Long-Term Contracts and Predicting Performance in MLB

Drew Goldstein

Duncan Thomas, Primary Advisor
Peter Arcidiacono, Secondary Advisor
Michelle Connolly, Seminar Instructor

Honors thesis 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

First and foremost, I would like to thank Dan Kantrovitz and the Oakland A's for allowing me access to their proprietary dataset beyond my summer with the team for the purpose of this study. I would also like to thank the three faculty members who steered me through this process; Michelle Connolly for the initial help in developing my topic and preparing the background of my paper, Peter Arcidiacono for helping me reason about the relationship between baseball and economic analysis, and Duncan Thomas for sponsoring an independent study and guiding me towards the completion of my thesis. I'd also like to credit the students from my seminar class for challenging my original topic and understanding baseball well enough to give me valuable feedback on my ideas. I owe my roommate Kyle a special thanks for the late night research deliberations that propelled me towards my most significant results, Leah for putting up with me throughout the process, and lastly Alex for making me write a thesis in the first place.

Abstract

In this paper, I examine whether MLB teams are capable of using players' past performance data to sufficiently estimate future production. The study is motivated by the recent trend by which teams have increasingly signed long-term contracts that lock in players for up to ten seasons into the future. To test this question, I define the "initial years" of a player's career to represent a team's available information at the time of determining whether or not to sign him. By analyzing the predictive ability these initial years have on subsequent performance statistics, I am looking to answer whether—and if so for how long—teams can justify signing players to long-term contracts with guaranteed salaries. I also compare the results of the predictive tests with actual contract data to determine the per-dollar returns on these deals for different types of contracts.

I conclude from my analysis that a player's past performance does in fact provide sufficient insight into his future value for teams to make informed decisions at the time of signing a contract. Teams are able to better predict the future production of potential signees by examining their consistency and relative value in the initial seasons of their careers. Furthermore, the results from examining the contract data coincide with my findings on performance; teams and players arrive at salaries for long-term contracts that divide the future risk between the two parties. The returns on long-term contracts are thus demonstrated to be higher than for short-term contracts, as the overall value of longer deals compensates teams for the associated higher annual salaries.

Keywords: Baseball, Predicting Performance, Forward Contracts

JEL Codes: Z2, Z22, Z23

Introduction

In the years since *Moneyball*[1], MLB teams have increasingly turned to advanced data analytics to help guide their decision making process for acquiring and trading players. Despite this trend, though, teams appear to still be guilty of signing players to massive, long-term contracts that imprudently reward past performance without properly accounting for the player's value over the life of the deal. Some of this is the result of teams continuing to weigh old biases—which will be explored later on in this paper—that are unsupported by the data. Given that most contracts in MLB are fully guaranteed, there are significant financial consequences to signing a player who suffers a long-term injury or a steep decline in performance. Teams must account for these risks and the difficulties they impose on the heart of statistical analysis in baseball. Variations in performance impact the capacity of team management to predict future performance based on information at the time of the contract. While teams include other factors into their decision analysis including an assessment of their own ability to train and improve players they acquire, this capacity is best approximated by assessing the predictive power of past performance. In order for teams to make educated decisions on whom to sign and for how long, they need to be able to rely on projection models that accurately quantify the return on their investments. In behaving like profit-maximizing firms, win-maximizing teams should seek to avoid committing a significant sum of money to an asset without having a clear idea of its future worth. This is especially important in a professional sport where salaries are constantly climbing on both an annual and aggregate basis. As I will describe in my analysis, this study is motivated

[1] Michael Lewis published *Moneyball* in 2003 and popularized the Oakland A's pioneering use of unconventional statistics to assess trading and signing players in Major League Baseball.

by the suspected downward trend in annual returns (in terms of performance) for longer-term contracts.

After the 1976 removal of the “reserve clause,” by which owners controlled their players’ rights indefinitely, MLB owners and the MLB Players Association agreed on a new form of free agency that is largely still in place today. Players are kept under team control for the first six years of their careers with the ability to appeal to an arbitrator for a raise after each season starting in their fourth year. Once a player leaves team control and hits free agency, he is able to market his services to any of the thirty teams in the league and select the most appealing offer (Reuter). In free agency, both teams and players alike have economic motivations for seeking long-term deals. A longer contract enables teams to lock in a player whose performance they expect to improve without having to renegotiate their salary and compete against other suitors. On the other hand, players can benefit from a multi-year contract that will guarantee salary into the future even in the event of injuries or a decline in performance. As a result, players are willing to accept an offer slightly below their expected value in exchange for passing along some of the risk onto the teams. In practice, it follows that most long-term contracts are awarded to top-tier players who have had consistent success and can use their bargaining power to lock-in future compensation and protect themselves from the consequences of an injury. Teams sign these deals with an expectation of how the player should perform over the life of their contract, and that expectation is derived from evaluating his prior production.

While this type of analysis is not normally associated with baseball players, it is no different from a venture capitalist buying stake in a company or a private investor purchasing equity. Investors must decide how to allocate a fixed amount of capital based on their expected payout, which mirrors the uncertainty facing baseball teams. Teams are forced to guarantee their salary commitments at the start of a contract, but they have no idea of knowing how exactly their players—like stocks or startups—will perform. In finance, investors approach this situation by performing projections on how they expect their target assets to perform. Baseball teams similarly must use their available information at the time of the contract to determine how much they are willing to pay. While actual teams have more information—including but not limited to additional performance data—at their disposal, a key part of their decision making involves estimating the value of an athlete over the life of a guaranteed contract.

This paper examines the mechanics behind statistical analysis in baseball by assessing the predictability of player value over a longer horizon. It is fairly straight forward to predict a player's statistics in the following season, but the ability to carry that analysis out further is less evident. For example, the two most robust, public websites for baseball statistics—fangraphs.com and baseball-reference.com—only publish projections at the beginning of a season for that season itself. Since estimating performance becomes increasingly difficult as the forecast window extends further into the future, teams looking at the value of offering long-term contracts to players must be wary of potential decline in performance before locking in their salary commitments.

I hypothesize that teams are not able to sufficiently predict a player's future production at the time of signing a contract, and that they therefore cannot justify signing players to long-term contracts above a certain threshold. Starting with one season and moving forward, there must exist a point where teams can no longer confidently estimate the value of having a player on its roster; if that happens after x years, then it follows that they should never give out a contract today for more than x seasons.

To test this hypothesis, this study estimates a series of regressions that assess the predictability of a collection of recent statistics on a gradually expanding time horizon. The first step will be using auto-regressions over a series of seasons to measure how predictive past performance is of future performance in subsequent seasons. For each of the offensive and pitching metrics, a cutoff will be set in terms of the number of future seasons after which these statistics can no longer be used as reliable predictors. Reliability will then be determined by an examination of the statistical significance of the lagged variables in the regression, which will be interpreted as the portion of future performance explained by past performance.

Teams continue to sign far-reaching contracts up to ten years in length, which is at odds with the expected results of the reliability analysis that teams cannot rely on the past performance information at their disposal beyond a period shorter than ten years. There are two potential explanations for this trend: either teams are offering extended contracts to entice players to sign even beyond the window of reliability, or they are taking other factors into consideration beyond performance. These will be explored through combining performance and contract data, which will enable a more thorough analysis of the returns of guaranteeing salary far into the future. Ultimately this paper will propose a limit on how far in advance teams *should* be wiling to sign their players, along with an explanation for why that "limit" appears to be repeatedly violated.

Literature Review

While most economic literature in baseball deals with assessing the value of performance and examining how it is reflected in player salaries, my research focuses on the increasingly prominent trend in Major League Baseball of the signing of huge, long-term contracts. Gerald

Scully (1974) was the first economist to evaluate whether or not baseball players were being paid their marginal revenue product. Their findings—that players were drastically underpaid overall—led to a complete overhaul of MLB's aforementioned "reserve system" and ushered in what has ultimately become the current free agency system that is meant to pay players according to their true market value. Nevertheless, present-day research remains centered around this idea of finding inefficiencies in the baseball player market that still exist. Still, as the sport becomes increasingly reliant on data-driven projections, research has expanded to incorporate a wider range of themes.

A more recent set of academic papers on baseball has placed the relationship between contracts and performance back into the spotlight. The intricacies of the baseball free agency market that include secondary considerations for signing players—beyond who has the best stats—have paved the way for newer studies that dig deeper into this relationship. Some, like Stankiewicz (2009) and Krautmann and Oppenheimer (2002), have taken the approach of analyzing contract "shirking," whereby the players who sign longer, guaranteed deals are more inclined to underperform while those who are on the verge of entering free agency over-perform. Stankiewicz's study found evidence against shirking, instead concluding that player performance improves for multi-year contracts. Their results do not suggest causation, however, as they failed to control for the fact that most players who sign longer contracts are better players in the first place. Krautmann and Oppenheimer's analysis avoids this error by controlling for player quality and finds that players do exhibit "shirking" and perform better in the year before the end of their contracts. The discrepancy in the results of these two analyses highlights the importance of factoring in player quality when measuring the difference in performance across different types of contracts. Maxcy's (2008) examination of risk premiums in baseball furthers this claim.

Maxcy found that long-term contracts do not serve as a compensating wage differential and on average carry higher annual salaries than shorter deals. The study's results reflect the trend in MLB for longer-term contracts to go to better-performing players.

Outside of academia, statisticians and analytics-leaning columnists have considered exactly how teams use available information to make educated guesses on potential signees' future production. Nate Silver (2005) postulated different ways to measure future player performance and discount it back to today's dollars. His research and that of Dave Cameron (2014) have laid the groundwork for determining the market value of players according to adjusted measures of their past performance that are meant to be more reflective of how their statistics will trend into the future. Their research is guided by the fact that teams should theoretically only pay up to the value of a player's expected future production. Therefore, building a model of market value aims to simulate the projections made internally by professional teams when deciding how much to offer a player. While others have followed Silver and Cameron and developed projection models that are widely available online, few go beyond estimating a single season into the future. Furthermore, the implications of these models have yet to be reexamined in an academic setting to connect projected future performance with player salaries.

Theoretical Framework

My study aims to answer the question of whether past performance is a sufficient measure of performance in subsequent seasons for teams to justify signing players to long-term contacts. As discussed in the introduction, my hypothesis is that fluctuations in value across a player's career will cloud the predictive ability of his initial performance. This would make it impossible for teams to model a player's future value, and as a result they would be making salary commitments without a significant window into the return on their investments. This trend is expected to be reflected in the contract data, as teams would be agreeing to longer deals that ultimately fail to create as much value as shorter deals do on an annual basis.

This hypothesis will be tested by setting up an analysis of player performance using Advanced Value Matrix (AVM) as a proxy for a player's cumulative contribution to a team's wins over a gradually increasing period of time. AVM, the result of a product developed over twenty years ago by AVM Systems, has been used by the Oakland A's to create a more complete accounting of players' contributions to any given play (Lindbergh). AVM differentiates itself from traditional baseball statistics in that it records the process of a play rather that the potentially misleading results. The difference is most evident in the case of a home run robbery, in which a defensive player leaps and catches a ball that would've otherwise been a home run but instead becomes an out. Traditional statistics would penalize the offensive player with committing an out and credit the defensive player with a simple put-out, a meaningless statistic that tells downplays significance of such an extraordinary play. Thanks to its parent company planting observers at each game and using a field diagram to track the speed and trajectory of a ball, though, AVM would properly reward the hitter for a hit that on average is expected to produce a run, and credit the outfielder for saving a run.

Most studies examining past performance would instead focus on Wins Above Replacement (WAR). WAR is meant to be a measure of the additional wins that a player provides for his team versus a "replacement level" player, which is the average output of a player who makes the league minimum salary. While WAR and AVM both create a picture of a player's value in terms of his individual contribution to the team, the former still relies on purely traditional statistics and is not as predictive of future results as the latter. Conducting the analysis in this paper using WAR would create an inaccurate picture of the predictability of performance from one year to the next, since the very statistics that it relies on are not indicative of a player's actual value. By instead using AVM, my analysis will focus on a predictive view of a player's performance rather than past results. Now that the entire MLB subscribes to a more advanced, internal version of AVM known as Statcast, this will also more closely reflect the methodology used by Major League teams than studies that are restricted to publicly available data like WAR.

My analysis will be done under the assumption that teams are signing players in order to maximize wins, subject to salary constraints, which ignores for the purposes of this analysis the confounding factors between wins and revenue. Teams will use the best available data, which I will proxy for with a player's past performance, to predict future performance. Since this in turn will affect the team's win total for the season, they should thus only make decisions in the time frame for which they can estimate their returns. The results of the regressions will determine how far into the future that is, which should yield an arbitrary cut-off after which they should stop committing salary. This lies on the assumption that past performance becomes a weaker indicator of a player's value as its used to project further into the future of a player's career, which will also be tested.

To augment the analysis, actual salary data from MLB will be incorporated to determine whether the hypothesis is supported in practice. Combining player value and pay to measure their AVM value in per-dollar terms can answer whether or not teams are indeed suffering declining returns for longer contracts. Finding that long-term contracts are worth less on an annual basis would support the original hypothesis as a direct consequence of not being able to predict future performance.

If the results are in line with the predictions above, it follows that teams are seemingly making systematic mistakes in signing players. This would require an examination of what other potential information teams are taking into account when making their decisions. Proxies of these factors would be included as variables into further regressions that examine the further determinants of pay, and could include:

1) Teams could be placing a premium on re-signing players that were already on their teams, which appeals to fans who want their favorite players to stay. This could be tested by including a dummy variable for whether a player is signing a contract to return to his current team in a regression of salary on performance.
2) Teams could also be paying a premium for players who are generally more popular, which is more correlated to past performance than projected future performance. This could also be tested with a regression that would instead include a proxy for a player's popularity, such as the number of all star votes he has received in his career.
3) Due to the need to fill a roster with players at each position, teams might pay above-value for players at certain positions in order to fill out their lineups. This could be tested by including position (or position group) within the hitters' analysis to see which positions are paid premiums.

Alternatively, the data could reject the hypothesis and suggest that there is in fact a way to predict future performance. If that is the case, then the core of the discussion will center around which components of past performance indicate future success. Identifying these factors would expand the study beyond a general conclusion of predictability; the findings would aid teams in choosing which players would provide better value over the course of a long-term contract. This would be supported by the inclusion of the contract data, which by rejecting the hypothesis should instead show a trend in which longer contracts exhibit higher annual value.

Data

This study incorporates data from two main sources that cover the past ten seasons (2006-2015) of Major League Baseball. The primary data set is the Oakland A's proprietary database of AVM statistics for player performance that they have collected. While the A's were the only team with access to the data, the data is league-wide and therefore contains the results for every player in the Major Leagues from 2006 through 2015.[2] With the aforementioned recent development of Statcast, the A's AVM dataset is a perfect example of the proprietary data that teams have at their disposal when making decisions.

The AVM data is broken down into two separate datasets: one on hitting statistics for offensive players and one on pitching statistics for pitchers. Each of these is at the player/season-level, such that every data point represents an individual player's cumulative performance over the course of a given season. The data include descriptive statistics as well, notably a player's age, handedness, individual identification number, and team ID. The Oakland A's AVM statistics are constructed such that a player's performance can be measured in terms of the number of runs he contributes offensively (for hitters) or saves defensively (for pitchers). Since the average hitting and pitching value under AVM are set around 0, the data can be interpreted as

[2] I received exclusive access to this data through my relationship with the A's front office that began after my internship with the team in the summer of 2014. The A's developed this dataset by subscribing to an advanced tracking system known as AVM (Advanced Value Matrix) far before the rest of MLB even thought to keep pace. While the league has now instituted a similar service known as Statcast for all 30 MLB teams to use, the A's early start provided them with an advantage over the rest of the league as their data goes further back than any other team's.

a player's marginal contribution in runs vs. the league average that he provides to his team. Given that the average number of runs scored per game in MLB is around 4, the A's use a rule of 8 AVM runs/win in order to convert the AVM values into a measure of wins. In doing so, the A's have internally demonstrated that computing the total expected wins for an upcoming season using AVM is a strong predictor of the team's actual win total.

The two panels in Table 1 below summarize the AVM statistics as they are laid out in a year-by-year fashion for the purposes of this analysis, the first panel for offensive players and the second for pitchers.

Table 1: Summary Statistics for Hitting/Pitching Performance Data

Year-by-Year Hitting Statistics						
	# Obs.	Min	Median	Max	Mean	Std. Dev
Year 1	1555	-40.8	-1.4	62.8	-1.1	7.7
Year 2	1204	-30.0	-1.9	52.3	-1.3	9.0
Year 3	939	-32.4	-1.8	62.4	-0.7	10.1
Year 4	736	-33.8	-1.6	69.3	0.2	11.7
Year 5	563	-28.5	-1.3	61.2	1.3	12.0
Year 6	434	-23.0	-1.3	63.7	1.4	12.2
Year 7	315	-27.7	-0.5	70.3	2.4	12.3
Year 8	228	-23.5	-0.2	80.2	2.5	12.0
Year 9	141	-19.8	-0.3	51.1	3.2	13.2
Year 10	65	-21.9	-0.1	45.1	2.4	12.2

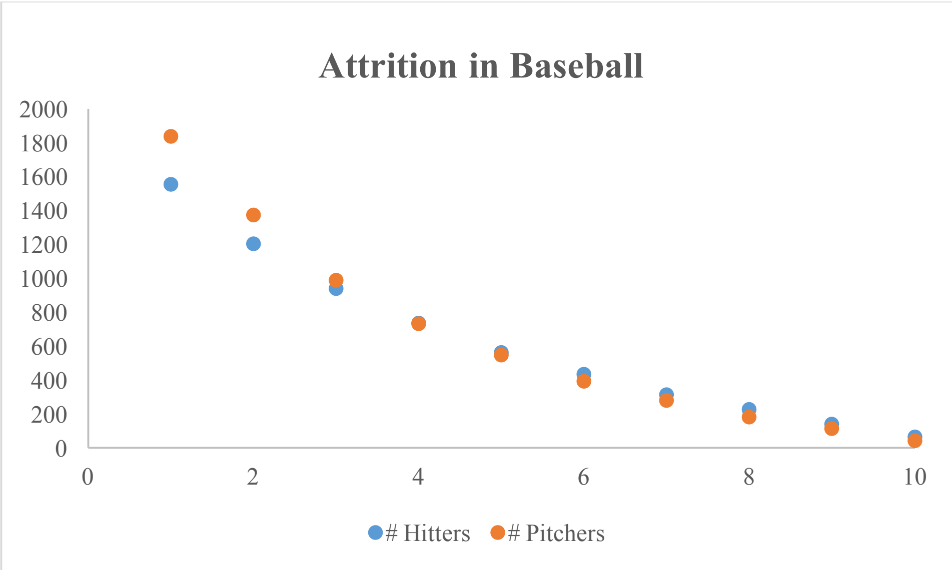
Year-by-Year Pitching Statistics						
	# Obs.	Min	Median	Max	Mean	Std. Dev
Year 1	1836	-30.6	-0.8	40.2	-0.5	6.5
Year 2	1371	-30.0	-0.7	42.7	0.2	8.2
Year 3	988	-28.6	-0.5	48.9	0.4	8.9
Year 4	733	-32.0	-0.5	51.0	1.1	9.9
Year 5	547	-29.7	0.0	36.3	0.9	10.6
Year 6	393	-27.5	-0.3	57.2	0.9	10.5
Year 7	281	-32.6	-0.7	41.4	0.3	10.3
Year 8	181	-36.3	-1.5	52.0	-0.8	11.0
Year 9	113	-34.7	-1.7	38.5	-2.4	10.6
Year 10	41	-27.4	0.0	35.3	1.0	12.7

It's important to note from the first panel that the average hitting value contributed by offensive players trends up in later years as players remaining in the league this late into their careers are generally more valuable. Despite this, the large range indicated by the minimum and maximum columns and the rising standard deviation each year indicate that there still remain players with stand-out seasons of poor performance even in years 8, 9, and 10. The second panel of Table 1 suggests that pitchers follow a similar pattern, although the average value is more consistent from one year to the next due to the higher volatility for a given pitcher as compared to a hitter in baseball.

Both the hitting and pitching datasets are highly skewed in that the mean value for any given year is higher than the median as a direct result of the contrast between high-performers and players who have short stints in MLB before exiting the league. Due to this high degree of turnover, there is a significant number of players who perform below the league average before being pushed out of the league. In my analysis, I account for the confounding effects of player turnover by reworking my analysis for only the players who remain longer than three years, which I explain later in the "Results/Discussion" section.

Better players generally have much longer careers. This trend reflects a potential bias in the data in which a higher degree of variance will be explained by players exiting the league in the initial years of data. Additionally, when making a comparison between contract values at different lengths, the return on long-term contracts will be overstated since longer contracts are generally only going to higher performing players. These two effects could confound the intended understanding of how player performance trends generally from one year to the next few years. Figure 1 below illustrates this trend and the severity of the decline in the number of observations from one year to the next for both hitters and pitchers.

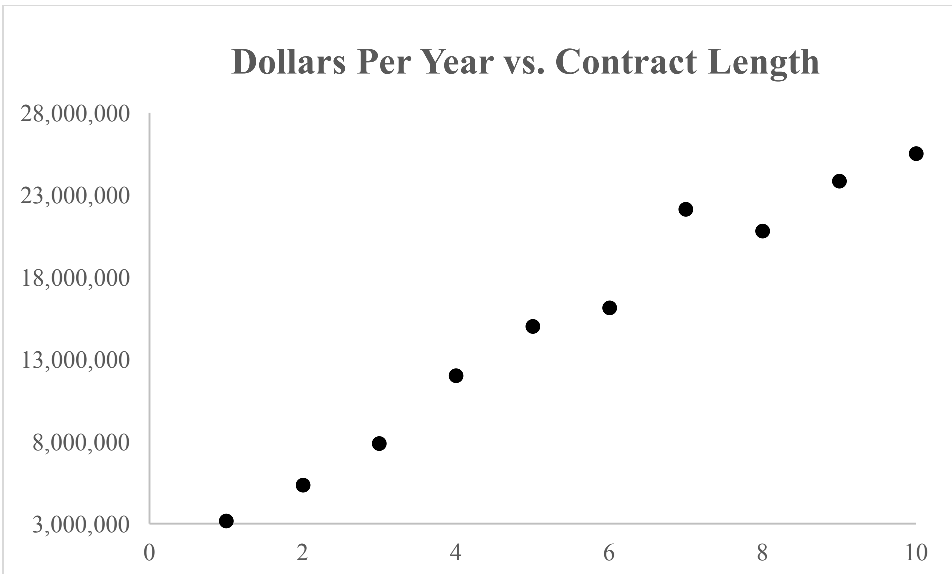
Figure 1: Players Exiting MLB



The second dataset comes from free agency contract signings reported by ESPN.com, which conveniently dates back to the 2006 season as well. ESPN's database includes any contract signed through the free agent process and contains both the length of the contract and the player's total compensation. While this represents the most expansive collection of contract signings available—the A's were not able to produce a more comprehensive dataset to be used in this paper—there are still a few significant limitations. One of the main drawbacks is that it omits contract extensions (when a team restructures a deal with one of its current players to lengthen his time under contract) and contract buyouts (when a team pays off a current player to terminate the contract early). There also is no mention of signing bonuses or incentive-based bonuses, two clauses present in many MLB contracts, and the source does not include the year-by-year breakdown of how the total salary is paid out over time. Despite these limitations, the contract signing data provides sufficient information to glean how much the free agency market values a given player and for how long they are willing to pay them.

The contract data demonstrates the trend in MLB for longer-term contracts to also carry higher annual salaries. This most likely arises from the stronger bidding competition among teams for higher caliber players, who can demand both more money and longer deals. This relationship is illustrated in Figure 2 below, as there is a high correlation between annual contract value and contract length in years.

Figure 2: Increasing Annual Contract Value for Longer Contracts



While longer deals carry more value, though, they are not nearly as common as short-term contracts. In contrast to annual value, the number of contracts signed falls for longer contracts, as indicated by Table 2 below.

Table 2: Far More Short-Term than Long-Term Contracts

Contract Data		
Contract Length	# Contracts	$/Year
1	646	3,144,679
2	222	5,355,957
3	88	7,853,250
4	42	12,000,000
5	24	15,000,000
6	14	16,100,000
7	11	22,100,000
8	3	20,800,000
9	1	23,800,000
10	3	25,500,000

These contracts are included in the analysis to be analyzed versus performance data. Therefore, the previous AVM datasets were merged with ESPN's contract data to compare player performance and pay across different contract lengths.

Results and Discussion

To answer whether past performance can sufficiently predict future performance, a simple set of regressions was used to measure the significance of the initial years of a player's career in predicting his value in subsequent seasons. In order to do so, a player's "initial years" need to be defined such that they could be used as the independent variables in testing their significance in predicting a player's value in his fourth, fifth...tenth seasons. Given that the longest careers modeled within the datasets span ten years, using the first three years of a player's career allows for testing the predictability of performance up to seven years into the future. To justify this selection, a probit model and a linear probability model were used to determine the marginal effects of each of a player's first three seasons of performance on the binary outcome of a career lasting at least four seasons.

Table 3: Regressing T = (1 if career >= 4 seasons, else 0) on Performance in Years 1-3

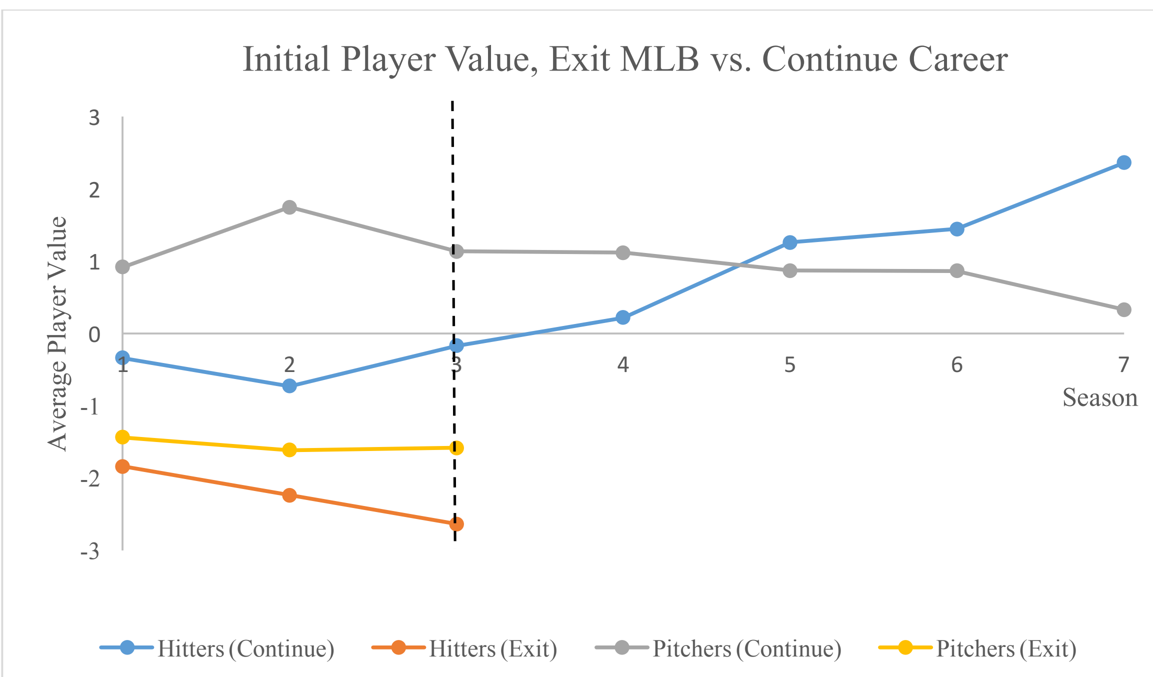
Hitters			
		LPM	Probit
Hit1	β	0.12	0.15
	t-test	*0.66*	*0.76*
Hit2	β	0.03	0.04
	t-test	*0.15*	*0.22*
Hit3	β	0.34	0.39
	t-test	*2.00*	*2.18*
	Sample Size	939	
	R^2, pseudo-R^2	0.01	0.01
	P > F, P > chi^2	0.02	0.01

Pitchers			
		LPM	Probit
Pitch1	β	0.31	0.33
	t-test	*1.50*	*1.59*
Pitch2	β	0.27	0.30
	t-test	*1.44*	*1.59*
Pitch3	β	0.46	0.53
	t-test	*2.66*	*2.91*
	Sample Size	988	
	R^2, pseudo-R^2	0.02	0.02
	P > F, P > chi^2	0	0

The results in Table 3 above for hitters (left) and pitchers (right) both suggest that, when controlling for age, the first three years of a player's career are indicative of whether or not he will continue to play beyond those initial years. *Hit1* denotes an offensive player's value in runs in year 1, *Hit2* in year 2, etc., while *Pitch1* would be a pitcher's value in runs saved in year 1. The chart demonstrates that the three initial seasons are jointly significant predictors under both a linear probability model (LPM) and a probit model (Probit). As expected, the most recent year of performance, *Hit3* for hitters and *Pitch3* for pitchers, is the strongest predictor in the regression and passes the t-test for significance in both models for both datasets. While overall 47.3% of hitters and 39.9% of pitchers played at least a fourth season, those percentages were higher among higher-ranked players. Consistent across the two models, the results suggest that a three-run increase in a hitter's third year value and a two-run increase in a pitcher's third year value lead to about a percentage point higher likelihood of continuing a career beyond three seasons. This finding justifies the selection of three years for a player's initial performance because they are jointly significant indicators of a player's career length, and thus contain insight into a player's future.

Furthermore, Figure 3 below demonstrates the significant difference in initial performance between players who continue their Major League careers and those whose careers end after three seasons. It is clear from the graph that players who exit the league suffer from declining player value from one season to the next whereas those who remain reach average and above average value by year 3 for hitters and pitchers, respectively. The portion of the figure to the right of the dashed line at year 3 represents the average performance of hitters and pitchers who continue their careers. The trend demonstrates that hitters, who previously exhibited negative average player value, rise significantly after their initial seasons while pitchers maintain a slightly positive average value. The results follow the intuition that players whose careers end prematurely are those with lower initial success, which discourages teams from resigning them.

Figure 3: Breakdown of which Hitters/Pitchers Exit MLB



Having selected three years as the "initial years" of a player's career, the aforementioned autoregressive model could be used to test the central question of this study. These regressions of past performance on subsequent value were carried out using the following equations for hitters and pitchers:

$$Hit_i = a_0 + b_1 Hit_1 + b_2 Hit_2 + b_3 Hit_3 + b_4 Age + e$$

$$Pitch_i = a_0 + b_1 Pitch_1 + b_2 Pitch_2 + b_3 Pitch_3 + b_4 Age + e$$

where Hit_i represents a hitter's run-value in season i and $Pitch_i$ represents a pitcher's run-value in season i. The regressions were repeated as a time series with the initial years of a player's career acting as predictors of Hit_i and $Pitch_i$ for i = 4,5,6...10.

Table 4: Results of Regressions on Subsequent Seasons

Hitting					Pitching				
Year	R^2	Significant?	P-value	#	Year	R^2	Significant?	P-value	#
4	0.46	3, 2, 1	0	736	4	0.20	3, 2	0	733
5	0.37	3, 2	0	563	5	0.19	3, 2	0	547
6	0.29	3, 2	0	434	6	0.13	3, 2	0	393
7	0.27	3, 2, 1	0	315	7	0.08	3	0	281
8	0.26	3,2	0	228	8	0.05	-	0.05	181
9	0.20	2	0	141	9	0.04	-	0.31	113
10	0.18	-	0.02	65	10	0.12	-	0.30	41

Table 4 above demonstrates the expected trend where the predictive power of past performance starts to decline as it is used to estimate value further out into the future of a player's career. Each row is a separate regression, where the three initial seasons are used as the to predict a given future year of performance. The significant column indicates which of the three initial years of performance pass the t-test for significance in each regression, i.e. "3,2" in the row for the Hit5 regression suggests that performance in years 3 and 2, but not 1, were significant predictors. For hitters, the P-value column suggests that the predictive power of a player's initial seasons is a significant indicator of his subsequent value for up to six years (Hit4 through Hit9). This window is even shorter for pitchers, as the initial years of a player's career are unable to predict player value for more than four season. The R^2 column denotes the percentage of variance in each year of future value explained by variance in the initial seasons of player performance. The R^2 trends down for predicting performance further into the future as a result of more variance and confounding effects for each additional year that elapses. It follows that teams should be wary of signing long-term contracts, especially for pitchers beyond seven years, as their information is shown to become increasingly unreliable.

It is possible that the declining significance is actually due to sample size rather than the predictive power of the initial seasons, so to test this the regressions were repeated using only players who played at least seven full seasons.

Table 5: Repeating Regressions on Players with 7+ Seasons

Hitting					Pitching				
Year	R^2	Significant?	P-value	#	Year	R^2	Significant?	P-value	#
4	0.51	3, 2, 1	0	454	4	0.25	3, 2	0	281
5	0.35	3, 2, 1	0	454	5	0.21	3, 2	0	281
6	0.31	3, 2, 1	0	454	6	0.13	3	0	281
7	0.26	3, 2, 1	0	454	7	0.08	3	0	281
8	0.26	3, 1	0	337	8	0.05	-	0.05	181
9	0.21	2	0	217	9	0.04	-	0.31	113
10	0.15	-	0	111	10	0.12	-	0.30	41

While Table 5 above shows that repeating the regressions demonstrated slightly improved predictive ability from the hitting statistics, the results once again suggest that the initial seasons of a player's career are not sufficient to predict the future entire horizon of performance. Table 6 below expands upon the initial regressions by breaking down the explained variance into what portion is explained by each independent variable. The third year of performance for hitters and pitchers is the most significant indicator of future performance, but its predictive power generally falls from the earlier to the later years.

Table 6: Partial Correlations of Variables in Performance Regressions

Partial Correlation Chart							
	Hitting				Pitching		
	Hit3	Hit2	Hit1		Pitch3	Pitch2	Pitch1
Hit4	0.35	0.26	0.15	Pitch4	0.31	0.16	0.01
Hit5	0.36	0.16	0.02	Pitch5	0.33	0.15	-0.51
Hit6	0.19	0.23	0.05	Pitch6	0.24	0.13	-0.01
Hit7	0.13	0.20	0.12	Pitch7	0.18	0.07	0.04
Hit8	0.28	0.10	0.06	Pitch8	0.10	0.09	0.04
Hit9	0.10	0.18	0.03	Pitch9	0.07	0.02	0.09
Hit10	0.13	0.11	0.04	Pitch10	0.11	0.09	-0.15

Given that there are several contracts exceeding these lengths—with hitters receiving deals up to ten years long and pitchers up to seven—it follows that teams are in fact signing players beyond the period in which teams can gauge their value.

The results from this initial testing seem to support the hypothesis in that looking at past performance alone is not sufficient to predict future performance. As the model attempts to project further into a player's career, the statistical significance of the performance variables falls as they become less reliable. This conclusion, however, operates under the assumption that teams would only be willing to sign players if performance is consistent from one year to the next and thus early years are predictive of later years. In practice, though, players with more inconsistent performance can be more valuable than consistent players and more sought after by MLB teams. This leads to a secondary hypothesis, that an inconsistent player holds more future value than a consistent player and thus teams can better inform their decisions by factoring in consistency.

Following the intuition above, the players in both the hitting and pitching datasets were segmented into two groups based on their level of consistency. At first, this was done based on the player's standard deviation in value as compared to the average standard deviation of the entire set of players over their entire careers. This set the cut-off at a standard deviation of 5.86 runs for hitters and 6.22 runs for pitchers. Under this criteria, there are 301 hitters and 315 pitchers listed as consistent versus 435 hitters and 418 pitchers listed as inconsistent. For hitters, inconsistent players are shown to have higher average value (3.55 runs vs. -2.69 runs), more seasons in their career with at least +8 runs[3]/year (2.38 seasons vs. 0.20 seasons), and slightly fewer seasons with below -8 runs/year (1.40 seasons vs. 1.42 seasons) than consistent players. Similarly, inconsistent pitchers hold an advantage over consistent pitchers in terms of average value (1.62 runs vs. -0.28 runs) and seasons with at least +8 runs/year (1.95 seasons vs. 0.51 seasons), although they do have a higher number of seasons below -8 runs/year (1.31 seasons vs. 0.39 seasons). The above results still hold when only taking into account years 4-10 of both hitters' and pitchers' careers, which suggests that the better performance of inconsistent players is not confounded by the more variable early parts of their careers.

In line with the original hypothesis, though, teams are making predictions based on already observed performance. Thus at the time of deciding whether or not to sign a player, teams should only have information on the consistency of a player in the seasons prior. Redoing the analysis above by grouping players into one set of consistent and one set of inconsistent performers based on their value in the initial three years of their careers allows for the simulation of the situation above. Once again, inconsistent hitters are demonstrated to be far more valuable than consistent hitters, with the upside of a variable performer outweighing the possibility of a down year. Under this new definition of consistency, the average value of inconsistent hitters,

[3] The Oakland A's use the ratio 8 AVM runs:1 Win Above Replacement to convert their metrics into a measure of wins

3.15 runs, is shown to be significantly higher than that of consistent hitters, -1.35 runs. For pitchers, though, the results are less clear as the average value of consistent pitchers is higher (0.44 runs vs. 0.09 runs) despite inconsistent pitchers posting slightly more seasons of greater than 8 runs/year (0.87 seasons vs. 0.67 seasons).

To test the significance of these findings, average player value in seasons 4 through 10 was regressed on player consistency in seasons 1 through 3. In following the patterns above, this test again mirrors a team's actual decision making in that it aims to predict future performance with the information available at the time of a potential player signing. Again controlling for age, separate regressions for hitters and pitchers both followed the model:

$$AvgValue_{Years4+} = a_0 + b_1 Consistency + \ b_2 Age + \ e$$

where $AvgValue_{Years4+}$ represents a player's average value starting in season 4 and *Consistent,* is equal to 1 if a player is a part of the "consistent" group and 0 if he is a part of the "inconsistent" group based on seasons 1-3.

Table 7: Regressions of Avg. Value on Consistency for Hitters, Pitchers

Hitters		
Consistent	β	-3.82
	t-test	*-5.67*
Age	β	-0.25
	t-test	*-2.61*
	Sample Size	736
	R^2	0.05
	P > F	0

Pitchers		
Consistent	β	-0.04
	t-test	*-0.08*
Age	β	-0.21
	t-test	*-2.69*
	Sample Size	733
	R^2	0.01
	P > F	0.03

The results from Table 7 above reveal that while consistency is a strong indicator of future player value for hitters, it is a less effective predictor of future value for pitchers. The coefficient of -3.82 on *Consistent_Years13* in the regression for hitters suggests that an inconsistent player's average value is nearly half a win (4 runs) better than a consistent player's, controlling for age. The t-test on this coefficient of -5.67 suggests that the finding is strongly significant for hitters compared to the insignificant, near-zero value in the t-test on the same coefficient for pitchers. This suggests that teams can find additional value when signing long-term contracts by looking for hitters with more inconsistent performance statistics in their initial seasons.

To determine whether consistency is predictive of future performance over and above a player's initial value, player hitting value from seasons 1-3 was included into the regression with consistency, yielding the model below:

$$AvgValue_{Years4+} = a_0 + b_1 Consistency + b_2 Hit_1 + b_3 Hit_2 + b_4 Hit_3 + b_5 Age + e$$

The results in Table 8 below demonstrate that Consistency is still a significant indicator even for a given set of initial performance value:

Table 8: Including Initial Hitting Value in Consistency Regression for Hitters

Hitters					
	Consistent	Hit1	Hit2	Hit3	Age
β	-1.13	0.06	0.24	0.34	-0.42
t-test	*-2.11*	*1.81*	*6.9*	*10.98*	*-5.58*
Sample Size	736				
$\mathbf{R^2}$	0.43				
P > F	0				

While the most significant variables in the above regression are the hitting values in the two most recent seasons, a player's consistency grouping still significantly predicts his average future performance. The negative coefficient suggests that players with inconsistent initial performance are more valuable than those with consistent performance. The joint test on these variables suggests that a player's initial performance combined with his consistency grouping are jointly significant predictors of his future average value.

The results above from segmenting players by consistency that suggest there is a higher return to inconsistent hitters contradict the original method of looking for consistency as a prerequisite for estimating future performance. Still, this form of analysis fails to account for the confounding effect of having different quality players grouped together. A potential bias could be underlying the examination of inconsistent vs. consistent players in that the ceiling for player value is much higher in the positive direction than the negative direction. If a player's performance spikes too far down, he will likely have trouble resigning and his career will come to an end; a player who significantly outperforms his average, though, will be able to sign a longer contract and thus continue to contribute to the averages measured above. It follows that the players must also be grouped by their average performance level.

Hitters and pitchers were split into four separate tiers based on their average value over the same initial three-year period that consistency was based on. At the bottom, tier #1 represents players below the 1st quartile in average value while at the top tier #4 represents the top 50 hitters and top 50 pitchers. The two tiers in between, tier #2 and tier #3, represent the remaining players with negative average value and positive average value, respectively. Segmenting the previous analysis of consistent vs. inconsistent players using these tiers produced different results for both hitters and pitchers at different levels of player quality. Within the lower two tiers, consistent players carry higher average value than inconsistent players due to the majority of spikes in performance being downward for these groups. For the two higher tiers, though, more inconsistent players once again hold an advantage in average value over consistent players. These results are illustrated in Table 9 below, with the average value denoted for each combined grouping of consistency and tier. It's clear that consistency is indicative of less negative value for both hitters and pitchers in the lower tiers, while inconsistency produces incrementally higher value for the higher tiered players. This is most likely caused by underperforming players having more spikes in the negative direction, while inconsistent higher performing players can have occasionally even more valuable seasons.

Table 9: Average Value of Players split by tier, consistency

	Hitters					Pitchers			
Tier	1	2	3	4		1	2	3	4
Consistent	-7.68	-2.33	4.72	23.05		-5.40	-1.11	3.35	16.45
Inconsistent	-9.23	-2.28	6.04	26.05		-7.43	-1.00	3.69	18.00

In order to measure the significance of tier and consistency groupings in predicting future value, another set of regressions was performed by adding a similar binary variable for *tier* into the previous model. Since a player's tier and consistency are both measured over their first three seasons, it is possible that the two are correlated and would thus pose multicollinearity issues in such an analysis. However, the VIF (Variance Inflation Factor) tests on the two variables revealed that with a VIF = 1.03, they are barely correlated and will not undermine the results. As expected, the results mirrored the previous finding that consistency, when also controlling for tier, is still not a significant predictor of value for pitchers. However, the results for hitters in Table 10 below suggest that even with *Tier* included in the regression, a player's consistency still provides insight into future performance.

Table 10: Tier/Consistency Regressions for Hitters

Hitters		
Consistent	β	-1.04
	t-test	*-2.79*
Tier	β	8.86
	t-test	*41.92*
	Sample Size	736
	R^2	0.72
	P > F	0

It's important to note that the coefficient of tier in the above regression is far greater than that of consistency. Grouping players by different ranges of performance, as *Tier* does, is sure to provide insight into the average value of a player into the future as the best players are normally those with better starts to their careers. A player's tier, however, is widely available information that teams likely already account for in their contract signings and thus pay more for. Understanding the predictive power of consistency, on the other hand, provides potentially more value to teams in that it is not as widely tested and taken advantage of by front offices.

Furthermore, combining the AVM datasets with the ESPN contract data allowed for an examination of how different contract types play out in practice. Under the original hypothesis that teams were mistakenly overpaying players with long-term deals, the contract data was expected to show a decline in value/year as contract length increased. However, the higher value/year of longer contracts outweighed the similarly higher annual salaries on those contracts. The "returns" on contracts were measured by taking a player's average value (hitting value for hitters and pitching value for pitchers) over the length of the contract and dividing it by the player's annual salary. The results were then multiplied by 1,000,000 to determine the average runs/year per millions of dollars, which is presented below. As demonstrated in Table 11, teams saw their highest returns on contracts longer than five years for both hitters and pitchers, which mirrors the similar results from the player value regressions that also went against expectations.

Table 11: Return on Contracts by Contract Length

Hitters		Pitchers	
Contract Length (Years)	Avg. Value (Runs/Year/$)	Contract Length (Years)	Avg. Value (Runs/Year/$)
1	-1.76	1	-0.81
2	-1.02	2	0.36
3	0.20	3	0.10
4	0.13	4	0.07
5	0.08	5	0.00
6	0.48	6	0.49
7	0.48	7	0.16
8	0.39		
9	0.62		
10	0.36		

Combining the datasets also allowed for segmenting the contract results by the same groupings used for the performance value. The table above suggests that teams are, in fact, seeing higher returns on longer contracts, and it follows that they could already be incorporating these factors into their judgments of whom to sign. Table 12 below demonstrates that the return on player contracts is better for inconsistent players and higher tiered players, such that the average value of such contracts again outweighs the higher costs.

Table 12: Return on Contracts by Tier, Consistency

Hitting Contracts by Tier			
Tier	Avg. Runs/Year/$	Avg. Runs/Year	Avg. $/Year
1	-4.2	-8.3	$2,973,629.00
2	-1.7	-2.3	$3,183,455.00
3	0.9	4.5	$7,469,013.00
4	1.1	11.7	$14,300,000.00

Hitting Contracts by Consistency			
	Avg. Runs/Year/$	Avg. Runs/Year	Avg. $/Year
Consistent	-1.807	-2.03	$3,648,416.00
Inconsistent	-0.841	0.585	$6,005,693.00

To assess whether the difference in these results is significant, additional regressions were performed to predict Avg. Runs/Year/$ based on a hitter's tier or consistency grouping and their outcomes are summarized in Table 13 below.

Table 13: Regressions of Value/$ on Tier and Consistency

ValuePerDollar vs. Tier		
Tier	β	2.28
	t-test	*10.50*
Age	β	-0.01
	t-test	*-0.26*
	Sample Size	490
	R^2	0.19
	P > F	0

ValuePerDollar vs. Consistency		
Consistent	β	-0.94
	t-test	*-2.49*
Age	β	-0.05
	t-test	*-1.03*
	Sample Size	544
	R^2	0.01
	P > F	0.02

The outputs above support a player's tier and consistency grouping being significant predictors of a player's per dollar value over the life of a contract. A player's tier group has a particularly strong positive correlation with value per dollar, such that a higher tier boosts that player's expected value by nearly 2.5 runs per dollar. As for consistency, inconsistent players hold higher value per dollar than consistent players, which suggests there is value to be found for teams by signing players with inconsistent value in their initial seasons.

Conclusion

The results of testing the predictive ability of the performance data do not support the hypothesis; instead of revealing an inefficiency in the MLB player market, they provide evidence that teams are actually making informed decisions on signing players to long-term deals. Bidding between teams for star players was expected to resort in teams overcommitting salary at the end of their contracts to entice them to join their teams. However, the per-dollar value of those players in their final seasons actually exceeded that of players being signed to shorter contracts. While teams have difficulty estimating player value for the final years of such long-term deals, they can sufficiently estimate the average value over the entire life of the contract. It follows that teams are rewarded for signing high-performing players to the extended contracts that they earn from having better bargaining power. This reflects positively on the recent trend of teams signing more long-term contracts with high annual salaries that motivated this paper. As teams continue to gravitate towards these deals, the increased competition for star players should further increase the length and amount of salary that they must commit to sign them. Eventually this may lower the returns on these contracts and reverse the findings of this study, but for now the results suggest that long-term contracts are worth the guaranteed salary that teams are allocating.

Furthermore, by splitting players into groups based on their initial performance values and consistency, this study essentially serves as a guide for determining which players are worth signing to long-term deals. While a player’s initial performance alone is not sufficient to predict value in a given season far into the future, understanding where that player’s value compares to the rest of the league and examining his consistency provide a significant window into his expected future performance.

Perhaps this study's most insightful result stems from combining player performance data and contract data to compute the returns on baseball contracts in an analogous way to determining the return on an investment in financial studies. Studies of this kind are rarely incorporated into academic analyses of baseball. While contract returns were included in this study as a means of evaluating and confirming the results from regressions on player performance, future research could provide a more detailed examination of which types of player contracts provide the most value to baseball teams. Since contracts also vary in more than just their average annual value and length—many include performance incentives and scaled salaries that gradually rise from one season to the next—an extended study could also incorporate these details to see how they affect the returns on player contracts.

With MLB player payrolls reaching nearly $300 million for the richest teams and the average steadily climbing over the past few seasons, a better understanding of player value and the return on various types of player contracts is worth further research and could significantly impact major decisions made by these teams. The key takeaway from this study—that there is predictive power in a player's past performance that can be measured against contract value—suggests that continued analysis will yield results that could be acted upon to improve team performance as a whole.

References

Cameron, Dave. "Win Values Explained: Part One | FanGraphs Baseball." FanGraphs Baseball. N.p., 20 Jan. 2009. Web. 06 Oct. 2016.

Farrell, A. M., Kadous, K. and Towry, K. L. (2008), Contracting on Contemporaneous versus Forward-Looking Measures: An Experimental Investigation*. Contemporary Accounting Research, 25: 773–802. doi:10.1506/car.25.3.5

Krautmann, A. C., and M. Oppenheimer. "Contract Length and the Return to Performance in Major League Baseball." Journal of Sports Economics 3.1 (2002): 6-17. Web.

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

Lindbergh, Ben. "Before Beane." *Grantland*. ESPN, 24 July 2015. Web. 31 Mar. 2017.

Maxcy, Joel. "Motivating Long-term Employment Contracts: Risk Management in Major League Baseball." Managerial and Decision Economics Manage. Decis. Econ. 25.2 (2004): 109-20. Web.

Reuter, Tim. "The Economics Of Major League Baseball Free Agency: Start It Earlier To Achieve Fiscal Sanity." *Forbes*. Forbes Magazine, 01 Sept. 2014. Web. 21 Jan. 2017.

Scully, Gerald W. "Pay and Performance in Major League Baseball." The American Economic Review 64.6 (1974): 915-30. JSTOR. Web. 06 Oct. 2016.

Silver, Nate. "Baseball Prospectus." Baseball Prospectus. N.p., 25 Aug. 2005. Web. 06 Oct. 2016.

Stankiewicz, Katie. "Length of Contracts and the Effect on the Performance of MLB Players." The Park Place Economist 17.1 (2009): n. pag. Digitalcommons.iwu.edu. Illinois Wesleyan University. Web.