12-2014

The Path to the Top: How Career Paths Differ Between College and Non-College Major League Baseball Players

Gerald Arnette III
Clemson University, garnett@g.clemson.edu

THE PATH TO THE TOP: HOW CAREER PATHS DIFFER BETWEEN COLLEGE AND NON-COLLEGE MAJOR LEAGUE BASEBALL PLAYERS

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
Economics

---

by
Gerald Nesmith Arnette III
December 2014

Accepted by:
Dr. Raymond D. Sauer, Committee Chair
Dr. Curtis J. Simon
Dr. Scott L. Baier

ABSTRACT

In route to the Majors, there are two major pathways. Talented players may either sign with a Major League Team to start their careers in the in the Minor Leagues, or play NCAA Baseball. Using ordinary least squares (OLS), the career differences between NCAA and Minor League players will be examined. Major League players whose careers began after 1985 are included and are divided into quintiles by the player's career mean. Dividing the players into quintiles increases the accuracy of the analysis because the career path varies according to player ability. The College Players are separated into eras with particular emphasis placed on those who played college after 1999. This paper confirms previous analysis that career paths differ among the OPS quintiles and finds that significant career path differences exist between college and non-college Major League Players. There are also small differences between the different college eras in the effect on a player's productivity profile. Overall, former NCAA baseball players retire and peak a little later in age but have shorter careers.

DEDICATION

I dedicate this Thesis to my loving parents who encouraged me to pursue a master’s degree and thank them for providing the support and means necessary.

ACKNOWLEDGMENTS

I want to thank my advisor, Dr. Raymond Sauer, for offering a topic for me to explore this summer that eventually became this thesis. Your guidance made this possible and you made it enjoyable. Thank you, Geddings Barrineau, for all your Stata advice and teaching me the basics of writing code. I also want to thank my committee members, Drs. Simon and Baier, for their helpful and insightful comments during the revision process.

TABLE OF CONTENTS

Table of Contents (Continued)

LIST OF FIGURES AND TABLES

CHAPTER ONE

INTRODUCTION

Ever since Bill James and the founding of SABR, baseball has been the focus of much statistical analysis. This paper wishes to add to their work by analyzing the differences between the career path of former NCAA baseball players, and those who signed directly out of high school. First, a broad overview of the methodology and justification for its reasoning. Major League players are separated into quintiles in order to confirm prior quintile analyses and discover new relationships. The final section will demonstrate and examine how the career paths differ between college and non-college players within the same quintile.

Major League Baseball is not only America's Pastime, but also the first major sports league to develop an extensive minor league system. Teams knew that in order to be successful good players would have to be identified, acquired and developed appropriately. The minor leagues are correctly seen as the first and last important step along this path to the Majors, with NCAA baseball playing a minor role in the development of professional talent. Professional football and basketball have traditionally relied on the college ranks to serve as the equivalent of minor league baseball and develop professional talent. However, the NBA is experimenting with a developmental league to better prepare the most talented players for the professional game. The Developmental league still is not the prescribed method for reaching the NBA. College basketball is still the dominant path. The NFL is not testing out a minor league system, but wholly relies on NCAA Football to procure their football talent.

All of the major sports leagues do differ with their college version in significant ways, but baseball is interesting because it has a league that largely bypasses college. The Minor League baseball players play with exactly the same rules as their Major League peers, but a slightly different version exists in NCAA baseball. This makes for interesting comparisons between those players who signed out of high school and those who chose the college route.

Baseball is unique in that it has two major pathways to reach the highest level, College and Minor League baseball. Those players who do not decide to begin their Minor League careers immediately after high school generally sign with a college team. Once a player decides to attend college, he must remain at his school until at least his junior season, at which point he may then sign a major league contract. Unlike professional Basketball and Football, these players do not immediately play in their professional league at the highest level. These former NCAA athletes typically spend some period of time in the Minor Leagues. This paper is going to examine the differences between the two paths to the Majors and their effect on a player’s career.

CHAPTER TWO

PRODUCTIVITY METRICS AND PROFILES

Productivity Metrics

Only offensive players are examined. Pitchers are not only primarily compensated for their pitching, but their main contribution to a team's winning is on the defensive side of the game. In half of major league baseball, the pitcher does not even log official at-bats. He has a designated hitter that bats in his stead. Including pitchers in the sample and gauging their career performance based on their batting skills would lead in inaccuracies and only muddy the findings.

Commonly used hitting metrics are batting average (BA), slugging percentage (SLG), runs batted in (RBI) and on-base percentage (OBP). The hitting metrics that correlate more highly with total bases are better. More total bases lead to more runs and this in turn leads to more winning. Since winning is the point of the game, production statistics that correlate most strongly with this logic are preferred.

Batting average is the percentage of time a hitter gets on base out of his official at-bats. It roughly measures how often a player hits his way safely to at least first. This is not the most accurate metric in measuring total bases and therefore, runs produced. Batting average treats all hits equally, and does not account for hits that are for more than one base. Slugging percentage is better because it weighs hits by the number of bases produced. A homerun generates more bases and counts more under slugging percentage than a single. However neither of these metrics measures a crucial way to get on base without hitting, walks.

Walks happen when the pitcher throws four balls, before either a playable hit, or a strike out. Counting walks as a measurement of a hitter's production, instead of just a pitcher's mistake, makes for much more accurate offensive analysis. Omitting this statistic is leaving out a key way for batters to earn bases without generating outs. Offensive production can be broadly stated as the ability to get on and advance bases without generating outs. Outs are key because they are the primary barrier to scoring runs and therefore winning games (Hakes and Sauer, 2007).

On-base percentage (OBP) simply measures how likely a player is to get on base safely, discounting sacrifice hits. OBP evaluates how likely a player is to get on base out of his trips to the plate. It is much better than batting average simply because it incorporates walks in its measurement.

On-base plus slugging (OPS) is the best metric to measure a hitter's production because it takes the likelihood of a player to simply arrive at first safely (OBP) and adds hits weighted by the number of bases generated (SLG). This is a simple, accurate metric to estimate hitting production and evaluate offensive production. It is also the single, best indexed hitting statistic that most highly correlates with run production (Lewis, 2003)

What about other offensive metrics, like RBI? RBI is not a reliable hitting metric. It is quite noisy and a significant amount of the variance can be attributed to a hitter's place in the batting order of the lineup. Some hitting spots are better suited to drive in runs than others, rendering RBI an inaccurate metric of an individual hitter's skill and production (James, 1988). Also, it is harder to isolate the individual's contribution to winning with RBI because there are multiple offensive players involved. A great hitter may have lower RBI's simply because his teammates are not productive hitters and cannot get on base. Moreover, this reasoning still holds that on-base percentage and slugging percentage are the two metrics that most highly correlate with creating runs and winning.

Productivity Profiles

As players age, their productivity increases until a certain maximum point, and declines after. Logically this makes sense because players would not retire otherwise and continue to increase production.

As time progresses, MLB players gain experience and knowledge of the game, increasing their skill. More knowledgeable players are preferred to those with less baseball knowledge, ceteris paribus. They are better able to predict what their opponents will do and adjust their strategies accordingly. Playing in more games exposes players to an increasing variety of situations, and more opportunities for learning. They also gain intuitive knowledge that is difficult to convey and quantify, but nonetheless increases his offensive skill and production. This "mental aging" combined with physical prowess is what leads to increases in the production of players.

The alternative force at play is physical aging. Older players' bodies cannot handle the physical toll of the grueling MLB season. These players do not recover from games as quickly, and they are more likely to have a serious injury. This leads to lower bat speed and less hitting production toward the end of his career. They also play fewer games because of the slower recovery time. In short, older players are not as productive when they play and play fewer games due to physical aging.

These two factors lead to more of a quadratic regression regarding offensive production because mental aging dominates the earlier career stages before physical aging takes over.

This is largely similar to what happens in the general workforce. Workers gain knowledge through experience and increase productivity. Then physical aging takes over and productivity levels decline. The average worker peaks between 45-55 before declining. Aubert and Crépon (2003) demonstrated that the average worker in France improves in productivity until about the age of 45. Then the productivity curve flattens out until finally the worker loses productivity and declines. This holds true across sectors.

Human Capital Theory also weighs in. The earnings of major league players are discounted by the training costs. This training is costly and should be viewed as an investment since it is undertaken to increase the player's future ability and therefore income. (Becker, 1968) This happens early in the career, but the players do earn a premium later in their careers. Viewing employees as capital investment projects yields important insights into wage and careers.

Human Capital Theory also explains firm behavior by the teams. Teams prefer younger players, ceteris paribus. They have a higher possibility of improving to become a great player and also younger players contribute more to winning than they are paid, mainly because of the collective bargaining agreement. This in part, explains why teams are so eager to dump players in the bottom quintile because there are other, younger players who have a higher chance of being good.

This paper is in the vein of Krohn (1983), Albert (1999) and Fair (2008). A player's career profile will be examined using ordinary least squares (OLS) regression and a quadratic analysis. However, it will differ by separating the hitters into quintiles to produce a more accurate model. Hakes and Turner (2009) showed there are biases that occur when all of the players are lumped together and examined as one group.

Players that differ in ability do not have the same productivity curve. They have different intercepts, slopes and peaks. Marginal players enter MLB around the same time as other players, but only improve slightly before retiring well before better players. They peak slightly above replacement level before flaming out soon after. Top quintile players both peak and retire later than their counterparts. Their findings are substantial enough to warrant the separation of the players in order to more accurately describe productivity profiles in Major League Baseball.

Hitters are separated into quintiles according to their OPS indexed by each position's mean. This prevents some positions (i.e. catchers) from being overly represented in the bottom quintiles and other positions (outfielders for example) in the top quintiles. Some positions are more valued for their defensive skills because some positions are more difficult to play in the field. Generally this applies to catchers, second basemen and shortstops. First and third basemen and outfielders are expected to have more offensive production in part, because they are easier to play defensively. Since teams make these distinctions and differ offensive expectations according to the position, it makes sense to separate the players into quintiles according to their position means. This also matches the methodology of Hakes and Turner.

CHAPTER THREE

DATA OVERVIEW AND BASIC OPS REGRESSION

Data and Variables

As in Hakes and Turner, only seasons with at least 130 official at-bats are analyzed[1]. Some players have abbreviated, highly productive seasons that are out of line with his normal offensive production and not indicative of his ability. Leaving these seasons in the sample would only muddy the findings. The converse is also prevented via this minimum threshold. Hitters with less than three Major League seasons are dropped from the sample. This drops the most marginal players who don't have substantial careers in the sample from being examined for the effects of OPS change and college experience. All years since 1985 are in the sample. Players who were active and their seasons prior to 1985 are also included in order to better gauge the progression of their careers. Not including these prior seasons would skew the data not accurately weighing the mean of their offensive production and not track the change in OPS over the course of a career. There are over 7600 player-seasons examined that correspond with over 950 player careers.

The summary statistics for the quintiles tell a story (Table 1). The top quintile has an OPS that is nearly 300 points above that of the average player in the bottom quintile. These elite players are also 150 points above the mean OPS in the sample. The average observed OPS in the sample is 10% higher than the mean for that player's career. This

[1] All data comes from Sean Lahman's Baseball Database and are available at http://www.seanlahman.com

matches the narrative set forth by Hakes and Turner that the top players peak later and retire well before reaching replacement level. The top players retire before their OPS dips to their original OPS percentage. Either teams are not willing to pay for these hitters to continue playing, or the players retire because they do enjoy playing the game at the new, lower point on the productivity curve. Either way, these players have more productive years but do not use them. The only quintile that performs well below its career mean are those in the bottom quintile. The second through fourth quintiles range from 104% the player's mean to 96.4%. They have an average OPS observation that is only 89% of their career mean. This is also matches the findings in Hakes and Turner that the marginal players debut at a certain OPS and do not much improve. Their OPS generally goes down from where they debut. The bottom quintile takes about half as many official at-bats as the top players. This is expected because teams want to minimize the plate appearances for marginal players.

OPS indexed by season also demonstrates that the top and bottom quintiles differ the most from the ones immediately next. The top quintile of players hit nearly 20% over the mean OPS for that season and 15% more than those players in the second quintile. The fourth quintile's OPS is 10% higher than the bottom quintile, the second largest difference. The top players are extraordinarily good compared to the middle quintiles and the bottom players extraordinarily bad.

The effects of differing bat eras in college baseball will also be tested to see if training with different equipment impacts the productivity of college players in their

MLB careers. The three NCAA eras test are the aluminum bat era (before 1986), weight-limit era (1986-1999), and the BESR era (1999-2011).

RYdummy is the dependent variable that corresponds with the last season played by the specific player. If that particular season is the last, then that observation will have one for retirement year. All other seasons are coded zero. Each player that stopped play before 2012 should have a season where RYdummy is one.

Playerops uses the ops for a specific season and is indexed by the mean performance of the player over the course of his career for the. A mean OPS is found for each player in the sample so that within career variation is accounted for. Comparisons with career means are preferable since players differ in abilities, so examining the career drop off against the mean of the entire sample will not yield as much information as the drop off against that player's career.

CollegeExp is a dummy variable that measures college-playing experience. This variable is coded 1 for players with college experience and 0 if a player signed with a major league team out of high school. Approximately 40% of the observations have college experience. There are differences between the quintiles regarding the number of college players.

Minoryears is a rough estimate of the number of years spent in the minor leagues. All players drafted out of high school are assumed to be drafted 18 years after their birthdate, and college players are drafted in their final season of amateur baseball. Minoryears is the difference between the earliest season with at least 130 at bats and the year drafted.

BESR is a dummy variable that is coded 1 if the MLB player played in college between 1999 and 2011. Bat exit speed ratio (BESR) is a performance standard for college baseball bats that made hitting far more difficult. These newer bats limited the amount of "pop" and also raised the minimum weight for the bats. This resulted in college bats more closely resembling MLB wooden bats than the previous aluminum bats. College offensive statistics dropped substantially after these new bats introduction. This college era most closely resembles major league hitting of any tested in this sample. The percentage of players that played during this era is 6.8%.

The interaction variables with age-squared and both BESR and CollegeExp are included to see the effects that those dummy variables have on OPS when physical aging takes over. If these coefficients are positive, then the playing in college lessens the aging effect and results in longer careers.

Activeplayers is an indicator variable that identifies the players who may be active currently. This is included for identification purposes only.

General Age and Age^2

Regressing player ops indexed by the mean of his career shows that the peak performance is between the ages of 29 and 30 (Table 2). This is a later peak than previous studies on the subject. It does largely follow what previous analysis has found. The r-squared and therefore the explanatory power of this regression is quite low at 3.1%. This is not surprising since age is only a small part behind the growth or decline of OPS. Overall hitting ability and the expected OPS for a given position also explain a player's

OPS for that season. However, both age and age-squared are statistically significant at the .1% level (Table 3).

One should note that this could be because players with fewer than three seasons with at least 130 at-bats are dropped from the sample. Including players with only one or two observable seasons would drag the peak age of players down. AAAA players are not likely to play later on in their careers.

Player OPS Graphs

Since Hakes and Turner did not index players according to their mean, this will add to their analysis. The graph shows that players reach their career mean at about 26 with the best players peaking higher and later than their peers (Figure 1). The peak is still only around 3% more than their much higher average, with marginal players hardly producing above theirs. Interestingly, Moneyball has a narrative that roughly speaks to this (Page 193). Billy Beane is trying to dump a player, but will not go through with the deal because the player is 26. Although these are pitchers, this does show that General Managers do have a good idea about what a player's output by his age and according production.

This graph more clearly shows the differences between the quintiles career paths (Figure 2). The bottom quintile does not have the parabolic shape of the other quintiles. These players start with lower OPS averages and do not much improve, but simply continue playing at a low level. The top two quintiles improve greatly over their careers.

Even the third and fourth quintiles exhibit the typical parabolic shape that most other career path studies have found.

The quintiles also on average, differ in the number of seasons played in the Minors. The top players spend closer to five seasons in the minor leagues and the bottom quintile closer to six (Table 4). The largest separation is between the first and second quintiles, of about half of a season. The other half season difference occurs between the second and fifth quintiles. This suggests that teams are eager to promote the best players to the Majors, but not as eager to promote marginal and even quite good players.

CHAPTER FOUR

COLLEGE ERAS AND REGRESSIONS

College Bat Regulations

College Baseball is divided into three eras. Unlike MLB, the bats in College Baseball have changed substantially over time. The major aluminum bat production companies will innovate and improve the "pop" of the bats until the NCAA intervenes. Coaches and NCAA administrators complain that the game of college baseball is too offensive and lower offensive production by regulating the aluminum bats. Eventually the bat companies innovate so much within the regulations that new ones must pass to have the desired offensive output.

Aluminum bats were first introduced into the college game in 1974. The bat companies continued to innovate until the NCAA passed the first round of major regulation in 1986. Bats now had a lower weight limit. Lighter bats with the same amount of "pop" are preferred because this increases the speed at which the player can swing. This "weight-limit" era limited offensive production for the first few years after it was enacted. However, the bat companies still continued to innovate and improve the bats hitting.

Much of these innovations focused on enhancing the "trampoline effect" and enlarging the "sweet spot." (Wilson, 2008) The trampoline effect occurs because aluminum bats are hollow, so the metal compresses at contact and pops outward with the ball. Wooden bats do not react to contact like this because they are not hollow and do not have the trampoline effect (USGS, 2004). Aluminum bats make hitting easier by increasing the speed and distance that the ball travels, leading to more runs scored. The sweet spot was also enlarged during the period, making solid contact with the ball easier. In wooden bats, there is a small area where the batter can hit the ball well and have a high chance of getting on base. This area is larger on aluminum bats and allows for poorer contact to create base hits. These two effects greatly increased offensive production in the college game throughout the 1990's, exemplified by the 21-14 score in the 1998 NCAA championship.

The next round of NCAA regulation created the Ball-Exit-Speed-Ratio (BESR) standard. Following the 1998 season (where many scoring records were broken), the NCAA created a higher minimum weight that varied according to the length of the bat. It also regulated the size of the barrel (larger barrels are preferred because it allows for a greater margin of error for the hitter), and limited the speed at which the ball could come off the bat. All of these regulations had the intended effect and greatly lowered offensive production (Russell, 2014).

The Runs scored per game drop with every new round of bat regulation (Figure 3) and the drop off is over about four seasons before runs per game bottom out. When there is a dramatic rise in offensive production over the course of around three seasons, the NCAA steps in and limits offensive production. When Runs per game approaches seven, this is when the NCAA has more strictly regulated the bats.

Bat regulation is also evident in statistics on home runs per game (Figure 4). Each round of regulation happens around a peak in home run production, and the regulations drop home runs substantially. The regulations lower home run production faster than with runs per game. The BESR regulations changed the home runs per game dramatically and over the course of two seasons. This supports the narrative that bat companies continue to innovate to help players increase their offensive production.

Unsurprisingly, NCAA batting averages also decrease with new bat regulations (Figure 5). The declines mirror the previous figures and take place over a few seasons before reaching a new equilibrium. The declines in batting averages demonstrate that bat regulations are effective at increasing the difficulty of hitting and lowering offensive production.

Outside of the obvious difference in the bats, college players play fewer games compared with their minor league counterparts. Minor leaguers play anywhere from 135-144 games a season (Baseball-reference.com), while college the college season maxes out at around 70 games. This will test the hypothesis that players only have a limited number of games that their bodies can handle. Fewer games played from 18-22 may translate to less stress and perhaps more games at the end of one's career.

General College Regression

The college variables are not statistically significant for the overall sample (Table 5). All are slightly negative. The r-squared for the entire sample is 3.4%. This contrasts with the r-squareds for the entire sample, which are all higher. The quintile with the highest r-squared is the second, and is nearly double that of the overall sample. The higher r-squared does not correspond with having the most significant variables, This is reserved for the bottom quintile.

The bottom quintile is where all of the college interaction variables are statistically significant. Former college players in this quintile have lower OPS and age over much faster than their counterparts. The average college player in the sample ages about 10% faster than his non-college peers, but for the bottom quintile, the aging more than doubles. Interestingly, the age-college interaction variable shows that these marginal players improve their OPS due to age much faster than other quintiles.

Age and College Interaction Variables

The impact of college experience on OPS is generally negative, but offset positively with age (Table 6). The effects are largest on the third and fifth quintiles. There are also positive interactions between age and players with college experience during the BESR period.

The college age interaction variable is positive and statistically significant. All quintiles are positive except the top quintile, which is negative, but not statistically significant. The age-squared variable is negative across all quintiles except for the top. The quintiles vary in significance with some not significant at all.

Interestingly the BESR interaction variables mirror the college experience variables but with larger coefficients. Overall none of these interaction variables are significant, but the second quintile seems to hold these variables significant. Overall there does not seem to be a large difference between the way that BESR players and overall college players have in their careers.

Basic Retirement Year Regression

The basic retirement-year regression that does not differentiate between the college eras has no statistically significant college interaction variables (Table 7). The only significant variables for the overall sample are indexplayerops, age, age-squared and playeractive. Playeractive is included in order to account for bias because these players are still active and have not retired. The lack of retirement could skew the data, so the identifier is included.

Again, the bottom quintile is most different because age and age-squared are not statistically significant. Age and age-squared may not factor since these players generally debut at a low OPS and burn out quickly. Marginal players do not have the nice, quadratic productivity profile where age is a significant factor. The OPS coefficient for the worst players is much larger than the other quintiles, telling that OPS is what most effects retirement for marginal players. Age does not matter as much since a slight decline in OPS forces these players out of Major League Baseball.

Retirement Year Regression

A strong inverse relationship is estimated between a player's OPS and the likelihood of retirement (Table 8). A decrease in OPS relative to his career mean corresponds with an increased chance of retirement. This is the most statistically significant of the variables and the largest coefficient. A one-percent decline in OPS when compared to his career average increases the likelihood of retirement by roughly 53%.

The college playing variable is strongly negative and statistically significant. College experience decreases the likelihood of retirement by about 9.8%.

The impact of the BESR variable is evident only in the fifth quintile of players. The impact on retirement is negative with a positive, offsetting age-interaction variable. Playing college baseball after 1999 has a positive effect on retirement age. Combining the BESR variable with the college experience reveals that playing baseball post-1999 increases retirement chances by about 10%. This is a vastly dissimilar outcome than the overall college variable.

However, the make up of the BESR players are different than those found in the college variable. The BESR players who have retired come primarily from the bottom quintile of players and are much worse than average college players. Two-thirds of these players come from the bottom quintile, while the majority of college players are in the third quintile. Whether or not this is indicative of a larger trend that college players are now worse on average, or an anomaly is not as clear.

CHAPTER FIVE

DISCUSSION OF RESULTS

Do college players debut later than their minor league counterparts? (Table 9) Yes, college players reach their first year above rookie status about a year later than those drafted right out of high school. This one-year gap persists throughout the quintiles. Interestingly, there is no discernable difference between the quintiles and all begin their major league career around the same age (24 for minor league players and 25 for college players). This does show that college is an approximate substitute for the Minor Leagues; otherwise the college and minor league players would spend the same number of years playing minor league baseball. This is not the case. Some amount of time is required to adjust to the slightly different major league game and this is the cause of the one-year gap.

Naturally, this gap is also observed by the different amounts of time spent playing baseball in the Minors. College players spend only two years less time in the Minor Leagues than their counterparts, further proving that NCAA baseball is an imperfect substitute for the Minor Leagues baseball.

The gulf continues regarding number of years spent in the Minor Leagues (Table 4). There is a consistent gap of about two seasons in the Minors between college and non-college players across the quintiles, even though the lowest quintile spends an extra season in the minors compared to the top. College players spend less time in the minors on average and about two seasons less than those in their ability quintile.

The average length of a career is between nine and ten seasons (Table 10). There is not much difference between the length of a college player's career and non-college players, but college players do have shorter careers. As expected, there is variation between the quintiles with better hitters playing longer. The top quintile's average career is around three seasons longer than those of the bottom quintile. The largest gap in the number of seasons played between college and non-college is in the top quintile. It is a difference of nearly a season.

Are college players proportionally represented throughout the quintiles? No. There are slight differences between the quintiles (Table 1). The top quintile has the lowest percentage of college players at 36%. This suggests that Major League teams are quite good at identifying talent and induce the best players to sign out of high school with lucrative offers. The second, fourth and fifth quintiles have percentages approximate to 43% of the overall sample. The third quintile contains the second lowest percentage of college players. There is not much of a difference between the second through fifth quintiles, but again the top quintile proves to be unique.

College players peak at different ages than their counterparts, and this holds across the quintiles (Table 2). There is a persistent gap between college and non-college players, with college players peaking later. Within the top quintile, there is a gap of nearly two years between the peak productivity age in college and non-college players. This gap changes between the quintiles, but the college players always have their maximum season later than their counterparts. The delay differs between the quintiles and there is no discernable trend. The middle quintile has the largest split at 2.1 years.

However, there is not as large of a gap between the overall average of the quintiles and that found in other papers. This is probably because the hitters in this sample are being measured against the mean offensive production of each individual career.

The gap in career peaks cannot be entirely explained by the fact that college players debut later in age than their counterparts (Table 11). In the top, third and fifth quintile, the divide is larger between peaks than between debut ages. The second quintile has a small difference, but the fourth quintile has almost a year difference that is opposite all the other quintiles. This is the only quintile where the difference between college and non-college debut ages is greater than the peak performance gap. College players usually peak somewhat later than their peers, even accounting for the difference in debut age, but the fourth quintile tells the opposite story. However, the difference between the peak ages is largely explained by the difference in when a career begins between college and non-college players. Using the player-seasons, instead of the regression tells a slightly different story (Table 12). The largest differences occur between the lower quintiles with college players peaking later. There is no real variation between college and non-college players in the top two quintiles, and may even be slightly negative.

There is also a difference between in the retirement ages between college and non-college players (Table 13). Most of the variation is between quintiles with the top players retiring nearly two years after those in the bottom quintile. Outside of the top quintile, there is a gap of about one year that persists between former NCAA players and their counterparts. The average retirement age for a college player is about one year later than those who signed directly out of high school.

Conclusion

The paths to Major League Baseball differ substantially between Minor League and college players. These differences lead to observable outcomes among players in Major League Baseball.

The different eras in college baseball are evident, regulations have a clear impact on scoring, and bat companies continue to innovate to make hitting easier at the college level. This translates into small differences in the trajectory of major league careers.

This research suggests that if a prospective professional baseball player values education, playing NCAA baseball is a reasonable alternative to signing directly out of high school. Your Major League career is marginally shorter, but your likelihood of retirement is less than your counterparts. This paper demonstrates that playing college baseball is not costly, meaning a player is not significantly disadvantaged by playing college baseball. He enters the majors about a year later and plays about one fewer season on average. He does peak later in his career. However, neither is a college substantially advantaged compared to his peers by playing college baseball. NCAA baseball does not develop high school players into superstars; it is just a reasonable alternative to the Minor Leagues.

However, having one year less of MLB experience can be costly for the best players. The additional season added to a top player's career amounts to millions of dollars in earnings. This likely explains why college experience is lowest in the top quintile of players.

Further Research

Now that NCAA has mandated new bat regulations that have made hitting even harder, perhaps this will lead to further change in how college players adjust to the major leagues. In 2011, the new BBCOR standards lowered homeruns and runs per game more in one season than any of the other major bat regulations. Offensive production continues to fall to levels not seen since the introduction of aluminum bats. Logically one can argue that these players will have an easier time adjusting to the Majors, since the bats are more similar. More seasons and careers will have to transpire before reasonable deductions can be formulated.

APPENDICES

APPENDIX A

Figures

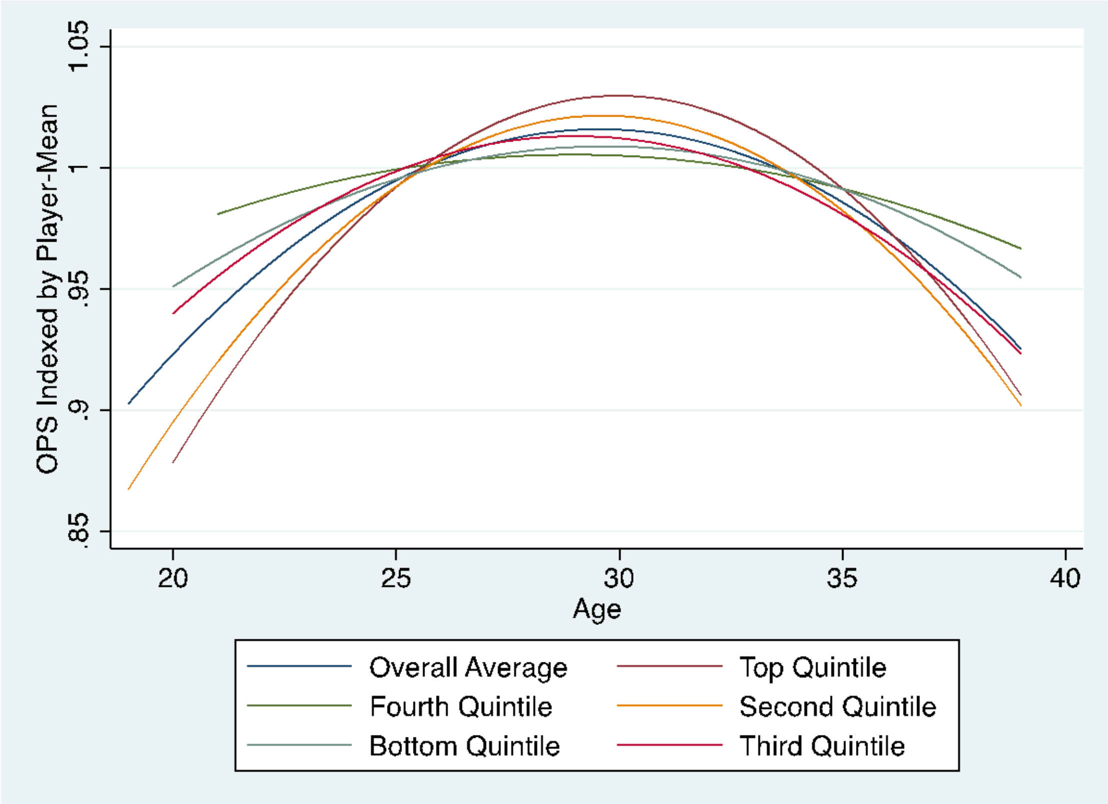

Figure A-1: Productivity Profile Indexed by Player Mean

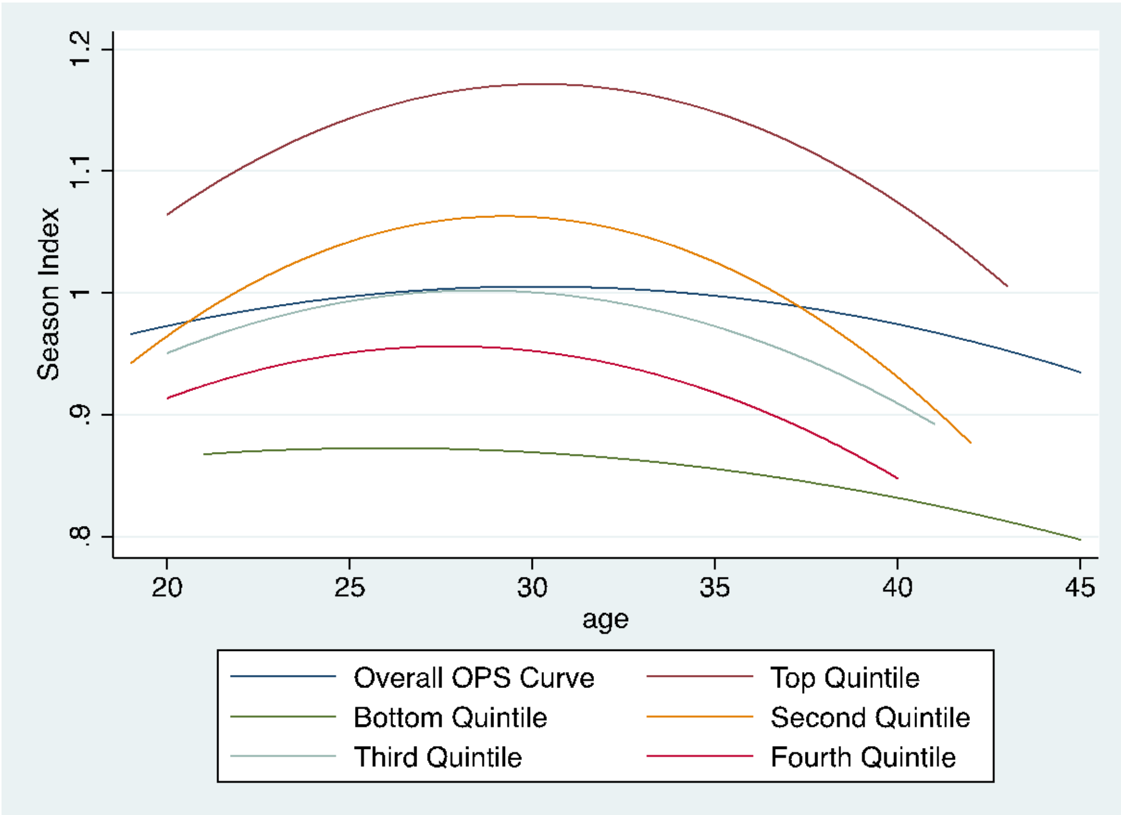

Figure A-2: Productivity Profile Indexed by Season Average

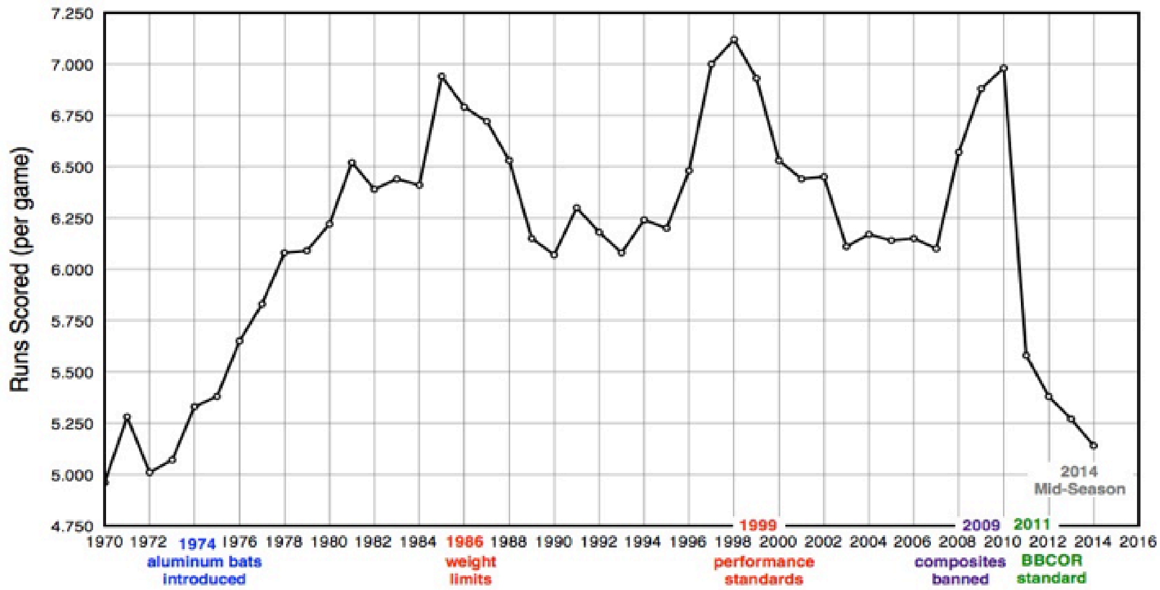

Figure A-3: NCAA Runs Scored per Game

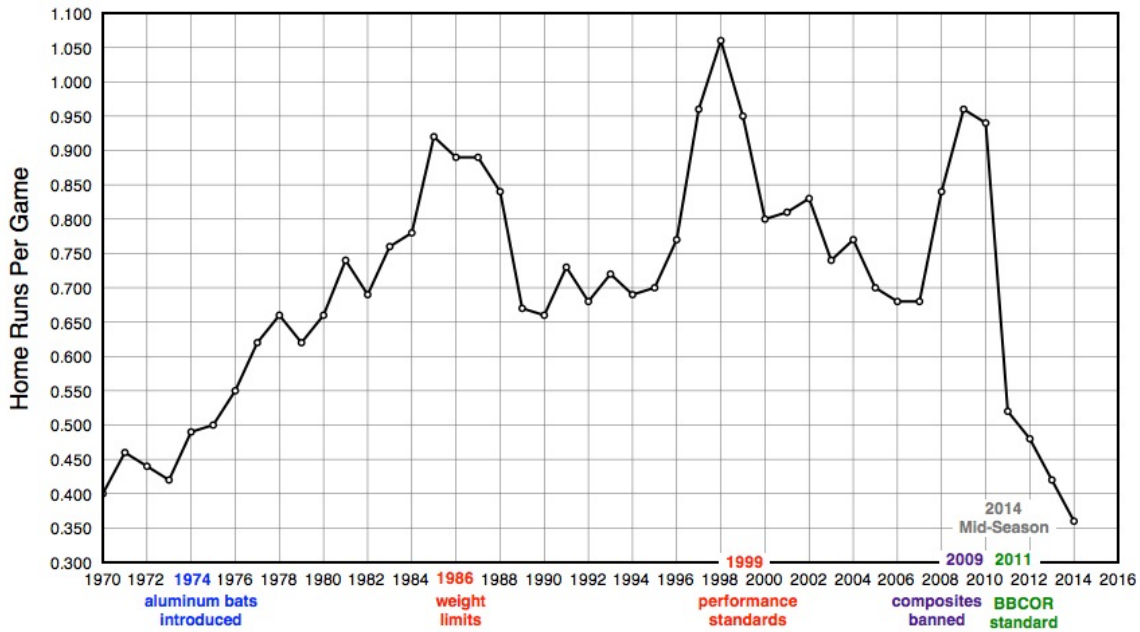

Figure A-4: NCAA Home Runs per Game

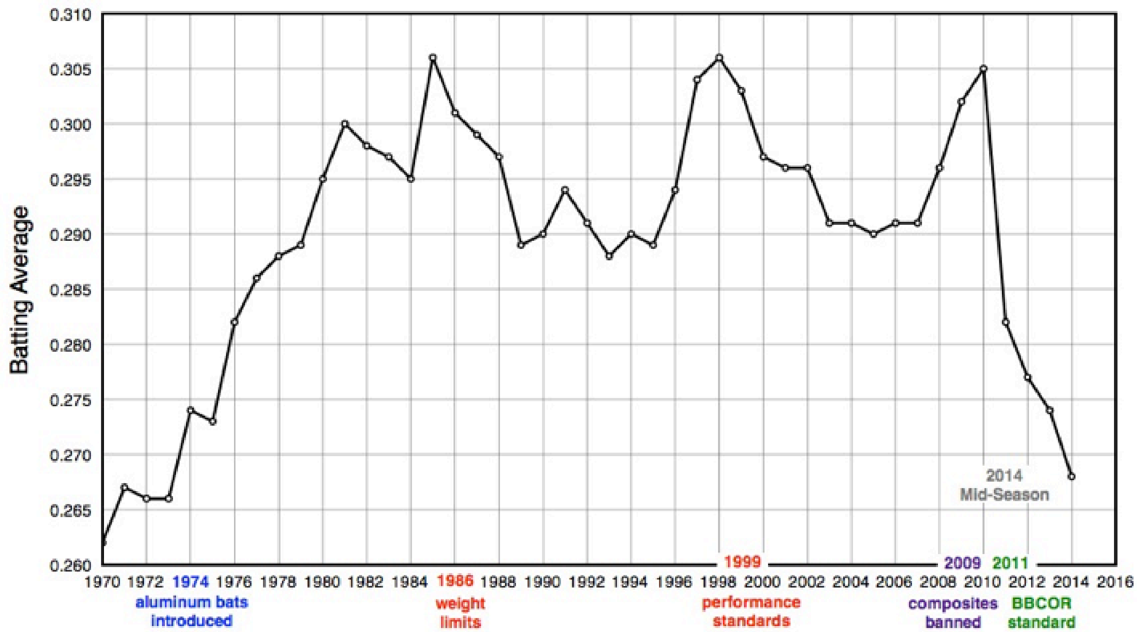

Figure A-5: NCAA Batting Average by Season

APPENDIX B

Tables

Summary of Statistics by Quintile						
VARIABLES	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall
OPS	0.911	0.806	0.75	0.698	0.615	0.756
Index OPS	1.1	1.042	1.004	0.964	0.89	1
age	29.27	29.37	29.15	29.12	29.08	29.2
yearid	2002	2002	2002	2001	2001	2002
At-Bats	461	419.6	395.6	354.3	278.1	381.7
RY-Dummy	0.0177	0.0564	0.0813	0.1115	0.2021	0.0938
Season OPS	1.195	1.061	0.9921	0.929	0.822	1
College Exp	0.3639	0.4219	0.3961	0.4337	0.4311	0.4093
BESR	0.0767	0.0702	0.0597	0.0643	0.0702	0.0682
Minor Years	4.7	5.22	5.3	5.44	5.65	5.03
Observations	1,525	1,524	1,525	1,524	1,524	7,622

Table B-1: Table of Summary Statistics.

Peak Using Position Quintiles				
	College	No College	Difference	Overall Peak
Top 20	31.47	29.45	2.02	30.04
Second 20	30.277	28.99	1.287	29.47
Third 20	30.2	28.055	2.145	28.86
Fourth 20	29.029	28.6	0.429	28.71
Bottom 20	29.76	28.03	1.73	28.99

Table B-2: Quintile Peaks Using Positions

Player OPS Age Regression						
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall
age	0.0422***	0.0474***	0.0412***	0.0386***	0.0405***	0.0560***
	0.00681	0.00535	0.00527	0.0052	0.0058	0.00351
agesq	-0.000703**	-0.000803**	-0.000711**	-0.000670**	-0.000699**	-0.000945***
	0.000113	8.86E-05	8.80E-05	8.65E-05	9.66E-05	5.84E-05
Constant	0.478***	0.355***	0.419***	0.418***	0.316***	0.187***
	0.102	0.0799	0.0778	0.0773	0.0861	0.0521
Observations	1,525	1,524	1,525	1,524	1,524	7,622
R-squared	0.025	0.054	0.046	0.045	0.038	0.034

Notes:

Standard Errors are Highlighted

*** p<0.01, ** p<0.05, * p<0.1

Table B-3: Simple Age-OPS Regression

Minor League Stint Between the Quintiles					
	Top 20	Second 20	Third 20	Fourth 20	Bottom 20
Overall	4.7	5.22	5.3	5.44	5.65
College	3.29	3.96	4.13	4.27	4.45
No College	5.5	6.1	6.08	6.31	6.48
Difference	2.21	2.14	1.95	2.04	2.03

Table B-4: Years Spent in the Minor Leagues

Player OPS Regression						
VARIABLES	Top Quintile	Second Quintile	Third Quintile	Fourth Quintile	Bottom Quintile	Overall
age	0.0487***	0.0443***	0.0378***	0.0401***	0.0284***	0.0547***
	-0.00812	-0.00664	-0.00647	-0.00654	-0.0071	-0.00432
agesq	-0.000830***	-0.000767***	-0.000670***	-0.000699***	-0.000506***	-0.000929***
	-0.000136	-0.000111	-0.000109	-0.00011	-0.000119	-7.26E-05
agecoexp	-0.00815	0.0181	0.0133	-0.00123	0.0351***	0.00748
	-0.0145	-0.0118	-0.0118	-0.0116	-0.013	-0.00779
age2coexp	0.000184	-0.000255	-0.000181	2.98E-05	-0.000560***	-0.000109
	-0.00024	-0.000193	-0.000195	-0.000191	-0.000215	-0.000128
collegeExp	0.0885	-0.305*	-0.231	0.00746	-0.532***	-0.128
	-0.219	-0.178	-0.177	-0.175	-0.196	-0.117
Constant	0.395***	0.411***	0.481***	0.398***	0.495***	0.212***
	-0.12	-0.098	-0.0945	-0.0956	-0.104	-0.0635
Observations	1,555	1,548	1,524	1,509	1,486	7,622
R-squared	0.037	0.061	0.049	0.046	0.043	0.034

Table B-5: Basic College-OPS Regression

OPS Regression with College						
VARIABLES	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall
age	0.0457***	0.0455***	0.0347***	0.0410***	0.0298***	0.0550***
	0.00829	0.0065	0.00652	0.00643	0.00706	0.00429
agesq	-0.000780***	-0.000784***	-0.000618***	-0.000714***	-0.000528***	-0.000932***
	0.000139	0.000109	0.00011	0.000108	0.000119	7.20E-05
agecoexp	-0.0136	0.0156	0.0336***	0.00466	0.0437***	0.0250***
	0.0164	0.0127	0.0119	0.0124	0.0133	0.00817
age2coexp	0.000256	-0.000221	-0.000498**	-5.61E-05	-0.000691***	-0.000377***
	0.000267	0.000207	0.000196	0.000202	0.000218	0.000134
collegeExp	0.189	-0.259	-0.550***	-0.0867	-0.669***	-0.409***
	0.251	0.193	0.179	0.188	0.2	0.124
age2besr	-5.22E-05	-0.00232**	-0.00149*	-0.000374	0.00108	-0.000813
	0.00129	0.00105	0.000804	0.00079	0.00103	0.00055
agebesr	0.00316	0.121**	0.0757*	0.0117	-0.0739	0.0292
	0.0707	0.0582	0.0459	0.0455	0.06	0.0312
besr	-0.0688	-1.565*	-0.944	-0.0382	1.226	-0.176
	0.962	0.804	0.651	0.651	0.865	0.441
Constant	0.437***	0.393***	0.527***	0.385***	0.475***	0.205***
	0.123	0.096	0.0952	0.0941	0.104	0.063
Observations	1,525	1,524	1,525	1,524	1,524	7,622
R-squared	0.035	0.069	0.064	0.055	0.053	0.045

Standard Errors are Highlighted

*** p<0.01, ** p<0.05, * p<0.1

Table B-6: College Era-OPS Regression

RY Dummy Regression						
VARIABLES	Top Quintile	Second Quintile	Third Quintile	Fourth Quintile	Bottom Quintile	Overall
indexplayerops	0.038	-0.0343	-0.182*	-0.16	-0.438***	-0.499***
	0.0417	0.0829	0.0967	0.11	0.113	0.0301
age	-0.0674***	-0.0897***	-0.0495**	-0.0719**	0.0328	-0.0316***
	0.0133	0.0218	0.0246	0.0281	0.0312	0.0115
agesq	0.00126***	0.00175***	0.00118***	0.00162***	4.14E-05	0.000862***
	0.000223	0.000364	0.000416	0.000475	0.000524	0.000193
agecoexp	0.0167	0.0115	-0.0327	-0.0161	0.0141	-0.00991
	0.0236	0.0381	0.0443	0.0493	0.0571	0.0205
age2coexp	-0.00035	-0.0002	0.00055	0.000372	-4.45E-05	0.000239
	0.000389	0.000625	0.000732	0.000809	0.00094	0.000337
collegeExp	-0.192	-0.177	0.454	0.142	-0.413	0.0695
	0.355	0.575	0.664	0.744	0.857	0.308
playeractive	-0.0144*	-0.0524***	-0.0887***	-0.122***	-0.237***	-0.0962***
	0.00749	0.0133	0.0156	0.019	0.0244	0.00737
Constant	0.863***	1.207***	0.719**	0.985**	-0.343	0.794***
	0.196	0.32	0.359	0.408	0.46	0.167
Observations	1,555	1,548	1,524	1,509	1,486	7,622
R-squared	0.064	0.109	0.137	0.178	0.234	0.161

Table B-7: Basic College Retirement Year Regression

Retirement Year Dummy Regression						
VARIABLES	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall
indexplayerops	0.0731*	-0.02	-0.196**	-0.245**	-0.412***	-0.504***
	0.0419	0.0842	0.098	0.108	0.111	0.0303
age	-0.0602***	-0.0961***	-0.0401	-0.0780***	0.0279	-0.0317***
	0.0135	0.0216	0.0251	0.0273	0.0307	0.0115
agesq	0.00112***	0.00186***	0.00101**	0.00174***	0.000107	0.000860***
	0.000226	0.00036	0.000424	0.00046	0.000516	0.000193
age2coexp	-3.96E-05	-0.000409	0.000502	0.000486	0.000399	0.000154
	0.000432	0.000675	0.000748	0.000843	0.000944	0.000353
agecoexp	-0.00183	0.0252	-0.0282	-0.0235	-0.00881	-0.00217
	0.0266	0.0415	0.0455	0.0517	0.0575	0.0216
collegeExp	0.0792	-0.405	0.35	0.249	-0.156	-0.0986
	0.406	0.631	0.684	0.785	0.864	0.327
age2besr	0.00378*	-0.00226	0.00242	-0.00244	-0.0104**	-0.00145
	0.00209	0.00343	0.00306	0.00331	0.00446	0.00146
agebesr	-0.194*	0.116	-0.139	0.128	0.581**	0.0718
	0.114	0.19	0.175	0.191	0.259	0.0826
besr	2.486	-1.443	2.038	-1.587	-7.880**	-0.783
	1.556	2.628	2.478	2.728	3.735	1.167
playeractive	-0.0168**	-0.0581***	-0.0981***	-0.126***	-0.270***	-0.111***
	0.00804	0.0142	0.0166	0.02	0.0253	0.00787
Constant	0.726***	1.285***	0.605*	1.147***	-0.274	0.806***
	0.2	0.316	0.366	0.396	0.452	0.167
Observations	1,525	1,524	1,525	1,524	1,524	7,622
R-squared	0.063	0.115	0.135	0.185	0.244	0.164

Table B-8: College Era Retirement Regression

Age when rookie status is lost						
Player-year Average	Overall Average	Top Quintile	Second Quintile	Third Quintile	Fourth Quintile	Bottom Quintile
College	25.13	24.43	25.06	25.19	25.38	25.48
No College	23.8	23.29	23.84	23.8	24.01	24.18
Difference	1.33	1.14	1.22	1.39	1.37	1.3
Player Average						
College	25.48	25.42	25.72	25.53	25.42	25.37
No College	24.15	24.25	24.28	24.28	24.17	23.92
Difference	1.33	1.17	1.44	1.25	1.25	1.45

Table B-9: Age when 130 At-bats are Reached

Career Length Between College and Non College			
	College	No College	Difference
Top 20	11.02	11.82	-0.8
Second 20	9.63	10.04	-0.41
Third 20	9.4	9.41	-0.01
Fourth 20	8.75	9.11	-0.36
Bottom 20	8.27	8.44	-0.17
Overall	9.36	9.81	-0.45

Table B-10: MLB Career Length

Difference in Peak versus Gap of Debut Age					
	Top Quintile	Second Quintile	Third Quintile	Fourth Quintile	Fifth Quintile
Peak Gap	2.02	1.287	2.145	0.429	1.73
Age Gap	1.14	1.22	1.39	1.37	1.3
Difference	0.88	0.067	0.755	-0.941	0.43

Table B-11: Difference in Quintile Peaks Minus the Difference in Debut Age between College and Non-college Players

Peak using Seasons						
	Top 20	Second 20	Third 20	Fourth 20	Bottom 20	Overall
College	28.58	28.5	29.19	28.83	28.44	28.89
No College	28.64	28.7	27.99	27.91	27.83	28.27
Difference	-0.06	-0.2	1.2	0.92	0.61	0.62

Table B-12: Peak Estimation Using Seasons

Retirement Age of Players By Quintile				
	College	No College	Difference	Overall
Top 20	35.36	35.22	0.14	35.27
Second 20	34.82	34.08	0.74	34.38
Third 20	34.58	33.49	1.09	33.93
Fourth 20	34.24	33.36	0.88	33.74
Bottom 20	33.83	32.82	1.01	33.23

Table B-13: Retirement Age of Players

LITERATURE CITED

Albert J (1999) Bridging different eras in sports: comment. J Am Stat Assoc 94:677–680

Aubert, P., & Crépon, B. (2003). Age, Wage and Productivity: Firm Level Evidence. *Center for Research in Economics and Statistics, 363*, 75-119.

Becker, G. (1962). Investment In Human Capital: A Theoretical Analysis. *Journal of Political Economy*.

Minor League Baseball History | Baseball-Reference.com. (n.d.). Retrieved October 8, 2014

Fair RC (2008) Estimated age effects in baseball. J Quant Anal Sports 4:11

Hakes, J., & Sauer, R. (2007). The Moneyball Anomaly and Payroll Efficiency. *International Journal of Sport Finance, 2*(4), 177-189.

Hakes, J., & Turner, C. (2009). Pay, productivity and aging in Major League Baseball. *Journal of Productivity Analysis,* 61-74.

James, B. (1988). *The Bill James Historical Baseball Abstract*. New York: Villard Books

Krohn GA (1983) Measuring the experience-productivity relation- ship: the case of Major League Baseball. J Bus Econ Stat 1:273– 279

Lewis, M. (2003). *Moneyball: The Art of Winning an Unfair Game*. New York: W.W. Norton.

Russell, D. (2014, June 12). Non-Wood Bats: Performance Trends in College Baseball. Retrieved October 9, 2014.

United States Geological Survey. 2004. Baseball Bat Innovations. Minerals in Sports. Washington, DC

Wilson, D. (2008, October 17). Effort to Get Metal to Act Like Wood. *The New York Times*. Retrieved October 12, 2014