Major League Baseball Team Building: Should Franchises Focus on the Amateur Draft or Free Agency?

Matthew Duran

Abstract:

This paper analyzes the relationship between the Major League Baseball amateur draft and team winning percentages from 2005 to 2018. The MLB is considered a free agency dominated league, and studies have proven that the voluntary movement of players has altered competitive balance. Large market teams exert financial pressure on smaller market teams by buying away their best players. To counteract this effect, the league instituted the amateur draft to balance the playing field, affording small market teams the ability to acquire better players on favorable terms. However, success rates of finding impactful players through the draft are low. To test the statistical significance of the amateur draft, I performed a two-stage least squared regression analysis that models the impact that draft success and payroll have on winning percentages. Understanding that drafting ability is not quantifiable, I use modern scouting theories to develop my first stage regressions and predict the rate at which teams find undervalued players. The second stage regressions formally test whether consistent draft success or free agency has a more statistically significant effect on winning. Even with significant investments in scouting and player development, this study refutes the theory that the amateur draft significantly impacts competitive balance.

ECON 4911 Sports Economics

Sunil Gulati

April 22, 2019

Section I: Introduction

Major League Baseball (MLB) is America's favorite pastime. The country has indulged in watching the larger-than-life personalities of Babe Ruth and Micky Mantle, who produced prodigious records that are still currently chased today. Whereas the modern generation of players, such as Mike Trout and Aaron Judge, not only dominate the competition but transcend the sport into the world of sponsorship and marketing. The business side of baseball has similarly transitioned from an afternoon escape for families to a lucrative business in which owners operate their teams like CEOs of major companies. As the chief executives of their respective teams, owners face a simple economic problem of maximizing their profits while minimizing their costs. In order to solve this problem, owners need to field the best possible teams that they can with the hopes that the increase in wins will lead to higher revenues. Profit maximization theory holds for many owners, but there are famous examples of large market owners who fielded winning teams at any cost, like George Steinbrenner. It is no coincidence that his franchise, the New York Yankees, has won the most championships in professional sports. Commissioners have noticed this inequity and have implemented policies, such as the luxury tax and compensatory draft picks (for teams that lose players in free agency), that attempt to neutralize the spending advantages that larger markets have over smaller markets. While fans may love watching and debating the historical accomplishments of dynasties, increased parity is better for business.[1]

Major League Baseball seeks to promote parity in the league through the First-Year Player Draft, which takes place every June. The draft order is determined by the reverse ranking of the prior year's winning percentage among all MLB teams, allowing the opportunity for the worst teams to select the best possible players available. Not only do these teams get the best players, but they get them on very

[1] Rodney Fort, "Sports Economics: Making Sense of It All," in *Sports Economics* (Pearson Education, 2014), 13. Fort cites Simon Rottenberg's Uncertainty of Outcome Hypothesis which states that fans care about the competitive balance in professional sports, preferring for their team to win close games over blowouts.

favorable terms.[2] Each year, thirty teams select players through fifty rounds of selections, which translates into fifteen hundred players drafted overall.[3] Teams have historically been bad at predicting which talented young prospects will make it from the draft room to the major league clubhouses. Unlike the National Basketball Association and National Football League, where players are drafted directly into the clubs in the league, MLB sends ninety-nine percent of its players to its farm system for a few years to mature physically and hone their skills. The delay in player development makes drafting players difficult, and it takes years for the "big club" to enjoy the payoff of its top prospects on its bottom line, if they ever get to.

Naturally, teams pay close attention to the draft because of the importance of finding young, affordable talent. More than any other sport, baseball games are inherently measurable and rich with statistics; every at-bat and pitch are cataloged for future reference. Throughout the draft's history, picking players was left mainly to a scout's gut feeling, leading to mixed results. With the recent emphasis on business optimization in baseball front offices, teams have questioned why they subject their draft processes to such a high degree of human error. In recent decades, books like *Moneyball* and *Astroball* have explained how performance statistics can be used by a team's front office to optimize the process of building through the draft.[4] This literature summarizes how sabermetricians have shifted conventional thought on the use of Major League Baseball statistics. Statistics are no longer fodder for nostalgic debates on which players had better careers. Instead, they are valuable data points to be used in regression analysis and statistical inference.[5] Small market baseball teams, like the Oakland Athletics and Houston Astros, have garnered heightened attention for their sabermetric philosophies which emphasize building

[2] It is well documented how Major League Baseball suppresses the wages of these minor league players, who make far below the federal poverty level on an annualized basis.
[3] In 2012, the MLB draft was cut from fifty to forty rounds.
[4] Written by Michael Lewis and Ben Reiter respectively.
[5] *Sabermetricians* are practitioners in the industry that take a detailed statistical approach to evaluating player performance.

through the draft. Both organizations have been able to achieve high winning percentages despite significant payroll disadvantages.

This paper seeks to empirically test the null hypothesis, which is that teams that draft better tend to have higher winning percentages, against the alternative hypothesis, which is that other variables like payroll—and by extension free agency—have a more causal effect on winning. The way I propose to test this hypothesis is through a two-stage least squared regression analysis that models the impact of draft success and payroll on winning percentage. Beginning with Section II, I provide an overview of the valuation of players through the determination of signing bonuses, the dependent variable of my predictive model that generates an *ex post* assessment of a player. Section III provides previous economic analyses of the effect that free agency has had on winning percentage, which are tangential arguments that illustrate the alternative hypothesis of this paper. Section IV analyzes current talent scouting theories that are the foundation of my models and includes logic regarding which performance metrics are the most predictive variables. Section V covers the data set and my methodology, which grades the thirty major league teams over a ten-year period from 2005 to 2015. Section VI discusses the results of my hypothesis test and ultimately determines whether drafting ability or free agency (tested via team payroll as a proxy) is more statistically significant. The last section addresses the limitations of the study and the opportunities for further research.

Section II: Determination of Player Worth

In the build-up to the MLB draft each June, teams deploy all of their scouting resources to scour the United States to find players that could contribute to a winning ballclub in the future. There are major financial incentives at play because the teams have exclusive rights to sign the player that they draft at a considerable discount on what he could truly be worth, and can retain him for almost a decade.[6] The scouting pipeline offers the opportunity to obtain a strong set of potential big league players at affordable prices. Naturally, there is a fierce debate in the industry regarding the merits of major draft-eligible groups: high school seniors, college juniors and seniors, and recently, international players.[7] Once these players are drafted, teams offer the players exorbitant bonuses which entice them to sign by compensating them for their opportunity costs of higher education and higher paying jobs. The determination of signing bonuses is not a simple task because numerous factors could possibly hide the player's actual ability.

The process of determining a valuation of the player is no different from valuing a stock. When banking analysts value AAPL stock, they project the company's future cash flows, discount them to the present time using an appropriate risk rate, and sum them up to yield the stock price. In his 1974 paper *Pay and Performance in Major League Baseball*, Gerald Skully argues that general managers face the same task when valuing their players. Although this paper featured a wage analysis of players during the Reserve Clause Era, a time of restricted player movement, the theory still applies to my more modern study. Skully claims that each player should be paid according to their marginal revenue product (MRP), meaning that owners need to determine beforehand how the player's future performance will indirectly impact overall revenues, and then discount these cash flows to the present time.[8] To determine MRP, Skully's hypothesis states that fan attendance and revenue positively respond to team wins, while players have a direct

---

[6] Noah Streib, Stephen J. Young, and Joel Sokol, "A Major League Baseball Team Uses Operations Research to Improve Draft Preparation," *Interfaces 42*, no. 2 (2012): 119-120.

[7] International players are not subject to the MLB draft and go through an auction type bidding for their services. These players remain in the analysis since they sign for comparable money and at similar ages as high school and college players.

[8] Gerald W. Skully, "Pay and Performance in Major League Baseball," *The American Economic Review*, no. 6 (1974): 918.

impact on winning. The MRP is the product of these marginal effects, which is the theoretical upper bound on what an owner will pay.[9]

When ball clubs use the theory that players should receive their marginal revenue product, owners and players face off in direct negotiations seeking to steer the terms of the deal in one or the other's favor. Different players have different leverage opportunities in these negotiations, such as the overall number where they were selected in the draft and the alternative opportunities for employment or education should they forego entry into the MLB. In *The Determination of Bonuses in Professional Sports*, Gary Neil Ross sets up a utility maximization function with risk-averse players and risk-neutral owners. He observes this dual problem and examines the incentives on both sides of the negotiating table. On the owner's side, if a player's expected net return to the club is greater than zero, the owner will be willing to offer the player a bonus. Because players bear most of the risk in this scenario, they will want to be compensated for various risk factors, like injury, retirement, and educational opportunities.[10] Training costs are a critical variable in this maximization exercise. Because baseball players require a few years in the minor leagues, owners factor these costs into the player's probability of success and deduct these costs from the bonus amount. Training costs are directly related to player skill, which allows higher draft picks, who typically require less time in the minor leagues, to obtain higher signing bonuses.[11]

In my models, I extend Skully and Ross's economic theory on player valuation. Because minor league and major league salaries vary among the terms of each contract, I use a player's signing bonus as the metric of what a player is worth. Each team spends months analyzing a player's probability of success in professional baseball and attaches a value to the player's skill in the form of his signing bonus. In Section V, I run regressions using a player's observed performance statistics as key independent variables and

---

[9] Ibid., 919.
[10] Gary Neil Ross, "The Determination of Bonuses in Professional Sports," *The American Economist* 19, no. 2 (1975): 43.
[11] Ibid., 45.

determine that player's predicted signing bonus on an *ex post* basis. These predicted valuations will measure each team's drafting ability by examining how accurately they were able to assess each player's value.

Section III: Previous literature

Prior to 1975, the Reserve Clause severely restricted player movement throughout the league. For the most part, if a team drafted a specific player, it was expected that he would stay with the team for his entire career. This rule was the subject of many anti-trust lawsuits, with the most famous being the Kurt Flood case in 1969. Flood was traded from the St. Louis Cardinals, where he spent the first twelve years of his career, to the Philadelphia Phillies. He refused to report to the Phillies, emphatically stating that he was not "a piece of property to be bought and sold irrespective of [his] wishes."[12] Although Flood lost his high-profile case, he set the groundwork for the Reserve Clause's repeal seven years later, which granted players the ability to consider offers from multiple clubs at the same time. Many sports and labor economists use 1976 as a base year to study the before and after effects of the Reserve Clause's repeal and determine the effects that free agency has had on the spread of talent in Major League Baseball.

In his 1999 paper, *Free-Agency and the Competitiveness of Major League Baseball*, Craig Depken examines whether the distribution of team wins was affected by the voluntary movement of free agents during the 1920 to 1996 period. Depken hypothesizes that free agency adversely affects the competitiveness in the league overall because the ability of big market teams to purchase all of the best players in the league limits competition over time. However, because teams are businesses and have rational owners, they seek to maximize their profits by adhering to the Coase Theorem, which states that

[12] Mary Craig, "Chained to the Game: Professional Baseball and the Reserve Clause, Part Two," Accessed March 5, 2019. http://www.beyondtheboxscore.com/2017/6/10/15766702/curt-flood-mlbpa-reserve-clause-free-agency.

teams will only hire additional labor if their marginal revenue product equals their wage rate.[13] Depken's econometric model regresses the standard deviation of winning percentages on five important independent variables, all of which are dummy variables. These include free agency, league expansion, integration, designated hitter and time-lagged parity measures. The result of this econometric analysis rejects the Coase Theorem and deduces that "the distribution of talent [across the league] determines the quality of teams and thus the distribution of wins across teams."[14]

In 2003, Peter Fishman extended Depken's model in *Competitive Balance in Free Agency in Major League Baseball*. Fishman similarly tests Ronald Coase's logic in a study of the league from 1950 to 2001. He models winning percentage as a function of eight control variables for free agency, including the size of the free agent pool, the number of years since free agency began, the presence of the amateur draft, the percentage of the population that plays the game, and the short and long term effects of league expansion. This regression is more robust than its predecessor because it transforms a simple dummy variable for free agency into more explanatory quantitative variables. Furthermore, Fishman includes the presence of the amateur draft in his econometric model as a dummy variable, which is a key addition. With all of these controls in his regression, the coefficient on the variable for the size of the free agent pool is positive and significant at the 5% alpha level. Thus, Fishman concludes that free agency does have an effect on competitive balance and states that "this positive coefficient supports the owner's argument that free agency harms competitive balance."[15]

Craig Depken and Peter Fishman's respective econometric models and results are the bases of my alternative hypothesis, which is that free agency has a significant causal effect on competitive balance in Major League Baseball. Peter Fishman concludes his paper stating that "previous research into the effect

---

[13] Craig A. Depken, "Free-Agency and the Competitiveness of Major-League Baseball," *Review of Industrial Organization* 14, no. 3 (1999): 206.
[14] Ibid., 210.
[15] Peter Fishman, "Competitive Balance and Free Agency in Major League Baseball," *The American Economist* 47, no. 2 (2003): 89.

of free agency on competitive balance has used a dummy variable … [and] the result of this approach has been that most researchers have found the coefficient of this dummy variable to be insignificant."[16] Interestingly, Fishman's model solely controls for the amateur draft and concludes that it has a positive effect by allowing worse teams to draft better players. I extend Fishman's logic regarding dummy variables by empirically testing the amateur draft with a two-stage regression. First, I use performance statistics and signing bonuses to determine which teams drafted better between 2005 and 2015. Second, I use the draft results to test against free agency.

Section IV: Scouting Theory

Depken and Fishman's conclusions, which I discuss in this paper's previous section, provided the motivation to examine the true effects that the MLB amateur draft has on competitive balance. Every June, the MLB draft takes center stage and is the most crucial part of a general manager's schedule. Each team has its eyes on the future, and for teams to sustain success, conventional wisdom says that they need to consistently draft high quality players. Understanding that drafting well is essential to sustained success, it is important to analyze how players are found and evaluated. In the industry, there are three main drafting theories, all of which are at odds with one another. They are the eye-test (employed by full-time scouts), the Moneyball model (driven solely by statistics), and the Astroball model (a blend of two).

Ever since the game of baseball organized into a structured league with teams and owners, scouts have been employed to find players that will contribute to winning teams. Scouts are the old-guard of baseball tradition and believe that the key to finding a great player is by driving thousands of miles and staying in low-quality hotels. They watch over two hundred games a year hoping to find one diamond in the rough. Scouts evaluate players during warm-ups, during competitive games, and during pregame and

[16] Ibid., 90.

postgame workouts. Throughout this evaluation process, scouts analyze players' build (height and weight) and determine how many "tools" each player possesses, such as the ability to run, throw, field, hit, and hit for power. Each scout has his own historical database in his mind where he houses memories about every player he has seen. Scouts use their historical experiences to find a comparable major league player to the amateur being scouted. The scout's job is to project how and whether seventeen to twenty-year-old amateurs can throw or hit a ball with millions of dollars on the line. In *Moneyball*, Michael Lewis describes Oakland Athletics general manager Billy Beane's philosophy of drafting players. He begins by describing the infatuation scouts have for high school players, especially high school pitchers, because they aren't close to the player they will become when they grow into their bodies in four to five years. However, studies on the draft have determined that these players are the riskiest and seldom make it to the big leagues.[17] After the draft, if you listen to scouts defend their recent high school pick, they customarily will say "you need to look at the guy. *Imagine* what he might become."[18] Scouts' affinity for high school players are misguided, and their habit of poorly predicting the future abilities of such players forms Billy Beane's opinion that scouts are the problem with the system. Beane's beliefs have led to an overhaul of how teams operate in the context of selecting players.

Billy Beane never truly forgave scouts for what turned out to be an incorrect projection of his own career. He was a "can't miss," "five-tool" future superstar, who fizzled out in the New York Mets farm system. When he became the general manager of the Oakland Athletics in 1997, he vowed that he would "rip away from the scouts the power to decide who [could] become a professional baseball player."[19] In essence, he wanted to remove the human element from player evaluation and transition to a data-driven system. Beane's first significant hire as general manager was Harvard graduate Paul DePodesta, an economics major who was adept at running regressions and using statistical inference to understand

[17] Michael Lewis, *Moneyball: The Art of Winning an Unfair Game* (W.W. Norton, 2003), 16.
[18] Ibid., 32.
[19] Ibid., 18.

patterns in data. Beane and DePodesta focused on gathering a large sample size of relevant data, which enabled them to obtain a more accurate understanding of a player's abilities. Their philosophy was rooted in the sabermetric field that baseball historian Bill James created in the seventies with his annual *Baseball Abstract*. Bill James noticed that the game kept copious notes on statistics, but solely used them for debate rather than statistical analysis. James was one of the first sabermetricians to use player statistics to run regressions and found that some of the more famous statistics did not have causal effects on winning percentage. He steered his philosophy away from batting average (number of hits / by number of at-bats), which was not an accurate measure of hitting skill, and created a new formula named runs created ((hits plus walks) x (total bases) / (at-bats + number of walks)). This metric calculates the total share of a team's runs scored that could be attributed to a specific player.[20] Billy Beane then dug deeper, evaluating the ability of hitters to control the strike zone, called on-base percentage, and the ability to hit for power, called slugging percentage.[21] Sabermetricians today use the summation of on-base percentage and slugging percentage (OPS). With these innovations, Beane turned around the Athletics franchise, but skeptics remained because the team's regular season success has not translated into postseason success. Criticisms point to Beane's robotic way of analyzing players and his removal of scouts from the equation as major reasons why the Athletics have not won a championship under this philosophy.

With the advent of Beane's data-driven analysis in baseball, small market teams all over the league have attempted to integrate the Moneyball philosophy into their own club's culture. The small market and historically dreadful Houston Astros were able to tweak Beane's formula and produce a championship, which was a win for sabermetricians everywhere. When Jeff Luhnow left the storied St. Louis Cardinals, an organization traditionally known for its excellent scouting department, and took the reins as general manager for the Houston Astros, he wanted to combine solid scouting with Beane's data-

[20] Ibid., 76.
[21] Ibid., 58.

driven analysis. In *Astroball*, Ben Reiter comments that *Moneyball* portrays scouts as antagonists, while Luhnow framed the equation including both scouting reports and performance measures.[22] Luhnow finds the value in every data entry, and thus combined Beane's underlying philosophy with the "recently overlooked source of information: humans."[23] Furthermore, he believes that scouts have prophetic information from their experiences and observations. His team developed algorithms that included scouts' immeasurable "gut-feelings" concerning a player's mechanics and personality. Luhnow and his team graded their scouts by analyzing the players that the scouts selected, relative to the projections that the scouts made at the time. This analysis effectively produced a weighting system that would be applied to a scout's opinion during the draft process.[24] This synthesis of qualitative and quantitative variables has become the new sabermetric model, and is one with more predictive power that can yield results, as evidenced by the Houston Astros World Series championship in 2017.

Currently, there is a push and pull dynamic between old school scouts and data-driven general managers, and there is no consensus philosophy that dominates the baseball landscape. On the surface, each team appears to rely on all three philosophies, but they tend to give more weight to the one that they believe in the most. However, following the release of Michael Lewis's *Moneyball* and Ben Reiter's *Astroball*, teams have begun to embrace machine learning and data analysis. Because we do not know which philosophy that each team employs, other than the Athletics and Astros, I used all of them and developed three models. The first relies on qualitative-scout based theory. The second relies on Billy Beane's data-driven Moneyball theory. The third relies on Jeff Luhnow's synthesis of quantitative and qualitative variables. Using these models, I can determine how well each of the thirty teams find undervalued amateur players in the draft. I then average out the results, compensating for not knowing the particular theory used by each team.

---

[22] Ben Reiter, *Astroball: The New Way to Win It All* (Crown Archetype, 2018), 28.
[23] Ibid., xiv.
[24] Ibid., 49-50.

Section V: Data Analysis and Empirical Models

The data set that I have gathered contains statistics and other biographical information for players drafted from 2005 to 2015.[25] The sample size is large with a total of 5,454 players, 2,723 of whom are hitters, and 2,731 of whom are pitchers. The data set does not contain every round for the ten drafts during the analyzed time period. I collected data only for players for whom a record of a signing bonus was available. As stated in Section II, signing bonuses are the initial valuation of a player by his team, and this number will be compared with the *ex post* valuation that is predicted according to his performance history. The regression models use a player's career minor and major league statistics, beginning when the player signs his contract and ending in 2018. One issue with the study is that the minor league system is challenging to evaluate because of the delay in player skill development. I accounted for this delay needed for players when deciding on the time period of the study, cutting it off at 2015. However, there may still exist inaccurate valuations for players drafted later in the period since they have a smaller sample size of statistics compared to players selected in 2005, which could potentially have thirteen years of data. For the purpose of this analysis, hitters and pitchers are split up because no statistic exists that can normalize the performance of these different groups of players.

The ultimate goal of these econometric regressions is to determine which teams in Major League Baseball are the best at drafting undervalued players. These results will provide a meaningful metric that will be regressed on winning percentage and resolve the outcome of the study. The first stage of the hypothesis test is the development of six models, including three for hitters and three for pitchers. These models logarithmically regress signing bonuses on various independent variables, according to the

[25] Gary Cohen, "Minor League Baseball," Accessed February 10, 2019. http://www.thebaseballcube.com/minors/. This is the website where I received all of my data for players signed between 2005 and 2015. Gary Cohen is an aggregator that helped me obtain signing bonuses and minor and major league statistics.

scouting theories outlined in Section IV. Through the use of these models, a grade is calculated for each of the thirty teams by parsing their draft history and evaluating their ability to pay less than marginal revenue product for players that performed above expectation. The second stage of the hypothesis test will take place in Section VI using the draft success metrics developed here.

The first model for hitters is located in Table 2, and was developed according to qualitative scouting theory:

$$\log(bonus) = \alpha + \beta_1 height + \beta_2 weight + \beta_3 age + i.location + i.position \quad (1)$$

where height is measured in inches, weight is measured in pounds, and age is measured in years. The dummy variables (i.location and i.position) control for where the player was drafted from and what defensive position he plays by taking the value of one if a criterion is met and zero otherwise. This regression is based on crude scouting theory, which doesn't include any performance metrics that can more accurately predict signing bonuses. When evaluating players, scouts focus first on the eye-test and determine if a player's height and weight indicate he is an athlete that can be successful at the major league level. These two variables are expected to have a positive effect on how much a player is worth, as more physically mature players can better handle big league competition. The height and weight variables have beta coefficients that correspond to an expected increase in a player's signing bonus by 7.9% and 0.56% respectively and are both statistically significant at the 1% alpha level. Furthermore, a player's age should have a negative effect on signing bonus, as older players should have less leverage in negotiations because they don't have as much room for growth. The coefficient on age corroborates this theory. Holding all else equal, if a player's age is increased by one year, his salary is expected to decrease by 30.4% and the coefficient is statistically significant at the 1% alpha level. Lastly, the model can only explain 20.7% of the variation in signing bonus, which is the lowest among the hitter models, leaving a lot of room for error.

The second model for hitters is located in Table 3, and was developed according to Billy Beane's Moneyball theory:

$$\log(bonus) = \alpha + \beta_1 milb_{OPS} + \beta_2 milb_{RC} + \beta_3 age + \beta_4 mlb_{OPS} + \beta_5 mlb_{RC} + \beta_6 overall \qquad (2)$$

where on-base percentage plus slugging percentage (OPS) and runs created (RC) are standard performance metrics for hitters, and overall is the number in the draft where a player was selected. This regression uses Billy Beane's standardized metrics that better assess a player's true value. OPS is the ability for players to extend innings, coupled with the ability to hit for power, and should have a positive effect on signing bonus. These OPS coefficients are 1.34 and 0.678 and are both significant at the 1% alpha level. However, minor league OPS has a larger magnitude than major league OPS in part due to the fact that many players in the sample didn't make it all the way to the major leagues. Runs created is another standard metric of determining the share of runs that a player contributes to a team's aggregate total and should have a positive effect. The runs created betas both have a statistically significant positive effect, albeit small in magnitude. Lastly, the overall pick with which a player was drafted should have a negative causal effect on signing bonuses, as players drafted lower have less skill and leverage. The regression coefficient corroborates this theory. Holding the other variables constant, as the ranking for a player increases by one, the expected signing bonus goes down by 0.2% and the coefficient is statistically significant at the 1% alpha level. Billy Beane's theory has more predictive power than the previous model, as all of the performance metrics explain 24.29% of the variation in signing bonuses.

The last model for hitters is located in Table 4, and was developed using Jeff Luhnow's Astroball theory:

$$\log(bonus) = \alpha + \beta_1 milb_{OPS} + \beta_2 height + \beta_3 mlb_{OPS} + \beta_4 age + \beta_5 overall + i.location + i.position \quad (3)$$

This model builds upon the previous two by applying a more nuanced approach to the regression analysis. Mirroring Luhnow's scouting theory, I eliminated the runs created and weight variables from the model and only used the statistically significant metrics that carry a high magnitude. Out of all of the variables in the regression, OPS still has the most significant causal effect, and the minor league version (β_1 = 1.53) has a larger magnitude than its major league counterpart (β_3 = 0.69). For the others, all else being equal, age continues to be the most significant qualitative variable, producing a negative causal effect that decreases predicted signing bonuses by 33.8% for a one-year increase in age. Lastly, 38.54% of the variation in signing bonuses can be explained through the combination of the significant qualitative variables, such as age and height, and standardized performance metrics, such as OPS, which yields a more effective model.

With these three models in place for hitters, I applied the same logic and techniques to the data set for pitchers, and came up with the following regressions located in Tables 5, 6, and 7:

$$\log(bonus) = \alpha + \beta_1 height + \beta_2 weight + \beta_3 age + i.location \quad (4)$$

$$\log(bonus) = \alpha + \beta_1 milb_{so9} + \beta_2 milb_{bb9} + \beta_3 age + \beta_4 mlb_{so9} + \beta_5 mlb_{bb9} + \beta_6 overall \quad (5)$$

$$\log(bonus) = \alpha + \beta_1 milb_{so9} + \beta_2 height + \beta_3 mlb_{so9} + \beta_4 age + \beta_5 overall + i.location \quad (6)$$

The only changes made were the inclusion of pitcher specific performance metrics, such as strikeouts per nine innings (so9) and walks per nine innings (bb9), and the removal of positional dummy variables (because all pitchers play the same position). When building these regressions, I initially used earned run average, which is the most popular standard statistic to judge a pitcher. However, the results were statistically insignificant prompting me to switch to "so9" and "bb9." These variables are important in scouting because the number of strikeouts and walks per nine innings are proxies for a pitcher's ability to get batters to swing and miss and ability to control the strike zone. The common thread among these three regressions is that all of their independent variables have a significant causal effect, but the magnitude of each one is negligible. These results illustrate the fact that there is no consensus standard metric to judge a pitcher's performance. In *Moneyball,* Michael Lewis notes that in judging pitchers, front offices "preferred their own subjective opinion [instead] of minor league pitching stats. [They] were flawed, enough to encourage uncertainty."[26] These results suggest that the econometric analysis performed is more applicable to hitters than pitchers, unless teams discover more descriptive variables that can better predict pitcher performance. Despite their flaws, these last three models can be applied to all the teams in the sample to estimate their ability to identify undervalued pitchers.

With the development of these six models, a predicted value was produced using each regression for all hitters and pitchers in their respective samples. Each player's predicted signing bonus was then compared to the actual one paid upon signing a contract. Given each player's career statistics and initial valuation, the three outcomes of these residual calculations are: the player was paid accurately, the player was overvalued, or the player was undervalued. In the sample, numerous players had residuals barely above and below the cutoff. To produce a more worthwhile analysis, I defined an accurate evaluation to be a residual between plus or minus one. Therefore, for a player to be undervalued, he had to have an *ex post* predicted signing bonus above the cutoff and higher than what he actually received, and vice versa for an overvalued player. Below are three examples following the methodology above:

[26] Lewis, *Moneyball*, 240.

Last name	First name	Draft year	Draft team name	Age	Log bonus	Moneyball residual	Scouts residual	Astroball residual	Valuation
Duran	Matt	2011	New York Yankees	18	12.721	0.389	-0.211	-0.116	accurate
Nava [27]	Daniel	2008	Boston Red Sox	24	0	-12.735	-10.851	-11.779	undervalued
Cole	Gerrit	2011	Pittsburgh Pirates	21	15.895	3.254	3.783	3.019	overvalued

With these estimates for each player, I counted the number of under valuations for each team and divided it by the total number of players that the team drafted. After completing these calculations, I averaged out the success rates for each team across the three models obtaining a standardized metric for how well each major league club drafted hitters and pitchers over the ten-year time period. These results are the core metric used in the second stage regression in Section VI and are located in Table 8.

Section VI: Hypothesis Test Results

Looking at the results of the first stage regressions in Table 8, the top five teams at drafting undervalued hitters over the time period were the: Baltimore Orioles, Los Angeles Dodgers, Boston Red Sox, Texas Rangers, and Toronto Blue Jays. With an analysis of average winning percentage from 2005 to 2015, the respective ranks of these top five drafting teams were: 29, 4, 2, 7, and 15.[28] According to average payroll data from 2005 to 2018, the Red Sox and Dodgers were ranked two and three respectively, while

[27] Daniel Nava famously signed for one dollar with the Red Sox in 2008 and was an integral part of their 2013 championship run.

[28] Rodney Fort, "Rods Sports Business Data," Accessed February 20, 2019. http://umich.app.box.com/s/41707f0b2619c0107b8b/folder/320021673. I downloaded a winning percentage data set for all 30 MLB teams from 2005-2018.

the Orioles, Rangers and Blue Jays were all within the top fifteen.[29] Furthermore, over this time period, the New York Yankees had the highest average payroll and winning percentage, while having a draft rank of only seventeen. These numbers provide conflicting inferences and call into question whether drafting ability or payroll has a more causal effect on winning.

Table 9 identifies the second stage regressions that formally test this question by running simple univariate panel regressions of winning percentage on drafting success (each with different data sorts) and payroll, and are as follows:

$$winning\ percentage = \alpha + \beta_1 hitter\_hit\_rate \quad (7)$$

$$winning\ percentage = \alpha + \beta_1 pitcher\_hit\_rate \quad (8)$$

$$winning\ percentage = \alpha + \beta_1 payroll \quad (9)$$

Regression equations 7 and 8 both yield statistically insignificant results. The coefficient on hitter draft success is high in magnitude. A one basis point increase in position player draft success ("hitter_hit_rate") corresponds with a twenty-three-basis point increase in winning percentage but has a t-statistic of 1.43. The results for pitcher draft success ("pitcher_hit_rate") are even more paltry, yielding a t-statistic of 0.68. For completeness, I performed these panel regressions on different sorts of the sample data (high school, college, and international players) and each regression yielded statistically insignificant results. The only noteworthy observation that can be formed from the data is that compared to high school hitters and pitchers, college players have larger t-statistics, 1.43 to 1.30 for hitters and 0.31 to 0.02 for pitchers. Although each variable is insignificant, these results indicate that college players are safer bets than their high school counterparts. Regression 9 provides the most interesting outcome. As a team's payroll increases by one dollar, we expect the winning percentage to increase by 5.45×10^{-10} basis points. The

[29] Ibid. In this database I also downloaded a payroll data set for all 30 MLB teams from 2005-2018.

payroll coefficient has a t-statistic of 5.31, which is overwhelmingly significant at the 1% alpha level. To illustrate the importance of these results, consider the following scenario. Assume a small market team signs a few major free agents and its payroll balloons from $50 million to $110 million. This $60 million increase will yield an additional 5.297 wins ($60 million x 5.45×10^{-10} x 162 games). These six additional wins could catapult a bad team into playoff contention. The results of these regressions do not validate the theory that drafting better significantly causes teams to have higher winning percentages. Therefore, we can reject the null hypothesis and say that team payroll–and free agency by extension–has a statistically significant relationship with winning.

Section VII: Conclusion

Many sports economists believe that Major league Baseball is a free agency driven league, in which big market teams, like the New York Yankees and the Boston Red Sox, flex their financial muscle and poach the best players that smaller market teams develop. Conventional wisdom says the First-Year Player Draft is thought of as an equalizer, providing small market teams the ability to draft good players and retain them on team friendly contracts. I posit that the amateur draft does not promote the parity that it is designed for and performed multiple regressions to test the causal effect that drafting has on winning. There are many confounding variables that affect players once they are drafted by a particular team, and many of these players do not make it to the MLB level. Even Oakland Athletics general manager Billy Beane questions the effectiveness of the draft explaining that "the draft is nothing but a crapshoot, we take fifty guys and we celebrate if two of them make it. In what other business is two for fifty a success? If you did that in the stock market, you would go broke."[30]

[30] Lewis, *Moneyball*, 17.

Although my study revealed significant results favoring my alternative hypothesis, there were a few limitations. With my qualitative model, I took a crude approach at incorporating scouting variables into the regression. A more refined approach would include grades for the scouts themselves, based on their drafting history for each Major League team and include specific positions that they specialize in identifying. In addition, the inclusion of dummy variables for location in the United States all yielded insignificant results, while it is known that players drafted from certain areas tend to perform better than others. Furthermore, experimentation with nonlinear regressions for better players, which attempt to measure factors like the superstar effect, could yield more robust conclusions because these players perform on a more exponential basis. Lastly, the most significant limitation of the study is the time period analyzed because it may be too close to the present time. Performing an *ex post* study on players that are still in the major leagues makes it challenging to compare with players that have finished their careers.

Even though this study proved that the draft has an insignificant effect on competitive balance from 2005 to 2018, this hypothesis could be extended into further research. The need for cost-effective impact players is a recent objective for teams and owners because the payroll gap between smaller and larger markets has widened substantially. The data suggest that optimal drafting strategies for major league teams should focus on the game-ready college position players and pitchers rather than high potential high school players that are far from finished products. Furthermore, the data analytics revolution, started by Billy Beane in 2003 and refined by Jeff Luhnow in 2017, is only approximately twenty years old. The theories and models outlined in this paper could be applied to different samples to test the effects of drafting on winning percentage over time. I posit that a time series analysis beginning in 1965, the inception of the amateur draft, will show an increase in the magnitude of the t-statistics of drafted players. The t-statistic for drafted position players is currently 1.43, which is not that far off from the necessary 1.65 needed for statistical significance. In a few years, when sabermetrics and data analytics are more widely adopted by the league, it is possible that the results of this study could change.

Table 1
Variable List and Descriptions

Variable	Description
Log_bonus	Bonus data is from 2005 to 2015. Log(signing bonuses) made data inference easier since players are paid large dollar amounts.
milb_OPS/MLB_OPS	Standardized assessment of hitter ability. Formula: On Base % + Slugging %.
milb_RC/MLB_RC	Standardized assessment of hitter's contribution to team wins. Formula: (Total Bases x (Hits + walks)) / (At Bats + walks).
Age	Age of players when they signed their contracts and played their first game.
Overall	U.S. players: pick that the player was drafted. International Players: overall = 0.
Heightinches	How tall the player was when he signed.
Weightlbs	How much the player weighed when he signed.
i.location	Dummy variables for which state/country the player was signed out of.
i.position	Dummy variables for position players. Pitchers are all in one category.
milb_so9/MLB_so9	Standardized assessment of pitching skill. Formula: ((9/Innings Pitched) * Strikeouts).
milb_bb9/MLB_bb9	Standardized assessment of pitching control. Formula: ((9/Innings Pitched) * Walks).
hitter_hit_rate/pitcher_hit_rate	The rate at which teams draft undervalued players. Formula: (# undervalued players) / (# players drafted by a team).
payroll	Payroll data is from 2005 to 2017 for all thirty major league teams. The payroll data reflects the salaries of the 25-man active roster.
winning percentage	Winning percentage data is from 2005 to 2017 for all thirty major league teams. Formula: (# wins / 162 games).

Table 2
Hitters Scout Model
Dependent variable: *log_bonus*

Regressor (Standard Error Below)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Heightinches	.0822** (.014)	.0798** (.401)	.057** (.015)	.073** (.014)	.079** (.015)
Weightlbs	—	.0005 (.0012)	.006** (.001)	.003* (.001)	.0056** (.0013)
Age	—	—	-.180** (.019)	-.310** (.022)	-.3040** (.022)
i.location	—	—	—	83 dummy variables	83 dummy variables
i.position	—	—	—	—	12 dummy variables
intercept	6.36** (.999)	6.44** (1.034)	10.56** (1.032)	12.64** (1.033)	15.17** (.381)
***F-statistics testing the null hypothesis: population coefficients on the following regressors are all zero*: (p-value below)**					
Heightinches	36.16 (.000)	26.93 (.000)	14.74 (.000)	26.03 (.000)	26.42 (.000)
Heightinches, Weightlbs	—	18.48 (.000)	30.77 (.000)	27.14 (.000)	37.81 (.000)
Heightinches, Weightlbs, Age	—	—	41.18 (.000)	78.51 (.000)	78.91 (.000)
i.location dummies	—	—	—	39.56 (.000)	91.70 (.000)
i.position dummies	—	—	—	—	4.31 (.000)
Regression summary statistics					
$\bar{R}^2$	.0140	.0137	.0760	.1681	.1783
R^2	.0144	.0144	.0770	.1941	.2070
Regression RMSE	1.3462	1.3464	1.3032	1.2365	1.2289
n	2723	2723	2723	2723	2723

*Notes: Heteroskedasticity-robust standard errors are given in parentheses under estimated coefficients, and p-values are given in parentheses under F-statistics. The F-statistics are heteroskedasticity-robust. Coefficients are individually statistically significant at the +10%, *5%, **1% significance level.*

Table 3
Hitters Moneyball Model
Dependent variable: *log_bonus*

Regressor (Standard Error Below)	(1)	(2)	(3)	(4)	(5)	(6)
milb_OPS	2.596** (.297)	1.73** (.401)	2.84** (.384)	1.859** (.363)	1.582** (.366)	1.34** (.354)
milb_RC	—	.0010** (.0002)	.0011** (.0002)	.0003 (.0002)	.0005+ (.0010)	.0005* (.0002)
Age	—	—	-.2214** (.0195)	-.2196** (.019)	-.2181** (.0187)	-.169** (.018)
MLB_OPS	—	—	—	1.049** (.114)	.8665** (.0004)	.678** (.125)
MLB_RC	—	—	—	—	.0008* (.0003)	.0007* (.0003)
overall	—	—	—	—	—	-.002** (.0001)
intercept	10.547** (.198)	10.974** (.245)	14.427** (.414)	15.043** (.376)	15.17** (.381)	14.76** (.365)
***F-statistics testing the null hypothesis: population coefficients on the following regressors are all zero*: (p-value below)**						
Milb_OPS	79.83 (.000)	18.61 (.000)	54.45 (.000)	26.21 (.000)	18.66 (.000)	14.33 (.000)
milb_OPS, milb_RC	—	57.10 (.000)	107.68 (.000)	22.27 (.000)	21.23 (.000)	18.05 (.000)
Milb_OPS, milb_RC, Age	—	—	85.27 (.000)	50.26 (.000)	48.15 (.000)	31.00 (.000)
Milb_OPS, milb_RC, Age, MLB_OPS	—	—	—	98.07 (.000)	70.55 (.000)	46.79 (.000)
Milb_OPS, milb_RC, Age, MLB_OPS, MLB_RC	—	—	—	—	78.93 (.000)	52.89 (.000)
Milb_OPS, milb_RC, Age, MLB_OPS, MLB_RC, overall	—	—	—	—	—	112.40 (.000)
Regression summary statistics						
$\bar{R}^2$	.0341	.0403	.1349	.1711	.1733	.2412
R^2	.0345	.0410	.1358	.1723	.1749	.2429
Regression RMSE	1.3312	1.326	1.2598	1.2332	1.232	1.179
n	2720	2720	2720	2720	2720	2720

*Notes: Heteroskedasticity-robust standard errors are given in parentheses under estimated coefficients, and p-values are given in parentheses under F-statistics. The F-statistics are heteroskedasticity-robust. Coefficients are individually statistically significant at the +10%, *5%, **1% significance level.*

Table 4
Hitters Astroball Model
Dependent variable: *log_bonus*

Regressor (Standard Error Below)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
milb_OPS	2.503** (.289)	.760* (.313)	1.966** (.313)	1.704** (.302)	1.378** (.277)	1.53** (.276)
Heightinches	.0743** (.014)	.073** (.013)	.071** (.013)	.068** (.012)	.063** (.012)	.075** (.013)
MLB_OPS	—	1.089** (.103)	1.087** (.103)	.877** (.097)	.720** (.087)	.690** (.087)
Age	—	—	-.2182** (.018)	-.169** (.017)	-.340** (.021)	-.338** (.021)
overall	—	—	—	-.0018** (.0001)	-.002** (.0002)	-.00024** (.0001)
i.location	—	—	—	—	83 dummy variables	83 dummy variables
i.position	—	—	—	—	—	12 dummy variables
intercept	5.187** (.987)	6.309** (.988)	9.811** (.939)	9.632** (.902)	14.12** (.888)	12.91** (.981)
***F-statistics testing the null hypothesis: population coefficients on the following regressors are all zero*: (p-value below)**						
Milb_OPS	75.03 (.000)	5.88 (.154)	39.46 (.000)	31.65 (.000)	24.79 (.000)	30.90 (.000)
milb_OPS, Heightinches	60.18 (.000)	18.85 (.000)	37.24 (.000)	33.35 (.000)	29.74 (.000)	33.58 (.000)
milb_OPS, Heightinches, MLB_OPS	—	78.07 (.000)	114.32 (.000)	88.43 (.000)	67.61 (.000)	71.02 (.000)
milb_OPS, Heightinches, MLB_OPS, Age	—	—	109.60 (.000)	76.45 (.000)	116.43 (.000)	108.31 (.000)
milb_OPS, Heightinches, MLB_OPS, Age, overall	—	—	—	144.43 (.000)	175.49 (.000)	167.60 (.000)
i.location dummies	—	—	—	—	50.95 (.000)	106.53 (.000)
i.position dummies	—	—	—	—	—	1.83 (.0504)
Regression summary statistics						
$\bar{R}^2$	.0455	.0892	.1813	.2488	.3605	.3626
R^2	.0462	.0902	.1825	.2502	.3810	.3854
Regression RMSE	1.3233	1.2927	1.2256	1.174	1.083	1.081
n	2720	2718	2718	2718	2718	2718

*Notes: Heteroskedasticity-robust standard errors are given in parentheses under estimated coefficients, and p-values are given in parentheses under F-statistics. The F-statistics are heteroskedasticity-robust. Coefficients are individually statistically significant at the +10%, *5%, **1% significance level.*

Table 5
Pitchers Scout Model
Dependent variable: *log_bonus*

Regressor (Standard Error Below)	**(1)**	**(2)**	**(3)**	**(4)**
Heightinches	.072** (.013)	.0769** (.401)	.062** (.014)	.067** (.014)
Weightlbs	—	-.0011 (.0012)	.002 (.001)	.00002 (.0012)
Age	—	—	-.143** (.016)	-.2210** (.017)
i.location	—	—	—	83 dummy variables
intercept	6.85** (.981)	6.67** (.995)	10.08** (1.033)	11.47** (1.08)
***F-statistics testing the null hypothesis: population coefficients on the following regressors are all zero*: (p-value below)**				
Heightinches	29.96 (.000)	29.21 (.000)	18.91 (.000)	22.67 (.000)
Heightinches, Weightlbs	—	15.53 (.000)	15.80 (.000)	13.21 (.000)
Heightinches, Weightlbs, Age	—	—	36.74 (.000)	69.02 (.000)
i.location dummies	—	—	—	45.36 (.000)
Regression summary statistics				
$\bar{R}^2$	.0111	.0109	.0503	.1239
R^2	.0114	.0116	.0513	.1499
Regression RMSE	1.3005	1.3006	1.2745	1.2241
n	2731	2731	2731	2731

*Notes: Heteroskedasticity-robust standard errors are given in parentheses under estimated coefficients, and p-values are given in parentheses under F-statistics. The F-statistics are heteroskedasticity-robust. Coefficients are individually statistically significant at the +10%, *5%, **1% significance level.*

Table 6
Pitchers Moneyball Model
Dependent variable: *log_bonus*

Regressor (Standard Error Below)	(1)	(2)	(3)	(4)	(5)	(6)
milb_so9	.057** (.015)	.058** (.015)	.064** (.015)	.033* (.017)	.032* (.015)	.027+ (.015)
milb_bb9	—	.003** (.0007)	.002** (.0006)	.003** (.001)	.003** (.001)	.004** (.001)
Age	—	—	-.141** (.015)	-.150** (.015)	-.153** (.015)	-.133** (.014)
MLB_so9	—	—	—	.056** (.013)	.032* (.015)	.025* (.013)
MLB_bb9	—	—	—	—	.071** (.022)	.061** (.019)
overall	—	—	—	—	—	-.001** (.0001)
intercept	11.76** (.123)	11.74** (.125)	14.52** (.308)	14.79** (.307)	14.83** (.305)	14.81** (.295)
***F-statistics testing the null hypothesis: population coefficients on the following regressors are all zero*: (p-value below)**						
milb_so9	13.86 (.0002)	14.14 (.0002)	18.11 (.000)	4.04 (.044)	4.23 (.0399)	3.37 (.067)
milb_so9, milb_bb9	—	11.32 (.000)	10.61 (.000)	5.33 (.0049)	5.05 (.0064)	7.34 (.000)
milb_so9, milb_bb9, Age	—	—	34.72 (.000)	37.03 (.000)	38.10 (.000)	32.47 (.000)
milb_so9, milb_bb9, Age, MLB_so9	—	—	—	32.97 (.000)	30.42 (.000)	25.38 (.000)
milb_so9, milb_bb9, Age, MLB_so9, MLB_bb9	—	—	—	—	40.42 (.000)	31.77 (.000)
milb_so9, milb_bb9, Age, MLB_so9, MLB_bb9, overall	—	—	—	—	—	69.09 (.000)
Regression summary statistics						
$\bar{R}^2$	.0055	.0056	.0459	.0803	.0894	.1504
R^2	.0059	.0063	.0469	.0816	.0911	.1523
Regression RMSE	1.3043	1.3043	1.2776	1.2544	1.2481	1.2056
n	2730	2730	2730	2730	2730	2730

*Notes: Heteroskedasticity-robust standard errors are given in parentheses under estimated coefficients, and p-values are given in parentheses under F-statistics. The F-statistics are heteroskedasticity-robust. Coefficients are individually statistically significant at the +10%, *5%, **1% significance level. Interestingly, ERA and WHIP (sabermetric stats for pitching) weren't significant. These variables were dropped from the regressions in favor of so9 and bb9.*

Table 7
Pitchers Astroball Model
Dependent variable: *log_bonus*

Regressor (Standard Error Below)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
milb_so9	.066** (.015)	.038* (.017)	.042* (.016)	.037* (.016)	.050** (.015)
Heightinches	.078** (.013)	.075** (.013)	.074** (.013)	.079** (.013)	.061** (.012)
MLB_so9	—	.051** (.013)	.055** (.013)	.044** (.012)	.035** (.009)
Age	—	—	-.150** (.015)	-.123** (.014)	-.247** (.016)
overall	—	—	—	-.002** (.0001)	-.002** (.0001)
i.location	—	—	—	—	83 dummy variables
intercept	5.88** (1.01)	6.23** (1.01)	9.22** (1.02)	8.84** (1.001)	12.48** (1.03)
***F-statistics testing the null hypothesis: population coefficients on the following regressors are all zero*: (p-value below)**					
milb_so9	18.72 (.000)	4.95 (.026)	6.36 (.011)	5.72 (.017)	11.75 (.000)
milb_so9, Heightinches	23.53 (.000)	16.96 (.000)	17.60 (.000)	20.56 (.000)	16.47 (.000)
milb_so9, Heightinches, MLB_so9	—	22.35 (.000)	25.28 (.000)	24.75 (.000)	22.01 (.000)
milb_so9, Heightinches, MLB_so9, Age	—	—	41.39 (.000)	36.34 (.000)	75.09 (.000)
milb_so9, Heightinches, MLB_so9, Age, overall	—	—	—	81.81 (.000)	114.91 (.000)
i.location dummies	—	—	—	—	482.65 (.000)
Regression summary statistics					
$\bar{R}^2$	.0184	.0465	.0918	.1565	.2743
R^2	.0191	.0476	.0931	.1580	.2963
Regression RMSE	1.2959	1.2772	1.2465	1.2013	1.1143
n	2730	2730	2730	2730	2730

*Notes: Heteroskedasticity-robust standard errors are given in parentheses under estimated coefficients, and p-values are given in parentheses under F-statistics. The F-statistics are heteroskedasticity-robust. Coefficients are individually statistically significant at the +10%, *5%, **1% significance level.*

Table 8
Team Drafting Ability and League Rank From 2005-2015

	Team Ability to Find Undervalued Hitters		Team Ability to Find Undervalued Pitchers	
Team Name	Average Hit Rate	Final Rank	Average Hit Rate	Final Rank
Arizona Diamondbacks	13.79%	18	9.86%	27
Atlanta Braves	13.41%	19	10.00%	26
Baltimore Orioles	23.47%	1	10.61%	25
Boston Red Sox	20.80%	3	13.69%	20
Chicago Cubs	10.03%	28	14.08%	18
Chicago White Sox	15.53%	16	15.38%	11
Cincinnati Reds	11.83%	25	13.83%	19
Cleveland Indians	10.49%	26	15.56%	10
Colorado Rockies	12.08%	24	9.41%	28
Detroit Tigers	12.79%	22	14.58%	14
Houston Astros	15.74%	14	12.89%	23
Kansas City Royals	13.11%	20	15.00%	13
Los Angeles Angels	16.91%	11	14.58%	15
Los Angeles Dodgers	22.57%	2	19.15%	5
Miami Marlins	15.73%	15	17.41%	7
Milwaukee Brewers	19.58%	7	13.33%	22
Minnesota Twins	16.49%	12	8.33%	30
New York Mets	16.30%	13	17.60%	6
New York Yankees	15.10%	17	13.33%	21
Oakland Athletics	8.77%	29	21.03%	3
Philadelphia Phillies	17.56%	9	23.93%	2
Pittsburgh Pirates	10.28%	27	9.06%	29
San Diego Padres	12.79%	21	17.23%	8
San Francisco Giants	8.71%	30	12.66%	24
Seattle Mariners	12.30%	23	24.91%	1
St. Louis Cardinals	19.73%	6	14.49%	16
Tampa Bay Rays	18.48%	8	15.73%	9
Texas Rangers	20.54%	4	15.33%	12
Toronto Blue Jays	20.15%	5	14.19%	17
Washington Nationals	17.14%	10	20.08%	4

Notes: Predicted signing bonuses are determined using data from 2005 to 2015 in the three models for hitters and pitchers. Once a player has a predicted ex post value using observed statistics, a residual is calculated by subtracting the predicted value from the observed value. Undervalued players have negative residuals, as they were paid less than their statistics determined they were worth. Hit rate is then calculated as the number of undervalued players divided by the total sum drafted by a team. Rank is determined by averaging out the hit rates predicted by all three models, for both pitchers and hitters.

Table 9
The Effect of Drafting Ability on Winning Percentage: Regression Results
Dependent variable: *winning percentage*

Regressor (Standard Error Below)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Groups tested	**All Hitters**	**All Pitchers**	**Team Payroll**	**H.S. Hitters**	**College Hitters**	**H.S. Pitchers**	**College Pitchers**	**All Intl. Players**
hitter_hit_rate	.230 (.160)	—	—	.135 (.103)	.126 (.087)	—	—	—
pitcher_hit_rate	—	.075 (.111)	—	—	—	.017 (.116)	.037 (.118)	—
payroll	—	—	5.45e-10** (1.03e-10)	—	—	—	—	—
International_players _hit_rate	—	—	—	—	—	—	—	.062 (.084)
intercept	.464** (.021)	.489** (.018)	.447** (.212)	.484** (.013)	.481** (.014)	.497** (.014)	.495** (.018)	.486** (.019)
***T-statistics testing the null hypothesis: population coefficients on the following regressors are zero*: (p-value below)**								
hitter_hit_rate	1.43 (.151)	—	—	1.30 (.193)	1.43 (.151)	—	—	—
pitcher_hit_rate	—	.68 (.497)	—	—	—	.02 (.8813)	.31 (.754)	—
payroll	—	—	5.31 (.000)	—	—	—	—	—
International_players _hit_rate	—	—	—	—	—	—	—	.74 (.459)
Regression summary statistics								
Number of groups	30	30	30	30	30	30	30	30
Number of observations	420	420	390	420	420	420	420	420

*Notes: Heteroskedasticity-robust standard errors are given in parentheses under estimated coefficients from the panel regression, and p-values are given in parentheses under T-statistics. The T-statistics are heteroskedasticity-robust. Coefficients are individually statistically significant at the +10%, *5%, **1% significance level. After doing a surface level analysis of hitters and pitchers, I broke the data sets down into high school, college, and international players to see if there were differing effects in the data.*

Bibliography

Cohen, Gary. “Minor League Baseball.” *The Baseball Cube*. Last modified 2019. Accessed February 20, 2019. http://www.thebaseballcube.com/minors/.

Craig, Mary. “Chained to the Game: Professional Baseball and the Reserve Clause, Part Two.” *Beyond the Box Score*. Last modified June 10, 2017. Accessed March 5, 2019. http://www.beyondtheboxscore.com/2017/6/10/15766702/curt-flood-mlbpa-reserve-clause-free-agency.

Depken, Craig A. "Free-Agency and the Competitiveness of Major League Baseball." *Review of Industrial Organization* 14, no. 3 (1999): 205-17. http://www.jstor.org/stable/41798858.

Fishman, Peter. "Competitive Balance and Free Agency in Major League Baseball." *The American Economist* 47, no. 2 (2003): 86-91. http://www.jstor.org/stable/25604283.

Fort, Rodney. “Rods Sports Business Data.” *Rods Sports Business Data*. Last Modified 2018. Accessed February 20, 2019. http://umich.app.box.com/s/41707f0b2619c0107b8b/folder/320021673.

Fort, Rodney. “Sports Economics: Making Sense of It All.” In *Sports Economics,* 13-15. Pearson Education, 2014.

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

Reiter, Ben. *Astroball: The New Way to Win It All*. Crown Archetype, 2018.

Ross, Gary Neil. "The Determination of Bonuses in Professional Sports." *The American Economist* 19, no. 2 (1975): 43-46. http://www.jstor.org/stable/25603016.

Scully, Gerald W. "Pay and Performance in Major League Baseball." *The American Economic Review* 64, no. 6 (1974): 915-30. http://www.jstor.org/stable/1815242.

Streib, Noah, Stephen J. Young, and Joel Sokol. "A Major League Baseball Team Uses Operations Research to Improve Draft Preparation." *Interfaces 42*, no. 2 (2012): 119-30. http://www.jstor.org/stable/41472741.