Signing Bonuses & Subsequent Productivity

Predicting Success in the MLB Draft

Bobby Hubley

Department of Economics
Haverford College
Advisor: Anne Preston
Spring 2012

This study examines the ability of Major League Baseball organizations to successfully value and project the future productivity of players selected in the amateur draft. To do so, the relationship between player valuations (i.e. signing bonuses) and future productivity is investigated. Productivity is measured using three different metrics: Wins-Above Replacement, the probability of making a Major League Appearance, and the probability of becoming an All-Star. The results suggest that holding constant round & placement in round, elevated draft pick compensation significantly influences the likelihood of making the Major Leagues, and to a much lesser extent, player productivity once there. A supplementary analysis reveals that while teams are somewhat successful in their attempts to project future productivity, they are not necessarily efficient in their allocation of signing bonus expenditures.

TABLE OF CONTENTS

INTRODUCTION…………………………………………………………………………...2

LITERATURE REVIEW………………………………………………………………6

DATA…………………………………………………………………………………10

EMPIRICAL PROCEDURES……………………………………………………...19

ANALYSIS AND RESULTS…………………………………………………………23

CONCLUSION……………………………………………………………………..31

REFERENCES……………………………………………………………………..33

APPENDIX…………………………………………………………………………34

INTRODUCTION

It is widely recognized that, within financial markets, risk is associated with the potential for higher returns. Hiring practices by employers provide one context in which risk plays a significant role. Often times, firms are faced with a pool of potential workers that vary not only in productivity, but in how well they signal this productivity to employers. In such instances, rational firms are tasked with valuing a worker, determining proper hiring costs, and predicting future productivity; all with the goal of selecting the worker with the highest expected value. One area, where this process is readily observable, is the professional baseball labor market, and more specifically, the market for amateur talent. This study will examine the ability of MLB organizations to predict the subsequent productivity of potential laborers in the market for MLB draft picks.

The first-year player draft, also known as the Rule 4 Draft, is Major League Baseball's primary mechanism for assigning baseball players, from high schools, colleges, and other amateur baseball clubs, to its teams. The draft order is determined based on the previous season's standings, with the team possessing the worst record receiving the first pick. When a player is selected by a team, he is offered a minor league contract, and if he chooses to sign, an additional bonus as compensation. These additional signing bonuses were an issue within the MLB draft that were rarely discussed during the first 25 years of its existence (1964-1989), as draft pick compensation remained at tolerable levels. However, this is no longer the case. Over the last two decades, average signing bonuses for first round picks have increased by over 600 percent, from roughly $300,000 in 1989 to just below $2,000,000 in 2009, after taking inflation into account.[1] Consequently, this 20 year period provides a relevant framework for investigating the question of whether MLB organizations act rationally in the market and are efficient in their valuation and compensation of prospective talent. The aim of this paper is to investigate over this twenty year period (1989-2009) whether increases in signing bonuses reflect a "bidding up" of player valuations or generally accurate assessments of player talent. That is, are the values of signing bonuses successful predictors of future success for MLB draft picks? Through attempting to answer such questions it can be determined whether MLB organizations are able to assess talent accurately in the amateur draft.

Professional baseball organizations spend millions of dollars annually towards acquiring information on the high school and college prospects they consider drafting. Each organization has an extensive network of scouts, located throughout the country, whose jobs are to seek out and evaluate prospective talent. The majority of these scouts are former players whose experience and knowledge of the game help them in the field and enhance the credibility of their information. This information collected from scouts is assembled, analyzed, and ultimately influences the players a team chooses to select. Although there are other factors which determine the ultimate level of a player's signing bonus, if MLB teams act rationally and use this information correctly, this would lead one to believe that, on average, players who receive signing bonuses of a large magnitude will turn out to be more productive than those players who receive modest signing bonuses. Such a phenomenon, if it exists, would support the notion that signing bonuses are more representative of player valuation than they are of underlying market pressures.

[1]The primary reason a team is motivated to pay a player a high signing bonus is that the skills and future rents a draft pick has to offer need to be secured. There must be incentives for the player to sign a contract; if the player refuses to sign, the team will have wasted its draft pick and added no value to its organization. Very talented players will demand high signing bonuses, not only because of their skills but because of signing bonuses given to players of "similar" skill in previous years. These players are projected potentially to generate revenue for teams; it is extremely important that they are offered large enough signing bonuses to ensure their participation.

The diversity of amateur talent in the Major League Baseball draft provides an opportunity to further the analysis by investigating a number of interesting questions worthy of empirical study. It is important to understand that baseball players come in two distinct varieties- pitchers and position players. A pitcher's future success is often associated with more randomness compared to that of a position player. Not only are they more injury-prone, but their natural performance is subject to much more fluctuation from game to game and season to season than position players' performance. In this sense, position players are safer investments as their futures are theoretically easier to project. Thus, one might expect signing bonuses to be better indicators of future productivity for position players than for pitchers. Similarly, there are two primary sources from which draft picks originate; high school and college. Analogous to the relationship between non-pitcher and pitcher, the productivity of high school draftees is considered relatively harder to project than the productivity of their college counterparts. College prospects may have been scouted for up to four years longer than high-schoolers, and may have up to four additional years of coaching and training to further development. It would therefore be reasonable to assume that the quantity and quality of information on these prospects would exceed that on high school players. Hence, one would expect a stronger relationship to exist between signing bonus and subsequent productivity for collegiate prospects.

Again, the focus of this research is to examine how successful MLB organizations are at evaluating talent in the amateur draft by analyzing the relationship between signing bonuses and future productivity[2]. Given the amount of information decision-makers in the MLB have access to, it is natural to wonder how successful organizations are at finding productive players in the

[2] A supplementary analysis will also be conducted to examine the efficiency MLB organizations exhibit in their expenditures on draft compensation.

draft. That is do draft pick compensation and ensuing productivity reflect accurate valuations of prospective talent?

LITERATURE REVIEW

There has been little or no research attempting to answer such a question in the context of the Major League Baseball Draft. However, Cade Massey and Richard Thaler (2005), Hendricks et al. (2003), and Berri et al. (2009) have examined decision making processes in the NFL draft and provide some interesting insight on what influences decision-making in the NFL and more broadly, how successful professional sports organizations are at valuing prospective talent.

Massey and Thaler's (2005) study tests whether there is evidence of biases in judgment and decision making on behalf of NFL teams in their annual player draft. Their analysis focuses on the relationship between players' market value and the *surplus value* they provide to the drafting team. Specifically, surplus value is determined by comparing the predicted monetary value for veteran free agents playing the same position and of the same performance level (ProBowler, Everyday Starter, Bench Player, etc.) to the drafted player's compensation costs. Examining player data from 1991-2001, Massey and Thaler find that surplus value rises throughout the first round, reaching its maximum early in the second round, whereupon it declines monotonically for the remainder of the draft. Simply put, decision makers in the NFL overvalue top picks in the draft.

This is not to say however, that NFL teams are unskilled in selecting players. Massey and Thaler observe, using any metric, that performance declines steadily throughout the draft and that players taken in the beginning of the draft perform better than those taken later. However, this decline is not steep enough to be consistent with the very high compensation costs of top draft picks. In other words, although first-round draft picks have a higher expected performance value than those taken later, they also have higher salaries, so in terms of performance per dollar these top picks are less valuable. For instance, they find that the first overall pick is riskier and has an expected surplus value lower than any pick in the second round.

The authors find evidence that decision makers in the NFL act in a manner contrary to what one would expect of rational market participants. According to Massey and Thaler, "a team blessed with the first pick could in principle, through a series of trades, swap that pick for four or more picks in the top of the second round, *each* of which is worth more than the single pick they gave up."(Massey and Thaler 2005, p.29). However, such a strategy is rarely implemented; in many cases a premium is paid by teams to "trade-up" for the less profitable first-round selection. NFL teams, due to a combination of non-rational expectations and the mispricing of players, often overestimate the value of top draft picks.

Others have investigated how the strength and robustness of information about a prospect, or a certain group of prospects, can influence an organization's decision making in the amateur player draft. Hendricks, DeBrock, and Keonker (2003) find that where a player is drafted reflects different kinds of uncertainty on behalf of teams about his future productivity. Specifically, the aim of the authors' research is to compare the relationship between ex ante hiring patterns and ex post productivity in the NFL Draft for workers who signal their productivity strongly to employers and for those who do not. In the context of this study, players who participate at a Division 1 school, where there is increased access to better coaching and training facilities, are considered to be workers who have strong signals. Division II and Division III athletes, whose performance is harder to evaluate, are considered to signal their productivity weakly. Hendricks et al. hypothesize that if differences in uncertainty about future player performance influence teams ex ante evaluations of potential draft picks (as given by their draft position), then there should be a difference in observed career performance for Division 1 athletes (strong signalers) and non-Division 1 athletes (weak signalers). To examine this, the author's collect all necessary data on roughly 5,000 players drafted between 1979 and 1992. A number of measures were used to evaluate player productivity, including, career length, percentage of years a player appeared in a Pro Bowl, and whether or not a player appeared in a single NFL game.

The authors found that if a team is faced with a decision between two high-caliber players at the beginning of the draft (high risk, high cost) they tend to be risk-averse and select the Division I athlete who signals future productivity more strongly. However, in later rounds of the draft, Hendricks et al. find that the reverse is true; Players from Division II and III schools are overvalued relative to those from top programs. Such a pattern emerges even in light of the fact that these players from non-Division I schools do not have significantly longer or stronger careers than players from Division I programs in this part of the draft. The main finding here is that Non–Division IA athletes (riskier, harder to project) are drafted more often at the bottom of the draft and less often at the top than one would expect based on their realized career lengths and pro-bowl appearances. Such results imply, as did those of Massey and Thaler's analysis, that within the market for NFL draft picks, decision makers are subject to certain tendencies which result in an inaccurate valuation of player talent.

Similarly, Berri & Simmons (2009) argue that there is a weak correlation between teams' evaluations of players on draft day and subsequent performance in the NFL. They claim that factors which influence a player's position in the draft are unrelated to future NFL performance. To substantiate these claims, Berri & Simons examine performance data for all quarterbacks drafted in the top 250 picks between 1970 and 2007 who played in at least one NFL game. To measure player productivity, the authors rely on statistics such as QB Score, Net

Points, Wins Produced, and the NFL's QB Rating. Upon examining the relationship between draft position and subsequent NFL performance, Berri and Simons discover that on a per-play basis, quarterbacks chosen with picks 11-50, as well as 51-90, outperformed those quarterbacks with the highest valuations (picks 1-10).[3] The authors then examine the relationship between the factors which determine where a player is selected in the draft and future performance in the NFL. NFL teams, in evaluating talent, rely primarily on two measures: statistics from a player's college career and the results from the NFL Scouting Combine where players take medical exams as well as physical and psychological tests. Not surprisingly, they find that taller, smarter, faster quarterbacks who play at Division I schools are more likely to be selected higher in the draft. More importantly however, they find that these physical characteristics strongly associated with a quarterback's draft position provide little predictive power in forecasting future productivity. Berri & Simmons' study highlights the discrepancy that exists between professional sports teams' valuations, the intuition behind these valuations, and their success at selecting quality talent in the draft.

These three studies discussed, taken collectively, highlight the immense difficulty associated with the process of evaluating workers in the uncertain environment of professional sports drafts. As mentioned previously, aside from the analysis conducted here, little or no formal research has been concerned with this process in the context of the MLB draft. With the wealth of information available to decision makers in Major League Baseball, and given the reasonable assumption that signing bonuses at least partially constitute a team's valuation of a player, increases in signing bonuses should be associated with increases in future player productivity. This reasoning forms the basis of this paper.

[3] In terms of aggregate measures, quarterbacks chosen the earliest in the draft outperform those who are taken later. When the per-play measures are taken into consideration, the aggregate measures imply that where you are selected in the draft impacts how much you play, but not how well you play.

DATA

To investigate this question, data on draft information, signing bonuses, and player performance statistics were collected for each player selected in the first ten rounds of the MLB draft between 1999 and 2009 and for those selected in the first round between 1989 and 1998.[4] Data was not collected after the 2009 season because enough time for skill development had to be given to players in order to ensure an accurate assessment of their productivity. Information on approximately 3,850 players was obtained. However, due to the failure of certain draftees to sign a contract after being selected, some observations had to be excluded. Similarly, players whose signing bonus information could not be obtained were omitted from the dataset.

When a player is drafted he is bound to the team that chooses him for three seasons in the minor leagues. After three years, a player must either be on a team's 40-man MLB roster, or he is eligible for what is known as the Rule 5 draft. In the Rule 5 draft, a player can be drafted by another organization for $50,000. However, that player must be on a team's 25-man active roster for the entire next season or he is offered back to the original team for $25,000. A player not on the 40-man roster and not taken in the Rule 5 draft remains under contract with his current organization. Thus, any player who entered into the Rule 5 draft and was selected by another team is omitted from the analysis. Because this research aims to examine the effect that a player's signing bonus has on the rents captured by the drafting team, any player who switched teams either before or after reaching the Majors (but before free agency) via a trade, was also excluded.[5] The final data set includes 3,370 of the roughly 6500 players selected from 1989-2010.

[4] Information on all player statistics was obtained from Baseball-Reference.com. For the 2002 and 1998 Drafts, only data on the first 100 picks (approximately 2.5 rounds) could be obtained.

[5] Once a player reaches the Majors, he must record six years of Major League service before he is eligible for free agency and can sign with another team. Thus any player who is traded before this 6 year cutoff is eliminated from the analysis because he no

Given that this research aims to examine the relationship between expenditures in the draft and future productivity, information on compensation as well as data on a number of performance measures was collected for the players under examination. To examine teams' valuations and projected productivity of players included in the analysis, player signing bonuses were collected for all those drafted during the period of interest[6].

Subsequent player performance is measured using three different statistics: Wins-Above Replacement (*WAR*), Probability of making an MLB Appearance (*MLAPP*), and the probability of appearing in an All-Star Game (*ALLST*). Wins Above Replacement is a statistic that measures the number of wins a player contributes to his team above what a hypothetical replacement player, or minor leaguer, would produce.[7] Specifically, WAR is to be measured in three different ways. The first measure of WAR is a player's Cumulative WAR. This is simply the sum of a player's annual WAR values during the period of his career under examination. The second measure of WAR, Discounted War (*DWAR)* is a weighted version of Cumulative WAR. When a team selects a player in the draft, the hope is that he will reach the Major Leagues and have an impact *sooner* rather than later. Thus, teams value present performance over future performance. *DWAR* accounts for this preference by discounting future WAR values using an 8% discount rate. The final measure of WAR used is Average WAR (*avgwar*). Average WAR is defined as a player's Cumulative WAR divided by his years of MLB service. Average WAR is included to prevent situations where a particularly large fluctuation in performance for a single year leads to a misrepresentation of one's "true" productivity.

---

longer provides rents for the team that drafted him. Players who remain with the drafting team for the first six years of their Major League career and are traded after, need not be excluded from the analysis.

[6] Data on signing bonus figures were obtained from Baseballguru.com, baseballamerica.com, and thebaseballcube.com

[7] Therefore, Minor League players cannot be assigned a value for WAR. In order to avoid diluting the results, these players who never reached the Major leagues, and who therefore, have no record of player productivity, are excluded from some of the analysis. This will be discussed further when the methodology is presented.

The second measure of productivity, *MLAPP*, is a binary variable equal to one if a player ever successfully reaches the Major Leagues and has a total of at least one Major League appearance throughout the course of his career. Similar to *MLBAPP*, *ALLST* is a dummy equal to one if a player appears in at least one All-Star Game throughout the course of his career. While *ALLST* and *MLAPP* may seem to be crude measures of productivity as other studies using a similar approach have noted (primarily because many other factors other than skill can influence these statistics), they nonetheless are measures of how successful a player is over the course of his career. Furthermore, the results of the regressions for these dependent variables will allow a comparison to those from the *WAR* analysis, serving as a check on the effectiveness of *WAR* as a productivity measure.

In addition to the production measures, a number of control variables include the year in which a player was drafted, the round in which a player was drafted, what pick number (within the round) he was selected with[8], a player's draft origin (i.e. high school or college)[9], position (pitcher vs. position player), years of MLB service, and games played at the Major League level. In Major League Baseball, or any sport for that matter, if a player is on the field more often, he has increased opportunities to add value to his team. Thus, player WAR values are in part determined by the amount of playing time one receives.[10] The inclusion of the latter two

[8] This variable will be called *rd_group* and is split into four individual dummy variables: *rd_group1, rd_group2, rd_group3, rd_group4.* Respectively, these variables represent players selected with picks 1-10, 11-20, 21-30, and 31+, within each round. An alternative model using *roundpick* as a continuous variable, instead of a series of dummies, was also estimated. The fit of each model is compared using an F-test. The F-test does not reject the hypothesis imposed by the constraints when the four dummy variables are used. The adjusted R^2 of the two models (one with the continuous *rdpck*, one with the four dummies) are comparable and the coefficient on the variable of interest (*sibo)* remains relatively unchanged. See Appendix A2 for further discussion.

[9] If a player was drafted out of college, controls are included to account for whether the player attended a Division 1, Division 2, or Division 3 institution. Players who attend Div 1 institutions are more likely to have access to better coaching and training facilities than Div II, DivIII and high school athletes. Thus, Division 1 players may have a higher probability of reaching the Majors as it will likely take less time for their skills to develop. As this could have an effect on a player's future productivity, it is important to control for these differences in player origin. These controls are denoted *div1, div2, div3, and hspck.*

[10] For instance, players who recorded six years of MLB service have a cumulative WAR that is roughly 32 times greater than those who only record 2 years of Major Leave Service and 8 times greater than those who play with the team that drafted them for 3 years. As these variables greatly influence a player's WAR, it is important that they are controlled for.

controls, *servicetime* and *g* account for these differences. Table 1 shown below lists the descriptive statistics for each of the variables included in the empirical analysis.

Table 1
Descriptive Statistics of Variables in Empirical Specification (n=3370)

Variable	Description	Mean (s.d.)	% Dummy equal to 1	% Dummy equal to 0
Compensation:				
1. *Sibo*	Signing Bonus (in 2010 dollars, millions)	$0.587(.809)		
Player Productivity:				
2. *war*	Cumulative WAR (n=1130)	3.131(6.41)		
3. *avgwar*	Average WAR (n=1130)	0.569(1.120)		
4. *dwar*	Discounted WAR (n=1130)	2.659(5.92)		
5. *Mlapp*	Dummy for MLB Appearance		33.56*	66.44
6. *allst*	Dummy for All-Star Game appearance		3.65	96.35
Controls:				
7. *rnd*	Round selected in	4.67[11]		
8. *Year*	Year drafted	2003		
9. *hspck*	Dummy for highschool draft pick		34.9	65.1
10. *div1*	Dummy for Division I draft pick		63.8	36.2
11. *div2*	Dummy for Division II draft pick		1.1	98.9
12. *div3*	Dummy for Division III draft picks		0.2	99.8
13. *pospl*	Dummy for position player		48.93	51.07
14. *servicetime*	Years of MLB service(n=1130)	1.238(2.087)		
15. *g*	Number of games played in MLB (n=1130)	140.77(293.075)		

*For instance, 33.56% of players in the sample reached the Major Leagues, 3.65% were All-Stars, etc.

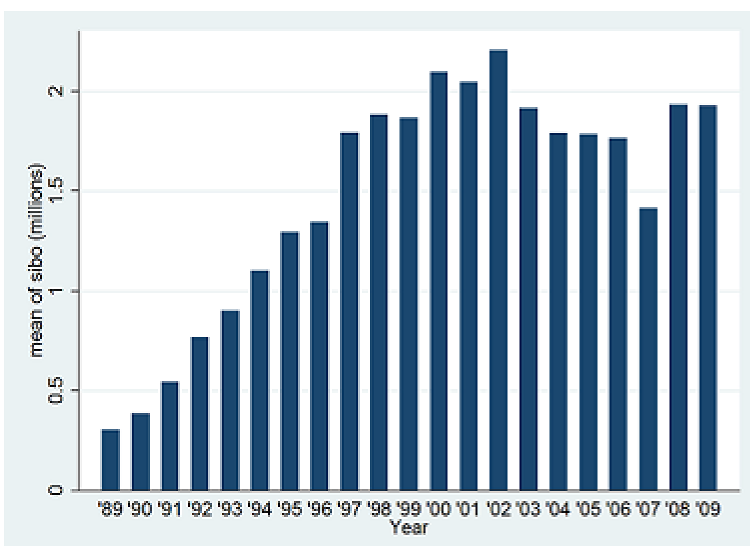
As can be seen in the first row of Table 1, compensation for drafted players averaged roughly $600,000 yet exhibited a great degree of variability with minimum and maximum signing bonuses of $0 and $7.6 million respectively. In a similar fashion, during the period of their career under examination, players on average contributed approximately 3 wins above what a replacement level player would contribute (row 2), with values ranging widely from -4.2 to

[11]The reason the mean for *rnd* is less than 5.00 is because for the years 1989-1998 only data on players selected in the first round were obtained. Consequently, there were a greater number of players in the data who were drafted between 1999 and 2009. This resulted in the mean of *year* being slightly higher than 2000.

35.8.[12] Given WAR's scale[13], it seems as if the mean contribution of players over the course of their early career is small, but appropriate. The difficulty associated with reaching the Major Leagues and being very productive once there is evidenced by the statistics presented for *MLAPP* (row 5) and *ALLST* (row 6): Approximately one-third of all players drafted are skilled enough to participate in the MLB (n=1130) with less than 4% becoming All-Stars (n=123).

In light of such information highlighting that the majority of draft picks do not provide significant value, the large observed variance in player compensation creates an interesting situation requiring further discussion. It is clear that some players receiving significantly high compensation do not always provide significantly high returns. So the question must be asked, are these signing bonuses successful indicators of a player's future success? When attempting to answer such a question, some historical perspective is necessary. Figure 1 highlights the evolution of player signing bonuses over the past 20 years.

Figure 1: Evolution of First Round Signing Bonuses, 1989-2009 (measured in 2010 dollars)



[12] Note these are the numbers for a player's Cumulative WAR, although similar variance is observed for Discounted War and Average War.

[13] Single player season WAR values are scaled as follows: 8+ - MVP Quality; 5+ - All-Star Quality; 2+ - Starter; 0-2 – Reserve; <0 - Replacement Level. Scale obtained from baseballreference.com

Signing bonuses have risen significantly throughout the past two decades, even after taking inflation into account. Average signing bonuses for first round picks stood at just $300,000 in 1989 steadily rising to a maximum of $2.2 million in 2002 whereupon they generally leveled off around $1.75 -1.85million.[14] Although a portion of this increase can surely be attributed to the significant growth in revenues over the past two decades, it nonetheless demands more attention than it has received. Especially when one considers the amount of money that MLB organizations spend on player research and scouting (in 1989, expenditures on scouting and player development averaged 12.6% of revenues for MLB teams), the question of whether these signing bonuses for unproven talent are justified comes to mind.

Upon closer examination of the data, some interesting trends emerge, providing insight into the way MLB teams approach the draft process. As mentioned earlier, players selected in the annual MLB draft can fall into a number of different categories (i.e. high school, college, position players, pitchers etc.). Regardless of which category a player falls into, every team in the draft is ultimately tasked with attempting to predict the future productivity of these unproven high school and college athletes. Previous research has shown that this may be more difficult for some groups of players compared to others. For instance, it has been concluded that high school draft picks on the whole, are more risky than their college counterparts. It has been found that college hitters, within the first few rounds, provide the most return (in terms of value added) than any other type of draft pick, with college pitchers providing more value than either high school hitters or pitchers, however by a much smaller margin.[15] Table 2, shown below, lists by round the percentage of players who successfully reached the Major Leagues and further, what percentage of high school and college draft picks were successful in doing so.

[14] These figures are all measured in 2010 dollars.
[15] James, Bill. *The New Historical Baseball Abstract.* New York: Free Press, 2001. Print.

Table 2- Percentage of Draft Picks Reaching MLB by Round, Draft Origin, and Position

Round	% Reaching Majors	% of HS Draft Picks Reaching Majors (n)	% of College Draft Picks Reaching Majors (n)	% of Position Players Reaching Majors (n)	% of Pitchers Reaching Majors (n)
1	0.64* (794)	54.7 (362)	71.99 (432)	67.11 (374)	61.43 (420)
2	0.426 (345)	32.12 (165)	52.22 (180)	42.69 (171)	42.53 (174)
3	0.344 (326)	30.58 (121)	36.59 (205)	32.75 (171)	36.13 (155)
4	0.26 (277)	25 (112)	32.73 (165)	30.77 (130)	28.57 (147)
5	0.26 (277)	22.54 (71)	27.18 (206)	27.82 (133)	24.31 (144)
6	0.199 (281)	13.64 (88)	22.8 (193)	17.73 (141)	22.14 (140)
7	0.16 (275)	7.79 (77)	19.19 (198)	19.7 (132)	12.59 (143)
8	0.153 (275)	16.13 (62)	15.02 (213)	16 (125)	14.67 (150)
9	0.104 (269)	8.2 (61)	11.06 (208)	9.46 (148)	11.57 (121)
10	0.155 (251)	10.53 (57)	17.01 (194)	14.52 (124)	16.54 (127)
Total	33.56 (3370)	31.55 (1176)	34.64 (2194)	33.96 (1649)	33.18 (1721)

*For instance, between 1989-2009, 64% of first round draft picks successfully reached the Majors. 198 of 362(54.7%) high school draft picks reached the Majors, while 311 of 432 (71.99%) college draft picks reached the majors in the first round

It should become clear that a similar trend has emerged in the data here, at least in the first few rounds of the draft. In Rounds 1-4, where the percentage of players drafted from college roughly equaled the percentage of those selected from high school[16], college draft picks were significantly more likely to make the Major leagues compared to their high school counterparts. With the exception of Round 8, this trend continues to appear in later rounds. It is natural to then wonder how this data matches up with the signing bonuses allocated to these two groups. Table 3 shows the average signing bonuses for high school and college draft picks selected in rounds 1-10.

[16] See appendix (Table A1) for table containing full breakdown of percentages by Round. Note that Table 2 only reports the number of high school and college players selected in each round, not the percentage of draft picks that were selected from high school and college in each round.

Table 3: Average Signing Bonuses by Round for High School and College Draft Picks (bonuses measured in 2010 dollars)

Round	High School	College	HS/College Ratio	Difference
1	$1,594,786.00	$1,523,815.00	1.05	$70,971
2	$788,050.80	$723,072.80	1.09	$64,978*
3	$574,197.50	$450,524.50	1.27	$123,673***
4	$402,430.50	$272,769.80	1.48	$129,660.7***
5	$285,730.30	$201,231.00	1.42	$84,499.3***
6	$277,947.90	$146,680.70	1.89	$131,267.2***
7	$218,766.80	$118,060.60	1.85	$100,706.2***
8	$183,495.80	$82,482.75	2.22	$101,013.05***
9	$150,829.00	$81,198.05	1.86	$69,630.95***
10	$146,473.40	$56,841.07	2.58	$89,632.33***

*Difference Significant with 90% confidence
***Difference Significant with 99% confidence

As shown above, throughout rounds 1-10 signing bonuses for high schools draft picks exceed those given to college graduates, and interestingly, the ratio between the two grows in magnitude as one progresses later into the draft. This is quite surprising, given the results highlighted in Table 2. MLB teams appear to pay a premium for high school talent throughout all rounds, especially in later rounds. For instance, in round 10, for high-schoolers who account for roughly one-sixth of players who make the major leagues, average signing bonuses are more than double those of collegiate players. While it is expected that high school players will receive some sort of premium in their signing bonuses as they demand compensation for passing up the opportunity to go to college, the magnitude of the disparity, especially in later rounds, is something that calls for further inquiry.

The trends discussed above provide an interesting backdrop for analyzing the relationship between signing bonuses and subsequent productivity in Major League Baseball. As signing bonuses have grown significantly in magnitude over the past two decades, it is important to determine whether these increases in spending are justified by increases in productivity.

EMPIRICAL PROCEDURES

Measuring Player Productivity Using WAR

In order to capture the effect of draftee signing bonuses on future productivity, the following model is estimated using OLS:

$$WAR_i = \alpha_i + \beta_1\, sibo_i + \beta_2\, rnd_i + \beta_3\, servicetime_i + \beta_4\, g_i + \beta_5\, year_i + \beta_6\, rd\_group_i + \beta_7\, div1_i + \beta_8\, div2_i + \beta_9\, div3_i + \beta_{10}\, hspck_i + \beta_{11}\, pospl_i + \varepsilon_{i,} \qquad \textbf{(1)}$$

where WAR measures player productivity, *sibo* measures a player's signing bonus, *rnd* controls for the round a player was drafted in, *servicetime* measures how many years of major league service a player has recorded, and *g* estimates games played at the Major League level; *year* and *rd_group* are series of dummy variables indicating the draft year and where a player is drafted within each round, respectively; *div1, div2, div3,* and *hspck* are dummy variables indicating a player's draft origin. The first three of these indicate whether a player was drafted from a Division I, Division II, or Division III College, respectively. The last, *hspck*, is a dummy variable equal to 1 if a player was a high school draft pick. Finally, *pospl* is a binary variable equal to 1 for non-pitchers and 0 for pitchers. ε is an error term with mean zero and constant variance.

Three different measures of WAR are used in this analysis; Cumulative WAR, Discounted WAR, and Average WAR. Cumulative WAR is simply the sum of a player's yearly WAR values over the period of his career under examination. Discounted WAR (*dwar*) is a weighted version of Cumulative WAR that values present performance over future performance. As mentioned previously, the intuition behind using discounted WAR is that a team would rather have an all-star season from a player now as opposed to three or five years into the future. To convert a player's future production into present production a discount rate of 8% is used.[17] Similarly, in order to account for situations where Cumulative WAR values can be biased based on performance in a single year, Average WAR is used as the third measure of wins-above-replacement. It is simply defined as a player's Cumulative WAR divided by his years of MLB service. The inclusion of these two additional measures of WAR not only serves the aims of this research, but indicates whether any discovered effects, if significant, are robust.

Lastly, it is important to note that players who stayed in the Minor Leagues throughout the course of their career are not assigned a WAR value. Therefore, the above model(s) were estimated for only those players who successfully reached the Major Leagues.

Measuring Productivity Using MLB Appearance and All-Star Game Appearance

Whether or not a player successfully reached the Major League level is another metric that can be used to evaluate the subsequent productivity of MLB draft picks. Although it is possible that one could reach the Majors and provide very little value to his organization, the fact that a player has progressed to baseball's highest level is most definitely indicative of a player's talent. In order to investigate the extent to which a draftee's signing bonus influences his probability of reaching the Major Leagues, the probit regression shown below is estimated for all 3,370 players drafted between1989 and 2009:

$$MLAPP_i = \alpha_i + \beta_1\, sibo_i + \beta_2\, rnd_i + \beta_3\, year_i + \beta_4\, div1_i + \beta_5\, div2_i + \beta_6\, div3_i + \beta_7\, hspck_i + \beta_8\, pospl_i + \beta_9 rd\_group_i + \varepsilon_i \qquad \textbf{(2)}$$

[17] Other studies examining player productivity in Major League Baseball through WAR have used an 8% discount rate. See (http://www.baseballprospectus.com/article.php?articleid=4291)

As mentioned earlier, *MLAPP* is a dummy variable equal to 1 one if a player records at least one Major League appearance over the course of his career. The second generated model is specified as **(1)** is, except games (*g*) and years of Major League service (*servicetime*) are left out.[18] A similar regression is estimated using *ALLST*, a dummy variable equal to 1 if the player is ever an all-star, as the dependent variable in order to isolate the effect of signing bonus value on the likelihood of appearing in an All-Star game. Similar to the WAR models, the All-Star model is estimated only for players who reached the Major League level.[19]

Comparison of High School and College Draft Picks

The data provide opportunities to make a number of interesting comparisons within the sample of players used in this analysis. One such comparison is between college and high school draft picks. As mentioned previously, there is evidence suggesting that high school draftees can be considered riskier investments than their college counterparts. Given the increased signing bonuses allocated to high school players in the draft in spite of such evidence, I investigate whether the excess compensation represents overvaluation or a more efficient allocation of resources (i.e. a more accurate assessment of talent and projected productivity). By incorporating the interaction, *sibo*hspck*, into the above models one can observe whether signing bonuses are better at predicting subsequent productivity for high school draft picks compared to those from college. The coefficient on this interaction, if significant, will reveal whether MLB organizations are differentially successful at devoting their resources efficiently and predicting the future success of college and high school players.

---

[18] Obviously it would not make sense to include these two variables. The amount of games and years of service *at the major league level* hold no explanatory power in estimating the probability of a player making the Major Leagues. Furthermore, in all models *rnd* (variable controlling for what round a player was drafted in) is included as a continuous variable instead of a series of ten dummy variables. A F-test comparing the overall fit of the two models could not reject the hypothesis imposed by the constraints of the model using round as a continuous variable (*rnd*). Similarly, the model using the continuous *rnd* variable provided a better fit than the model using the dummy variables. . (Adjusted R^2 = 0.4871> 0.4755). Results and discussion can be found in A2 of the appendix.

[19] Players who never reached the MLB are excluded because without participating at the Major League level, it is impossible to earn a spot on the All-Star roster. This restricts the sample to 1,130 observations.

Comparison of Pitchers and Position Players

A similar comparison can be made between pitchers and non-pitchers. By including the interaction *sibo*pospl* in our model, it will become evident whether or not there exists a difference in the ability of MLB organizations to value pitchers in the draft compared to position players.

ANALYSIS & RESULTS

As mentioned above, the first model (and its two variations) examines the effect of signing bonus on three different measures of future WAR. Additionally, interaction terms were included to examine if MLB teams experience varying success in projecting future productivity for players of different types, as evidenced by their respective signing bonuses. Results are shown below[20]:

Table 4: OLS Regression Explaining WAR, Discounted WAR, and Average WAR variables[22]

n=1130	Cumulative WAR			Discounted WAR			Average WAR		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
sibo[21]	0.6967*** (.1745)	0.8394*** (.2122)	0.5197*** (.2018)	0.5814*** (.1629)	0.7179*** (0.1981)	0.5718*** (.1885)	0.1495*** (.0346)	0.1908*** (.0419)	0.1363*** (.0399)
sibo x pospl		-0.3116 (.2635)			-0.2982 (0.2461)			-0.0903* (.0522)	
sibo x hspck			0.2869 (.2769)			0.0263 (0.2586)			0.0361 (.0549)
	*indicates significance at 90% confidence; **indicates statistical significance at 95% confidence; ***indicates significance at 99% confidence								

Columns **(1), (4), and (7)** show the results for the main effect of signing bonus on future productivity in each of the three specifications. The significant coefficients on *sibo* suggest that there is a relationship between draft compensation and future performance, although the effect is quite small, if not negligible. A one standard deviation increase in signing bonus is associated with a 0.12standard deviation increase in Cumulative WAR, 0.11 s.d. increase in Discounted WAR and a 0.13 s.d. increase in Average War. This works out to 0.75, 0.63, and 0.16 wins-

[20] Full Results for the Cumulative WAR specification are reported in Table A3 of the Appendix. Complete results for the Discounted WAR and Average WAR models are reported in Tables A4 and A5 respectively. Note in columns (2), (5), and (8) the coefficient in Row 1 represents the effect of signing bonus on future productivity for pitchers. The effect for position players is equal to the sum of the coefficients in Rows 1 and 2 of these columns. Thus, the difference between the effects for position players and pitchers is shown in row 2. The same applies to results shown in Rows 1 & 2 of columns (3), (6), and (9) except in this case row 1 shows the effect for college draft picks.

[21] To find a model providing the best fit for the data, two models were originally estimated; one using the natural log of signing bonus and the other using real dollar figures for signing bonuses (2010 dollars). The model using the transformed signing bonuses reported an $R^2 = 0.4979$ while the latter model reported an $R^2 = 0.5017$. The stronger fit is primary reason signing bonuses are measured in dollars throughout the analysis.

[22] Signing bonuses are scaled in millions. Thus, as seen in Row 1(column 1), a $1 million increase in signing bonus is associated with a 0.697 increase in cumulative WAR.

above-replacement respectively. This fact that the effect is so small is not very surprising. There are countless non-quantifiable factors that influence a player's chances of reaching the Majors and their subsequent success if they are fortunate enough to participate at that level. That any significant effect is found is itself surprising and reinforces the notion that signing bonuses are to a certain extent, manifestations of player valuations; MLB organizations are at least somewhat successful in their projections of prospective talent.[23]

To investigate further whether teams exhibit an increased ability to accurately project a certain group of draft picks over another, interaction terms were included (Rows 1 and 2). Columns **(2), (5),** and **(8)** show, for each specification of WAR, the results of the model after including an interaction between signing bonus and position (*sibo x pospl)*. The coefficient on this interaction provides some insight as to whether MLB organizations are differentially successful in valuing prospective pitchers and position players. The coefficients on *sibo* and *sibo x pospl* indicate that the effect of signing bonus on future productivity does in fact differ between position players and pitchers, but with the exception of the AVG WAR model, these differences are insignificant.[24] Thus, increases in signing bonuses are associated with larger increases in Average WAR for pitchers compared to position players, although again the observed effect is very small.[25] The existence of this difference (in favor of pitchers) is somewhat surprising given that a pitcher's ultimate success is usually associated with more randomness. It could be

[23] A similar regression was estimated using career games played at the MLB level (g) as the dependent variable, as it can be considered yet another measure of player productivity. It is important to note there are two types of pitchers; Starting pitchers and Relief pitchers. Because Relief pitchers generally appear in more games than Starting pitchers, it is important that pitcher type is controlled for. However, the data collected does not provide the information necessary to divide the sample of pitchers into these two sub-groups. Thus, this model was estimated for position players *only.* The specification and results of this model are reported in Table A6 of the Appendix. Signing bonus does not appear to be a successful predictor of games played at the major league level.

[24] This is not to say that the effects for pitchers and position players, *individually*, are statistically insignificant. The relationship between signing bonuses and productivity for pitchers and for position players is statistically significant at the 1 percent level.

[25] Specifically, a one standard deviation increase in signing bonus is associated with a 0.09 s.d. increase AVG WAR for position players but only a 0.08 s.d. increase in AVG WAR for pitchers. This translates to 0.2 and 0.1 Wins Above Replacement respectively. These numbers represent extremely small, if not negligible effects. Thus while this difference may be statistically significant, it is not necessarily economically significant.

possible that over time, because pitchers represent the more "risky" investment, MLB teams have recognized this, and corrected any inefficiencies in the market by learning to better evaluate pitcher prospects and more accurately project their future productivity. However, further research outside the scope of this paper would be required to confirm this supposition.

Columns **(3)**, **(6)**, and **(9)** allow for a similar comparison to be made between high school and college draft picks. The coefficients (rows 1 & 3) in these columns indicate that signing bonuses are successful indicators of future productivity for both high school draft picks and college draft picks, but no significant difference in the magnitude of the effects for these groups exists.[26] For instance, a one standard deviation increase in signing bonuses for college draft picks is associated with a 0.63 increase in cumulative WAR, while a similar increase in compensation for high school draft picks is associated with a 0.94 increase in cumulative WAR. The insignificant margin between the two effects reveals that MLB employers do not have a significantly better ability to value high school draft picks over college draft picks, *in terms of productivity at the Major League level*. However, it could be the case that a significant difference will exist when other productivity measures are used to evaluate player performance.

The results of the second model, which examines the effect of compensation in the draft on a player's probability of reaching the Major Leagues, can help us investigate if such a claim is true. Results with and without the inclusion of interactions are shown below in Table 5:

[26] It is important to note the effects of signing bonus on future productivity for high school and college draft picks separately, are statistically significant at the 1 percent level. It is the difference between the two effects that is statistically insignificant.

Table 5: Probit Regression with Marginal Effects, MLAPP dependent variable[27]

n=3370	MLAPP (Dummy variable equal to 1 if player reached Major Leagues)		
	(1)	**(2)**	**(3)**
Sibo	0. 1188*** (.0156)	0.1036*** (.0185)	0.1573*** (.0225)
Sibo x pospl		0.0333 (0.0229)	
Sibo x hspck			-0.0659*** (.0249)

***Indicates significance at 99% confidence

The results of the second model suggest that there is a significant relationship between signing bonus and the probability of making an MLB appearance. As seen in column (1), a one standard deviation increase in signing bonus increases the likelihood of a player appearing in at least one Major League game by 9.6%.[28] One could argue that such a finding is caused by the fact that players who are chosen earlier in the draft (theoretically the most talented players) demand the highest signing bonuses. However, in the analysis the round in which a player is selected as well as his overall draft pick number is controlled for.[29] The results provide further evidence for the claim that MLB teams are able to value prospective talent in the draft; players who receive higher compensation have higher probabilities of providing positive returns on initial investments for drafting teams. The results in column (2) indicate that increases in signing bonuses have a positive effect on the likelihood of appearing in a Major League game for both pitchers and position players individually, but no significant difference in the magnitude of these effects is apparent. An examination of column (3) however, reveals some more interesting trends. The negative coefficient on *sibo x hspck* indicates that increases in signing bonuses are

[27] Signing bonuses scaled in millions. Thus, a $1 Million increase in signing bonus increases the probability of making the major leagues by 11.8% (Row 1, Column 1). Full Results can be found in Table A7 of the Appendix.
[28] This value is simply determined by multiplying the coefficient in (1) by the standard deviation of signing bonus.
[29] This essentially allows for comparisons to be made between players selected with the exact same draft pick in the draft. When one considers this, the results are even more striking. The fact that the magnitude of the effect is 9.6% *after controlling for draft position* significantly strengthens the belief that MLB teams are successful in evaluating talent and projecting productivity in the draft (i.e. players who receive higher compensation outperform those who receive lower signing bonuses).

associated with larger increases in the probability of success for college draft picks. Specifically, a one s.d. increase in signing bonus for college players *selected at identical positions in the draft,* is associated with an 11.8% increase in MLB appearance likelihood, while the effect for high school draft picks is just 7.94%. As indicated in Table 5 (Column 3), these effects are significant with 99 percent confidence. These results, taken together with those from Table 1, suggest that teams have an ability to better value college prospects in terms of selecting players that have the potential to reach the big leagues but a harder time predicting how players will perform once they reach the major league level.

A similar model was estimated to examine whether or not signing bonuses influence a draftee's chances of making an All-star Appearance. The results in Table A8 (appendix) indicate that there is no significant relationship between signing bonus and All-Star appearance(s). However, this is not very surprising. Of the 3,370 draft picks included in the sample, only 123 (approximately 4%) became All-Stars within their first six years of Major League Service. Moreover, in most years, the majority of players who make up All-Star Team rosters are veteran players; rarely do rookies or amateurs get selected to the All-Star game. However, such a trend can more likely be attributed to the All-Star voting process rather than an objective evaluation of player productivity. [30]

The results described above suggest that increased draft pick compensation influences the odds of making the big leagues (controlling for draft round and pick), and to a much lesser extent, his subsequent productivity once there. Teams seem to value highly (and compensate them accordingly), players who have an above-average ability to reach the Major leagues.

[30] For instance, the starting line-up for both the American and National leagues are determined solely by fan voting. This accounts for roughly one-third of all players voted to the All-Star game. Such a player-selection method is often criticized because a situation could occur in which a majority of the players selected originate from teams that have large fan bases. Other criticisms claim that being selected for the All-Star game is a "popularity" contest rather than a proper award for successful performance. In either case, it is of little surprise that signing bonus has little weight in determining a player's overall chances of being voted into the all-star game.

Furthermore, these valuations are more accurate for college draft picks compared to those selected out of high school. Projecting success for players once they are at the Major League level however seems to be a more challenging task.

Testing the Efficiency of MLB Signing Bonus Expenditures

The evidence presented suggests that MLB teams are at least, marginally successful in their evaluations and projections of prospective talent in the draft. When either WAR values or the probability of making an MLB appearance are used as dependent measures of productivity, the results prove significant. Thus to a certain extent, it is evident that increases in signing bonuses are associated with increases in productivity. This trend reveals that, overall, MLB decision-makers make sound investments, but it does not provide information on how efficient MLB teams are in allocating signing bonuses. For instance, players selected in Round 1 have both the highest average signing bonuses and the highest mean productivity compared to players selected in all other rounds. But does the magnitude of the increased compensation received by these players match accordingly with the excess productivity they provide? That is, on a per dollar basis, are first round draft picks more productive than players chosen in later rounds? To examine how the "value" of draft picks change as one progresses later into the draft, and to gain a greater understanding of how efficient MLB teams are in draft expenditures, a simple regression examining the relationship between Average WAR-per-dollar[31] (measured in millions) and Draft Round[32] is estimated. Although data was collected for players selected in

[31] Again, Average WAR is simply defined as a player's Cumulative WAR value divided by his years of Major League service. Average WAR is used over Cumulative WAR in order to account for the fact that players have careers of varying length (players with longer careers generally will have higher Cumulative WAR statistics than those with shorter careers).

[32] In this simple model, draft round is estimated using a series of dummy variables (with each one corresponding to a different round). The omitted group is Round 1. Another model was estimated using a continuous variable *round.* The results of this regression report similar insignificant results. See table A9 of the Appendix. Furthermore, both models are run only on players who reached the Major Leagues (n=1130). The reason for this is that Minor League players do not have a recordable WAR. It would not be appropriate to include players who do not reach the Majors (and give them a WAR of 0) because WAR can take on a negative value for Major league players.

rounds 1-10, the analysis presented here is conducted only on players selected in the first *five* rounds[33]. Results are shown below:

Table 6: OLS Regression Explaining Average War-Per-Dollar (millions): Rounds 1-5

Average WAR/Dollar (millions)	Coefficient	Standard Error
round_2	-.01013	2.67
round_3	-0.1122	2.96
round_4	-0.2746	3.38
round_5	0.2716	3.58
constant	0.7913	0.097

The results indicate that round does not have a significant impact on the Average WAR provided per $1 Million spent on signing bonuses[34]. As seen in Table 6, the insignificance of these results implies that teams do not get more "bang for their buck" at any point throughout the first five rounds. Compared to first round draft picks (reference group), players selected in the third or fifth round do not appear to be any more "valuable" (in terms of WAR-per-dollar). Nonetheless, a more rigorous analysis outside the scope of this research would be required to verify whether teams are "efficient" in their draft spending.

[33] In later rounds (6-10), the odds that a player reaches the Major Leagues are very small. For instance, in round 6 only 20% of players drafted between 1989-2009 reached the Majors. By round 9, this figure drops to 10%. Since so few players from these rounds make the Majors, those who do, tend to be extremely talented (i.e. have high WAR values). Thus, inclusion of these observations could lead to selectivity bias and a misrepresentation of the results. This conjecture is supported when all *ten* rounds are included in the analysis. See Appendix A10 for results of this regression and further discussion.

[34] A third model that only examined the first round was also estimated. Results and discussion can be found in Table A11 of the appendix.

CONCLUSION

In Major League Baseball, or any professional sports league, the amateur draft provides a great platform for economic inquiry. Over the past 20 years, the MLB draft has experienced unprecedented increases in signing bonuses, sparking much discussion, but little research. The question of how economically efficient MLB teams are in the draft has fueled this study and its attempts to determine whether major inefficiencies exist in the market for MLB draft picks. More specifically, this research seeks to examine whether there exists a significant relationship between signing bonus compensation and future productivity for players selected in the amateur draft, holding constant where a player is selected in the draft.

There is statistical evidence to conclude that draft pick compensation predicts subsequent performance in Major League Baseball. There exists a small, yet significant, positive relationship between signing bonuses and a player's WAR value. The effect remains significant when other variations of WAR, namely Discounted WAR and Average WAR, are used as dependent variables. Furthermore, the relationship between signing bonus and Average WAR is stronger for pitchers compared to position players, suggesting MLB signing bonuses represent more accurate valuations of talent for these groups. A significant relationship also exists between signing bonuses and the likelihood a draft pick reaches the Major League level. I find that the relationship between signing bonus and the probability of MLB appearance is strongest for college draft picks. However, no significant differences are found in the effects for pitchers and position players. These effects of signing bonus on WAR and MLB appearance taken together suggest that teams have the ability to better value prospects in terms of selecting players that have the potential to reach the big leagues but a more difficult time predicting how those players will perform once they reach the major league level (yet they are still somewhat successful in doing so). For instance, *holding constant the exact round and pick a player is selected in(with)*, a one s.d. increase in signing bonus results in a 9.6% increased likelihood of reaching the Major League level. However, the effect of signing bonus on future WAR was of a much smaller magnitude; a one s.d. increase in signing bonus is associated with a 0.75 increase in Cumulative WAR. This result might be expected. To believe that MLB teams can predict perfectly the future productivity of draft picks, one would have to disregard the randomness inherent in professional sports. The fact that *any* effect was found between signing bonuses and subsequent productivity after controlling for where in the draft a player is selected strengthens the notion that Major League Baseball organizations are accurate in the valuations of prospective talent. Increases in signing bonuses appear to be justified through increases in productivity.

REFERENCES

Berri, David J.; and Simmons, Rob. "Catching a Draft: On the Process of Selecting Quarterbacks in the National Football League Amateur Draft." *Journal of Productivity Analysis* 35(2011):37-49

Hendricks, W., DeBrock, L., & Koenker, R. "Uncertainty, Hiring, and Subsequent Performance: The NFL Draft." *Journal of Labor Economics*, 21,4 (Oct. 2003), pp. 857-886.

James, Bill. *The New Historical Baseball Abstract.* New York: Free Press, 2001. Print.

Massey, Cade and Thaler, Richard H. "The Loser's Curse: Overconfidence vs. Market Efficiency in the National Football League Draft." *Social Science Research Network* no. 697121(2010).

Quirk, James P. *Pay dirt: The Business of Professional Team Sports.* Princeton, New Jersey:Princeton University Press, 1992. Print.

APPENDIX

A1: Percentage of Players Drafted from HS versus College, by round

Round	% of Players Drafted from HS	% of Players Drafted from College
1	0.46	0.54
2	0.48	0.52
3	0.45	0.55
4	0.4	0.6
5	0.26	0.74
6	0.31	0.69
7	0.28	0.72
8	0.23	0.77
9	0.23	0.77
10	0.22	0.78

A2: F-test Results Comparing Fits of Models

General Specification of the Constructed F (for all F tests conducted)

$$F = \frac{(RSS_r - RSS_{ur})/\,m}{RSS_{ur}\,/\,(n-k)}$$

Where: r = the restricted model (coefficients are assumed to be equal to a certain value)
ur = the unrestricted model (coefficients can take any value)
m = the number of restrictions in the restricted model
n = the number of observations
k = the number of parameters estimated in the unrestricted model

The first F-test conducted compares the fits of two models that differ only in the way draft round (the round in which a player is selected) is controlled for. The first model controls for it by including a continuous *round* variable measuring round number. This is the restricted model as simply including *round* as an independent variable results in certain constraints being placed on the independent variables used. Specifically, it is assumed, and is the null hypothesis that the effect of draft round on productivity is the same for all ten rounds (effect for rd 1 = rd2 = rd3…=rd10). Thus there are nine restrictions in this model. The second model controls for round by using a series of ten dummy variables, namely round_1, round_2, …, round_10. The model using round dummies represents the unrestricted model as the effect of draft round on productivity can differ for each individual round (Alternative Hypothesis)

If the null hypothesis is true (that the effect of draft found on productivity is the same for all ten rounds), then using the unrestricted model (round dummies) does not help to more accurately explain the variation in player productivity. At the 5% level, the critical value of F with m degrees of freedom in the numerator and n-k degrees of freedom in the denominator is 1.88. In this first case, $F=0.509 < 1.88$. Therefore, the null hypothesis cannot be rejected. This implies that using a series of round dummies does not provide a significant increase in fit for the model. As a result, *round* is included in all models as a continuous variable.

A similar test was conducted to examine the best way to control for where, *within each round*, a player was drafted. There are again two options. The first is a series of 64 dummy variables each corresponding to a different pick within round (the most players selected in any round between 1989-2009 was 64). This is the unrestricted model as the effect for every round pick can take different values. The alternative is to include a variable called *rd_group*, which as described earlier, is split into only four dummy variables: *rd_group 1, rd_group2, rd_group3, rd_group4*. Each of these dummy variables correspond players selected with picks 1-10, 11-20, 21-30, and 31+, within each round. This alternative model is subject to 60 restrictions. Here, the critical value of F, at the 5% level, is approximately 1.3. In this case, $F= 0.8607 < 1.3$. Thus, the unrestricted model using a series of 64 dummy variables does not provide a significantly better fit for the data. As a result, the variable *rd_group* was incorporated into all models.

In both instances, the "restricted" models are chosen. In light of this, it is important to note that no significant differences (in terms of the coefficients estimated) exist between the results of the restricted and unrestricted models .

A3: OLS Regression , Cumulative WAR dependent variable (full results)

Cumulative WAR	**Coefficient**	**Standard Error**
sibo	0.000000697***	1.75E-07
rnd	.0188092	0.0759586
g	0.0108137***	0.0005068
div1	-0.030397	1.447022
div2	1.814606	2.702856
div3	1.262172	4.859657
hspck	-0.0245355	1.470143
pospl	-1.942137***	0.3083505
servicetime	0.5369976***	0.1020934
rd_group2	-0.3399724	0.3577597
rd_group3	-0.351253	0.3664886
rd_group 4	-0.940163	0.5674617

*indicates significance at 90% confidence
**indicates statistical significance at 95% confidence
***indicates significance at 99% confidence

A4: OLS Regression, Discounted WAR dependent variable (full results)

Discounted WAR	**Coefficient**	**Standard Error**
sibo	0.000000581***	1.63E-07
rnd	0.0271247	0.0709146
g	0.0097992***	0.0004734
div1	-0.4339204	1.350915
div2	1.400584	2.523342
div3	-2.658175	4.536919
hspck	-0.7423624	1.372517
pospl	-1.946071***	0.2878777
servicetime	0.4519022***	0.0953456
rd_group2	-0.2770513	0.3342065
rd_group3	-0.4996306	0.3424265
rd_group 4	-0.8694453	0.5298618

***indicates significance at 99% confidence

A5: OLS Regression, Average WAR dependent variable (full results)

Average WAR	**Coefficient**	**Standard Error**
sibo	0.000000149***	3.46E-08
rnd	0.0069281	0.0150463
g	0.0017934***	0.0001004
div1	-0.147781	0.2866335
div2	0.3907298	0.5353956
div3	0.4290299	0.9626259
hspck	-0.0545868	0.2912135
pospl	-0.3265755***	0.0610797
servicetime	0.1303464***	0.0202232
rd_group2	-0.0699594	0.0708669
rd_group3	-0.0617914	0.072596
rd_group 4	-0.2463193	0.1124057

***indicates significance at 99% confidence

A6: OLS Regression- Dependent Variable: Games Played at Major League Level (estimated for position players only)

(n=559)	**Coefficient**	**Standard Error**
sibo	0.0000294	0.0000182
rnd	9.7564	7.7639
hspck	37.6313	31.4832
servicetime	152.6019***	9.5771
rd_group2	-0.9795	35.619
rd_group3	-70.6068***	36.7671
rd_group4	-17.606	59.2201

*** indicates significant at 99% confidence

As seen in Row 1 above, signing bonus is not a successful predictor of the number of games a position player participates in during his first six years of MLB service. This is not so surprising. There are non-quantifiable team specific factors, such as roster depth, which can significantly influence the number of games a position player participates in. Additionally, player position influences the number of MLB appearances one records. Dummies for individual position would satisfy this concern but the data obtained does not provide the necessary information. Injuries also add an element of randomness that cannot be captured. Nonetheless, in spite of the insignificant results here, when other measures of productivity are used in the analysis, signing bonus appears to be a successful predictor of future productivity.

A7: Probit Regression with Marginal Effects, MLAPP Dependent Variable (full results)

MLAPP	Coefficient	Standard Error
sibo	.000000119***	1.56E-08
rnd	-.039438***	0.0043164
div1	0.1012369	0.0627975
div2	0.026766	0.1269946
div3	0.0234866	0.2444156
hspck	-.0493964	0.0655185
pospl	0.0234866	0.0171994
rd_group2	-0.0291256	0.0215487
rd_group3	-0.0130095	0.021927
rd_group4	-.0423381	0.0353574

***indicates significance at 99% confidence

A8: Probit Regression with Marginal Effects, ALLST Dependent Variable (full results)

	ALLST (Dummy variable equal to 1 if player selected to an All-Star Game)		
	(a)	**(b)**	**(c)**
Sibo	1.61e-08 (1.02e-08)	2.09e-08 (1.21e-08)	2.05e-08 (1.17e-08)
Sibo x pospl		-1.15e-08 (1.53e-08)	
Sibo x hspck			-1.16e-08 (1.56e-08)
rnd	-.0014578 (.0048786)	-.0015937 (.0048815)	-.0010958 (.0048855)
servicetime	.0067444 (.0064371)	.0065135 (.0064388)	.0067135 (.0064221)
g	.0000179 (.0000319)	.0000204 (.000032)	.0000181 (.0000318)
hspck	.015837 (.0196422)	.0173882 (.019812)	.0292716 (.0273144)
pospl	.0035761 (.0201178)	.0145523 (.0249361)	.0049009 (.0201641)
rd_group2	.0141426 (.0236419)	.0144394 (.0236696)	.0147559 (.023653)
rd_group3	.0070995 (.0242043)	.0071573 (.0241899)	.0074732 (.0242056)
rd_group4	-.0268773 (.0309845)	-.0265256 (.0310199)	-.0255104 (.0313308)

A9: OLS Regression Explaining Average WAR/ $Million (Rounds 1-5) (n=921)

	Coefficient	**Standard Error**
round	0.00031	0.0557
constant	0.75773	0.1326

As mentioned earlier, estimating Average WAR/$Million using round number as an independent variable rather than a series of dummy variables yields insignificant results. There does not appear to be any significant relationship between Average WAR/$Million and draft round.

A10: OLS Regression Explaining Average WAR/$Million (Rounds 1-10) (n=1130)

AVGWAR/Dollar (millions)	**Coefficient**	**Standard Error**
round_2	-.0101	2.666
round_3	-0.112	2.964
round_4	-0.275	3.379
round_5	0.272	3.576
round_6	0.533	3.998
round_7	20.069***	4.462
round_8	8.948**	4.559
round_9	18.128***	5.512
round_10	-2.838	4.718
constant	0.791	1.259

**Indicates significance at 95%
*** Indicates significance at 99%

As mentioned earlier, the majority of players who made the major leagues between 1989 and 2009 were drafted within the first 5 rounds. Only 209 of the 1130 major league players in the sample were selected in rounds 6-10. Thus, it is not surprising that the coefficients for round 7, 8, and 9, are significant. Although it appears as if players drafted in rounds 7-9 are more productive on a per-dollar basis than first round draft picks, these effects are likely capturing the performance a few players who happened to be very successful. For this reason, these observations were excluded from the analysis shown in Table 6 of the text.

A11: OLS Regression Explaining Average WAR/$Million by Round Pick (1st Round Only)

(n=415)	**Coefficient**	**Standard Error**
roundpick	-0.002	0.013
constant	0.911	0.214

In the table above, *roundpick* refers to the pick number (1-30) of each player drafted in the first round. The insignificant coefficient on this variable indicates that player productivity per dollar spent does not significantly differ as one moves later into the first round. The first overall pick does not appear to provide any more(less) value (in terms of WAR/dollar) than the 30th pick. Using a series of 30 dummy variables in the model (with each corresponding to a different pick number) instead of the continuous *roundpick* variable yielded similar, insignificant results.