MARKET EFFICIENCY OF STATISTICS IN MAJOR LEAGUE BASEBALL:

A LOOK AT THE *MONEYBALL* HYPOTHESIS TEN YEARS LATER

by

Seth Lawrence Rubin

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Honors Bachelor of Science in Economics with Distinction

2013

MARKET EFFICIENCY OF STATISTICS IN MAJOR LEAGUE BASEBALL:

A LOOK AT THE *MONEYBALL* HYPOTHESIS TEN YEARS LATER

by

Seth Lawrence Rubin

Approved: ______________________________

Dr. Charles Link, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: ______________________________

Dr. Daniel Brown, Ph.D.
Committee member from the Center for Applied Demography & Survey Research

Approved: ______________________________

Dr. Meryl Gardner, Ph.D.
Committee member from the Board of Senior Thesis Readers

Approved: ______________________________

Michael Arnold, Ph.D.
Director, University Honors Program

ACKNOWLEDGMENTS

I first and foremost would like to thank Dr. Link for all of his assistance and support. He supported me in finding a way to combine my passion for baseball and economics to write this senior thesis. I am truly grateful for all of his help.

Dr. Brown was extremely helpful in updating the database. He helped me particularly with putting all of the salary and statistics information together and with the regressions and without his support this thesis would not have been possible. He really pushed me and challenged me along the way.

I would like to thank Dr. Gardner, my third reader, for all of her support and help in planning the group presentations and for her feedback.

I also need to thank Alex Gilchrist, Zefan Wang, Kaijing Wang, and Ricky Shum for helping to input the salary information into the access database.

Also, I would like to thank Eddie Epstein for his generosity in taking his time to help me edit my paper and for taking his time to talk to me about his research.

I want to thank my friends and family for their continued support, especially Lindsey Calabro and Cirrus Foroughi, who supported me in pursuing my goal of successfully completing a senior thesis.

Finally, I would like to thank the Undergraduate Research Program who were willing to support me and help me with any questions I had throughout the process to make this possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF EQUATIONS

ABSTRACT

Moneyball (2003), written by Michael Lewis, is about the Oakland Athletics' path to success in the early 2000's focusing on the 2002 season. This senior thesis investigates the impact of *Moneyball*. Namely, the hypothesis that after *Moneyball* more weight has being given to the statistics of on-base percentage and slugging percentage when it comes to evaluating players to determine their salaries, as well as other statistical measures. To evaluate this hypothesis, regressions were estimated to improve upon those by Hakes and Sauer (2005), who looked at the impact of on-base percentage and slugging percentage on salary. Next on-base percentage, slugging percentage, on-base plus slugging percentage, weighted runs created, win shares, and wins above replacement were evaluated to determine their impact on the natural log of salary. Models were first estimated individually for the time period of 1996 to 2002 and then for 2003 to 2011. Then standard tests were conducted to determine if the changes in the coefficient estimates were significant. The results are evident that on-base percentage, slugging percentage, on-base plus slugging percentage, weighted runs created, win shares, and wins above replacement are more strongly related to salary in the evaluation of Major League Baseball player salaries since the release of *Moneyball.*

Chapter 1

INTRODUCTION

In 2002, the Oakland Athletics shocked the Major League Baseball world. After winning 91 games in 2000 and 102 games in 2001, they had just lost their top three players to free agency and were expected to be non-contenders in 2002.

> First baseman Jason Giambi had left for the Yankees for $120 million over seven years. Outfielder Johnny Damon had gone to the Red Sox for $32 million over four years. Closer Jason Isringhausen had signed with the Cardinals for $28 million over four years. The $33 million the three players would make each year was just $5 million less than the entire team (Lewis 22-23).

Despite the loss of their three key players, they still were able to win 103 games, tying for the league lead in victories in 2002.

The Oakland Athletics, with the third lowest payroll in all of Major League Baseball, did not have a large budget for salaries, making them unlike most teams. According to former Oakland General Manager Sandy Alderson: "We had new owners who weren't going to spend any money…" (Lewis 58). The owners kept this promise and therefore the A's had to find other ways to evaluate and value players and other ways to win. They did this through evaluating talent based on their ability to get on base and score runs, which is the major premise in *Moneyball* (2003).

The Athletics tied for the league lead in wins because they were the most efficient team in Major League Baseball, spending "half a million dollars per win" (Lewis XIII). The New York Yankees won the same amount of games, but did it with a payroll three times as large as the A's. The Athletics were able to be efficient through their ability to evaluate a player's value better than any other team by using statistical analysis.

While the A's were not the first team to use statistical analysis to evaluate players, they were the first to receive national attention for doing so, especially from author Michael Lewis. The idea of applying statistical analysis to Major League Baseball began when Bill James (1977) wrote his *Baseball Abstract*. He saw that the measurements of a player's skill were flawed because as he put it, "it should be obvious that the purpose of an offense is not to compile a high batting average," but rather "a hitter should be measured by his success in … trying to create runs" (Lewis 76). Beginning in 1986, Eddie Epstein was able to use statistical analysis working in baseball operations full-time for two teams over a ten year periods (Epstein). Despite his success, it was not until the success of Beane and DePodesta in the Oakland front office which led to the overwhelming popularity of using statistics. This allowed the A's to become the most "*efficient*" team in Major League Baseball (Lewis 123).

The success of the Oakland Athletics led to other teams using statistical analysis to evaluate players. *Moneyball* chronicled how they took advantage of the fact that the ability to get on-base was undervalued. In this research, the significance and value of major hitting performance measures that are correlated with success on the ball field and scoring runs, in relation to salary is evaluated. The key question in the research is as follows. Did the impact of *Moneyball* lead to changes in the variables MLB teams use to determine player salaries?

The plan of the paper is as follows. Chapter 2 provides a brief survey of literature. Chapter 3 describes the methodology underlying the paper. Chapter 4 includes the empirical results, and Chapter 5 includes the conclusion.

Chapter 2

LITERATURE REVIEW

2.1 Hakes and Sauer

The most widely cited papers on the topic of the influence of *Moneyball* by Michael Lewis were released by Hakes and Sauer (2005, 2007). As noted, *Moneyball* examined how the Oakland Athletics were able to take advantage of information that was undervalued by other teams when evaluating player worth. It was through their innovative methods of evaluating players through the use of statistical analysis that brought change to the methods of evaluating Major League Baseball players. Hakes and Sauer used econometric techniques to test whether there was a change in the mindset of the front offices of Major League Baseball after *Moneyball*.

2.1.1 An Economic Evaluation of the *Moneyball* Hypothesis

In 2005, Hakes and Sauer introduced their first paper on this topic titled, "An Economic Evaluation of the *Moneyball* Hypothesis." They tested for an effect of *Moneyball* by looking at the timeframe of 2000 to 2004, to cover the time prior to and post *Moneyball*.

To perform this analysis Hakes and Sauer estimated individual regressions for the years 2000 to 2004 to test the impact of on-base percentage and slugging percentage from the prior year on the natural log of salary for the current season. They used a sample of all position players (non-pitchers) with at least 130 at-bats. Plate appearances, the ability to negotiate one's contract through arbitration eligibility or free agency, and position were included as control variables.

They found that slugging percentage was more significant and valuable in its returns to salary for 2000 to 2003 but in 2004 the coefficient for on-base percentage is significant and larger than that of slugging percentage. This led them to conclude that the market was no longer inefficient in 2004 but that further research was required to see if this change in market valuation was permanent.

2.1.2 The *Moneyball* Anomaly and Payroll Efficiency: A Further Investigation

Hakes and Sauer (2007) published a second research paper looking at the impact of *Moneyball* on the labor market of Major League Baseball. To conduct their research they focused on two separate regression equations that relate to the "Hitting" aspect of the game. In the first equation, "Hitting" is a function of the variables on-base percentage and slugging percentage, as was the case in their previous study. In the second equation, "Hitting" is a function of three variables called, "Eye," "Bat," and "Power."[1] To perform an analysis, the data was separated into four separate time periods (1986 to 1993; 1994 to 1997; 1998 to 2003; and 2004 to 2006). The main purpose was to focus on the impact of *Moneyball* during the 2003 season to see if it had an impact on the subsequent time period in the evaluation of players.

They estimated pooled regressions for each of their four time periods and 21 yearly regressions for 1986 to 2006. In each case the dependent variable was the

---

[1] $Eye = \frac{Base\ on\ Balls + Hit\ By\ Pitch}{Plate\ Appearances}; Bat = \frac{Hits}{At\ Bats}; Power = Isolated\ Power = \frac{Total\ Bases - Hits}{At\ Bats} = \frac{(Singles) + (Doubles * 2) + (Triples * 3) + (Home\ Runs * 4) - Hits}{At\ Bats}$

natural log of salary. The independent variables of interest were the hitting statistics. Position, plate appearances, contract and year were included as control variables. Contract is a dummy variable for whether the player is a free agent or arbitration eligible and year is a "set of indicators to control for annual fixed effect due to changes in overall player demand, and other season-specific phenomenon" (Hakes and Sauer 2007).

The first set of results verified the hypothesis that the return to slugging percentage was significant and constant in its value to determining salary. The other result was that on-base percentage was not significant and less valuable to determining salary from 1986 to 2003. However, in 2004 the inefficiency in the market was corrected and on-base percentage became 1.5 times more valuable than slugging percentage. The individual yearly regression results indicated that slugging percentage was valuable and significant for every year studied. However, on-base percentage was only significant at the 1% level in 2004 and 2005 but was then insignificant again in 2006. This led to the conclusion that 2006 could have either been an aberration or that Major League Baseball market had overvalued on-base percentage in 2004 and 2005.

In the second regression estimated by Hakes and Sauer, they evaluated the significance of their variables *Eye*, *Bat*, and *Power* in the four time periods. Their results showed that the ability to hit for power and get base hits was significant and important for 1986 to 2006. The ability to get on base via walks (*Eye*) was only significant and important for 1998 to 2006. When evaluating the yearly regressions, the ability to get on base via hits or the ability to hit for power were consistently statistically significant determinants of salary. However, *Eye* was only significant for 1997 to 2000 and for 2004 to 2005. Similar to on-base percentage, the ability to walk in 2004 may have been overvalued after the release of *Moneyball*, with the market correcting itself in 2006.

Based on the studies by Hakes and Sauer it is difficult to conclude whether Major League Baseball corrected itself in the way it valued players after *Moneyball* or if it was a short term anomaly. Therefore, the current paper will improve upon the studies of Hakes and Sauer to reach conclusions on whether *Moneyball* changed the way Major League Baseball teams evaluated players through the use of statistical analysis and whether on-base percentage became more significant and important in terms of determining salary.

Chapter 3

Methodology

3.1 Data

Through working with Dr. Link, I was able to work with an extensive data set that he has compiled since 1984. Since the original research was completed, the data set has been maintained and updated by Dr. Link and Dr. Brown, with the assistance of graduate and undergraduate students. Over the past nine years, the data set has been the basis of two PH.D dissertations, seven Masters papers and three undergraduate senior theses. Other research has shown their source as public information available on websites that contain data on Major League Baseball Salaries. The salary data in the current thesis is based on what are called the Joint Exhibits. This document is put together by Major League Baseball. The Joint Exhibits contain detailed information on players' contracts including the actual amount that players were paid whereas available public information regarding contracts may be erroneous. Therefore, it is the most reliable and accurate data on Major League Baseball salaries. Each player per year has to be inserted into a database with their player ID, their major league experience, and their contract information which includes the year of the contract and the year it started, as well as any buyouts or options in the contract. The salary amount is then calculated by adding the base salary plus the signing bonus prorated over the length of the contract.

In addition to the salary information, the database is updated with the most recent statistics to match up with the salary database. Specifically the statistics Wins

Above Replacement and Runs Created were downloaded from FanGraphs.com. Win Shares, a statistic developed by Bill James was downloaded from www.billjamesonline.com.

Once the data was uploaded into Stata, it was evaluated using the summary statistics and scatterplots to look for potential outliers in the data for all players who were eligible for free agency. The outliers were distinguished by their high statistics despite low salaries, leading to a potential skew in the analysis. They were removed via the use of a variable where only those with 130 at-bats would be included. This is the amount of at-bats that counts as a player completing a full season for rookies, and is also the number of at-bats chosen by Hakes and Sauer in their analysis as the minimum amount of at-bats to be included in the study.

3.2 Analysis

The overall question that Lewis poses in *Moneyball* is whether there were certain statistics that were undervalued and exploited by the Oakland Athletics that allowed them to win as many games as the Yankees, despite the major difference in payroll. Over the three seasons of 2000 to 2002, "Oakland A's had paid about half a million dollars per win," whereas "the Baltimore Orioles, for instance, or the Texas Rangers–paid nearly $3 million for each win" (Lewis XIII). By spending as little as one-sixth the amount as other teams per win, the Oakland A's proved that there was an inefficiency in Major League Baseball in regard to evaluating players.

The belief in *Moneyball* is that players were being misevaluated in Major League Baseball. While the A's focused on using statistical analysis to evaluate a player's performance, most other teams were not using these same techniques. *Moneyball* discusses this in the chapter titled "Field of Ignorance." In this chapter, James discusses how most scouts and general managers focus on a player's ability to hit and their amount of runs batted in (RBIs). He stated that instead a, "hitter should be measured by his success in that which he is trying to do, and that which he is trying to do is create runs." Following the advice of Bill James, Billy Beane and Paul DePodesta focused on using statistics to evaluate players to decide who could provide the most value to their team.

In my research, the main focus is on three statistics discussed in *Moneyball* that were utilized by the Oakland Athletics. These statistics are on-base percentage (OBP), slugging percentage, on-base percentage plus slugging percentage (OPS), and runs created. These are statistics most associated with creating runs and winning ball games. The hypothesis is that after the release of *Moneyball* these statistics have grown in magnitude as determinants of players' salaries.

3.3 Statistics

3.3.1 On-Base Percentage

The protagonist in *Moneyball* is Billy Beane. However, the beginning of the focus of on-base percentage began with former Oakland Athletics General Manager and current Mets General Manager Sandy Alderson. He believed that the whole team would be most successful if they could maximize the amount of times they got on base. This in turn would lead to more runs scored. When teams can maximize the runs they score they are expected to win more games. This chain of thought was transferred from Alderson to Beane when he was first hired in 1993 by the Oakland Athletics. Beane continued with this philosophy when he was hired to take over the position of General Manager in 1997.

Another major name in the front office of the Oakland Athletics at the time was Paul DePodesta. An economics major at Harvard, DePodesta was not the average front office employee. However, he had a passion for baseball and a mind to analyze it. During his studies he analyzed what made teams most successful at winning. He looked for correlations between performance statistics and winning percentage. It was no surprise that he "found only two, both offensive statistics, inextricably linked to baseball success: on-base percentage and slugging percentage" (Lewis 127). Despite his studies he found that "a player's ability to get on base – especially when he got on base in unspectacular ways – tended to be dramatically underpriced" (Lewis 128). This main thought is what allowed the Oakland Athletics to be so successful in maximizing success given their limited budget.

On-base percentage is easily calculated as the amount of times that a player gets on base divided by the number of times they appear at plate[2]. Therefore players with higher on-base percentages should be more highly valued because they help their teams win more. It is based on this notion that it is an important performance measure to focus on when researching the relationship between major league baseball salaries and different measures of hitting performance.

---

[2] OBP= $\frac{Hits+Walks+Hit\ By\ Pitch}{At\ Bats+Walks+Hit\ By\ Pitch+Sacrifice\ Flies}$

3.3.2 Slugging Percentage

The other statistic found by Paul DePodesta to be significantly correlated with winning, was slugging percentage which is a weighted measure of a player's hits, total bases divided by at bats[3].

3.3.3 OPS (On-Base Plus Slugging)

The addition of Slugging Percentage to On-Base Average creates another statistic that can be used to evaluate a player's performance. Despite how basic it was to add OBA and Slugging Percentage, "it was a much better indicator than any other offensive statistic of the number of runs a team would score" (Lewis 128). As seen in the prior section, DePodesta had found that the only two statistics highly correlated with winning were on-base percentage and slugging percentage. Therefore it should be expected that a combination of both would provide a better measure of success. Unfortunately, research shows that on-base percentage and slugging percentage are not equally weighted when it comes to creating runs. According to DePodesta's research, "an extra percentage point of on-base percentage was worth three times an extra point of slugging percentage" (Lewis 128). Despite the potential flaws of the weighting of the individual statistics of OBA and slugging percentage in calculating OPS, it is still a widely used gauge for success.

3.3.4 Weighted Runs Created Plus

Runs Created is a statistic that was created by Bill James to predict how many runs a player contributes to his team. James originally began looking at the problem

[3] Slugging Percentage = $\frac{1B+(2*2B)+(3*3B)+(4*HR)}{At\ Bats}$; where 1B=signles, 2B=doubles, 3B=triples and HR = Home Runs

back in 1979 in his *Baseball Abstract*. The goal as obviously stated by James, is that when a player comes to the plate they want to create runs. Unfortunately, back when this model was being developed, there was a lack of statistics available at his disposal. After evaluating team statistics which were more readily available he came out with his first equation for runs created[4]. Over time this equation has evolved with the availability of statistics and analysis techniques to allow for better predictive models. These have now enabled statisticians to use the model as a way to evaluate players as well.

Runs Created today is used as a method to evaluate the number of runs a player contributes to their team. Hitters are being paid to create runs for their teams which in turn helps contribute to games won for their team. Paul DePodesta did not use the Bill James' exact version of Runs Created, but rather amended it to make his own predictive model. Due to the lack of attention that Bill James received, the "Bill James Baseball Abstract," and the statistic of Runs Created was not used by Major League Baseball teams to evaluate player worth.

In this research, a more recent calculation used by baseball fanatics known as Weighted Runs Created was used instead of the computation used by either Bill James or Paul DePodesta. Weighted Runs Created allows for a computation of runs created and compares it to the rest of the players in the league. It is a function of their weighted on-base average. The statistic was developed in a book titled, "The Book: Playing the Percentages in Major League Baseball" by an author who goes by the alias

[4] Original Runs Created formula: $RC = \frac{(Hits+Walks)*Total\ Bases}{At\ Bats+Walks}$

of Tom Tango. He developed weighted on Base Percentage[5] which provides run values to each major event that allows a hitter to reach base (intentional walk, walk, hit by pitch, single, double, triple and home run). These run values are changed yearly to adjust the formula to the actual number of runs scored. This allows users of the statistic to really understand the value of the player by giving each positive outcome a value. The average wOBA for the league will be the same as that for the average league On-Base Percentage. Instead of using the simple math dividing the number of times getting on base by the amount of times going up to hit, it is scaled to give outcomes meaning and values. This is used to create the formula for weighted Runs Created (obtained from FanGraphs.com):

$$wRC = \left(\frac{wOBA - League\ wOBA}{wOBA\ Scale6} + \frac{League\ Runs}{Plate\ Appearances}\right) * Plate\ Appearances \quad (1)$$

The wOBA scale is what is used to get the wOBA to equal the average OBP for the league. This formula calculates the weighted Runs Created, which provides a better understanding of the contribution of a player to creating runs for their teams, compared to the older formulas developed by Bill James. Then weighted Runs Created becomes weighted Runs Created Plus when it is scaled to the league average which is set at 100. As a result the league average for wRC is going to be equal to 100.

---

[5] Original formula for
wOBA=$\frac{(.72*BB)+(.75*HBP)+(.9*1B)+(.92*ROBE)+(1.24*2B)+(1.56*3B)+(1.95*HR)}{Plate\ Appearances}$

BB= Walk; HBP=Hit By Pitch; 1B = single; RBOE = reached base on error; 2B = double; 3B=triple; HR = home run

[6] wOBA Scale = $\frac{Weighted\ On-Base\ Average}{League\ Average\ On-Base\ Percentage}$

Therefore:

$$wRC+=\frac{wRC}{league\ Average\ wRC}*100 \tag{2}$$

This makes it very easy to interpret the statistic. It is just the percentage of the individual's player weighted Runs Created compared to the league. This would be useful in making comparisons between players across the league.

I used weighted Runs Created as a statistic as a determinant of player salaries, despite it just being meant as a better way to evaluate a player's contribution. It is more advanced than On-Base Percentage and the original Runs Created formula. This allowed for the ability to judge a player's contributions to scoring runs on a better scale in general and for comparison purposes.

3.3.5 Win Shares

Win Shares was established by Bill James (2002) in his book titled *Win Shares*. Similar to wins above replacement, Win Shares looks at the contributions of a player to his teams wins throughout a season. The basic notion is that the sum of win shares for the players on an individual team should equal three times the amount of wins the team won because a win share is one-third of a win. His calculation takes wins as a result of hitting, fielding, base running and pitching. Win Shares is a valuable metric because it is a more complete evaluation of a player's actual contribution to team wins. Unfortunately, the calculation for win shares is quite long and difficult. Therefore, my research only focuses on the final result (win shares) rather than its component's. The data on win shares was downloaded from BillJamesonline.com.

Although Win Shares is not mentioned in *Moneyball*, the Bill James book was originally released during the period that I evaluate in the research. This is the 2002 season chronicled in *Moneyball*. Because it was released during the season chronicled in *Moneyball*, it will be interesting to see if its effect is larger post 2003. The release of "Win Shares" also happened to occur right before James was hired by the Red Sox as a consultant (Neyer). It will be interesting to see how valuable this new statistic became to teams.

3.3.6 Wins Above Replacement

One final performance measurement is Wins Above Replacement. It is very similar to win shares, but has recently taken over as the better predictor of a player's contribution to his team (FanGraphs.com). Wins Above Replacement is a statistic that evaluates the value contributed to a team by a player in terms of wins and compares their contribution to the statistical performance for a replacement player. This concept was introduced by Bill James who described replacement level, as a below-average player who is just good enough to make the majors (Epstein). Wins Above Replacement also evaluates the "wins" which must be replaced if a player becomes injured or leaves via free agency. The statistic is a function of the player's hitting, base running and fielding skills. The player's hitting is once again a function of wOBA, similar to RC+. However, in this calculation it needs to be adjusted to ball-park factors which are a function of the hits and runs the team scores at home versus on the road. The base running is based on a function of how well a player can get an extra base or steal a base because these help to create runs and therefore help teams win games. Finally, fielding is a function of the player's arm strength, their ability to turn double plays, their range to get the baseball and the number of errors the player makes.

These are all combined to get the number of wins the player contributes to his team. This is then compared to the individual player's replacement which calculates the player's wins above replacement (FanGraphs.com).

Despite the difficulty in calculating this statistic it is a good indicator of player performance and how much a player should be paid. One of the major discussions of *Moneyball* is how the Oakland Athletics needed to find players to replace Jason Giambi, Johnny Damon and Jason Isringhausen, all of whom left as free agents. Beane and DePodesta realized that to be successful they needed to find players who could make up for the number of wins they lost as a result of the defections. The bottom line is when players leave a team, they need to be replaced in terms of the wins they contribute to the team.

Wins above replacement is not a performance statistic that is discussed in *Moneyball*, unlike on-base percentage, OPS or Runs Created. It is a new statistic that is used by sabermetricians and calculates a result that is easy to understand, despite the complication in its original calculation. One of the main focuses of *Moneyball* was on the Athletics' ability to replace the players they lost to free agency. Therefore it is expected that Wins Above Replacement will be important factor in deciding how best to measure the contribution of players relative to others.

3.4 Econometric Analysis

The overall purpose of my research is to look at the impact of various performance measures on the salaries of Major League Baseball players. Specifically these measures and their impact on salaries are evaluated for contracts that began from1996 to 2002 (Pre-*Moneyball* Era) and contracts that began from 2003 to 2011 (Post-*Moneyball Era*). These time periods were also selected because they coincide with the collective bargaining agreements in Major League Baseball. The first agreement was from 1996 to 2002, signed after the Major League Baseball strike from 1994. For the 2003 season a new collective bargaining agreements was ratified. This new agreement allowed for greater revenue sharing between the wealthier and poorer teams and also put a luxury tax on those teams that spent over a specific threshold (Zimbalist). It was subsequently renewed in 2006, with minor changes. Therefore these two major time periods not only coincide with *Moneyball* but also with the Collective Bargaining Agreements, which may have also have had an impact on the way that players were compensated that cannot necessarily be measured.

Multiple regressions were estimated to complete this analysis. Only players who are eligible to be free agents at the time they signed their contracts are analyzed (players with six or more years of experience). In the regressions, the dependent variable is the natural log of salary set to real 2010 dollars. First examined is the Hakes and Sauer test where they specifically looked at OBA and Slugging Percentage yearly and their impact on salaries of Major League Baseball players. Next, the estimated regressions look at the five major performance measures (OBP, OPS, wRC+, Win Shares and WAR) in the periods prior to and post the release of *Moneyball*. Finally, a test was conducted to see if there was a statistically significant change in the performance measures from before 2003 to after 2003.

All regressions were estimated using robust standard errors. The first set of estimates was run using robust standard errors. The purpose of this is to ensure that the estimates of the coefficients are unbiased. The second batch was run clustering on team to remove the potential autocorrelation that could cause a bias in the standard errors for the regressions results.

3.4.1 Test of Hakes and Sauer

The first estimated equations replicated the Hakes and Sauer salary equations estimating the importance of on-base percentage and slugging percentage over the years of 1996 to 2011. These are also the statistics that DePodesta found to be most significant when it came to contributing to a team's overall winning percentage. As mentioned earlier, DePodesta found that "an extra-point of on-base percentage was worth three times an extra point of slugging percentage" (Lewis 128). Therefore, one would expect that the coefficient of on-base percentage should be greater than that of slugging percentage in a salary regression.

To complete this test, a model similar to that of Hakes and Sauer from both their research in 2005 and 2007 was estimated. The main difference between the two is that in my research, only players eligible for free agents are being evaluated. The dependent variable is the natural log of the salary in 2010 dollars. The independent variables are on-base percentage, slugging percentage, plate appearances (all three are from the year t-1), and a dummy variable for each position. Designated hitter is excluded to serve as the omitted category in the equation.

$$lnSal2010 = \beta_0 + \beta_1 OBP + \beta_2 Slugging + \beta_3 PAs + \gamma_i Position_i + \varepsilon \quad (3)$$

The purpose of this equation is to expand and improve upon the models originally estimated by Hakes and Sauer in both their 2005 and 2007 papers. This will provide a bigger window of time post *Moneyball* to see the true impact on these two performance measures year by year. Also they could not decipher if in 2006, the coefficient for OBA went back to being non-significant or an anomaly in the data. As already discussed, the results received may be different due to the fact that in my study, only players who were eligible for free agency are included in the sample.

3.4.2 Individual Timeframe Test

Pooled regressions were run to test the significance and the coefficients for each of the variables in the two timeframes (1996-2002 and 2003-2011). The players were separated based on the year their contract began. For players who signed contracts as free agents that began from 1996 to 2002 were in the first group and those whose contracts began in the time period of 2003 to 2011 were in the second group. To get the exact coefficients and significance for each of the two time periods, several regressions were run. The natural log of the salary measured in 2010 dollars is the dependent variables. The main independent variables are the two year averages of the performance measure of interest prior to the new contract signing date. In addition, a two-year average of at-bats prior to the current contract, experience, experience squared, position and year were included as control variables. Each model varied in the performance variable, *Hitting*.

$$lnSal2010 = \beta_0 + \beta_1 Hitting + \beta_2 ABs + \beta_3 Experience + \beta_4 Experience^2 + \gamma_i Position_i + \delta_i Year + \varepsilon \quad (4)$$

The variable *Hitting*[7] is used to represent the performance measure tested in the regression model.

The basis is that prior to 2002, teams focused more on physical attributes or on performance measures such as runs scored, batting average or Runs Batted In. However, these statistics are not actually the most relevant to winning based on the research from Paul DePodesta, as well as others. Therefore after the release of *Moneyball* and the success of the Oakland Athletics it is expected to see the

---

[7] *Hitting* =On-Base Percentage, Slugging Percentage, On-Base and Slugging Percentage, OPS, wRC+, Win Shares and WAR based on the model run.

coefficients for the performance measures to increase in magnitude and significance after *Moneyball* (after 2013).

Chapter 4

Results

4.1 Hakes and Sauer Test

Results from Hakes and Sauer (2007) are summarized in Table 1. These results will then be duplicated in Table 2 adjusting for the changes to improve upon the study by Hakes and Sauer. The two main variables of interest in the Hakes and Sauer yearly regression are On-Base Percentage and Slugging Percentage for the one year prior, whereas the new estimated regressions focus on these statistics one-year lagged to the signing of the contract because this is what players are evaluated on when their salary is determined. The coefficient of each in determining the natural log of salary is noted, as well as its significance.

Table 1: Results from Hakes and Sauer ("The *Moneyball* Anomaly and Payroll Efficiency: A Further Investigation")

Year	On-Base	Slugging Percentage
1986	0.8	1.03*
1987	1.41	1.82**
1988	0.34	1.51**
1989	1.36	1.99**
1990	-0.14	2.15**
1991	-0.36	1.76**
1992	-0.99	2.13**
1993	0.87	2.54**
1994	0.21	3.12**
1995	2.63*	2.45**
1996	-0.66	2.58**
1997	2.52*	2.17**
1998	1.81*	2.40**
1999	1.75*	2.42**
2000	2.53*	2.49**
2001	0.12	3.29**
2002	0.81	2.31**
2003	1.43	1.94*
2004	4.11**	2.32**
2005	3.64**	2.72**
2006	2.09	2.14**

** $p<0.01$, * $p<0.05$

Table 2: Coefficient Estimates for On-Base Percentage and Slugging Percentage using the test of Hakes and Sauer's Estimates (Robust)

Year	On-Base Percentage	Slugging Percentage
Average	0.967 (0.607)	4.737*** (0.269)
1996	-0.929 (1.745)	3.204*** (1.065)
1997	-0.515 (1.580)	4.067*** (0.714)
1998	2.818 (1.841)	3.989*** (0.904)
1999	1.187 (1.383)	5.342*** (0.639)
2000	-1.040 (2.119)	4.579*** (0.990)
2001	-1.506 (1.735)	4.860*** (0.936)
2002	2.193 (1.956)	4.367*** (1.021)
2003	4.069 (2.819)	3.602** (1.514)
2004	6.013*** (1.796)	3.572*** (0.935)
2005	-0.157 (2.340)	4.580*** (0.894)
2006	2.057 (1.991)	4.747*** (1.252)
2007	-1.251 (2.285)	4.872*** (1.044)
2008	1.798 (1.945)	5.131*** (1.015)
2009	1.866 (1.873)	5.706*** (1.079)
2010	3.449* (2.024)	4.519*** (1.137)
2011	6.509*** (1.956)	3.428*** (0.969)

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2 reports the results of the yearly regression estimates used to determine the natural log of salary for players with six or more years of experience in 2010 dollars. Slugging percentage is significant at the 1% level for every year except for 2003 when it is still significant at the 5% level. The results for on-base percentage are different than those obtained by Hakes and Sauer. From the results it can be inferred that from 2002 to 2004 Major League Baseball began to look more in depth at on-base percentage and became more widely adopted by the league as seen in it becoming significant in 2004 at the 1% level, the year after *Moneyball* was released. However it is then not significant again until 2010. This implies that on-base percentage was more highly valued than in these three years but not in the other years due to its lack of significance. Based on the yearly regressions underlying the senior thesis, one cannot conclude that *Moneyball* had an impact on the valuation of Major League Baseball players using on-base percentage as a measure of player performance. This is also not the best sample due to the limited sample size per year. Complete results for the yearly regressions are summarized in the appendix table 9.

4.2 Individual Timeframe Test

4.2.1 *Moneyball* Statistics (On-Base Percentage & Slugging Percentage)

Results are presented in Table 3 below looking at two of the major statistics discussed in *Moneyball*; On-Base Percentage and Slugging Percentage. The control variable of the two year average of at-bats at the signing of the most recent contract is also included in the regressions due to its importance in deciding how much playing time a player receives during the season. Other control variables include: experience, experience squared, position and yearly dummy variables. Results for these variables are included in appendix table 10, with the omitted variable being the year 1996 and 2003 and the position of designated hitter being the position.

Table 3: Coefficient Estimates of On-Base Percentage and Slugging Percentage on Natural Log of Salary

	On-Base Percentage		Slugging Percentage		Slugging and On-Base	
VARIABLES	1996-2002	2003-2011	1996-2002	2003-2011	1996-2002	2003-2011
On-Base	0.00838***	0.0129***			0.00266***	0.00629***
	(0.000583)	(0.000656)			(0.000700)	(0.000799)
Slugging			0.00588***	0.00740***	0.00504***	0.00547***
			(0.000302)	(0.000329)	(0.000392)	(0.000402)
At-Bats	0.00538***	0.00437***	0.00475***	0.00389***	0.00471***	0.00375***
	(0.000173)	(0.000158)	(0.000182)	(0.000167)	(0.000180)	(0.000149)
Constant	2.356***	1.265***	2.860***	2.572***	2.329***	1.299***
	(0.333)	(0.337)	(0.272)	(0.268)	(0.303)	(0.310)
Observations	1,153	1,154	1,153	1,154	1,153	1,154
R-squared	0.606	0.674	0.656	0.705	0.660	0.72
Estimated Difference Tests						
On-Base - Chi Squared	27.01***				11.94***	
Probability	0.0000				0.0005	
Slugging - Chi Squared			11.78***		0.56	
Probability			0.0000		0.4543	
At-Bats - Chi Squared	18.92***		12.13***		13.81***	
Probability	0.0000		0.0005		0.0002	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3 summarizes the salary regression results for the performance measures of on-base percentage and slugging percentage. The results summarize the impact of on-base percentage, slugging percentage and the control variable of at-bats for contracts signed from 1996 to 2002 and then again from 2003 to 2011. The first two columns include the results for on-base percentage as the lone performance measure and the second two columns include the results for slugging percentage as the lone performance measure, while the last two columns include both performance measures in the regression. These results show that on-base percentage and slugging percentage when estimated in equations separately were a significant and important measure in both time periods when it came to determining the salary of a player. It can also be seen that the two year average of at-bats a player has remains significant and valuable in determining how much a player gets paid. The more at-bats a player has indicates that the manager has more confidence in the player to produce. To determine whether the coefficients are significantly different in the two time periods an estimated equation test was used. To use the Suest function, first each regression equation must be run and stored in Stata without using robust standard errors. The suest function is then run and is able to estimate both equations using robust standard errors. Once each equation is estimated by the Suest function, a Chi-squared test is run to test the significance of the difference between the coefficient estimates. In the first equation where the lone performance variable of on-base performance is tested, a Chi-squared of 27.01 with a p-value of 0.00 is received. This means that on-base percentage is significantly more important in the time period of 2003-2011 in comparison to the time period of 1996-2002[8]. Similar results were received when the coefficients of slugging average were tested for significant difference for 1996-2002 and 2003-2011. When the suest test was run for slugging average a Chi-squared of 11.78 was received and with a p-value of 0.00[9]. The suest test also confirms the fact that the importance of an additional at-bat in determining salary is less important in the time period post *Moneyball*. In the test where the lone performance variable is on-base percentage, a Chi-squared of 18.92 with a p-value of 0.00 is received. When the lone performance variable is slugging percentage, the Chi-squared result for at-bats is 12.13 and a p-value of 0.00. Both of these Chi-squared results combined with the individual regressions show that the importance of at-bats in determining salary decreases from 1996-2002 to 2003-2011. This implies that with regards to salaries teams focused more on what players were able to accomplish at the plate as opposed to how many times they went to the plate.

Additionally, on-base percentage, slugging percentage and the two year average of at-bats are all statistically significant in the first and second time period when both are included as performance measures in the regression. On-base percentage, slugging percentage and at-bats are significant at the 1% level for both time periods. A suest test was run to see if the changes in on-base percentage and slugging percentage in this equation was statistically significant. The results were that the difference in on-base percentage resulted in a Chi-squared result of 11.94 and a p-value of .00 and the difference in slugging percentage from 1996-2002 to 2003-2011 had a Chi-squared result of .17 and a p-value of .6762. These results show that the difference in the importance of on-base percentage is significant and the change slugging percentage is statistically insignificant. This supports the hypothesis that on-base percentage was underutilized relative to slugging percentage prior to the release to *Moneyball* but is then more important than slugging percentage when used as a performance measure to determine player salary after the release of *Moneyball*. However, the change in the coefficient of at-bats is significant with a Chi-squared value of 13.81 and a p-value of 0.00[10]. The return to at-bats, despite remaining significant at the 1% level, is less valuable to salary most likely due to the importance of how a player performs when at the plate versus how many times they actually get up to hit.

Table 4: Coefficients of On-Base and Slugging Percentage on Natural Log of Salary using Interactive Variables

VARIABLES	On-Base Percentage Salary	Slugging Percentage Salary	Slugging and On-Base Salary
PostMoneyball	-1.638***	-0.737***	-1.408***
	(0.271)	(0.157)	(0.259)
On-Base	0.00797***		0.00245***
	(0.000568)		(0.000679)
On-Base x PostMoneyball	0.00472***		0.00373***
	(0.000820)		(0.00104)
Slugging		0.00565***	0.00491***
		(0.000282)	(0.000363)
Slugging x PostMoneyball		0.00166***	0.000466
		(0.000390)	(0.000513)
At-Bats	0.00530***	0.00472***	0.00470***
	(0.000165)	(0.000170)	(0.000169)
At-Bats x PostMoneyball	-0.000971***	-0.000874***	-0.000985***
	(0.000217)	(0.000224)	(0.000220)
Constant	2.622***	3.047***	2.523***
	(0.271)	(0.213)	(0.252)
Observations	2,307	2,307	2,307
R-squared	0.643	0.682	0.691

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Another way to test the difference in premium placed on specific performance measures in the time period of 2003-2011 is to run a regression with interactive dummy variables pooling the data from 1996 to 2011 into one model. This model was created using a variable (PostMoneyball) that equals one if the contract begins between the years 2003 to 2011. By interacting PostMoneyball with on-base and slugging percentage, the premium that teams place on these performance measures in determining player salary from 2003 to 2011 can be determined. In the model the interactive variables included were for the performance measure and at-bats. The control variables included were: experience, experience squared, position and year. The following regression was estimated:

$$lnSal2010 = \beta_0 + \beta_1 Hitting + \beta_2 HittingxPostMoneyball + \beta_3 ABs + + \beta_4 ABsxPostMoneyball + \beta_5 Experience + \beta_6 Experience^2 + \gamma_i Position_i + \delta_i Year + \varepsilon \quad (5)$$

Table 4 shows the results of the coefficients of on-base percentage and slugging percentage from estimating the model that was estimated three times. The first column summarizes the results with on-base average as the lone performance measure evaluated. The second column summarizes the results when slugging percentage is the lone performance measure. And the third column summarizes the results where both on-base percentage and slugging percentage were included as dummy variables. The resulting positive coefficients and t-statistics of the interactive variables with p-values of 0.00 in the first and second columns confirm the results received from the suest test, that on-base percentage and slugging percentage were both deemed to be more important when used as the performance measure in separate regressions. The model also confirms the fact that Major League Baseball placed a higher premium on on-base percentage in the second period but had no significant change in premium placed on slugging percentage. This can be seen in the results in the third column that the interactive variable of *On-BasexPostMoneyball* is positive and has a p-value of 0.00, meaning that the premium is significant at the 1% level but *SluggingxPostMoneyball* is insignificant with a t-statistic less than one. In all three estimated regressions, the decline in the premium placed on at-bats in the second period is significant at the 1% level. The interactive variable model tests confirm the results received using the estimated regression tests. Full results are summarized in appendix table 12.

Through the use of a t-test it was able to be determined whether on-base percentage was undervalued in the time period of 1996-2002 and whether there was a market correction from 2003-2011. Using a t-test to show if the coefficients for on-base percentage and slugging percentage were equal from 1996-2002, it can be seen that on-base percentage was undervalued relative to slugging percentage and was significant at the 5% level[11]. Then running a t-test for the coefficients for 2003-2011 on on-base percentage and slugging percentage again, it cannot be determine that on-base percentage is more valuable than slugging percentage however, based on the t-test they are now of equal value to salary[12]. Therefore base on these t-tests it can be concluded that on-base percentage was undervalued from 1996-2002 but then there was a market correction from 2003-2011.

4.2.2 Sabermetrics (OPS, Runs Created, Win Shares, WAR)

Results are presented in Table 5 looking at other statistics known as measuring performance which are used to evaluate players. The dependent variable is the natural log of salary. The independent variables are the performance measure and the control variables. The control variable of the two year average of at-bats is also included in the results due to its perceived importance. Other control variables are experience,

---

[11] 1996-2002 Test oba_2yravg1000=sa_2yravg1000, F(1, 1135) = 5.69, Prob > F = 0.0173

[12] 2003-2011 Test oba_2yravg1000=sa_2yravg1000, F(1, 1134) = 0.56, Prob > F = 0.4546

experience squared, position and yearly dummy variables. These are included in appendix table 11, with the year 1996 and 2003 being the omitted year and the position of designated hitter being the position. Full results are available in the appendix table 11.

Table 5: Coefficients Estimates for Sabermetrics on Natural Log of Salary

	OPS		Runs Created		WAR		Win Shares	
VARIABLES	1996-2002	2003-2011	1996-2002	2003-2011	1996-2002	2003-2011	1996-2002	2003-2011
OPS	0.00431***	0.00572***						
	(0.000210)	(0.000230)						
Runs Created			0.00936***	0.0133***				
			(0.000456)	(0.000591)				
WAR					0.193***	0.322***		
					(0.0105)	(0.0133)		
Win Shares							0.0730***	0.0966***
							(0.00380)	(0.00484)
At-Bats	0.00474***	0.00375***	0.00231***	0.000362	0.00436***	0.00244***	0.00282***	0.00132***
	(0.000178)	(0.000164)	(0.000252)	(0.000286)	(0.000189)	(0.000197)	(0.000242)	(0.000270)
Constant	2.080***	1.389***	5.269***	5.654***	5.457***	5.950***	5.619***	5.928***
	(0.290)	(0.276)	(0.261)	(0.259)	(0.268)	(0.267)	(0.259)	(0.265)
Observations	1,153	1,154	1,153	1,154	1,153	1,154	1,153	1,152
R-squared	0.658	0.720	0.643	0.695	0.614	0.705	0.646	0.691
Estimated Difference Tests								
OPS - Chi Squared	20.74***							
Probability	0.0000							
Runs Created- Chi Squared			28.06***					
Probability			0.0000					
WAR - Chi Squared					56.27***			
Probability					0.0000			
Win Shares - Chi Squared							14.93***	
Probability							0.0000	
At-Bat Chi Squared	17.02***		26.4***		38.34***		17.44***	
Probability	0.0000		0.0000		0.0000		0.0000	

Robust standard errors in parentheses

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

The results show that all of the more advanced sabermetric performance measures are statistically significant variables in the salary equations. As hypothesized, the tests show the significance of going from before and after *Moneyball* when using more advanced sabermetric measures of performance to evaluate the salary determination process in Major League Baseball. The most significant changes in the values of the coefficients are for Runs Created and Wins Above Replacement, which are both statistic that have been modified by sabermetricians to find the most reliable and accurate measures of the value of a player to his team. The difference between 1996-2002 and 2003-2011 for each performance measure is statistically significant in the suest test. Specifically the test yields a Chi-squared value of 20.74 for on-base plus slugging percentage, 28.06 for Runs Created, 56.27 for Wins Above Replacement and 14.93 for Win Shares. Each of these Chi-squared values has a p-value of 0.00 meaning that the difference in the importance of each performance measure is statistically significant. This result provides evidence teams are placing more weight on statistical analysis to evaluate a player's worth in terms of salary. As seen in the prior set of regressions, at-bats is significant at the 1% level for both time periods. However, the return to salary for at-bats has a statistically significant decrease from 1996-2002 to 2003-2011. In each regression the p-value associated with the Chi-squared for at-bats is 0.00, showing that the decrease is in fact statistically significant[13]. The potential explanation is similar to that of the decrease in the first set of regressions, where players are being valued more on what they do at the plate versus the amount of times the player takes an at-bat.

Table 6: Coefficients Estimates for Sabermetrics on Natural Log of Salary using Interactive Variables

VARIABLES	OPS Salary	Runs Created Salary	WAR Salary	Win Shares` Salary
PostMoneyball	-1.035***	-0.0321	-0.00169	-0.0847
	(0.198)	(0.104)	(0.103)	(0.105)
OPS	0.00416***			
	(0.000199)			
OPS x PostMoneyball	0.00144***			
	(0.000276)			
Runs Created		0.00895***		
		(0.000435)		
Runs Created x PostMoneyball		0.00398***		
		(0.000663)		
WAR			0.168***	
			(0.0111)	
WAR x PostMoneyball			0.154***	
			(0.0171)	
Win Shares				0.0682***
				(0.00369)
Win Shares x PostMoneyball				0.0290***
				(0.00593)
At-Bats	0.00472***	0.00240***	0.00444***	0.00291***
	(0.000168)	(0.000240)	(0.000179)	(0.000230)
At-Bats x PostMoneyball	-0.000990***	-0.00194***	-0.00204***	-0.00164***
	(0.000220)	(0.000344)	(0.000254)	(0.000340)
Constant	2.253***	5.392***	5.608***	5.726***
	(0.231)	(0.195)	(0.200)	(0.193)
Observations	2,307	2,307	2,307	2,305
R-squared	0.690	0.668	0.660	0.670

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6 summarizes the results for the sabermetric performance measures using interactive variable models. The results support the conclusions from estimated regression test that the sabermetric performance measures were more important in the post *Moneyball* period. The table shows that all of the interactive variables are significant at the 1% level. These results support the hypothesis that there was an increase in the premium placed on the performance measures of On-Base Plus Slugging, Runs Created, Wins Above Replacement, and Win Shares in the time period of 2003-2011. As expected, the interactive variables for the two year average of at-bats is significant and negative, showing that at-bats were a less important measure of determining salary after *Moneyball* in comparison to the time period of 1996-2002. Full results are summarized in appendix table 13.

RESULTS FOOTNOTES

[8]The following are the commands in Stata that are used to conduct the suest test for on-base average as the lone performance measure:

reg lnsalthesis oba_2yravg1000 ab_2yravg expcurrent exp2 catcher firstbase secbase ss thirdbase outfield y1997 y1998 y1999 y2000 y2001 y2002 if position>1 & premb==1 & fa1==1 & year>1995 & ab130==1

est store prembroba

reg lnsalthesis oba_2yravg1000 ab_2yravg expcurrent exp2 catcher firstbase secbase ss thirdbase outfield y2004 y2005 y2006 y2007 y2008 y2009 y2010 y2011 if position>1 & ab130==1 & year>1995 & fa1==1 & postmb==1

est store postmbroba

suest prembroba postmbroba, robust

test [prembroba_mean]oba_2yravg1000=[postmbroba_mean]oba_2yravg1000
test [prembroba_mean]ab_2yravg=[postmbroba_mean]ab_2yravg

[9]The following are the commands in stata that are used to conduct the suest test for slugging percentage as the lone performance measure:

reg lnsalthesis sa_2yravg1000 ab_2yravg expcurrent exp2 catcher firstbase secbase ss thirdbase outfield y1997 y1998 y1999 y2000 y2001 y2002 if position>1 & premb==1 & fa1==1 & year>1995 & ab130==1

est store postmbsa

reg lnsalthesis sa_2yravg1000 ab_2yravg expcurrent exp2 catcher firstbase secbase ss thirdbase outfield y2004 y2005 y2006 y2007 y2008 y2009 y2010 y2011 if position>1 postmb==1 & fa1==1 & year>1995 ab130==1

est store prembsa

suest prembsa postmbsa, robust

test [prembsa _mean]sa_2yravg1000=[postmbsa _mean]sa_2yravg1000
test [prembsa_mean]ab_2yravg=[postmbsa _mean]ab_2yravg

[10] The following are the commands in stata that are used to conduct the suest test for on-base average and slugging percentage, both a performance measures:

reg lnsalthesis oba_2yravg1000 sa_2yravg1000 ab_2yravg expcurrent exp2 catcher firstbase secbase ss thirdbase outfield y1997 y1998 y1999 y2000 y2001 y2002 if position>1 & premb==1 & fa1==1 & year>1995 & ab130==1

est store prembobasa

reg lnsalthesis oba_2yravg1000 sa_2yravg1000 ab_2yravg expcurrent exp2 catcher firstbase secbase ss thirdbase outfield y2004 y2005 y2006 y2007 y2008 y2009 y2010 y2011 if position>1 & postmb==1 & fa1==1 & postmb==1

est store postmbobasa

suest prembobasa postmbobasa, robust
test [prembobasa _mean]oba_2yravg1000=[postmbobasa _mean]oba_2yravg1000
test [prembobasa _mean]sa_2yravg1000=[postmbobasa _mean]sa_2yravg1000
test [prembobasa_mean]ab_2yravg=[postmbobasa _mean]ab_2yravg

[13]The following are the commands in stata that are used to conduct the suest test, where *performancemeasure* represents each sabermetric in a separate equation:

reg lnsalthesis *performancemeasure* ab_2yravg expcurrent exp2 catcher firstbase secbase ss thirdbase outfield y1997 y1998 y1999 y2000 y2001 y2002 if position>1 & premb==1 & fa1==1 & year>1995 & ab130==1

est store premb*performancemeasure*

reg lnsalthesis *performancemeasure* ab_2yravg expcurrent exp2 catcher firstbase secbase ss thirdbase outfield y2004 y2005 y2006 y2007 y2008 y2009 y2010 y2011 if position>1 & ab130==1 & year>1995 & fa1==1 & postmb==1

est store postmb*performancemeasure*

suest premb*performancemeasure* postmb*performancemeasure*, robust
test [premb*performancemeasure* _mean] *performancemeasure*
[postmb*performancemeasure*_mean] *performancemeasure*

test [premb*performancemeasure* _mean]ab_2yravg=[postmb*performancemeasure* _mean]ab_2yravg

Chapter 5

Conclusion

Based on the analysis and results, it can be concluded that *Moneyball* led to the use of statistical analysis for evaluating talent. This supports the theory that prior to *Moneyball* there was an inefficiency in baseball that led to a change in the front office player evaluation procedures. After the tremendous success of the Oakland Athletics in the early 2000s, teams started to take notice and copy their strategy. The Toronto Blue Jays hired J.P. Ricciardi after the 2001 season (Rubin), the Boston Red Sox hired Bill James at the end of 2002 (Neyer) and the Los Angeles Dodgers hired Paul DePodesta prior to the 2004 season ("DePodesta…"). They are all noted for the use of statistical analysis to evaluate player performance. This change in thought process is seen in the regression analysis comparing the *Hitting* statistics in the time periods of 1996-2002 and 2003-2011. The change in the significance and coefficients of these major sabermetric measures of player performance show the importance of these statistics evaluating players. There were significant jumps in the coefficient and significance in the statistics of on-base percentage, slugging percentage, slugging plus on-base percentage (OPS), Runs Created, Win Shares, and Wins Above Replacement. These conclusions supported by Chi-squared statistics that have p-values of 0.00 for the changes in the coefficients of the performance measures. This infers league wide adoption in the use of these performance statistics to value players since *Moneyball* and implies that actual performance.

There was a lack of evidence when both on-base percentage and slugging percentage are included to support the yearly significance of on-base percentage in a regression analysis with slugging percentage in the Hakes and Sauer test. This is most likely due to the correlation of the two percentages of .7153, meaning they are highly correlated. This is potentially why slugging percentage is so highly significant and on-base percentage varies throughout the time period of 1996 to 2011 in the yearly regressions.

In the pooled regressions that included both on-base percentage and slugging percentage as performance measures, on-base percentage became more important than slugging percentage as a way to determine player salary. This supports the notion that *Moneyball* had an impact when it came to evaluating on-base percentage versus slugging percentage. On-base percentage is worth more than slugging percentage when it comes to evaluating players which supports DePodesta and James who both believe that on-base percentage should be valued higher than slugging percentage. It is also supported from the regressions where the lone performance measure is on-base percentage, that it is highly significant in predicting salary and even more so after the release of *Moneyball*.

For future research it would be interesting to include different statistics such as RBI's and Run's with On-Base Percentage. It is also of interest to study the correlation between winning and the different performance measures used, because this has yet to be studied. This will give a better view of whether the most significant and highly valued performance measures are in fact correlated with winning the most games. One final area of study would be the use of a j-test on the performance measures due to their high collinearity. These studies will to further the study of the impact of *Moneyball* and statistical analysis on Major League Baseball.

This research can settle the debate of the true impact of *Moneyball* on the use of statistical analysis to evaluate Major League Baseball players. Limiting the sample to non-pitchers who are were free to sign contracts as free agents, allows the change in the importance of statistics in the valuation method to be seen from the time period before *Moneyball* (1996-2002) to the post-*Moneyball* time period (2003-2011). The results showed that all statistics were more important to the value of one's salary in the post-*Moneyball* time period and the difference is significant. This research also proves that on-base percentage was undervalued relative slugging percentage prior to *Moneyball*. From 2003-2011 it became equally important as slugging percentage in terms of evaluating talent, showing that there was a market correction. This shows that *Moneyball* and the Oakland Athletics really had a major impact on the way front offices in Major League Baseball operate.

REFERENCES

"Baseball-Reference.com - Major League Baseball Statistics and History." Baseball-Reference.com. N.p., n.d. Web. 20 Mar. 2013.

"DePodesta Plucked from A's Front Office." ESPN.com. ESPN, 17 Feb. 2004. Web. 20 Mar. 2013.

Epstein, Eddie. Personal interview. 22 Feb. 2013.

FanGraphs Baseball Statistics and Analysis. FanGraphs.com, n.d. Web. 21 Mar. 2013

Hakes, Jahn K., and Raymond D. Sauer. 2006. "An Economic Evaluation of the *Moneyball* Hypothesis." Journal of Economic Perspectives, 20(3): 173-186.

Hakes, Jahn K., & Sauer, Raymond D. (2007). "The *Moneyball* Anomaly and Payroll Efficiency: A Further Investigation." International Journal of Sport Finance, 2, 177-189

James, Bill. "Bill James Online." *BillJamesOnline.com*. Bill James, 17 Mar. 2013. Web. 21 Mar. 2013.

Link, Charles R., and Martin Yosifov. "Contract Length and Salaries Compensating Wage Differentials in Major League Baseball." *Journal of Sports Economics* 13.1 (2012): 3-19. *Contract Length and Salaries Compensating Wage Differentials in Major League Baseball*. Journal of Sports Economics, Feb. 2012. Web. 21 Mar. 2013

Meltzer, Josh. "Average Salary and Contract Length in Major League Baseball: Why Do They Diverge?" Thesis. Stanford University, 2005. *Economics.Stanford.edu*. Stanford Department of Economics. Web. 29 Jan. 2013.

Neyer, Rob. "Red Sox Hire Bill James as Advisor." ESPN.com:MLB. ESPN, 5 November 2002. Web. 17 March 2013.

Neyer, Rob. "Sabermetricians Slowly Being Added to the Inner Circle." *ESPN.com*. ESPN, 2002. Web. 21 Mar. 2013.

Rubin, Adam. "Mets Hire ESPN's J.P. Ricciardi." ESPN.com. ESPN, 2 Nov. 2010. Web. 20 Mar. 2013.

Tango, Tom M., Mitchel Lichtman, and Andrew Dolphin. "WOBA - Weighted On Base Average." Inside the Book. The Book: Playing the Percentages, n.d. Web. 20 Mar. 2013.

Appendix

Tables

Table 7: Summary Statistics for Variables for 1996-2002

Variable	Obs	Mean	Std. Dev.	Min	Max
lnsalthesis	1159	8.15181	1.156451	5.020491	10.31381
On-Base	1153	357.2988	40.6128	222.2222	479.1503
Slugging	1153	459.1102	83.98247	232.9193	775.1046
OPS	1153	816.409	115.9965	455.1415	1254.255
Runs Created	1153	141.9954	71.53953	13.84894	375.307
WAR	1159	2.096549	2.163089	-1.65	10.35
Win Shares	1153	15.02168	8.113077	1	43
At-Bats	1153	424.7441	131.7297	130	649.5
Experience	1159	9.7486	3.016769	6	22.03333
$Experience^2$	1159	104.1283	69.45939	36	485.4677
Catcher	1159	0.141501	0.348688	0	1
First Base	1159	0.107852	0.310327	0	1
Second Base	1159	0.112166	0.315706	0	1
SS	1159	0.105263	0.307025	0	1
Third Base	1159	0.099224	0.299091	0	1
Outfield	1159	0.381363	0.485931	0	1
y1996	1159	0.072476	0.259387	0	1
y1997	1159	0.092321	0.289604	0	1
y1998	1159	0.118205	0.322991	0	1
y1999	1159	0.136324	0.343281	0	1
y2000	1159	0.13805	0.345101	0	1
y2001	1159	0.144953	0.352205	0	1
y2002	1159	0.132873	0.339584	0	1

Table 8: Summary Statistics for Variables for 2003-2011

Variable	Obs	Mean	Std. Dev.	Min	Max
lnsalthesis	1158	8.117185	1.139911	5.847317	10.42808
On-Base	1154	343.5872	34.42677	228.3654	449.5413
Slugging	1154	441.9125	73.875	242.7184	649.1053
OPS	1154	785.4997	101.812	511.9492	1079.495
Runs Created	1154	128.9856	67.26323	14.03714	316.8819
WAR	1158	2.225864	1.925096	-0.95	8.05
Win Shares	1152	13.59028	7.631364	0.5	36.5
At-Bats	1154	414.6282	146.6333	130	687
Experience	1158	9.765755	3.019545	6	23.53333
$Experience^2$	1158	104.4798	69.22224	36	553.8177
Catcher	1158	0.146805	0.354064	0	1
First Base	1158	0.143351	0.350581	0	1
Second Base	1158	0.1019	0.302647	0	1
SS	1158	0.143351	0.350581	0	1
Third Base	1158	0.126079	0.332082	0	1
Outfield	1158	0.316062	0.465138	0	1
y2003	1158	0.060449	0.23842	0	1
y2004	1158	0.096719	0.295702	0	1
y2005	1158	0.120035	0.325142	0	1
y2006	1158	0.124352	0.330126	0	1
y2007	1158	0.126943	0.333053	0	1
y2008	1158	0.120898	0.32615	0	1
y2009	1158	0.11658	0.321058	0	1
y2010	1158	0.122625	0.328148	0	1
y2011	1158	0.111399	0.314762	0	1

Table 9: Correlation Matrix of Performance Measures by Time Periods

1996-2011

	OBA	SA	OPS	Runs Created	WAR	Win Shares
OBA	1					
SA	0.7153	1				
OPS	0.8635	0.9701	1			
Runs Created	0.74	0.7965	0.8318	1		
WAR	0.649	0.689	0.7226	0.8407	1	
Win Shares	0.7114	0.7346	0.7772	0.9354	0.8518	1

1996-2002 (Pre-*Moneyball*)

	OBA	SA	OPS	Runs Created	WAR	Win Shares
OBA	1					
SA	0.6967	1				
OPS	0.8546	0.9679	1			
Runs Created	0.7451	0.8022	0.8417	1		
WAR	0.6363	0.6815	0.7162	0.8186	1	
Win Shares	0.7288	0.75	0.7982	0.9309	0.8158	1

2003-2011 (Post-*Moneyball*)

	OBA	SA	OPS	Runs Created	WAR	Win Shares
OBA	1					
SA	0.7319	1				
OPS	0.8693	0.973	1			
Runs Created	0.7323	0.7862	0.8181	1		
WAR	0.7054	0.7164	0.7583	0.8832	1	
Win Shares	0.6876	0.7114	0.7487	0.9394	0.9108	1

Table 10: Coefficients Estimates for Hakes and Sauer Test in Equation 3 using Yearly Regression Estimates for the Natural Log of Salary for Player's Eligible for Free Agency for the Years of 1996 to 2011

Year	On-Base	Slugging	PA	Catcher	1B	2B	SS	3B	OF	Constant
Average	0.967	4.737***	0.00342***	0.716***	0.329***	0.401***	0.729***	0.455***	0.497***	3.494***
	(0.607)	(0.269)	(0.000122)	(0.101)	(0.0991)	(0.105)	(0.105)	(0.103)	(0.0962)	(0.192)
1996	-0.929	3.204***	0.00423***	0.0615	0.305	0.0408	0.553*	-0.107	0.00247	4.654***
	(1.745)	(1.065)	(0.000398)	(0.300)	(0.231)	(0.270)	(0.279)	(0.294)	(0.210)	(0.691)
1997	-0.515	4.067***	0.00387***	0.501	0.570*	0.107	0.809**	0.277	0.373	4.009***
	(1.580)	(0.714)	(0.000371)	(0.352)	(0.317)	(0.351)	(0.344)	(0.319)	(0.314)	(0.534)
1998	2.818	3.989***	0.00348***	0.330	0.255	0.316	0.284	-0.0118	0.165	3.203***
	(1.841)	(0.904)	(0.000372)	(0.282)	(0.246)	(0.288)	(0.306)	(0.300)	(0.253)	(0.568)
1999	1.187	5.342***	0.00317***	0.468**	0.309	0.494**	0.969***	0.605***	0.525***	3.069***
	(1.383)	(0.639)	(0.000385)	(0.226)	(0.197)	(0.239)	(0.212)	(0.207)	(0.201)	(0.406)
2000	-1.040	4.579***	0.00407***	0.892***	0.255	0.400	0.646**	0.655***	0.466**	3.840***
	(2.119)	(0.990)	(0.000408)	(0.233)	(0.259)	(0.246)	(0.267)	(0.242)	(0.226)	(0.585)
2001	-1.506	4.860***	0.00384***	0.972**	0.681	0.410	0.794	0.536	0.670	3.947***
	(1.735)	(0.936)	(0.000453)	(0.491)	(0.473)	(0.528)	(0.499)	(0.520)	(0.477)	(0.684)
2002	2.193	4.367***	0.00339***	0.344*	-0.179	0.0346	0.179	0.0966	-0.106	3.755***
	(1.956)	(1.021)	(0.000446)	(0.179)	(0.158)	(0.222)	(0.224)	(0.216)	(0.146)	(0.533)
2003	4.069	3.602**	0.00274***	1.217***	0.647*	0.575	0.956**	0.373	0.947***	2.839***
	(2.819)	(1.514)	(0.000542)	(0.407)	(0.359)	(0.439)	(0.423)	(0.450)	(0.358)	(0.609)
2004	6.013***	3.572***	0.00320***	0.353*	-0.227	0.0945	0.127	0.0865	0.167	2.670***
	(1.796)	(0.935)	(0.000372)	(0.195)	(0.189)	(0.143)	(0.159)	(0.119)	(0.106)	(0.484)
2005	-0.157	4.580***	0.00352***	0.322	-0.00986	-0.00371	0.182	0.366*	0.174	4.281***
	(2.340)	(0.894)	(0.000384)	(0.201)	(0.212)	(0.207)	(0.224)	(0.195)	(0.167)	(0.633)
2006	2.057	4.747***	0.00350***	0.460**	0.182	0.242	0.257	0.409*	0.206	3.396***
	(1.991)	(1.252)	(0.000503)	(0.225)	(0.273)	(0.226)	(0.281)	(0.218)	(0.217)	(0.561)
2007	-1.251	4.872***	0.00360***	0.109	-0.212	-0.312*		-0.517**	-0.0418	4.978***
	(2.285)	(1.044)	(0.000451)	(0.150)	(0.181)	(0.187)		(0.218)	(0.149)	(0.546)
2008	1.798	5.131***	0.00311***	0.323*	-0.374	-0.000817	0.461***		0.111	3.805***
	(1.945)	(1.015)	(0.000466)	(0.188)	(0.233)	(0.183)	(0.157)		(0.166)	(0.442)
2009	1.866	5.706***	0.00340***	0.319	-0.590**		0.229	-0.253	-0.200	3.568***
	(1.873)	(1.079)	(0.000434)	(0.237)	(0.257)		(0.233)	(0.241)	(0.246)	(0.507)
2010	3.449*	4.519***	0.00249***	-0.124	-0.541**	-0.274	0.0664		0.0355	3.963***
	(2.024)	(1.137)	(0.000440)	(0.237)	(0.227)	(0.252)	(0.231)		(0.227)	(0.697)
2011	6.509***	3.428***	0.00326***	0.431*	-0.210	-0.0944	0.294		0.241	2.857***
	(1.956)	(0.969)	(0.000392)	(0.238)	(0.303)	(0.232)	(0.231)		(0.216)	(0.640)

Robust standard errors in parentheses

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

On-Base, Slugging Percentage, Plate Appearance are from year t-1 to signing contract for players with at least 130 at-bats

Table 11: Coefficient Estimates for *Moneyball* Statistics on Natural Log of Salary

	On-Base		Slugging Percentage		On-Base and Slugging Percentage	
VARIABLES	1996-2002	2003-2011	1996-2002	2003-2011	1996-2002	2003-2011
On-Base	0.00838***	0.0129***			0.00266***	0.00629***
	(0.000583)	(0.000656)			(0.000700)	(0.000796)
Slugging			0.00588***	0.00740***	0.00504***	0.00547***
			(0.000302)	(0.000329)	(0.000392)	(0.000412)
At-Bats	0.00538***	0.00437***	0.00475***	0.00389***	0.00471***	0.00375***
	(0.000173)	(0.000158)	(0.000182)	(0.000167)	(0.000180)	(0.000164)
Experience	0.116**	0.0416	0.0883**	0.0259	0.0853**	0.0323
	(0.0508)	(0.0437)	(0.0407)	(0.0414)	(0.0417)	(0.0399)
Experience	-0.00673***	-0.00305	-0.00517***	-0.00187	-0.00522***	-0.00246
	(0.00234)	(0.00200)	(0.00183)	(0.00190)	(0.00188)	(0.00183)
Catcher	0.504***	0.468***	0.573***	0.555***	0.603***	0.624***
	(0.146)	(0.153)	(0.137)	(0.145)	(0.137)	(0.142)
1B	0.175	-0.106	0.207	-0.0368	0.192	-0.0714
	(0.142)	(0.150)	(0.132)	(0.143)	(0.132)	(0.141)
2B	-0.104	-0.135	0.283**	0.231	0.254*	0.200
	(0.149)	(0.149)	(0.144)	(0.143)	(0.143)	(0.140)
SS	0.305**	0.156	0.639***	0.453***	0.639***	0.480***
	(0.146)	(0.152)	(0.143)	(0.146)	(0.142)	(0.142)
3B	0.111	0.127	0.279*	0.143	0.286*	0.187
	(0.158)	(0.154)	(0.147)	(0.146)	(0.147)	(0.143)
OF	0.262*	0.134	0.322**	0.202	0.327***	0.228*
	(0.136)	(0.145)	(0.127)	(0.139)	(0.126)	(0.136)
y1997	-0.329***		-0.219***		-0.215***	
	(0.0789)		(0.0721)		(0.0723)	
y1998	-0.343***		-0.270***		-0.262***	
	(0.0800)		(0.0764)		(0.0752)	
y1999	-0.233***		-0.175**		-0.161**	
	(0.0717)		(0.0684)		(0.0679)	
y2000	-0.124*		-0.0819		-0.0663	
	(0.0752)		(0.0690)		(0.0695)	
y2001	-0.0484		-0.0188		-0.00777	
	(0.0757)		(0.0701)		(0.0704)	
y2002	0.00688		0.0196		0.0389	
	(0.0765)		(0.0713)		(0.0712)	
y2004		0.0226		-0.0126		-0.00814
		(0.109)		(0.103)		(0.101)
y2005		0.359***		0.273***		0.299***
		(0.100)		(0.0945)		(0.0926)
y2006		0.397***		0.343***		0.348***
		(0.104)		(0.0985)		(0.0966)

Table 11:Coefficient Estimates for *Moneyball* Statistics on Natural Log of Salary (Continued)

y2007		0.507***		0.453***		0.451***
		(0.106)		(0.100)		(0.0979)
y2008		0.526***		0.465***		0.471***
		(0.110)		(0.103)		(0.102)
y2009		0.530***		0.504***		0.497***
		(0.108)		(0.101)		(0.0993)
y2010		0.454***		0.434***		0.420***
		(0.109)		(0.104)		(0.102)
y2011		0.595***		0.578***		0.570***
		(0.107)		(0.103)		(0.100)
Constant	2.356***	1.265***	2.860***	2.572***	2.329***	1.299***
	(0.333)	(0.337)	(0.272)	(0.268)	(0.303)	(0.307)
Observations	1,153	1,154	1,153	1,154	1,153	1,154
R-squared	0.606	0.674	0.656	0.705	0.660	0.720

Robust standard errors in parentheses

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

On-Base, Slugging Percentage, At-Bats are average from years t-1 and t-2 from year contract began

Table 12: Coefficient Estimates for Sabermetrics on Natural Log of Salary

	OPS		Runs Created		WAR		Win Shares	
VARIABLES	1996-2002	2003-2011	1996-2002	2003-2011	1996-2002	2003-2011	1996-2002	2003-2011
OPS	0.00431***	0.00572***						
	(0.000210)	(0.000230)						
Runs Created			0.00936***	0.0133***				
			(0.000456)	(0.000591)				
WAR					0.193***	0.322***		
					(0.0105)	(0.0133)		
Win Shares							0.0730***	0.0966***
							(0.00380)	(0.00484)
At-Bats	0.00474***	0.00375***	0.00231***	0.000362	0.00436***	0.00244***	0.00282***	0.00132***
	(0.000178)	(0.000164)	(0.000252)	(0.000286)	(0.000189)	(0.000197)	(0.000242)	(0.000270)
Experience	0.0860**	0.0316	0.109**	0.0274	0.119***	0.0341	0.0921**	0.00762
	(0.0428)	(0.0399)	(0.0447)	(0.0438)	(0.0455)	(0.0429)	(0.0450)	(0.0422)
$Experience^2$	-0.00535***	-0.00240	-0.00639***	-0.00210	-0.00667***	-0.00195	-0.00572***	-0.00123
	(0.00194)	(0.00183)	(0.00203)	(0.00202)	(0.00206)	(0.00198)	(0.00205)	(0.00192)
Catcher	0.610***	0.622***	0.474***	0.475***	0.149	0.00413	0.153	0.127
	(0.137)	(0.142)	(0.139)	(0.136)	(0.148)	(0.144)	(0.136)	(0.143)
1B	0.184	-0.0683	0.104	-0.141	0.132	-0.140	0.0187	-0.133
	(0.132)	(0.141)	(0.135)	(0.135)	(0.146)	(0.143)	(0.134)	(0.141)
2B	0.216	0.209	0.0827	0.0451	-0.213	-0.397***	-0.269*	-0.274**
	(0.143)	(0.139)	(0.144)	(0.132)	(0.151)	(0.141)	(0.140)	(0.139)
SS	0.616***	0.485***	0.412***	0.363***	0.0813	-0.165	0.0898	-0.0486
	(0.141)	(0.142)	(0.141)	(0.138)	(0.148)	(0.143)	(0.137)	(0.142)
3B	0.277*	0.185	0.194	0.166	-0.0719	-0.284*	-0.0507	-0.114
	(0.148)	(0.143)	(0.148)	(0.138)	(0.155)	(0.146)	(0.146)	(0.144)
OF	0.324**	0.228*	0.285**	0.190	0.141	-0.0568	0.0908	0.0427
	(0.127)	(0.136)	(0.129)	(0.130)	(0.138)	(0.139)	(0.126)	(0.136)
y1997	-0.221***		-0.202***				-0.263***	
	(0.0730)		(0.0739)				(0.0734)	
y1998	-0.264***		-0.238***		-0.212***		-0.265***	
	(0.0750)		(0.0787)		(0.0795)		(0.0776)	
y1999	-0.159**		-0.157**		-0.192***		-0.172**	
	(0.0678)		(0.0702)		(0.0662)		(0.0689)	
y2000	-0.0629		-0.0660		-0.130*		-0.0929	
	(0.0696)		(0.0711)		(0.0667)		(0.0716)	
y2001	-0.00538		0.00595		-0.0604		-0.00231	
	(0.0707)		(0.0715)		(0.0661)		(0.0725)	
y2002	0.0458		0.0296		-0.0492		0.0322	
	(0.0713)		(0.0737)		(0.0684)		(0.0734)	
y2004		-0.00909		0.00161		0.0931		0.100
		(0.101)		(0.104)		(0.101)		(0.105)

Table 12: Coefficient Estimates for Sabermetrics on Natural Log of Salary (Continued)

y2005		0.296***		0.312***		0.380***		0.357***
		(0.0925)		(0.0935)		(0.0927)		(0.0941)
y2006		0.347***		0.371***		0.441***		0.397***
		(0.0966)		(0.0990)		(0.0959)		(0.0987)
y2007		0.450***		0.504***		0.592***		0.549***
		(0.0979)		(0.100)		(0.0986)		(0.101)
y2008		0.469***		0.473***		0.568***		0.539***
		(0.102)		(0.104)		(0.102)		(0.104)
y2009		0.497***		0.486***		0.602***		0.549***
		(0.0992)		(0.102)		(0.101)		(0.103)
y2010		0.420***		0.417***		0.509***		0.499***
		(0.102)		(0.104)		(0.101)		(0.104)
y2011		0.570***		0.532***		0.617***		0.596***
		(0.100)		(0.102)		(0.0996)		(0.102)
Constant	2.080***	1.389***	5.269***	5.654***	5.457***	5.950***	5.619***	5.928***
	(0.290)	(0.276)	(0.261)	(0.259)	(0.268)	(0.267)	(0.259)	(0.265)
Observations	1,153	1,154	1,153	1,154	1,153	1,154	1,153	1,152
R-squared	0.658	0.720	0.643	0.695	0.614	0.705	0.646	0.691

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
OPS, Runs Created, WAR, Win Shares and At-Bats are average from years t-1 and t-2 from year contract began

Table 13: Coefficient Estimates for *Moneyball* Statistics on Natural Log of Salary using Interactive Variables

VARIABLES	On-Base Percentage Salary	Slugging Percentage Salary	Slugging and On-Base Salary
PostMoneyball	-1.638***	-0.737***	-1.408***
	(0.271)	(0.157)	(0.259)
On-Base	0.00797***		0.00245***
	(0.000568)		(0.000679)
On-Base x PostMoneyball	0.00472***		0.00373***
	(0.000820)		(0.00104)
Slugging		0.00565***	0.00491***
		(0.000282)	(0.000363)
Slugging x PostMoneyball		0.00166***	0.000466
		(0.000390)	(0.000513)
At-Bats	0.00530***	0.00472***	0.00470***
	(0.000165)	(0.000170)	(0.000169)
At-Bats x PostMoneyball	-0.000971***	-0.000874***	-0.000985***
	(0.000217)	(0.000224)	(0.000220)
Experience	0.0769**	0.0556*	0.0571*
	(0.0334)	(0.0298)	(0.0293)
Experience Squared	-0.00488***	-0.00352***	-0.00383***
	(0.00153)	(0.00136)	(0.00134)
Catcher	0.478***	0.569***	0.610***
	(0.106)	(0.101)	(0.100)
1B	0.0393	0.101	0.0675
	(0.104)	(0.0987)	(0.0979)
2B	-0.111	0.266***	0.231**
	(0.106)	(0.103)	(0.102)
SS	0.223**	0.542***	0.549***
	(0.106)	(0.104)	(0.102)
3B	0.129	0.220**	0.241**
	(0.110)	(0.104)	(0.104)
OF	0.200**	0.273***	0.281***
	(0.101)	(0.0955)	(0.0948)
1997	-0.190**	-0.107	-0.104
	(0.0844)	(0.0781)	(0.0783)
1998	-0.200**	-0.152*	-0.145*
	(0.0851)	(0.0819)	(0.0809)
1999	-0.0853	-0.0529	-0.0398
	(0.0774)	(0.0746)	(0.0740)

Table 13: Coefficient Estimates for *Moneyball* Statistics on Natural Log of Salary using Interactive Variables (Continued)

2000	0.0292	0.0438	0.0583
	(0.0805)	(0.0756)	(0.0757)
2001	0.102	0.104	0.114
	(0.0823)	(0.0774)	(0.0775)
2002	0.160*	0.146*	0.163**
	(0.0824)	(0.0781)	(0.0778)
2004	0.299***	0.228***	0.222***
	(0.0725)	(0.0690)	(0.0678)
2005	0.510***	0.403***	0.410***
	(0.0712)	(0.0672)	(0.0660)
2006	0.571***	0.486***	0.479***
	(0.0784)	(0.0742)	(0.0728)
2007	0.667***	0.583***	0.570***
	(0.0812)	(0.0780)	(0.0757)
2008	0.685***	0.593***	0.588***
	(0.0864)	(0.0814)	(0.0805)
2009	0.688***	0.635***	0.616***
	(0.0855)	(0.0797)	(0.0787)
2010	0.613***	0.567***	0.542***
	(0.0876)	(0.0847)	(0.0826)
2011	0.757***	0.715***	0.696***
	(0.0856)	(0.0832)	(0.0811)
Constant	2.622***	3.047***	2.523***
	(0.271)	(0.213)	(0.252)
Observations	2,307	2,307	2,307
R-squared	0.643	0.682	0.691

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14: Coefficient Estimates for Sabermetrics on Natural Log of Salary using Interactive Variables

VARIABLES	OPS Salary	Runs Created Salary	WAR Salary	Win Shares Salary
PostMoneyball	-1.035***	-0.0321	-0.00169	-0.0847
	(0.198)	(0.104)	(0.103)	(0.105)
OPS	0.00416***			
	(0.000199)			
OPS x PostMoneyball	0.00144***			
	(0.000276)			
Runs Created		0.00895***		
		(0.000435)		
Runs Created x PostMoneyball		0.00398***		
		(0.000663)		
WAR			0.168***	
			(0.0111)	
WAR x PostMoneyball			0.154***	
			(0.0171)	
Win Shares				0.0682***
				(0.00369)
Win Shares x PostMoneyball				0.0290***
				(0.00593)
At-Bats	0.00472***	0.00240***	0.00444***	0.00291***
	(0.000168)	(0.000240)	(0.000179)	(0.000230)
At-Bats x PostMoneyball	-0.000990***	-0.00194***	-0.00204***	-0.00164***
	(0.000220)	(0.000344)	(0.000254)	(0.000340)
Experience	0.0572*	0.0664**	0.0757**	0.0482
	(0.0299)	(0.0321)	(0.0324)	(0.0315)
Experience Squared	-0.00386***	-0.00423***	-0.00434***	-0.00347**
	(0.00137)	(0.00147)	(0.00149)	(0.00144)
Catcher	0.614***	0.480***	0.127	0.152
	(0.100)	(0.101)	(0.107)	(0.0995)
1B	0.0665	-0.00310	0.0494	-0.0379
	(0.0982)	(0.0994)	(0.107)	(0.0987)
2B	0.217**	0.0734	-0.251**	-0.251**
	(0.102)	(0.102)	(0.108)	(0.0998)
SS	0.542***	0.389***	-0.00146	0.0267
	(0.102)	(0.103)	(0.107)	(0.0999)
3B	0.236**	0.190*	-0.133	-0.0711
	(0.104)	(0.104)	(0.110)	(0.103)
OF	0.281***	0.246**	0.0880	0.0780
	(0.0950)	(0.0958)	(0.103)	(0.0946)
1997	-0.110	-0.0982	-0.0118	-0.135*
	(0.0790)	(0.0790)	(0.0855)	(0.0784)

Table 14: Coefficient Estimates for Sabermetrics on Natural Log of Salary using Interactive Variables (Continued)

1998	-0.147*	-0.129	-0.135	-0.132
	(0.0805)	(0.0834)	(0.0886)	(0.0821)
1999	-0.0381	-0.0427	-0.101	-0.0365
	(0.0738)	(0.0753)	(0.0773)	(0.0741)
2000	0.0608	0.0516	-0.0294	0.0484
	(0.0758)	(0.0766)	(0.0780)	(0.0769)
2001	0.116	0.120	0.0400	0.135*
	(0.0777)	(0.0779)	(0.0780)	(0.0783)
2002	0.170**	0.149*	0.0589	0.175**
	(0.0778)	(0.0795)	(0.0797)	(0.0789)
2004	0.223***	0.208***	0.241***	0.313***
	(0.0679)	(0.0696)	(0.0690)	(0.0708)
2005	0.406***	0.383***	0.413***	0.431***
	(0.0659)	(0.0664)	(0.0686)	(0.0677)
2006	0.476***	0.463***	0.495***	0.489***
	(0.0728)	(0.0749)	(0.0740)	(0.0751)
2007	0.567***	0.586***	0.629***	0.630***
	(0.0757)	(0.0778)	(0.0785)	(0.0788)
2008	0.584***	0.554***	0.610***	0.620***
	(0.0804)	(0.0823)	(0.0817)	(0.0823)
2009	0.614***	0.575***	0.647***	0.636***
	(0.0786)	(0.0812)	(0.0814)	(0.0828)
2010	0.541***	0.508***	0.555***	0.587***
	(0.0826)	(0.0843)	(0.0831)	(0.0851)
2011	0.695***	0.629***	0.668***	0.689***
	(0.0810)	(0.0823)	(0.0816)	(0.0827)
Constant	2.253***	5.392***	5.608***	5.726***
	(0.231)	(0.195)	(0.200)	(0.193)
Observations	2,307	2,307	2,307	2,305
R-squared	0.690	0.668	0.660	0.670

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1