What Gets Paid? Analyzing the Major League Baseball Contract Market

Brian Pollack

Professor James Roberts, Faculty Advisor
Professor Michelle Connolly, Seminar Instructor

Honors theses submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University
Durham, North Carolina
2017

Acknowledgements

I would like to thank my thesis advisor, Professor James Roberts, for agreeing to help me with this project and for his advice throughout the year. Thank you to Professor Michelle Connolly for providing comments on my drafts and for making sure I stayed on track to produce a finished product. I would also like to thank my classmates in my thesis seminar for bearing with me and giving helpful feedback during my presentations. And lastly, thank you to my friends and family who have offered their constant support during this project, as they have for all my other endeavors.

Abstract

This paper aims to assess the efficiency of the Major League Baseball contract market in the past decade, given that teams are employing more analytical approaches to player evaluation. First, analysis of team-level data reveals the most important determinants of run scoring and run prevention, respectively. Models of player contract value, controlling for player-specific variables and environmental factors, then determine what is most significantly rewarded on the free agent market. Overall, teams have identified the individual skills that are most important and compensated them accordingly, and there is evidence to suggest teams are becoming smarter about this in recent years.

JEL Classification: D7, O3, Z2
Keywords: MLB, Free Agent Contracts, Innovation

Table of Contents

Introduction

In every industry, executives are constantly searching for more optimal ways of conducting operations to glean an advantage over their competitors. Sometimes, increased efficiency comes through improved technology, and other times, new ideas bring about better ways of doing things. While the effects of new technology are generally easier to measure, ideas are a product of human capital and theoretically represent an endless supply of innovation.[1] Ideas cannot always be seen, though, and so it can be sometimes difficult to pin down their precise impact.[2] For the most part, ideas are transportable, so they are often mimicked by competitors (when feasible) and good ideas, regardless of where they originate, spread quickly.[3] This is frequently what drives large-scale progress; if an idea works well, those who fail to incorporate it are at a competitive disadvantage and get left behind, while those who introduce it effectively can get a leg up on their peers.

Professional sports is a rare industry that allows for comprehensive analysis of new ideas and innovation. Meticulous statistical records are kept for every game, permitting a valuation of both individual and team performances that is uncommon in almost any other sector. The primary objective for all teams, regardless of the sport, is the same: to win.[4] On the field, sports are a zero-sum game (for every team that wins, another must lose), and so teams are in direct competition with each other over the resources (players) that help them achieve their goal. Thus,

[1] In a 2006 paper, Susanto Basu, John G. Fernald, and Miles S. Kimball are able to control for a variety of market conditions and isolate the effects of technological change on the economy, showing that short-run investment decreases when technology improves.
[2] In 2004, W. Scott Frame and Lawrence J. White attempted to review studies analyzing the impact of innovation in the financial sector, but found that empirical analysis of the issue was quite rare due to a lack of data.
[3] For instance, certain ideas fall under intellectual property jurisdiction and may be subject to patents or copyrights, so they cannot be legally replicated by others.
[4] While team owners do look to make money, disentangling the dual goals of winning and turning a profit is very difficult due to a lack of detailed data on bottom-line profits. Instead, this paper looks at teams through the perspective of their on-field products, with the paramount objective being to win games.

a novel idea or a more accurate way of evaluating player performance can be extremely beneficial for teams who develop these innovations first, and disastrous for teams who are slow to realize that better methods exist. This paper aims to examine the effects of a breakthrough shift toward data-based evaluations in Major League Baseball (a shift that has significantly changed how teams assess and utilize players) as a microcosm of industry-wide adoption of new techniques.

Every year, billions of dollars are spent by MLB teams on player salaries via free agent contracts.[5] With so much money being spent annually, it would behoove the teams and executives who pay these salaries to ensure they are spending this money wisely; that is, that the market for MLB salaries operates efficiently. However, in 2003, author Michael Lewis published *Moneyball,* a book that claimed, in a nutshell, that this was simply not true. By going behind the scenes with Billy Beane, the general manager of the Oakland Athletics, Lewis revealed how the Athletics had been taking advantage of gross market inefficiencies to acquire players at a fraction of their true value. Despite consistently running one of the lowest payrolls in baseball, the Athletics were able to exploit other teams' misunderstanding of how wins are produced and be one of the most successful teams in the league.

In the early 2000s, Athletics were the owners of an idea, a better way valuing players, that gave them a tremendous edge over other teams, but the rest of the league has since caught on. *Moneyball*'s widespread popularity has contributed to more teams employing analytical approaches and using data as an increasingly integral part of their decision-making process.[6]

[5] The total amount spent by all 30 teams in free agency first cleared $2 billion following the 2013 season, and climbed another $500 million two years later.
[6] A team of researchers at fivethirtyeight.com discovered that the number of quantitatively-focused analysts employed by teams had risen from 44 in 2009 to 156 in 2016—more than a 350% increase.

Given those advances, this paper aims to assess how well teams have used this new-age approach by studying the efficiency of the market for MLB salaries and contracts in the past decade.

MLB player salaries are determined by one of three methods, depending on the player's experience or service time; team-mandated, arbitration, and free agency. Younger players with less than three years of service time have no negotiating power over their salaries, which are unilaterally determined by the team but must be no less than the league-mandated minimum salary (which in 2015 was $507,500).[7] Players with more than three but less than six years of service time enter an arbitration process in which both the team and the player offer a proposed salary for the upcoming season. An arbitrator then chooses one of those figures to become the player's salary if a compromise is not reached between the two parties. After six years, players have the ability to become free agents and can sign a contract with any team for any number of seasons. Past studies of the MLB salary market have included all three of these buckets, even though the team-mandated group does not have any characteristics of a competitive market. I analyze salaries determined only through free agency and arbitration, where salaries can more accurately be represented as a function of market forces and negotiations between teams and players.

It is important to note that this paper is backward-looking, in the sense that it assumes past performance is the best indicator of future performance. The analysis does not try and incorporate projected future performance or aging curves, but rather evaluates how teams compensate players for what they have done in the past. It is a subtle distinction, but this paper utilizes descriptive models, not prescriptive models. The central goal is to explain what has

[7] One year of service time is 172 days on a major league roster. The MLB season consists of 183 days. A player must accumulate one day of service time to be eligible for the MLB minimum salary.

already happened in the free agent market, as opposed to attempting to advise teams how they should compensate players as they become free agents.

In the first part of my analysis, I utilize team-level single-season statistics to conduct my own evaluation of which skills and statistics have contributed most heavily to winning in the past 10 years. Past attempts to evaluate salaries have only looked at offensive statistics; I conduct separate analyses (at the team level) for offense and defense to create a more complete picture of what is most important to scoring runs (the objective of the offense) and preventing runs (the objective of the defense).

If the MLB contract market operates efficiently, the traits I identify in this first analysis are the ones for which teams should be paying more. After determining what drives run scoring and run prevention, I analyze how those factors are compensated in the contract market. Using 10 years of free agent data, I build separate models of total contract value for position players and pitchers. Inputs for the model include player-specific information such as recent individual statistics, length of contract, age and defensive position. These models allow me to assess what variables are the most significant determinants of a player's total contract value.

After controlling for environmental factors, I focus the interpretation of the results on a few statistics (which are different for position players and pitchers) that emerged in the preliminary analysis as being critical to run scoring and run prevention. These important statistics are indeed compensated significantly on the free agent market, an indication that teams are successfully evaluating the worth of free agents before signing them. Moreover, the compensation rates for these statistics are generally in line with their relative importance to run scoring and run prevention, a positive sign that teams are progressing toward a market equilibrium.

Literature Review

The first major attempt to rigorously evaluate the impact of advanced statistical analysis on baseball salaries was made by John K. Hakes and Raymond D. Sauer (2006). Using data from the 2000-2004 seasons, they consider two key offensive statistics, on-base percentage (OBP) and slugging percentage (SLG), and conclude that OBP is, as *Moneyball* posited, a more important contributor to winning.[8] Hakes and Sauer look at all salaries from the 2000-2004 seasons and conclude that SLG is compensated at a higher rate than OBP, a finding consistent with the hypothesis that there are (or were, at that point) still large inefficiencies in the market. In 2004, one year after the book was published, an increase in OBP compensation is observed, which they view as potential evidence of the market converging toward equilibrium.

In their second attempt at analyzing the baseball salary market (2007), Hakes and Sauer added more depth to the statistics they choose as inputs for analysis. Recognizing the multicollinearity between OBP and SLG, they separate offensive skills into three variables: batting average, walk rate, and power.[9] For a batter, identifying pitches that are likely to be called balls, and avoiding swinging at them, is critical. Players with this skill, which is often informally referred to as "plate discipline", draw more walks by simply not swinging at pitches that are not strikes, and also become more effective hitters by swinging at pitches in the strike zone that are easier to hit. Plate discipline cannot be measured directly, so walk rate is used as a proxy for this skill, which tends to be undervalued because it is less obvious to an observer.

---

[8] OBP is the percentage of plate appearances in which a batter reaches base safely, and thus avoids making an out. SLG is the number of total bases a batter attains per at-bat. A hit is either a single (1 base), a double (2 bases), a triple (3 bases) or a home run (4 bases). Total Bases (TB) is the sum of these bases. Whereas OBP is a measure of how often a hitter reaches base, SLG is a measure of how powerful a hitter is.

[9] Batting average (BA) is the number of hits per at-bat. BA = H / AB. Walk rate is the percentage of plate appearances that end in a walk. Since a plate appearance must result in a hit, a walk, or an out, walk rate is equal to the difference between OBP and BA: Walk Rate = OBP – BA. A main takeaway of *Moneyball* was that OBP was highly undervalued relative to BA, making walk rate a critical underlying skill that teams were not valuing properly. Power is equal to the number of total bases per hit.

Using data from the 1986-2006 seasons, Hakes and Sauer demonstrate that power has been compensated at a fairly constant rate over time; compensation for walk rate has been low but volatile, with a large increase after 2003. This last finding supports the idea that, after the publication of *Moneyball,* teams began to place a higher value on plate discipline and walk rate, since compensation for this skill has increased over time.[10]

The most recent and rigorous attempt at evaluating the baseball contract market was undertaken by the team of Daniel T. Brown, Charles R. Link, and Seth L. Rubin (2015).The three authors investigate the question posed by Hakes and Sauer, but with the benefit of more recent data and more detailed interaction terms in their regressions. Their paper uses contract data from the 1996-2011 seasons, centering around the midpoint of 2003 when *Moneyball* was published. Brown, Link, and Rubin make several important methodological improvements upon the existing literature, many of which I adopt in some form into my own analysis.

First, they treat each of the experience-based salary buckets (team-mandated, arbitration, free agent) as separate data pools, recognizing that there are different market forces determining salaries for players in each of these buckets. Since team-mandated salaries do not reflect negotiations for any individual player, I ignore that group of player salaries entirely and instead focus on arbitration (where data collection was possible) and free agent contracts. Secondly, Hakes and Sauer create an average of a player's statistics from the last two seasons prior to the year in question as inputs for their analysis of determining salary. This average allows the model to incorporate a larger baseline of past performance (as most teams presumably do when considering players) as opposed to just relying on the most recent year's statistics. Building off

[10] Despite an apparent market correction for walk rate compensation, Hakes and Sauer note there is still a weak correlation between payroll and winning percentage—an indication that there may still be other inefficiencies to exploit.

this concept, I employ a similar average of past statistics that incorporates data from the three most recent seasons prior to the new contract.

Lastly, a more comprehensive pre- vs. post-*Moneyball* framework enables them to more accurately quantify changes in compensation for different statistics across time. By using interaction terms on key statistics to indicate whether the salary was pre-2003 or post-2003, Brown, Link, and Rubin were able to more definitely assess how contracts have changed in recent years. Their main conclusion was that there is sufficient statistical evidence to show OBP has been compensated at higher rates since the publication of *Moneyball*.

I plan to continue the existing research that focused on the relative contributions of OBP and SLG to winning and their compensations in player salaries. However, the current body of work on the topic focuses almost exclusively on offensive statistics; I aim to offer a preliminary and parallel analysis for pitchers and defensive statistics, in addition to the aforementioned offensive analysis.

Framework

My analysis occurs in two independent stages. The first stage determines which skills and statistics contribute most to run scoring (offense) and run prevention (defense), respectively, and the extent to which the impact of any of these underlying skills have changed over time.[11] These offensive and defensive regressions are conducted separately from each other.

A key assumption for the first stage approach is that any marginal run scored on offense is equally valuable as any marginal run prevented on defense. The objective of any baseball

[11] For instance, a takeaway from the first-stage analysis could be that OBP is more important to scoring runs than home runs, but that the impact of OBP on scoring runs (relative to home runs) has decreased in the last five years.

game is for a team to score more runs than its opponent; or, equivalently, to score more runs on offense than it allows on defense. A team wins the game if it accomplishes this objective, and loses if it does not. All wins and losses are weighed equally in the final standings, and there is no additional benefit to winning by one run as opposed to five runs (or, in parallel, losing by one run has the same impact on a team's final record as losing by five runs). I use runs instead of wins as the dependent variable for this analysis because wins are a function of both a team's offense and its defense; in order to isolate the effects of offense and defense, it is necessary to look at only runs scored and runs allowed.

The conclusions from this first stage will serve as a benchmark for the second-stage analysis, which determines which skills and statistics contribute most to the total salary of a player's free agent or arbitration contract. I model the real dollar value of a player's free agent (or arbitration) contract based on a weighted average of his past statistics, in addition to other variables such as length of contract, age, and defensive positon. This second stage of analysis considers my original research question. If teams are valuing players properly and allocating salary efficiently, then the characteristics that contribute more to winning (determined by the first-stage analysis) should be compensated at higher rates in free agent or arbitration contracts. Any noticeable gaps or differences between the stage one and two analyses could then be interpreted as a potential remaining market inefficiency.

In order to view salaries as a function of relevant market forces, I avoid looking at salaries from later years in the contract that are already locked in. Free agents often sign contracts for multiple years that secure salaries ahead of time. A player's salary in Year 4 of a contract is not actually a function of his statistics from Year 3 (or Year 2 or Year 1, for that matter), and so Year 4 salary should not be included in any regressions that aim to analyze the contract market for that season. I therefore only look at arbitration contracts (which are all one-year contracts) and free agent contracts that were signed in the offseason before the season in question. For example, a hypothetical analysis for the 2008 season would include all arbitration contracts and all free agent contracts that begin in 2008 and were signed after the 2007 season.

Empirical Approach

The first-stage analysis utilizes team-level offensive and defensive data. There are 30 teams in the league and data was collected for the 2006-2015 seasons, giving a total of 300 data points for both the offensive and defensive regressions. Each data point represents a team total for one season during the year in question. Runs scored is the dependent variable for the offensive regressions and runs allowed is the dependent variable for the defensive regressions. The independent variables for these regressions are team-aggregated statistics for the year in question. A single-season offensive regression looks as follows:

$$Runs\ scored_{it} = B_0 + B_1 * OBP_{it} + B_2 * SLG_{it} + error_{it}$$

Here, B_0 represents the intercept, while B_1 and B_2 represent the regression coefficients for OBP and SLG, respectively. The subscript *i* identifies which team the statistics are for, and the year is represented by subscript *t*.

Similarly, a single-season defensive regression looks as follows:

$$Runs\ allowed_{it} = A_0 + A_1 * Strikeout\_Rate_{it} + A_2 * Walk\_Rate_{it} + error_{it}$$

As with the offensive regression, the subscript *i* identifies which team the statistics are for, and the year is represented by subscript *t*. A_0 represents the intercept, while A_1 and A_2 are the regression coefficients for strikeout rate and walk rate, respectively.

These first-stage regressions can be run for each specific season (i.e. 2007, 2008, etc.), or for a group of seasons (i.e. 2007-2010, 2011-2014, etc.) as a way to track any changes in the value of certain statistics over time.

For the second-stage regressions, the data is on salaries determined by free agent and arbitration contracts. Specifically, the dependent variable is the natural log of the real total salary guaranteed over the life of the contract. The nominal salary values are indexed to 2007 dollars (the first year that contracts in the dataset began) using consumer price index (CPI) data to adjust for inflation. Independent variables include the weighted average of a player's previous statistics, the length of the contract in years, and other player-specific information such as age and defensive position.[12] For instance, the regression for all position players who signed free agent contracts beginning in 2008 looks as follows:

$$Log(Total\ Salary)_j = a + b * OBP_{2007,\ 2006,\ 2005} + c * SLG_{2007,\ 2006,\ 2005} + d * Length\_of\_Contract_j + e * Age_j + f * Position\_Dummy_j + error_j$$

The new subscript, j, is for the individual player.

These regressions can be run for all position players and pitchers who had new contracts in any individual season (i.e. all free agent position players in 2008, all free agent pitchers in 2008, all free agent position players in 2009, etc.) as well as groups of seasons (i.e. all position players from 2008-2010, all pitchers from 2008-2010, etc). A dummy variable for the year of the contract allows the analysis to control for any exogenous league-wide factors such as increasing salaries (the result of higher revenue and inflation) or an impending change in the collective bargaining agreement between the players' union and the league. For instance, the regression for all free agent contracts signed by position players from 2006-2010 looks as follows:

---

[12] Using OBP as an example, the weighted average is calculated as follows:
$OBP_{2007,\ 2006,\ 2005} = (3 * OBP_{2007} + 2 * OBP_{2006} + OBP_{2005})/6$

$$Log(Total\ Salary)_{jt} = a + b * OBP_{Year-1, Year-2,\ Year-3} + c * SLG_{Year-1,\ Year-2,\ Year-3} + d * Length\_of\_Contract_j + e * Age_j + f * Position\_Dummy_j + g * Year + error_{jt}$$

Again, the subscript *j* represents the individual player who signed the contract. The subscript *t* represents the year, which is also reflected in the *Year* dummy variable. Later on, another subscript is added for the team a player signs the contract with, accompanied by a team dummy variable. All pitcher free agent contracts are also modeled with regressions that are identical in form to the position player contract regressions, just using different individual statistics.

Data

All data I have collected and used for this paper are publically available online. For the first-stage analysis, team-level offensive and defensive statistics come from baseball-reference.com. The offensive statistics show how many runs a given team scored in a given season, as well as the number of doubles, home runs, stolen bases, etc., that the team accumulated during the year. I calculated certain statistics that were not explicitly found in the original dataset, but are easily derived from the existing statistics.[13]

The most important of these derived statistics is isolated slugging percentage (ISO). ISO is the number of extra bases (interpreted as anything more than a single) a hitter gains per at-bat, and is calculated as the mathematical difference between SLG and BA. For instance, an at-bat that

[13] For instance, singles were not given in the downloaded data. However, since all hits are either singles, doubles, triples, or home runs, I can calculate the number of singles a team hit, since singles = hits – (doubles + triples + home runs).

ends in a triple has an SLG of 3.0 and a BA of 1.0, and therefore an ISO of 2.0. It is similar, but slightly different, to the Power component of Hakes and Sauer's 2007 paper.

This is an important distinction, because whereas OBP and SLG both share a common component in BA, OBP and ISO share no computational factors. Any positive correlation between OBP and ISO is a result of the fact that hitters who get on base often are generally good hitters, and therefore are also likely to be proficient at getting extra-base hits. The reverse logic applies as well; hitters who have a high ISO tend to be good overall hitters and also accumulate high OBPs. Simply put, there is no mathematical overlap between OBP and ISO, which allows me to evaluate them separately as predictors of runs scored.

Across the entire league, run scoring has declined by more than 12 % in the last decade, from an average of 787 runs per team in 2006 to an average of 688 runs per team in 2015. Over this same period, the league-wide batting average, on-base percentage, slugging percentage, and isolated slugging percentage have all decreased.[14] Each of these trends, taken independently, are expected to contribute to lower run levels.

The team-level defensive statistics are in a similar format to the offensive data. For any given team in a given year from 2006-2015, the dataset contains the number of runs that team allowed, as well as its earned run average (ERA), number of strikeouts recorded, home runs allowed, etc.[15] As with the league batting data, there are several trends on the defensive side that contribute to decreased run levels in the past decade. Strikeout rate (the percentage of plate appearances that end in a strikeout) increased by more than 25% from 2006 to 2015, and walk

[14] See Table 1 in the Appendix for summary statistics on league offensive data.

[15] ERA is the number of earned runs a team (or pitcher) allows per nine innings. Runs can be unearned if a defensive error gives the hitting team extra opportunities to score, but the vast majority of runs are earned runs. Because the ERA formula includes runs, ERA will not be used as an independent variable in the first-stage analysis.

rate (the percentage of plate appearances that resulted in walks) and home runs allowed have also decreased significantly.[16]

For the second-stage analysis, all free agent contract data come from espn.com. Arbitration contracts were available for the 2015 and 2016 seasons from mlbtraderumors.com, which includes 115 position players and 86 pitchers. The free agent dataset consists of a total of 1,579 players (804 position players, 775 pitchers) who signed free agent contracts that started between 2007-2016. Salary data contains the total dollar value and years covered by the contract, as well as the player's previous team and the new team he signed the contract with. Using publically available CPI data, I indexed the nominal dollar values to 2007 (the first year of contracts in the dataset). Then, I took the natural log of these adjusted dollar values, since the range of total salaries is quite large across all contracts.

The statistics for each of player in the three seasons prior to their new free agent contract come from fangraphs.com. For a player who signed a contract beginning in 2007, the Year -1 data will be his 2006 statistics, the Year -2 data will be his 2005 statistics, the Year -3 data will be his 2004 statistics, etc. The statistics from these three seasons are weighted to form a single value for each statistical category, with greater weight placed on the more recent seasons. If a player missed an entire season due to injury (and thus no statistics were available from that year), that season was ignored in calculating the weighted average. I created a dummy variable to flag if a player missed one or more of the three seasons preceding the new contract to account for the increased risk a player in that situation would presumably carry moving forward.

Some of the position player contracts were minor league contracts, meaning a major league salary was not guaranteed. I exclude these players from the second-stage salary regression since a

[16] See Table 2 in the Appendix for summary statistics on league defensive data.

major-league salary is not guaranteed (and in many cases a salary figure was not provided). I also exclude players who signed MLB contracts after playing internationally, since statistics for those leagues are not as reliable or readily available. After making these cuts, I have free agent contracts for 534 position players and 484 pitchers across 10 years.

When incorporating dummy variables for defensive position in the second-stage salary regression, I utilized "position buckets". All other statistics I use in the regression are offensive statistics, but a player can provide both offensive and defensive value to his team. These "position buckets" are a way to approximate the defensive value a player who plays a certain position will provide to his new team. Since individual defensive statistics are not as precise or widely accepted as offensive statistics, the model cannot capture individual defensive skill apart from position. In other words, the "position bucket" treats all shortstops as the same (defensively) and cannot account for individual shortstops who are particularly strong or weak defensively.

There are a total of four "position buckets" in the position player analysis, which are represented with dummy variables: one for all outfielders (left fielders, center fielders and right fielders), one for all middle infielders (shortstops and second basemen), one for all corner infielders (third basemen and first basemen, plus designated hitters) and one for all catchers. Each category consists of similar defensive positions, such that players within a group could be expected to play another position in that group with reasonable competence. Catchers are in their own category because their defensive responsibilities are so unique, and players rarely switch to or from the position.

The pitcher data set contains nearly twice as many relief pitchers as starting pitchers.[17] A relief pitcher dummy variable is used in the free agent pitcher analysis to account for the different roles and contributions teams expect from relievers and starters (this can be thought of as the parallel for the "position buckets" from the position player analysis).

Dummy variables for the year that the contract began were included for each year from 2007 to 2016 to account for any league-wide fixed effects that may have affected a given year's free agent class. Similarly, dummy variables were included for all 30 teams to account for any team-specific characteristics (such as payroll, potential endorsement opportunities, etc.) that may come into play. Finally, I created a "same team" dummy variable to indicate whether a player re-signed a free agent contract with the last team he previously played for, in order to test whether continuity or familiarity is valued significantly in the free agent market.

Findings

Position Players

The first-stage analysis, which is meant to serve simply as a guide for the second-stage analysis, consists of the 300 team data points from the 2006-2015 seasons. Two key offensive statistics, OBP and ISO, are able to explain much of the variation in runs scored during this stretch. The coefficients on both variables are quite large in magnitude, positive, and statistically significant, indicating that teams with high OBP and ISO will score more runs than low-OBP and low-ISO teams. As demonstrated in Figure 1 below, both OBP and ISO are positive and statistically significant indicators for scoring runs.

[17] Teams typically use a five-man rotation of starting pitchers, but have seven or eight relief pitchers available in their bullpen at any given time. Starting pitchers appear in fewer games, but pitch a greater amount of innings, than relief pitchers.

Figure 1: MLB Offensive Regression (2006-2015)

MLB Offensive Statistics (2006-2015):
Dependent Variable = Runs Scored

	I.
	2006-2015
Year > 2010 Dummy	143 *
	(81)
On-base Percentage	4056 ***
(OBP)	(188)
OBP * Year > 2010 Interaction	-541 *
	(276)
Isolated Slugging Percentage	1595 ***
(ISO)	(123)
ISO * Year > 2010 Interaction	231
	(182)
Constant	-840 ***
	(57)
n	300
adjusted r-squared	0.8982

***** = significant at .01 level, ** = significant at .05 level, * = significant at .1 level**
Standard Errors in parenthesis

The positive signs on the non-interaction coefficients demonstrate that, from 2006-2010, OBP and ISO were both significant contributors to run scoring. The magnitudes of these coefficients show that, for the first five years in question, OBP is much more important for scoring runs than ISO. The signs of the interaction term coefficients (negative for OBP, positive for ISO) show that this difference in value has decreased slightly in 2011-2015 as compared to 2006-2010. Only the OBP interaction coefficient is significant at any level (at the 10% significance level) but even with this drop, OBP is still a much stronger predictor of runs scored than ISO. This is a notable observation because it means that the ability to get on base (and thus avoid making outs as a batter) is far more important for scoring runs than the ability to hit for power. Therefore, if the free agent market is operating efficiently (which is explored in the stage two analysis), OBP should be compensated at a higher rate than ISO.

The second-stage analysis contains 534 major league free agent contracts. The base model contains all player-specific information—a player's age, length of contract, past statistics, defensive position, and the "missed year" dummy variable. The second version of the model accounts for the circumstances surrounding the player when he signed a contract (year and team dummy variables). The dependent variable in all versions of the model is the natural log of the (inflation-adjusted) total salary guaranteed by the contract.[18] Columns III and IV show the results of Model 2 when it is run for only 2006-2010 (the first five years of contracts available) and 2011-2015 (the latter five years of contract data). This allows the model to pick up on any changes in the way certain traits or characteristics are being compensated, or, in other words, a market adjustment. Figure 2 below displays the results of both models.

[18] All versions of the model were also run using average salary (instead of total salary) as the dependent variable. See Tables 3 of the Appendix for position player output and Table 4 of the Appendix for pitcher output. Results remain.

Figure 2: Position Player Free Agent Regressions

Free Agent Batter Contracts (2007-2016): Dependent Variable = log (real Total Salary)

	I.	II.	III.	IV.
	Base Model: All Player Characteristics	**Model 2: Add environmental factors**	**Model 2a: Add environmental factors 2007-2011**	**Model 2b: Add environmental factors 2012-2016**
Constant	11.88 ***	11.65 ***	10.79 ***	11.17 ***
	(.42)	(.45)	(.62)	(.77)
Age	-.0031	.0027	.0004	-.004
	(.01)	(.02)	(.01)	(.02)
Years of Contract	.59 ***	.56 ***	.59 ***	.53 ***
	(.03)	(.02)	(.03)	(.04)
On-base Percentage	1.99 *	4.16 ***	3.30 ***	8.17 ***
	(1.05)	(1.0)	(1.25)	(2.01)
Isolated Slugging Percentage	3.69 ***	3.49 ***	4.04 ***	2.98 ***
	(.63)	(.61)	(.76)	(1.13)
Plate Appearances	.0033 ***	.0032 ***	.0034 ***	.0025 ***
	(.0002)	(.0002)	(.0003)	(.0004)
Outfield Dummy	-.11	-.17 **	-.33 ***	-.12
	(.09)	(.09)	(.12)	(.14)
Middle Infield Dummy	-.03	-.05	.002	-.16
(Catcher Dummy for Model 2a)	(.1)	(.1)	(.13)	(.15)
Corner Infield Dummy	-.20 **	-.20 **	-.29 **	-.25 *
	(.09)	(.09)	(.12)	(.14)
Missed Year Dummy	-.24	-.25	-.27	-.02
	(.19)	(.18)	(.26)	(.3)
Same Team Dummy		.098	.052	.21 **
		(.06)	(.62)	(.1)
Year of Salary Dummies		Some significant	None significant	Some significant
Individual Team Dummies		None significant	Few significant	None significant
n	534	534	307	227
adjusted r-squared	0.7803	0.8169	0.8147	0.8125

***** = significant at .01 level, ** = significant at .05 level, * = significant at .1 level**
Standard Errors in parentheses

Weighted plate appearances are statistically significant and positive at the .01 level in all iterations, which should come as no surprise; players who play more are, in general, going to be better players and therefore receive more lucrative contracts. With a few exceptions, the defensive "position bucket" dummy variables did not come up as significant, perhaps a nod to the range of individual defensive skillsets within a given bucket that the model could not account for.

In the base model, OBP and ISO are only significant at the .10 level, but once external factors are controlled for, they both become positive and significant at the .01 level. The magnitude of the respective coefficients reveal that OBP is compensated at a slightly higher rate than ISO, a finding consistent with the takeaways from the first-stage analysis. However, in column IV, it is clear that the compensation for OBP is much higher in the most recent five years than it was previously, and the compensation for ISO has decreased very slightly. Teams, it appears, have become more efficient in allocating their resources and have compensated OBP and ISO at rates that are more in line with their relative importance for scoring runs.

Surprisingly, Age did not turn out to be significant in any version of the model. This may be due to the fact that there was not a great deal of variation in the Age variable; more than 90 percent of all contracts were for players between the ages of 28 and 38. Players rarely have extended careers past their late-30s, and, that, combined with the fact that they can't sign a free agent contract for several years after reaching the majors, leaves a narrow range of ages for all free agent contracts.

Another possible explanation is that any decline in skills that players experience as they age should be reflected in a similar decline in their statistics and playing time. A separate version of the model that included upward and downward trend dummy variables for the player's statistics was run to try and capture the effects of underlying skills that are progressively changing.[19] Age continues to be not significant in this model, although the upward and downward trend dummy variables for plate appearances (playing time) were both significant at

[19] Here, an upward trend is defined, for a given statistic, where Year -1 > Year -2 > Year -3. Similarly, a downward trend is defined as a statistic where Year -1 < Year -2 < Year -3.

the .05 level.[20] As a player's skills begin to erode, he becomes less worthy of playing time and is therefore less valuable on the free agent market; similarly, if a player's skillset is improving and he continues to receive more playing time, he is more valuable. Age no doubt contributes as a conflating factor to these upward or downward trends, but it is the level of the player's remaining skill, not his age in and of itself, that drives his value.

Pitchers

Like the position player analysis, the first-stage analysis consists of the 300 team pitching seasons from 2006-2015. Although the data is aggregated at the team level, I included only strikeout rate and walk rate as independent variables for the first-stage analysis. These two statistics do not involve a team's defense in any way, and so they can be thought of as better reflections of a pitcher's underlying ability than other measures that can be susceptible to influences outside his direct control.[21] As seen in Figure 3 below, strikeout and walk percentage are significant predictors of run prevention.

[20] Full results of this version of the model can be found in Table 7 of the Appendix.

[21] Home runs are another statistic that does not involve the defense. Along with strikeouts and walks, these three are often referred to as "true outcomes" because they strip away luck and defense, allowing for a better assessment of pitcher ability. Home runs were not included in the first-stage analysis because they are too similar to the dependent variable (runs), but they will be included in the second-stage analysis as a potential predictor of contract value.

Figure 3: <u>MLB Defensive Regression (2006-2015)</u>

MLB Defensie Statistics (2006-2015):
Dependent Variable = Runs Allowed

	I.
	2006-2015
Year > 2010 Dummy	-110
	(103)
Strikeout Percentage	-2699 ***
	(264)
Strikeout * Year > 2010 Interaction	503
	(364)
Walk Percentage	2856 ***
	(488)
Walk * Year > 2010 Interaction	252
	(805)
Constant	984 ***
	(65)
n	300
adjusted r-squared	0.5792

***** = significant at .01 level, ** = significant at .05 level, * = significant at .1 level**
Standard Errors in parenthesis

Strikeouts, which are a good outcome for a pitcher because they record outs without involving the defense, unsurprisingly have a significant and negative non-interaction coefficient, meaning they are inversely related to runs allowed for the 2006-2010 period. Walks, which are poor outcomes for a pitcher because they put runners on base without giving the defense a chance to record an out, have a significant and positive coefficient over the same time period, since more baserunners generally leads to more runs. Initially, walk rate has a slightly stronger impact on run prevention than strikeout rate. Since neither of the interaction terms are significant at any level, there is no evidence to suggest that this relationship has changed in the past five seasons of data. Thus, the walk rate coefficient is a little larger in magnitude than the strikeout rate coefficient, although the thin margin implies that limiting walks is not vastly more (or less) important for preventing runs than compiling walks.

The second-stage pitcher analysis covers 484 free agent contracts. Similar to the position player regressions, the base version of the model contains all player-specific characteristics (age, length of contract, past statistics, position, "missed year" dummy variable), and the second version accounts for environmental circumstances such as year and team. In this case, the defensive "position bucket" is a dummy variable for relief pitchers. See Figure 4 below for full results.

Figure 4: Pitcher Free Agent Regressions

Pitcher Free Agent Contracts (2007-2016): Dependent Variable = log (real Total Salary)

	I.	II.	III.	IV.
	Base Model: All Player Characteristics	**Model 2: Add environmental factors**	**Model 2a: Add environmental factors 2007-2011**	**Model 2b: Add environmental factors 2012-2016**
Constant	12.95 ***	13.24 ***	11.77 ***	13.55 ***
	(.40)	(.45)	(.79)	(.68)
Age	.018 *	.018 *	.026	.0086
	(.0097)	(.01)	(.017)	(.01)
Years of Contract	.71 ***	.70 ***	.73 ***	.69 ***
	(.029)	(.03)	(.047)	(.04)
Innings Pitched	.0056 ***	.0055 ***	.007 ***	.0033 **
	(.0008)	(.0008)	(.001)	(.001)
Strikeout Rate	4.61 ***	4.17 ***	4.62 ***	3.12 ***
	(.63)	(.69)	(.96)	(1.1)
Walk Rate	-5.96***	-4.59 ***	-3.29 *	-4.59 **
	(1.27)	(1.3)	(1.93)	(2.1)
Saves	.03	.026	.026	.0066
	(.026)	(.027)	(.033)	(.06)
Home Runs Allowed	-2.97	-2.37	-4.67	2.64
	(3.4)	(3.5)	(4.77)	(5.9)
Relief Pitcher (RP) Dummy	-.31 ***	-.34 ***	-.23 *	-.53 ***
	(.087)	(.09)	(.13)	(.15)
RP/Saves Interaction	-.006	.002	.0028	.023
	(.027)	(.03)	(.03)	(.06)
Missed Year Dummy	-.23 ***	-.27 ***	-.30 **	-.14
	(.40)	(.09)	(.14)	(.12)
Same Team Dummy		.13 *	.13	.16
		(.07)	(.1)	(.097)
Year of Salary Dummies		Few significant	None significant	Few significant
Individual Team Dummies		Few significant	Few significant	None significant
n	484	484	254	230
adjusted r-squared	0.784	0.7982	0.7617	0.8122

***** = significant at .01 level, ** = significant at .05 level, * = significant at .1 level**
Standard Errors in parentheses

As expected, the coefficient on strikeout rate is significant and positive in all versions, while the coefficient on walk rate is significant and negative. These signs are flipped from those in the first-stage regressions, but that is because the dependent variable has changed; strikeouts have a negative impact on runs, but a positive impact on salary since the ability to accumulate strikeouts is a sign of a good pitcher (and vice versa for walks). The walk rate coefficient is slightly greater in magnitude than the strikeout rate coefficient in three versions of the model, but overall there is not a large enough difference in any version to conclude any meaningful differences. This is consistent with the first-stage findings, where walks were slightly more important than strikeouts in terms of run prevention, but not drastically so. Home runs allowed, though not included in the stage one analysis because of multicollinearity with the dependent variable, are surprisingly not significant in any version of this stage two analysis.

The relief pitcher dummy variable is significant and negative in all versions, even after accounting for the difference in innings pitched between starters and relievers. Innings pitched, like plate appearances for position players, are significant and positive since better pitchers are going to, on average, pitch more innings and receive higher salaries. The significance of the relief pitcher dummy variable could be explained by the fact that there are more relief pitchers than starters (nearly twice as many relievers as starters in the free agent dataset), which would, in turn, drive down the market price for relief pitchers.

Interestingly, the “missed year” dummy variable came up as significant and negative in several versions of the pitcher model, while it was never significant in the position player models. This is perhaps a nod to the inherent health risks associated with pitching. Repeatedly hurling a baseball overhand at very high speeds is not a natural motion for the human body and creates a great deal of stress on the elbow and shoulder areas. Because of this, pitchers are more susceptible to career-derailing injuries and the statistical significance of the "missed year" dummy variable can be interpreted as teams viewing pitchers with recent injuries as riskier investments moving forward. For whatever reason, teams seem to be much more risk-averse when it comes to signing pitchers than position players, even aside from injuries; Age is significant at the .10 level in versions 1 and 2 of the pitcher model, which hints at the idea that teams are hesitant to sign older pitchers, even after accounting for measured changes in ability.

There is a similar incongruence in the significance of the individual team dummy variables in the position player and pitcher models. For the position player contracts, only one team, the Seattle Mariners, was statistically significant at the .10 level, whereas 11 team dummy variables (more than one-third of all teams) were significant for the pitcher contracts.[22] All 11 teams had negative coefficients, but there were no unique or defining characteristics for this group of teams. These 11 teams who paid significantly less than others on the free agent market for pitchers had no clear geographic bias, and included organizations on both the high and low ends of the payroll distribution, as well as teams that were among the most and least successful in the sport.

The "same team" dummy variable was significant in only one version of the pitcher model (at the .10 level) and none of the position player models. There may be conflicting factors at play here, pulling the impact of the "same team" dummy variable towards 0. Theoretically, teams could benefit from re-signing players who are familiar with their organization and to their fan base; a positive coefficient here would indicate that continuity and consistency is valuable to teams. However, players may be willing to give teams a "hometown discount" in the sense that they are willing to take slightly less money to avoid moving to a new team in a new city, which would be

[22] The Washington Nationals, San Francisco Giants, Boston Red Sox, Toronto Blue Jays, Baltimore Orioles and Seattle Mariners were all significant at only the .10 level. The Miami Marlins, Atlanta Braves, Pittsburgh Pirates, San Diego Padres and Tampa Bay Rays were all significant at the .05 level.

expressed with a negative coefficient for the "same team" dummy variable. The fact that the variable was rarely significant means we cannot discern which of these factors are driving the outcome. While it is always desirable to glean economically significant conclusions from the output of models, a powerful or practical explanation for the statistical significance (or lack of) of some of the individual team and "same team" dummy variables may not be feasible given the parameters of the model.

Arbitration

As discussed above, arbitration represents the first time in a player's career that he has any say or negotiating power over his salary. Even though arbitration-eligible players cannot sign with a different team, their salaries still represent a result of market forces and allow for analysis on how certain traits are valued. Arbitration contracts were only available for the 2015-2016 seasons, so I limited this analysis to all contracts (both free agent and arbitration) from these seasons. Since all arbitration contracts are, by definition, one year contracts, I use average salary instead of total salary as the dependent variable for these analyses and exclude length of contract as an independent variable. The yearly dummy variables are dropped from this portion of analysis since there were only two years of data.

The findings from the arbitration portion are generally consistent with those found in the primary free agent analyses above.[23] For position players, the coefficients for OBP and ISO were once again positive and significant at the .01 level, with OBP compensated at a higher rate than ISO. For pitchers, the relief pitcher dummy variable was negative and significant at the .01 level,

[23] The full results of both analyses can be found in Tables 5 and 6 of the Appendix.

while strikeout rate and walk rate were also significant with coefficients of roughly equal magnitude and opposite signs (positive for strikeouts, negative for walks).

Arguably the most important result from the arbitration analysis comes from the arbitration dummy variable itself, which was negative and significant at the .01 level for all versions of both the position player and pitcher models. Whether it be because of qualities that less experienced players lack and the model does not account for (i.e. leadership, marketability, etc.) or simple wage suppression from the structure of the arbitration system, players earn far less per year with arbitration contracts than through free agency, holding ability and other factors constant.

Conclusion

This paper allows for a unique look at the industry-wide development of a new idea. These innovations are often difficult to quantify on a large scale, but thanks to the nature of record keeping and salaries in sports, MLB provides an excellent environment in which the progress of ideas can be measured. In recent years, teams have changed the way they approach evaluating players and adopted more statistically-inclined methods of analysis; the impact of this change and the extent to which it has affected salaries can be tracked through contracts signed in the free agent market.

Overall, the results are encouraging for the progression of MLB teams in better understanding what player skills should be valuable to them, and compensating those skills accordingly. Past research has shown that significant market inefficiencies existed through the beginning of the 21st century, most notably in the form of OBP being vastly undervalued relative to its importance to winning. Brown, Link, and Rubin determined that the publication of

Moneyball in 2003 marked a structural change in how teams evaluated players, and the results of this paper are consistent with that hypothesis. This paper looks exclusively at the post-*Moneyball* window and finds significant evidence that compensation for OBP on the free agent market has surpassed the compensation for ISO; teams are allocating their financial resources in line with what contributes most to winning. The fact that the OBP/ISO compensation ratio has increased in the last five years of data available can be viewed as further proof that more teams are adopting an analytical approach to player evaluation and are successfully incorporating objective analysis into their decision-making process.

This paper also explores the efficiency of the free agent market for pitchers, and finds similarly encouraging results. The direct impact of pitcher statistics on run prevention are a little harder to define than the impact of position player statistics on run scoring, likely due to the role that team defense plays in run prevention, which this model cannot fully account for. Individual defensive performance, which is becoming easier to quantify thanks to more precise and widespread tracking technology, is a significant source of value that position players can provide. This model attempts to proxy that value with the "position bucket" dummy variables, but as outlined earlier, these buckets do not allow for varying performance within a group. Even still, analysis of run prevention revealed that strikeout rate and walk rate are both important determinants of run prevention, and teams have identified this by compensating both at roughly equivalent rates on the free agent market.

Since the literature on pitcher compensation is much sparser than position player compensation, there is room for much deeper analysis on this subset of the free agent market. Specifically, home run rate (something that directly contributes to run prevention) was not compensated at a significant rate, a surprising result. This could be a persisting market inefficiency, but further investigation on the matter could help draw more definitive conclusions. On the position player side, a more comprehensive model of player value that combines both their offensive and defensive contributions could be another beneficial contribution to the field. It is generally accepted that position players provide the majority of their value on offense, but defense cannot be neglected entirely, and some players make most of their contributions there. The models in this paper do a good job of predicting total variation in position player contract value, but in order to fully account for all factors teams presumably consider, further study that can more precisely incorporate individual defensive value is needed.

Appendix

Table 1: League Batting Statistics, 2006-2015

Summary Statistics: MLB Team Batting, 2006-2015

Year	Runs	Batting Average	On-base Percentage	Slugging Percentage	Isolated Slugging Percentage
2006	786.63336	0.26926667	0.3364	0.4318	0.1625333
2007	777.40002	0.26793333	0.3357	0.42246667	0.1545333
2008	752.83331	0.26366667	0.33293333	0.41613333	0.1524667
2009	747.29999	0.26233333	0.3328	0.41756667	0.1552333
2010	710.26666	0.25726667	0.3253	0.40256667	0.1453
2011	693.59998	0.25493333	0.32046667	0.39883333	0.1439
2012	700.56665	0.2544	0.319	0.4052	0.1508
2013	675.16669	0.25336667	0.31746667	0.39633333	0.1429667
2014	658.70001	0.2511	0.31366667	0.3862	0.1351
2015	688.23334	0.2544	0.31666667	0.40456667	0.1501667

* All statistics represent the league average

Table 2: League Pitching Statistics, 2006-2015

Summary Statistics: MLB Team Pitching, 2006-2015

Year	Runs	Strikeout Rate	Walk Rate	Home Runs Allowed
2006	786.63336	0.1683876	0.0841603	179.53334
2007	777.40002	0.1707342	0.0851596	165.23334
2008	752.83331	0.1754478	0.0869404	162.60001
2009	747.29999	0.1796259	0.0888179	168.06667
2010	710.26666	0.1849503	0.0849844	153.76666
2011	693.59998	0.1862819	0.0810292	151.73334
2012	700.56665	0.1979245	0.0798094	164.46666
2013	675.16669	0.1986927	0.0791394	155.36667
2014	658.70001	0.2036598	0.076177	139.53334
2015	688.23334	0.2041151	0.0765565	163.63333

* All statistics represent the league average

Table 3: Position Player Free Agent Regressions w/ Average Salary as Dependent Variable

Free Agent Batter Contracts (2007-2016): Dependent Variable = log (real Average Salary)

	I.	II.	III.
	Model 2: Add environmental factors	**Model 2a: Add environmental factors 2007-2011**	**Model 2b: Add environmental factors 2012-2016**
Constant	12.08 ***	10.96 ***	11.53 ***
	(.38)	(.54)	(.65)
Age	.003	.0025	-.0026
	(.008)	(.01)	(.01)
Years of Contract	.20 ***	.22 ***	.19 ***
	(.02)	(.03)	(.03)
On-base Percentage	3.90 ***	3.17 ***	7.13 ***
	(.88)	(1.1)	(1.7)
Isolated Slugging Percentage	3.75 ***	4.19 ***	3.53 ***
	(.52)	(.66)	(.95)
Plate Appearances	.0031 ***	.0033 ***	.0024 ***
	(.0001)	(.0002)	(.0003)
Outfield Dummy	-.17 **	.019	-.11
	(.07)	(.11)	(.1)
Middle Infield Dummy	-.066	.02	-.16
	(.08)	(.1)	(.1)
Corner Infield Dummy	-.21 ***	-.28 ***	-.21 *
	(.07)	(.1)	(.12)
Missed Year Dummy	-.19	-.17	.0066
	(.15)	(.22)	(.23)
Same Team Dummy	.08	.04	.16 *
	(.05)	(.07)	(.09)
Year of Salary Dummies	Most significant	None significant	Some significant
Individual Team Dummies	None significant	Few significant	Few significant
n	534	307	227
adjusted r-squared	0.7246	0.7204	0.7018

***** = significant at .01 level, ** = significant at .05 level, * = significant at .1 level**

Standard Errors in parentheses

Table 4: Pitcher Free Agent Regressions w/ Average Salary as Dependent Variable

Pitcher Free Agent Contracts (2007-2016): Dependent Variable = log (real Average Salary)

	I.	II.	III.
	Model 2: Add environmental factors	**Model 2a: Add environmental factors 2007-2011**	**Model 2b: Add environmental factors 2012-2016**
Constant	13.88 ***	12.33 ***	14.07 ***
	(.42)	(.74)	(.60)
Age	.013	.022	.004
	(.009)	(.02)	(.01)
Years of Contract	.26 ***	.28 ***	.26 ***
	(.03)	(.04)	(.04)
Innings Pitched	.0057 ***	.007 ***	.0038 ***
	(.0007)	(.001)	(.001)
Strikeout Rate	4.21 ***	4.57 ***	3.19 ***
	(.62)	(.89)	(.95)
Walk Rate	-4.81 ***	-3.56 **	-4.76 **
	(1.23)	(1.79)	(1.86)
Saves	.019	.02	.0009
	(.02)	(.03)	(.06)
Home Runs Allowed	-3.16	-4.20	-.73
	(3.2)	(4.4)	(5.2)
Relief Pitcher (RP) Dummy	-.36 ***	-.26 **	-.55 ***
	(.08)	(.11)	(.13)
RP/Saves Interaction	.006	.006	.026
	(.02)	(.03)	(.06)
Missed Year Dummy	-.27 ***	-.29 **	-.14
	(.08)	(.12)	(.11)
Same Team Dummy	.01	.087	.16 *
	(.06)	(.09)	(.09)
Year of Salary Dummies	Few significant	Few significant	Few significant
Individual Team Dummies	Some significant	Few significant	Few significant
n	484	254	230
adjusted r-squared	0.6599	0.6163	0.6714

***** = significant at .01 level, ** = significant at .05 level, * = significant at .1 level**
Standard Errors in parentheses

Table 5: Position Player Free Agent/Arbitration Contracts (2015-2016)

Position Player FA and Arbitration Contracts (2015-2016): Dependent Variable = log (real Average Salary)

	I. Base Model: All Player Characteristics	II. Model 2: Add environmental factors
Constant	11.10 ***	11.96 ***
	(.36)	(.49)
Years of Contract	.17 ***	.16 ***
	(.03)	(.03)
On-base Percentage	6.20 ***	5.27 ***
	(1.17)	(1.31)
Isolated Slugging Percentage	4.20 ***	3.32 ***
	(.76)	(.89)
Plate Appearances	.003 ***	.0031 ***
	(.0003)	(.0003)
Outfield Dummy	.16 *	.22 **
	(.08)	(.09)
Catcher Dummy	.21 **	.29 ***
	(.10)	(.1)
Corner Infield Dummy	.019	.12
	(.09)	(.11)
Missed Year Dummy	-.43 ***	-.38 **
	(.17)	(.19)
Arbitration Dummy	-.26 ***	-.33 ***
	(.06)	(.07)
Individual Team Dummy		Few significant
n	195	195
adjusted r-squared	0.7648	0.7685

***** = significant at .01 level, ** = significant at .05 level, * = significant at .1 level**
Standard Errors in parentheses

Table 6: Pitcher Free Agent/Arbitration Contracts (2015-2016)

Pitcher FA and Arbitration Contracts (2015-2016): Dependent Variable = log (real Average Salary)

	I. Base Model: All Player Characteristics	II. Model 2: Add environmental factors
Constant	14.35 ***	14.45 ***
	(.35)	(.5)
Years of Contract	.22 ***	.22 ***
	(.04)	(.04)
Innings Pitched	.005 ***	.005 ***
	(.001)	(.001)
Strikeout Rate	3.41 ***	3.26 ***
	(.87)	(.9)
Walk Rate	-4.69 **	-3.60 *
	(1.9)	(2.1)
Saves	.083	.065
	(.3)	(.3)
Home Runs Allowed	0.049	-.28
	(5.5)	(5.9)
Relief Pitcher (RP) Dummy	-.48 ***	-.49 ***
	(.15)	(.17)
RP/Saves Interaction	-.06	-.04
	(.3)	(.3)
Missed Year Dummy	-.48 ***	-.45 ***
	(.12)	(.14)
Arbitration Dummy	-.33 ***	-.33 ***
	(.09)	(.09)
Individual Team Dummy		Few significant
n	179	179
adjusted r-squared	0.6971	0.6951

***** = significant at .01 level, ** = significant at .05 level, * = significant at .1 level**
Standard Errors in parentheses

Table 7: Position Player Free Agent Contracts w/ Trend Dummy Variables (2007-2016)

Free Agent Batter Contracts (2007-2016): Dependent Variable = log (real Total Salary)

	Model 2: Add environmental factors
Constant	11.87 ***
	(.45)
Age	.0021
	(.009)
Years of Contract	.54 ***
	(.02)
On-base Percentage (OBP)	3.89 ***
	(1.03)
Isolated Slugging Percentage (ISO)	3.56 ***
	(.61)
Plate Appearances	.003 ***
	(.0002)
Outfield Dummy	-.18 **
	(.09)
Middle Infield Dummy	-.05
	(.09)
Corner Infield Dummy	-.20 **
	(.09)
Downward OBP Trend	-.18 **
	(.07)
Upward OBP Trend	.04
	(.08)
Downward ISO Trend	.07
	(.07)
Upward ISO Trend	.005
	(.08)
Downward PA Trend	-.16 **
	(.06)
Upward PA Trend	.17 **
	(.08)
Missed Year Dummy	-.31 *
	(.18)
Same Team Dummy	.10
	(.06)
Year of Salary Dummy	Some significant
Individual Team Dummy	Few significant
n	534
adjusted r-squared	0.8222

***** = significant at .01 level, ** = significant at .05 level, * = significant at .1 level**

References

Basu, Susanto, Fernald, John G., and Kimball, Miles S. "Are Technology Improvements Contractionary?" American Economic Review. Vol. 96. No. 5. December 2006. 1418-1448.

Frame, W. Scott, White, Lawrence J., "Empirical Studies of Financial Innovation: Lots of Talk, Little Action?" Journal of Economic Literature. Vol. 42. No. 1. March 2004. 116-144.

Todd, Jeff. "2015-16 Free Agent Spending by Team to Date." February 1, 2016. https://www.mlbtraderumors.com/2016/02/2015-16-free-agent-spending-by-team-to-date.html

Arthur, Rob and Lindbergh, Ben. "Statheads are the Best Free Agent Bargains in Baseball." April 26, 2016. https://fivethirtyeight.com/features/statheads-are-the-best-free-agent-bargains-in-baseball/

Hakes, John K. and Sauer, Raymond D. "An Economic Evaluation of the *Moneyball* Hypothesis." The Journal of Economic Perspectives. Vol. 20. No. 3. Summer 2006. 173-186.

Hakes, John K. and Sauer, Raymond D. "The *Moneyball* Anomaly and Payroll Efficiency: A Further Investigation." International Journal of Sport Finance. 2007. 177-189.

Brown, Daniel T., Link, Charles R., and Rubin, Seth L. "*Moneyball* After 10 Years: How Have Major League Baseball Salaries Adjusted?" Journal of Sports Economics. 2015. 1-16.