determining if an unsuccessful defensive play is considered an error. The determination of whether or not a play is an error is made by a league appointed official scorer. The sample of Rule 10.12 replicated above contains four instances where the rule explicitly states that the official scorer must use their own judgment in determining whether a player is charged with an error. This suggests that the determination of the error statistic is subjective, depending on the opinion of a league appointed scorer. The rule attempts to create a formulaic statistic, as Rule 10.12, in its entirety, outlines the criteria for scoring nearly any play imaginable. However, errors are still dependent on judgment, leaving room for bias. It is for these reasons that both errors and fielding percentage serve as poor proxies for defensive ability. Both statistical measures, despite being among the only defensive data collected by Major League Baseball, contain too many sources of bias to used to judge a player's ability.

It is unlikely that my previous conclusions concerning both the fielding percentage and errors statistics are groundbreaking. Previous literature has examined this issue in further detail (see Kalist et. al, 2006), with results from a macro analysis of official scorer decisions indicating that official scorers have a bias towards the home team. That being said, these findings are important to note when considering the overall conclusions from my regression analysis concerning defensive statistics. This conclusion is that, broadly speaking, defensive ability is a significant determinant of player salary in Major League Baseball. This is shown by the results of the equations when the positionally adjusted variables of DWAR and DEF are added to the core equation. For each of these equations, with the exception of the two markets that include constrained negotiation processes, the addition of each of these two variables increases the adjusted r-squared value, along with yielding positive statistically significant coefficient estimates. These results indicate that, all else fixed, defensive ability has a positive impact on player salary, regardless of level of player bargaining power.

IX. Conclusions

From this research, I reached a variety of conclusions regarding the salary determination process in the MLB. First, when examining player salary, one must first examine player contracts. This is especially important when matching player performance data to be used to create a salary determination model. Players can engage in multi-year contracts, which means they will negotiate their salaries for multiple seasons at once. When analyzing the salary observations of players that sign a multi-year contract, it is important to only consider a player's performance prior to when the contract was signed. This means that if a player signs a two-year contract in 2012, his batting average during the 2012 season is not a determinant of his 2013 salary. These adjustments were made during my data collection process, and from this process I was able to collect data correctly that would yield statistically significant results.

Second, official MLB defensive statistics are not complete indicators of a player's defensive ability. This conclusion was explained in detail in the previous section, but it is worth mentioning again. The results of my regression analysis indicate that defensive ability is a significant determinant of player salary across all markets and levels of bargaining power. This was shown by the consistent statistical significance of both the DEF and DWAR variables in five of my seven sets of regression. In all seven sets of regressions, the fielding percentage variable was not statistically significant. In six of my seven sets of regressions, the errors variable was not statistically significant. In a model where the results indicate that defensive ability is an important determinant of player salary, both the fielding percentage and errors variables were insignificant.

Third and lastly, my hypothesized effect of defensive variables in the salary determination model held true. The positionally adjusted defensive variables in my model had a positive statistically significant effect on player salary for all cases where a conventional negotiation took place. I mentioned previously that during my literature review I examined 24 different salary determination models. Only eight of those models included a variable that is a measure of defensive performance. Of those eight models, four of them used fielding percentage as the defensive statistic of interest. In only one of the four models that used a fielding percentage variable were the results significant. These results suggest the defensive ability has largely been ignored during the construction of salary determination models. My results indicate that defensive ability cannot be ignored, in fact, failure to include a defensive variable in the salary determination model could result in omitted variable bias.

References

1. Banaian, King, and Orn Bodvarsson. 1998. "The Value of Arbitration Rights in Major League Baseball: Implications for Salaries and Discrimination." *Quarterly Journal of Business and Economics* 37 (1): 65-80.
2. Barilla, Anthony. 2002. "An Analysis of Wage Differences in Major League Baseball, 1985-1995." Dissertation, Department of Economics, Kansas State University, Manhattan.
3. Baseball Reference. 2015. *Player Stats.* Accessed December 2015. http://www.baseball-reference.com/players/.
4. Bradbury, John Charles. 2013. "What Is Right With Scully Estimates of a Player's Marginal Revenue Product." *Journal of Sports Economics* 14 (1): 87-96.
5. Brown, Daniel T. 2008. "Final Offer Arbitration in Major League Baseball: An Empirical Analysis of Bargaining." Dissertation, University of Delaware, Dover.
6. Fangraphs. 2015. *Defense.* Janurary. Accessed December 2015. http://www.fangraphs.com/library/defense/.
7. Giles, D. E. . 2011. "Interpreting dummy variables in semi-logarithmic regression models: exact distributional results." Ecnometrics Working Paper, Department of Economics, University of Victoria.
8. Gius , Mark P. , and Timothy R. Hylan. 1999. "Testing For The Effect Of Arbitration On The Salaries Of Hitters In Major League Baseball: Evidence From Panel Data." *Pennsylvania Economic Review 7.1* 20-35.
9. Haupert, Michael J. . 2007 . *The Economic History of Major League Baseball.* Edited by Robert Whaples. December 3. Accessed December 1, 2015 . https://eh.net/encyclopedia/the-economic-history-of-major-league-baseball/.
10. Holmes, Paul. 2011. "New Evidence Of Salary Discrimination In Major League Baseball." *Labour Economics* 18 (3): 320-331.
11. Horowitz, , Ira, and Christopher Zappe. 1998. "Thanks For The Memories: Baseball Veterans' End-Of-Career Salaries." *Managerial and Decision Economics* 377-382.
12. Kalist , David, and Stephen Spurr. 2006. "Baseball Errors." *Journal of Quantitative Analysis in Sports* 2 (4).
13. Lahman, Sean. 2015. *Download Lahman's Baseball Database.* Accessed January 2016. http://www.seanlahman.com/baseball-archive/statistics/.
14. Link, Charles, and Martin Yosifov. 2012. "Contract Length And Salaries Compensating Wage Differentials In Major League Baseball." *Journal of Sports Economics* 3-19.
15. MacDonald, Don N., and Morgan O. Reynolds. 1994. "Are Baseball Players Paid Their Marginal Products?" *Managerial and Decision Economics* 15 (5): 443-457.
16. Major League Baseball. 2011. *2012-2016 Basic Agreement.* Accessed January 2016. http://mlb.mlb.com/pa/pdf/cba_english.pdf.
17. —. 2016. *Official Baseball Rules-2016 Edition.* Accessed March 2016. http://mlb.mlb.com/mlb/downloads/y2016/official_baseball_rules.pdf.
18. Marburger, Daniel R. 1994. "Bargaining Power and the Structure of Salaries in Major League Baseball." *Managerial and Decision Economics* 15 (5): 433-441.
19. Marburger, Daniel R. . 2004. "Arbitrator Compromise in Final Offer Arbitration: Evidence from Major League Baseball." *Economic Inquiry* 42 (1): 60-68.

20. Marburger, Daniel R. . 1996. "Racial Discrimination And Long-Term Contracts In Major League Baseball." *Review of Black Political Economy* 25 (1): 83-94.
21. Miller, Phillip A. . 2000. "A Theoretical And Empirical Comparison Of Free Agent And Arbitration-Eligible Salaries Negotiated In Major League Baseball." *Southern Economic Journal* 67 (1): 87-104.
22. Palmer, Matthew C. , and Randall H. King. 2006. "Has Salary Discrimination Really Disappeared From Major League Baseball?" *Eastern Economic Journal* 32 (2): 285-297.
23. Pantuosco, Louis J., and Gary Stone. 2010. "Babe Ruth As A Free Agent: What The Old-Time Greats Would Earn In Today's Labor Market For Baseball Players." *American Economist* 55 (2): 154-161.
24. Pedace, Roberto, and Curtis Hall. 2012. "Home Safe: No-Trade Clauses And Player Salaries In Major League Baseball." *Industrial Relations* 51 (3): 627-644.
25. Rockerbie, Duane W. . 2009. "Strategic Free Agency in Baseball." *Journal of Sports Economics* 10 (3): 278-291.
26. Scully, Gerald W. . 1974. "Pay And Performance in Major League Baseball." *The American Economic Review* 64 (6): 915-930.
27. Scully, Gerald W. . 2004. "Player Salary Share And The Distribution Of Player Earnings." *Managerial and Decision Economics* 25 (2): 77-86.
28. Solow, John L., and Anthony C. Krautmann. "Leveling The Playing Field Or Just Lowering Salaries? The Effects Of Redistribution In Baseball." Recent Developments in the Economics of Sport. Volume 2. 138-149. n.p.: Elgar Research Collection. The International Library of Critical Writings in Economics, vol. 254. Cheltenham, U.K. and Northampton, Mass.: Elgar, 2011. EconLit. Web. 20 Sept. 2015.
29. Sommers, Paul M. . 1990. "An Empirical Note On Salaries In Major League Baseball." *Social Science Quarterly* 71 (4): 861-867.
30. Stone, Gary, and Louis Pantuosco. 2008. "Estimating Baseball Salary Equations from 1961-2005: A Look at Changes in Major League Compensation." *International Journal of Sport Finance* 228-238.
31. Yosifov, Martin. 2006. "Salary Determination and Contract Length in Major League Baseball." Dissertation, Economics, University of Delaware.