Who Is The American League MVP On a Per Dollar Basis?

The Honors Program
Senior Capstone Project
Student's Name: Patrick Decker
Faculty Sponsor: John Quinn
April 2013

Table of Contents

ABSTRACT

This study will look at the value of MLB players in the American League on a per dollar basis. It will be based on the 2011 statistics and will use a linear regression model to create a value for each player. The value will be based on a variety of statistics based upon position. For all players: age, team wins, and years in the league will be accounted for. For field players: games played, hits, runs, doubles, triples, home runs, runs batted in, walks, strike outs, stolen bases, on base percentage, slugging percentage, fielding percentage and errors will be accounted for as well. Pitchers will be based on stats on a per inning basis of: innings, strike outs, walks, earned run average, walks hits per inning pitched, quality starts, saves and wins above replacement. The general public sees players as stars based on their giant contracts, and who is paid the most. There are players in the league who play better on a per dollar basis that the famous stars who have massive contracts. I will discover who those players are.

INTRODUCTION

Throughout Major League Baseball there is news of incredibly long contract deals between team's General Managers and player's agents. We have seen multiple players receive over twenty million dollars per year to play baseball. These superstars, as they are deemed, add a certain value to their teams. My goal is to determine, based on actual performance, who adds the most value per dollar, in essence, who is truly the Most Valuable Player, or MVP.

With the Luxury Tax increasing and becoming an issue for teams as they try to sign superstars with large contracts, this idea of value per dollar is becoming more and more important. Only four teams have paid the luxury tax since 2003 including the New York Yankees, Boston Red Soxs, Detroit Tigers, and Los Angeles Angles. However it should be interesting as more teams have increased their payroll in the most recent offseason, such as the Toronto Blue Jays, Los Angeles Dodgers, and even the Los Angeles Angels. At this point for the 2013 season the luxury tax cap is 178 million dollars which means the team's payroll of player's average salary for that year plus bonus must not exceed 178 million dollars or the team will be taxed on the amount in excess. At this time, the penalty tax is 22.5% for first time offenders. If a team breaks the luxury cap twice, they must pay a 30% rate, and for a third time, a 40% rate. If a team does not break the luxury cap then their luxury tax drops to 17.5%, 25% or 40% depending on their previous five years payrolls. The threshold is expected to jump up to 189 million dollars for the 2014-2016 seasons. There is one team that is already over that mark.

Currently, the Yankees are looking at rates higher than 40% for their luxury tax because they have broken the threshold more than three times. This is why they are trying to put together a team with less than the threshold to try and decrease the tax they will be paying in this year. Other teams have already made moves to help stay below this threshold, most notably the Boston Red Sox's in last year's house cleaning where they traded away a large part of their team to reduce contracts for players that they believed were not adding value to the team.

Since the unexpected success of the Oakland Athletics in 2001 under Billy Beane using his Moneyball system, statistic evaluation of athletes has become more and more popular. At this point, every team in the Major League of Baseball is using some sort of statistic evaluation for their players. However, there is little research or few case studies available about this because teams formulas have become a guarded secret. This project will attempt to create a variable that can value players in a statistically significant way. In other words, I will try to create my own formula to evaluate players as the Major League teams do.

The MVP is an award bequeathed to the most notable player in the League, either American League or National League, over the course of the season. This award is determined by the performance of the players throughout the league; however, I will be looking at it in a different way. The definition of MVP I will be using is the player who adds the most value per dollar based on their performance. This means that I will have to determine what value is created by each player as well as what their salaries are. I will be looking at a sample of the 2011 American League using data only from relevant players. My definition of a relevant player is a field player who has played 100 games or a pitcher who has either started 25 games or made 40 appearances. I will have to complete a regression model to discover if my value variable is significant and then can determine who adds the most value per dollar for the league.

LITERATURE REVIEW

There have been thousands of studies of players in Major League Baseball up to this point in time. However, most if not all of them are done within baseball organizations and are proprietary to ensure success for each team. Statistics in baseball became an integral part of teams beginning in the early 2000's. It was first used by Billy Beane and the Oakland A's in 2001, then by four teams in 2002 consisting of the Oakland Athletics, Cleveland Indians, Toronto Blue Jays, and the Boston Red Sox. After 2003, there was a surge of interest in applying statistics to baseball due to the book written by Michael Lewis called Moneyball: The Art of Winning an Unfair Game.

This book detailed the successful season the Oakland A's had in 2001 and showed how statistics were vital to the success of teams in the Major League. Now a minimum of 30 teams use statistics with 15 to 20 of them relying heavily on those statistics. The Associated Press released this statement:

> The debate no longer centers on whether there is a place for statistical analysis in the game. Instead, it's how prevalent it should be. . . . While there are still holdouts, more teams each year rely more on the numbers, an inevitability considering the investments teams make into players. With so much money and prestige at stake, it's only natural that teams will continue to look for whatever kind of edge they can find. While some speculate that the advantage teams gain by using modern statistical analysis is decreasing as more teams get

> on board, James believes there will always be a new front that will benefit the most forward-thinking teams. (albanygovermentlawreview, October, 2012)

As can be seen from this quote, it spread like wildfire though the league. Those who did use statistics already were fairly upset with Billy Beane for allowing Lewis to basically write an outline for the rest of the league. Theo Epstein of the Boston Red Sox was quoted as saying "I can't believe Billy is letting him write this book. He's handing out the blueprint."(Art of winning...more unfair game) Moneyball started a league-wide phenomenon even though statistics had been in baseball for a long time.

The art of using statistics in baseball has been around well before 2000. Earnshaw Cook, an engineering professor at Johns Hopkins, published a book called Percentage Baseball in 1964, which showed some insightful perspectives about how statistics were useful in baseball. It is considered the first sabermetric work. In 1977, Bill James, a man who had no baseball knowledge, wrote his first book called, 1977 Baseball Abstract: Featuring 18 Categories of Statistical Information That You Just Can't Find Anywhere Else. This led to the start of statistics in baseball.

Baseball Prospectus is a site (baseballprospectus.com) which has all of the sabermetric statistics for individual players, teams, and the league. It breaks them down to very useful outputs. This site was founded in 1996 and was getting increasingly used until it exploded in popularity after the release of Moneyball. In 2005, Fangraphs.com launched to provide these data to the general public. At this point, many of these statistics are used by the general public for fantasy leagues.

An interesting article that was recently published was in Sports Illustrated titled, "The Art of Winning an (Even More) Unfair Game" (115,13). This is the account of the Boston

Red Sox with Theo Epstein from 2002 until 2010. This article discussed how now that everyone has cracked the code on sabermetrics, those who have more money once again have the upper hand over those teams with limited budgets. This article discusses how the Red Sox have the financial ability to pay for people who they believe, according to these statistics, have an edge, and they have the resources that many teams do not have. Because of this division occurring once again, teams will be on the lookout for the next statistic that will help them. This article shows the example of Adrian Gonzalez's acquisition and how the Red Sox used sabermetrics to discover his talent, then used their financial strength to close the deal.

The second part to this story is given in the Sports Illustrated article titled "Fall of the Red Sox". This recounts the story of the 2012 Boston Red Sox season and how after they fell and based on the statistics of how they were performing, let go of most of their roster. They released many of the key players who had provided success in the years previous and were no longer worth what they were being paid. They then traded these high priced players to the Los Angeles Dodgers.

The article "Predicting Run Production and Run Prevention in Baseball: The Impact of Sabermetrics" published in the International Journal of Business, Humanities and Technology in 2012 gives us a look at Sabermetric variables that have been created and which variable have been eliminated as significant to predicting salary. It mentions how batting average is a useless statistic compared to the ability to not create an out and get on base in On Base Percentage. It also discussed the correlation of On Base Percentage and Slugging Percentage and how they should each be weighted to create the best model. I will start with a 1 to 1 weighting but as I improve the equation, I will weigh the variables as best suited. This article also discusses the importance of Runs Batted In as a statistic. Finally, after searching the literature, it is apparent that there is a lack of models that compare player's offensive and defensive statistics as well as pitching which my model is looking to do.

The Business Studies Journal article "Determinants of Value and Productivity in a complex Labor Market: How Sabermetrics and Statistical Innovation Changed the Business of Professional Baseball" shows the reader how valuation was previously conducted previous to the advent of sabermetrics. It shows, as does the book Moneyball, the scouting process and how players were determined to be attractive or not. It also shows an example regression model of how it is possible some teams value players at this point and discusses the history of sabermetrics and how it has changed that process in a technology driven world. The article from the Albany Government Law Review also discusses this transition. "According to Yogi Berra, 'in baseball you don't know nothing.' Sabermetricians would argue that for the first one hundred years or so of professional baseball, Berra was right" (Frankel). This argument exemplifies the shift from a scout driven process to the statistical process that is used.

As far as looking toward where the future of sabermetrics is going, the article "Moneyball: Can Sports Statistics Save Political Studies?" gives the reader some insight. Sabermetrics, as the article details, has been put to use in other sports in different varieties of ways. It highlights the use of statistics in soccer in the English Premiership which is comical because this was my original idea of the capstone project. However, due to the private ownership of clubs, I was unable to complete that. This article goes on to discuss the possibility of using the same statistical studies in political studies. We have seen the start of this very recently with the 2012 presidential election analysis by Nate Silver. He has predicted the past few elections by how many electoral votes each candidate will receive and has been exceptionally accurate. This has given some insight into where the future of this industry could go.

As far as other directions in which this use of statistical analysis could be applied is the prediction of minor league players in the major league as well as players in college being successful in the professional leagues. This is discussed in the article "The speed of human capital formation in the baseball industry: the information value of minor-league performance in predicting major-league performance." However, this article goes on to explain how there is a large gap between performance in the minor leagues and the major league, and as a result, predictions have been ineffective. The inability to predict between leagues has caused problem in the college draft as well. The future of this industry on the baseball side would be figuring out a way to create a model that can predict between leagues and calculate players' values.

I also have Jason Scott's capstone project from last year. This project looked at the question of whether NFL players were overvalued or undervalued. It deals with a similar valuation of players based on much different statistics. However, it uses Minitab in the same way that I will be using Minitab to create a regression model. He also used an individual statistic factor that created a value for each player based on individual statistics.

My thesis will provide similar type information to that of a baseball organization. It will require me to look at each player in the American league, and give them each a value based on these sabermetrics. It is something that is done often in the organization, however, is not often available to the public. My thesis will hopefully be able to show at least part of how teams value players and decide who deserves what money. Those who are my most valuable players on a per dollar basis would be the players that teams would want to try and trade for.

There will be a compilation of websites I will have to use in order to find all of the statistics I will need. I will be using espn.go.com, mlb.com, and baseballprospectus.com for the individual statistics needed. I will be using usatoday.com and spotrac.com to collect the salaries and financial sides of the data. I will also be using the online Sabr Encyclopedia to define all my statistics. These will be the sites that I need to complete the data portion of my thesis.

VARIABLE ANALYSIS

I used four variables in my main regression model: Age, Team Wins, Experience, and finally the Player Value Variable. These are fairly simple variables except for the PVV. In this section I will analyze each variable and explain how they are used according to this experiment.

Age: This variable is the age of each player first and foremost. To create the actual value used in the regression model, 27 was divided by the players age as it was proven by Bill James in his book titled *The Bill James Historical Baseball Abstract* that the age 27 was when baseball players peaked in talent. This means that players who were older had a lower value added to the regression while younger players had more value to add because they still had time to their peak.

Team Wins: This variable was created using the team wins of the team that each player was on during the 2011 season. I created an average winning percentage based on the American League which was slightly above 50% which can be explained by interleague play. Each player's team wins was divided by the average and added to the model. This means that those players who played on better teams had more value.

Experience: This variable took into account how many years each player had been in the league. I took an average of each player in the American League which was then used by each player's experience divided by the average. This created more value for those veteran players who were in the league for many years. This may seem as if it is counteracting the Age variable and I agree. However, they measured two different statistics and therefore one may add value while another does not or vice versa.

Player Value Variable: The player value variable was a variable that I created which incorporated a plethora of other offensive and defensive or pitching statistics. It used these statistics to create a value that averaged to 1, so an under average value player had a PVV less than 1 while a over average player had a PVV higher than 1. Here are the statistics for Field Players first:

Runs: The number of Runs scored by each player divided by the average of all players.

Hits: The number of Hits made by each player divided by the average of all players.

Doubles: The number of doubles hit by each player divided by the average of all players.

Triples: The number of triples hit by each player divided by the average of all players.

Home Runs: The amount of home runs hit by each player divided by the average of all players.

Runs Batted In: The amount of runs each player brought in to score divided by the average of all players.

Base on Balls: The amount of walks received by each player divided by the average of all players.

Strike Outs: The amount of times a player struck out divided by the average of all players.

Stolen Bases: The amount of times a player stole second, third or home while base running divided by the average of all players.

On Base Percentage: This is the percentage of times a player reaches base either from a hit, walk, or hit by a pitch compared to the average of all players.

Slugging Percentage: This is the total number of bases for each hitter per at bat divided by the average of all players.

Fielding Percentage: This is the percentage of balls that each defender stopped and made a play on that came into their area. It takes into account errors and fielder choices.

Range: This statistic gives each player a range based on how much ground and field they can cover with a high fielding percentage.

This is a list of the Pitching statistics:

Innings: Number of Innings pitched by each pitcher.

Strike Outs: This is the number of strike outs each pitcher has on a per inning basis.

Walks: The number of walks given up by each pitcher on a per inning basis.

Earned Run Average: The number of runs allowed excluding runs caused by errors on a per 9 inning basis, or a complete game.

Walks Hits per Inning Pitched(WHIP): The amount of walks and hits a pitcher gives up per inning.

Wins Above Replacement: The statistic that shows how many more or less wins the team has compared to if that player did not play.

Saves: The amount of times a pitcher came into a close game and shut the team down to secure the victory.

All of these statistics were put into an equation to create a PVV for Pitchers and Field players that should be comparable.

METHODOLOGY

My original project began with the idea of creating a way to value soccer teams per dollar in order to evaluate if teams were more likely to win if they spent more money. Obviously, some things changed since that initial idea. After completing research with my research librarian, we could only find information for a couple different teams from many years ago due to the fact that Premier League English Teams are all privately owned for the most part and do not share their financial information. I worked with my research librarian in searching a variety of other sports to discover which sport had the most available information. We concluded that baseball had the most statistics as well as financials available. Proceeding from this, we came up with the idea of valuing players to figure out who should be the MVP based on a dollar amount as opposed to the best of the best based on statistics.

I generated a list of 104 relevant field players and 112 relevant pitchers totaling a sample size of 216 American League players. These players were selected based upon 100 game appearances for field players, 40 appearances with over 40 innings for relief pitchers, and 25 plus starts for starting pitchers. After compiling the list of players, I created a spreadsheet in Excel which tells the team for each player and then compiled all of the data I needed. I researched and found salaries, ages, team wins, years in the league and a variety of individual statistics which were used to create the player value variable. These included runs, hits, doubles, triples, home runs, runs batted in, walks, strike outs, stolen bases, on base percentage and slugging percentages as well as defensive statistics or fielding percentage, errors and range for the field players. For the pitchers, I compiled innings pitches, strike outs, walks, earned run average, walks hits per inning pitched, wins above replacement and saves. Once I had compiled all of these data in a large excel sheet, I could manipulate it to create my equations for each player. I then created averages for each variable to be used in creating value for each player. My equation looks like this:

$$\text{Salary} = \beta_1\text{age} + \beta_2\text{team wins} + \beta_3\text{years in the league} + \beta_4\text{player value variable}$$

For the age variable, I divided 27 by the player's age. This creates a variable that is higher younger players and lower for older players, as it is proven that 27 is the age players start to decline (James). For team wins, I used the average team winning percentage in the league and put that player's percentage over the average. This creates a higher factor for players who are playing on better teams. I did likewise with years in the league as those with more experience have higher value than those with less. Finally, I calculated the player value variable which was compiled of all the individual statistics. I created averages for each statistic and compared each player to the average, some having inverse relationships, such as strike outs which one would want less of. I summed all of these for field players first and ended up with 14 variables which I then divided by the amount of variables. In some cases, for designated hitters, they did not have fielding statistics and therefore had to be divided by 11 instead. This created a value. I then took the average of this value for every player in the league and divided by the average once again, doubly comparing these variables. This became the player value variable. For Pitchers, it was a little different as some pitched far more than others. Due to this, I calculated strike outs and walks per inning so each stat was comparable. I then completed the same process with some being inversely related such as walks per inning. I doubly compared these once again giving a final variable that was comparable between field players and pitchers.

There are multiple parts of the project that have changed at this point. I changed the criteria which was originally decided for relevant players by Professor Quinn and myself after looking at the data of players. I included errors in my player value variable originally, but after seeing how it affected the variable, it was taken out because it skewed the data. I still have fielding percentage included which calculated errors into it so I believe this is acceptable. I also dropped doubles and triples which created a more accurate variable because slugging percentage took these into account. This dropped my player's statistics to 11 variables instead of 14 as well. The most important change that brought a significant increase to the significance of this model was using the Log of the salary instead of just the salary. This, as Professor Olinsky informed me is used quite often in these types of models.

Once I had an equation I plugged the log of the salary into the regression models to try and make a valid formula to predict salary. I used this to verify that my PVV was a significant calculation of value. I made several adjustments along the way including those above to create an accurate model. Besides regressions and looking at the outliers, I also calculated covariance and finally stepwise regressions. After completing the different models, I realized how some of the outcomes were a little skewed towards the pitchers, so I completed another regression for just pitchers and for just field players.

The amount of players was also brought up as a factor of whether it was a significant number or not. I believe it is a significant sample size for the specificity of the project. Teams are likely to have larger databases with minor league players as well as players in the National League. However, I am simply doing a study of the American League from the year of 2011 of significant players, therefore not including minor league players who played only a few games. I believe this sample size is large enough for this specific project due to the specificity of it. The amount of scholarly articles was also brought into question. As there are few published articles on this due to the secrecy of teams to keep their formulas private, I found few pieces, however, I will be continuing to add articles that have appeared on this subject, especially those about Bobby Valentine's misuse of sabermetrics and how the Boston Red Sox's plummeted in comparison to previous years.

RESULTS

The initial test completed was a regression of Salary versus my four variables of team wins, experience, age and the player value variable (PVV). This showed an adjusted R-squared of 35.8% which falls outside the range of 40-60% I was hoping for, however, there were some problems.

```
The regression equation is
salary = - 2287727 - 2556529 age + 208387 team wins + 3373875 experience
         + 5162488 PPV


Predictor       Coef  SE Coef      T      P
Constant    -2287727  4452573  -0.51  0.608
age         -2556529  4087130  -0.63  0.532
team wins     208387   487555   0.43  0.670
experience   3373875   728502   4.63  0.000
PPV          5162488   968514   5.33  0.000


S = 4041921   R-Sq = 37.0%   R-Sq(adj) = 35.8%


Analysis of Variance

Source           DF           SS           MS      F      P
Regression        4  2.02638E+15  5.06595E+14  31.01  0.000
Residual Error  211  3.44713E+15  1.63371E+13
Total           215  5.47352E+15


Source      DF       Seq SS
age          1  1.10588E+15
team wins    1  3.95185E+13
experience   1  4.16811E+14
PPV          1  4.64175E+14
```

There were very high P values for both age and team wins which would mean they are not statistically significant. There were also very high coefficients. Due to this, I created a stepwise regression to only include significant variables.

```
Response is salary on 4 predictors, with N = 216


Step              1         2
Constant     278781  -4846080

experience  3775213   3771114
T-Value        9.21      9.78
P-Value       0.000     0.000
```

```
PVV                      5128830
T-Value                     5.35
P-Value                    0.000

S            4279716   4027780
R-Sq           28.39     36.87
R-Sq(adj)      28.05     36.28
Mallows Cp      27.9       1.5
```

As shown, there is a R-Sq (Adj) of 36.28% which is lower than the range I was looking for. We proceeded to review the data as there were a large number of unusual observations, all of which had to do with Age. Every one of the unusual observations was of an older player who would be considered an upper level player. It was concluded that this was caused by the fact that these players had received large contracts due to their performance earlier in their career and they loaded up their contracts and were now declining in performance. This was a side trend that was interesting to find from this data set. We also noticed an error in the data set that caused a large mistake in the data. When final numbers were put into the model, human error took its toll and led to misplacement.

In the corrected data, we saw a little better result when dealing with statistical significance, but not enough to create an accurate model. After discussions with Professor Olinsky and Professor Quinn, it was concluded that using the Log for the salary was the right method to further improve this model as this is what is done at the Major League Level. This is the regression model of LOG(salary) versus the four variables:

```
The regression equation is
LOG(salary) = 6.76 - 1.24 age - 0.126 team wins + 0.298 experience + 0.522 PVV


212 cases used, 4 cases contain missing values


Predictor       Coef  SE Coef      T      P
Constant      6.7649   0.4940  13.70  0.000
age          -1.2402   0.4011  -3.09  0.002
team wins    -0.1258   0.2230  -0.56  0.573
experience   0.29819  0.07137   4.18  0.000
PVV          0.52182  0.09811   5.32  0.000
```

```
S = 0.398205   R-Sq = 48.0%   R-Sq(adj) = 47.0%
Analysis of Variance

Source           DF       SS      MS      F      P
Regression        4  30.3437  7.5859  47.84  0.000
Residual Error  207  32.8234  0.1586
Total           211  63.1671


Source      DF   Seq SS
age          1  22.4749
team wins    1   0.0869
experience   1   3.2964
PVV          1   4.4855
```

This regression using the log of the salary shows us that all factors are significant except the team wins factor with a p value of .573. As expected, age is negatively related to the log(salary) while experience and the PVV are positively related. This leaves us with an adjusted R squared of 47% which falls within the initial range I wanted to show of 40-60%. We can see in the Histogram (Appendix C) how it looks like a normal curve. We can also see a smaller number of obscure observations in the versus order graph compared to the graphs compared to just the salary instead of log. The common theme of the unusual observations is that they are all older than 27 in age. This could mean that they are being paid for long term high paying contracts that they received after proving themselves while younger. This would explain why they are still receiving higher pay with less value added to the team. To remove the Team wins factor I ran a stepwise regression which increased the Adjusted R-Squared to 47.21%

```
Response is LOG(salary) on 4 predictors, with N = 212
N(cases with missing observations) = 4 N(all cases) = 216


Step            1      2      3
Constant    5.813  5.329  6.639

experience  0.481  0.481  0.297
T-Value     11.63  12.26   4.17
P-Value     0.000  0.000  0.000

PVV                0.484  0.511
T-Value             4.96   5.32
```

```
P-Value            0.000  0.000

age                       -1.23
T-Value                   -3.07
P-Value                   0.002

S           0.428  0.405  0.398
R-Sq        39.19  45.60  47.96
R-Sq(adj)   38.90  45.08  47.21
Mallows Cp   34.2   10.7    3.3
```

After looking through the variables and continuing to analyze, I adjusted the PVV to create a more statistically significant Valuation variable which was the goal all along.

```
The regression equation is
LOG SAL = 6.70 - 1.23 age - 0.117 team wins + 0.297 experience + 0.562 PVV2


212 cases used, 4 cases contain missing values


Predictor      Coef   SE Coef      T      P
Constant     6.7047    0.4903  13.67  0.000
age         -1.2266    0.3975  -3.09  0.002
team wins   -0.1172    0.2205  -0.53  0.596
experience  0.29682   0.07080   4.19  0.000
PVV2        0.56194   0.09920   5.66  0.000


S = 0.395027   R-Sq = 48.9%   R-Sq(adj) = 47.9%


Analysis of Variance

Source           DF       SS      MS      F      P
Regression        4  30.8654  7.7164  49.45  0.000
Residual Error  207  32.3017  0.1560
Total           211  63.1671


Source      DF   Seq SS
age          1  22.4749
team wins    1   0.0869
experience   1   3.2964
PVV2         1   5.0072

Stepwise Regression
Response is LOG SAL on 4 predictors, with N = 212
N(cases with missing observations) = 4 N(all cases) = 216


Step            1      2      3
Constant    5.813  5.287  6.588

experience  0.481  0.478  0.295
T-Value     11.63  12.29   4.18
P-Value     0.000  0.000  0.000
```

```
PVV2                     0.528  0.552
T-Value                   5.33   5.67
P-Value                  0.000  0.000

age                             -1.22
T-Value                         -3.07
P-Value                         0.002

S            0.428  0.402  0.394
R-Sq         39.19  46.48  48.79
R-Sq(adj)    38.90  45.97  48.05
Mallows Cp    38.1   10.7    3.3
```

After completing a regression and stepwise with the new PVV variable, we see the Adjusted R-Squared jumped up to 48.05% which is almost a complete percentage jump. This showed a significant valuation variable as we can see the P value of 0.000 which is exactly what I was looking for. We again saw the trend of unusual observations with older players. Once the regression model had proven to be successful and the PVV variable had proven to be statistically significant, I turned to the MVP portion of the project. I ranked each player based on their PVV which would equivalently be similar to the actual MVP calculations that are made by Major League Baseball to determine who wins the award. My variable turned out to be even more accurate than predicted as the player who was named number one based on statistics was Justin Verlander who won the MVP of the American League in 2011. I then proceeded to put these values to a per dollar basis by dividing by the salary. Here is a chart of the top outcomes:

All Players			
Rank	Name	PVV	Salary
1	Robertson D	1.505592	460450
2	Gonzalez G	1.261844	420000
3	Tomlin J	1.242634	417200
4	Santana, C	1.198258	416600
5	Ogando A	1.235498	430150
6	Hellickson J	1.195404	418400
7	Santos S	1.21832	435000
8	Harrison M	1.198707	428830
9	Walden J	1.143093	414000
10	Andrus, E	1.225212	452180
11	Trumbo, M	1.10883	414000
12	Pineda M	1.09844	414000
13	Masterson J	1.239138	468400
14	Feliz N	1.185618	457160
15	Pestano C	1.066798	414100
16	Avila, A	1.077392	425000
17	Holland G	1.24683	497150
18	Sale C	1.064691	425000
19	Nova I	1.068577	432900
20	Alburquerque A	1.213082	495000
21	Humber P	1.198458	500000
22	Joyce, M	0.988869	426500
23	Holland D	0.994527	431810
24	Bailey A	1.069468	465000
25	Gardner, B	1.214106	529500

David Robertson was the number one ranked player for value added per dollar according to my model. He was followed closely by Gio Gonzalez and Josh Tomlin. Eight of the top ten players projected as my most value added per dollar were pitchers, mainly relief pitchers. Carlos Santana was the first field player ranked 4th in the league in my model. The top two players David Robertson and Gio Gonzalez were considered top pitchers in the league in 2012 which makes me think my model is very accurate. At the bottom of the table were players with large contracts who would be considered stars in the league consisting of the likes of John Lackey, Vernon Wells, A.J. Burnett, Tori Hunter and Mark Texeira. These are the bottom five players and although they may not have the largest contracts, they certainly have higher ones and are older players. This goes back to the trend we saw in the regression study which showed older players who had received large contracts as star players.

However, Professor Olinski, Professor Quinn and I saw a pattern in the data as 20 out of the top 25 players were pitchers. This made us think that it could be possible even with the standardizing of pitchers valuation and field players valuation, that these were not even. Therefore, I completed two separate cases, one for field players, one for pitchers. This table shows the top field players in value added per dollar:

Field Players			
Rank	Name	PVV	Salary
1	Santana, C	1.198258	416600
2	Andrus, E	1.225212	452180
3	Trumbo, M	1.10883	414000
4	Avila, A	1.077392	425000
5	Joyce, M	0.988869	426500
6	Gardner, B	1.214106	529500
7	Escobar, A	0.981185	428000
8	Bourjos, P	0.944698	414000
9	Wieters, M	1.0122	452250
10	Jackson, A	0.982236	440000
11	Hosmer, E	1.120773	502500
12	Nunez, E	0.930604	419300
13	Arencibia, J	0.908352	417400
14	Smoak, J	0.900698	419000
15	Pennington, C	0.901362	420000
16	Moreland, M	0.905826	426000
17	Revere, B	1.022891	492500
18	Andino, R	0.869667	421500
19	Getz, C	0.913748	443000
20	Fuld, S	0.853912	418300
21	Boesch, B	0.873198	430000
22	Brantley, M	0.848503	421800
23	LaPorta, M	0.827912	431400
24	Valencia, D	0.825013	437500
25	Rodriguez, S	0.795354	428600

After completing the Field Players, I completed the case for the Pitchers shown here:

Pitchers			
Rank	Name	PVV	Salary
1	Robertson D	1.505592	460450
2	Gonzalez G	1.261844	420000
3	Tomlin J	1.242634	417200
4	Ogando A	1.235498	430150
5	Hellickson J	1.195404	418400
6	Santos S	1.21832	435000
7	Harrison M	1.198707	428830
8	Walden J	1.143093	414000
9	Masterson J	1.239138	468400
10	Feliz N	1.185618	457160
11	Pestano C	1.066798	414100
12	Holland G	1.24683	497150
13	Sale C	1.064691	425000
14	Nova I	1.068577	432900
15	Alburquerque A	1.213082	495000
16	Humber P	1.198458	500000
17	Holland D	0.994527	431810
18	Bailey A	1.069468	465000
19	Sipp T	0.969363	436800
20	Thompson R	0.922908	418000
21	Cecil B	0.814453	443100
22	Collins T	0.744092	414000
23	Schlereth D	0.73997	418000
24	Bard D	0.862951	505000
25	Romero R	1.594869	1000000

After seeing this difference between the Pitchers and Fielders in the outcomes of my valuation model, I decided to relook at the regression models for just pitchers and just field players. Here are the regression results for the Pitchers only:

```
The regression equation is
LOG sal = 6.62 - 0.873 Age - 0.337 Team Wins + 0.262 Experience + 0.463 PVV


109 cases used, 3 cases contain missing values



Predictor       Coef   SE Coef       T       P
```

```
Constant     6.6239   0.5999  11.04  0.000
Age         -0.8726   0.4848  -1.80  0.075
Team Wins   -0.3370   0.3043  -1.11  0.271
Experience  0.26176  0.08606   3.04  0.003
PVV          0.4627   0.1100   4.21  0.000


S = 0.388611   R-Sq = 43.3%   R-Sq(adj) = 41.1%

Response is LOG sal on 4 predictors, with N = 109
N(cases with missing observations) = 3 N(all cases) = 112


Step            1      2      3
Constant    5.812  5.384  6.323

Experience  0.379  0.377  0.251
T-Value      7.13   7.55   2.93
P-Value     0.000  0.000  0.004

PVV                 0.43   0.44
T-Value             3.93   4.06
P-Value            0.000  0.000

Age                       -0.88
T-Value                   -1.80
P-Value                   0.074

S           0.419  0.393  0.389
R-Sq        32.19  40.81  42.59
R-Sq(adj)   31.55  39.69  40.95
Mallows Cp   19.3    5.5    4.2
```

As we can see from the P values, team wins is not statistically significant and Age is just over the barrier of significance. When the stepwise regression was completed, we saw an Adjusted R-Squared of 40.95% which is lower than the adjusted R-Squared of our total players regression which was 48.05%. This means that the pitchers were not valuated as well as the field players which are shown here in the regression model of just the field players:

```
The regression equation is
LOG sal = 6.70 - 1.61 Age + 0.096 Team Wins + 0.375 Experience + 0.745 PVV


103 cases used, 1 cases contain missing values


Predictor       Coef  SE Coef      T      P
Constant      6.7032   0.7629   8.79  0.000
Age          -1.6126   0.6284  -2.57  0.012
Team Wins     0.0965   0.2925   0.33  0.742
Experience    0.3745   0.1121   3.34  0.001
PVV           0.7452   0.1813   4.11  0.000


S = 0.359054   R-Sq = 62.0%   R-Sq(adj) = 60.4%
```

```
Response is LOG sal on 4 predictors, with N = 103
N(cases with missing observations) = 1 N(all cases) = 104


Step             1      2      3
Constant     5.778  5.098  6.821

Experience   0.628  0.631  0.371
T-Value      10.72  11.48   3.34
P-Value      0.000  0.000  0.001

PVV                  0.68   0.76
T-Value              3.81   4.34
P-Value             0.000  0.000

Age                        -1.65
T-Value                    -2.67
P-Value                    0.009

S            0.392  0.368  0.357
R-Sq         53.24  59.18  61.92
R-Sq(adj)    52.78  58.36  60.77
Mallows Cp    21.5    8.2    3.1
```

As we see by the adjusted R-Squared of 60.77% here in the field players only, the valuation model is more accurate for the field players than it is for the pitchers. This would explain why there are mainly pitchers in the top 25 players of value added per dollar.

CONCLUSION

There were two main parts to this study. The first was creating a regression model that had a significant valuing variable, my PVV; and the second was discovering which players added the most value per dollar using this variable and salary. In the first part, I collected data and successfully showed over a 48% correlation of salary. My model therefore explained 48% of the salary for baseball players, however when broken up to only Pitchers or only Field Players, the model explained just over 40% and just over 60% correlation of salary respectively. By identifying these three variables and creating another variable we described almost 50 % of salaries which was right in the range of 40-60% that was expected in the beginning of the experiment.

The second part was developing a list of which players were the most valuable per dollar to teams. We noticed pitchers were of higher value than many field players and this may be due to the fact that less salary was explained by this model for pitchers than for field players. It was also noted that higher paid big name players who are supposed to be the best on the team were toward the bottom of the list for value added per dollar. This was a theme throughout the entire process. Many of these stars caused unusual observations in the regression model because their salaries were so high. What this means is that these stars play very well in the beginning of their career and then receive long term deals for large sums of money and then slowly start to age not earning what they are paid. This is also why we see lower paid players at the top of value added because they are usually new to the league and are proving themselves before they receive a larger contract.

Areas to Improve

The regression model explained 40% of Pitchers salary which could have been higher. The PVV for pitchers could have gone more in depth of higher and more complicated metrics. There are many different statistics that combine a variety of statistics similar to my PVV that could have been used in the model had I thought they would add more value than those used. This could have been done with Field Players as well. There was a popularity variable that was considered, but left out. This variable would have taken into account jersey sales as a guess of popularity which would have possibly explained ticket sales brought in by single players. Unfortunately, there are many problems with this idea in calculation theory. I do not have the sales numbers for hundreds of stores to calculate the sales of players jerseys as well we the fact that not all players jerseys are offered for sale. It may have served as a popularity number, however it would have been difficult to calculate ticket sales generated by a single player from this. That number would have to come directly from a baseball teams information which I did not have access to.

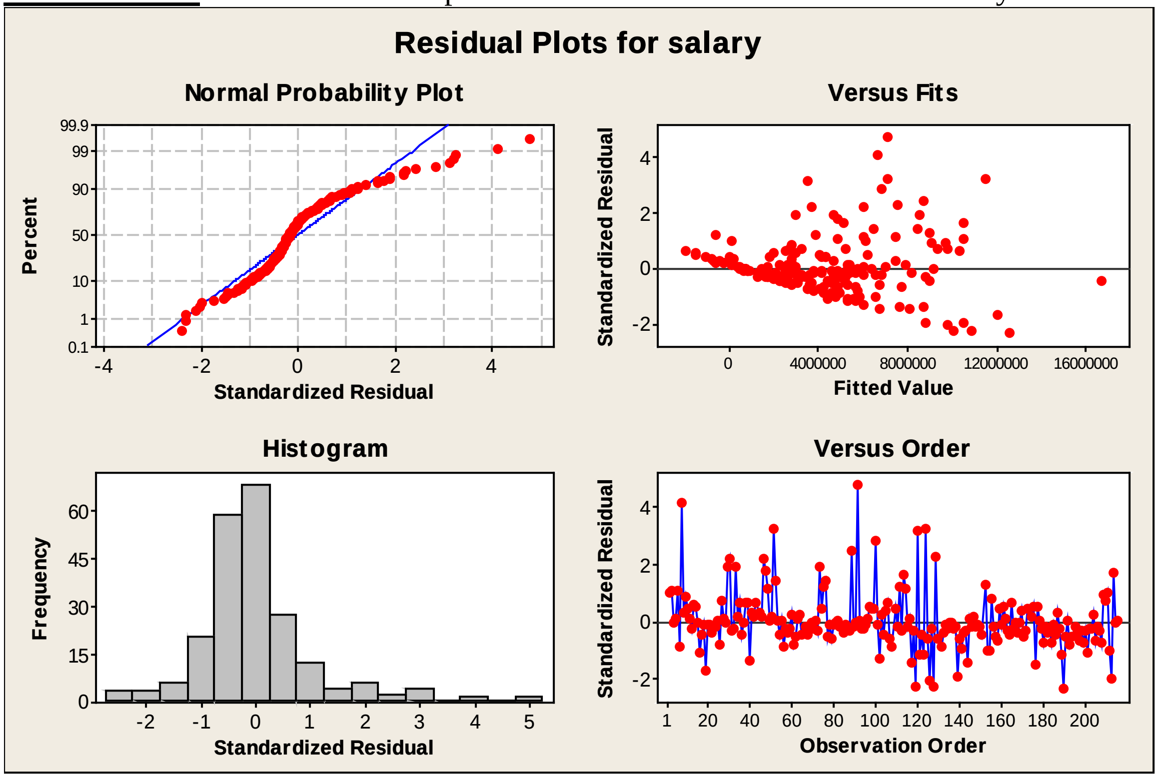
Other pieces that could have been improved were notation in my spreadsheets. There were multiple times that I became disoriented and had to go back and figure out what each column meant instead of having a title at the top of each column explaining what each statistic or calculation was.

Continuing Research

If I were to continue this project, I would compete a second year of data in attempt to see the outcome of that year and if these lower paid younger players received higher contracts or continued to produce as much value per dollar.

Another direction this research could be taken would be to apply this to a college league or the Minor League to see if it transferred over to Major League Baseball draft and see which players were picked first according to value. This would not allow for a value per dollar approach as these players do not receive a salary. This could be interesting for future research to see if predicting the number one draft pick is possible from just the PVV variable used in my model.

APPENDIX A – Four In One Graph of Residuals for Total Data Set on Salary



APPENDIX B – Unusual Observations for total data set on Salary

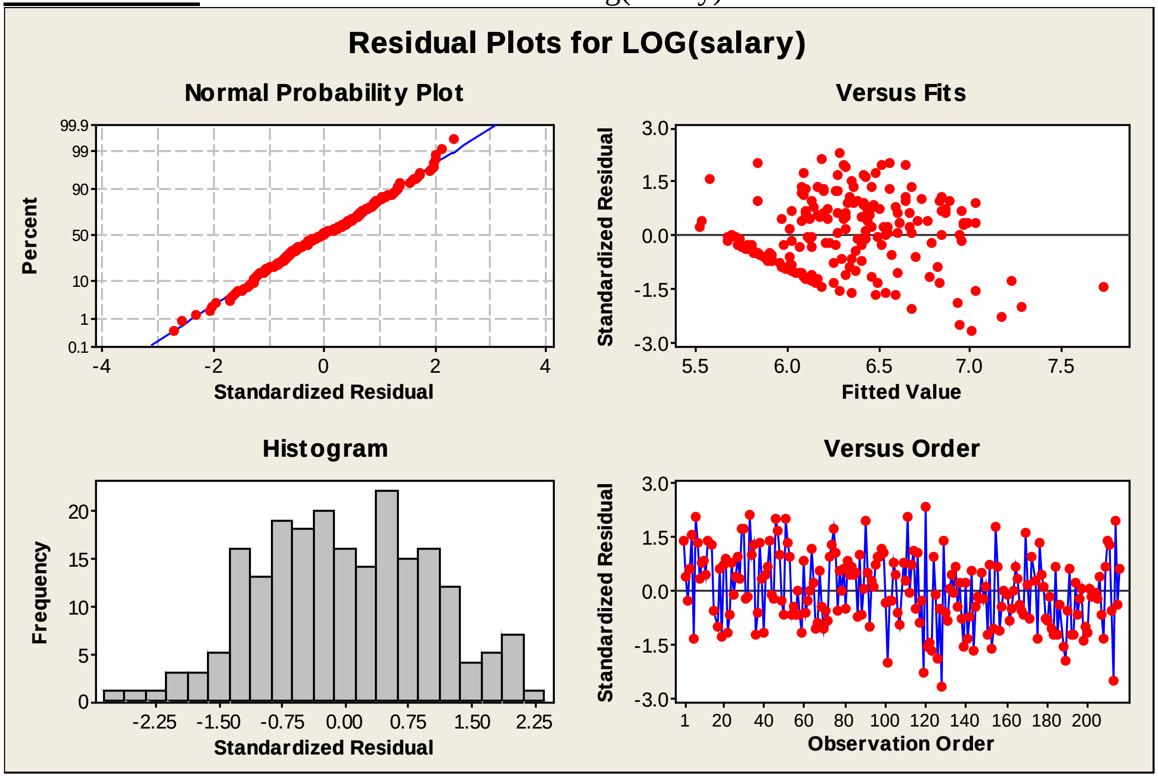
Unusual Observations

Obs	age	salary	Fit	SE Fit	Residual	St Resid
7	0.87	23125000	6640855	478213	16484145	4.11R
30	0.90	14857142	6048885	377819	8808257	2.19R
46	0.90	12500000	3716166	341332	8783834	2.18R
51	0.96	20000000	7056932	561767	12943068	3.24R
88	0.75	18500000	8700487	569078	9799513	2.45R
91	0.84	26187500	7109664	511364	19077836	4.77R
100	0.71	18000000	6813609	862565	11186391	2.84R
107	0.75	3000000	5302765	1503089	-2302765	-0.61 X
119	0.60	2000000	10863192	1220461	-8863192	-2.30RX
120	0.82	15950000	3518807	805892	12431193	3.14R
122	0.66	14911700	16765339	1250437	-1853639	-0.48 X
124	0.87	24285714	11534973	914025	12750741	3.24R
126	0.77	1500000	9835879	649806	-8335879	-2.09R
128	0.71	900000	10062374	691769	-9162374	-2.30R
129	0.79	16500000	7582353	761581	8917647	2.25R
165	0.96	12850000	10352296	1113943	2497704	0.64 X
190	0.66	3250000	12571851	951144	-9321851	-2.38R
211	0.75	3800000	69731	1344149	3730269	0.98 X
213	0.75	900000	8765544	1015493	-7865544	-2.01R

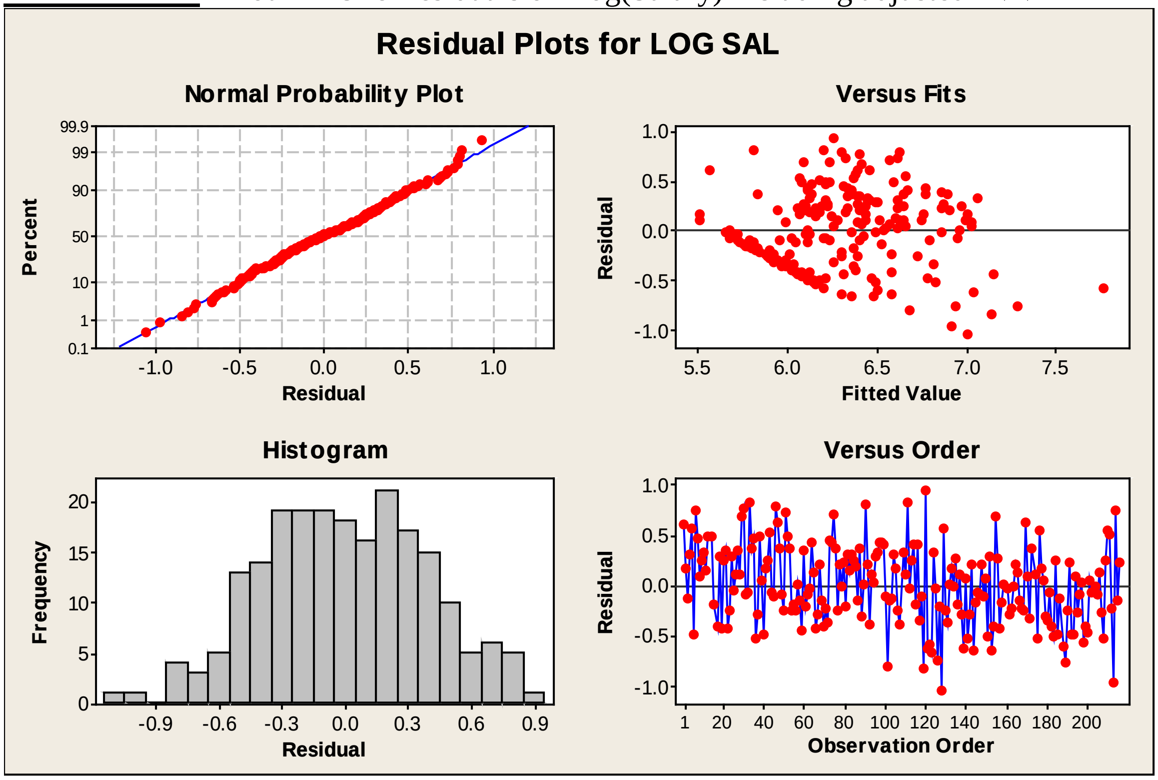
R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Note: All age variables are less than 1

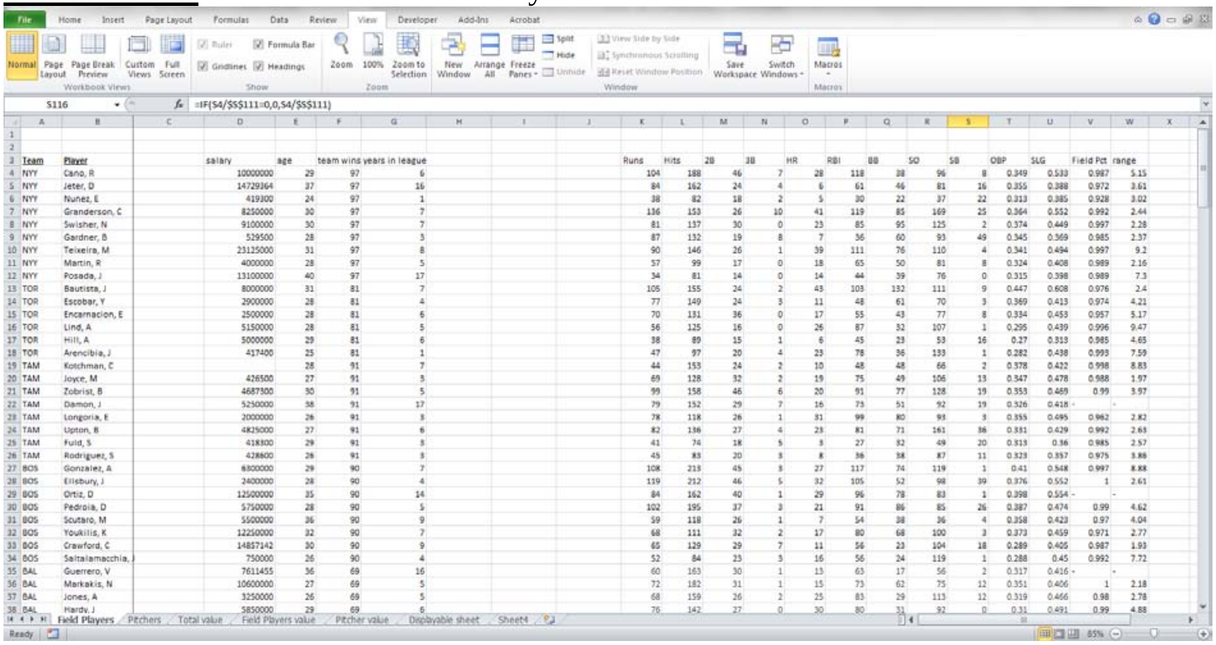
APPENDIX C – Four In One Residuals on Log(Salary)



<u>APPENDIX D</u> – Four In One Residuals on Log(Salary) including adjusted PVV



<u>APPENDIX E</u> – Screenshot of Field Players Statistics Table



	Team	Player	salary	age	team wins	years in league	Runs	Hits	2B	3B	HR	RBI	BB	SO	SB	OBP	SLG	Field Pct	range
4	NYY	Cano, R	10000000	29	97	6	104	188	46	7	28	118	38	96	8	0.349	0.533	0.987	5.15
5	NYY	Jeter, D	14729364	37	97	16	84	162	24	4	6	61	46	81	16	0.355	0.388	0.972	3.61
6	NYY	Nunez, E	419300	24	97	1	38	82	18	2	5	30	22	37	22	0.313	0.385	0.928	3.02
7	NYY	Granderson, C	8250000	30	97	7	136	153	26	10	41	119	85	169	25	0.364	0.552	0.992	2.44
8	NYY	Swisher, N	9100000	30	97	7	81	137	30	0	23	85	95	125	2	0.374	0.449	0.997	2.28
9	NYY	Gardner, B	529500	28	97	3	87	132	19	8	7	36	60	93	49	0.345	0.369	0.985	2.37
10	NYY	Teixeira, M	23125000	31	97	8	90	146	26	1	39	111	76	110	4	0.341	0.494	0.997	9.2
11	NYY	Martin, R	4000000	28	97	5	57	99	17	0	18	65	50	81	8	0.324	0.408	0.989	2.16
12	NYY	Posada, J	13100000	40	97	17	34	81	14	0	14	44	39	76	0	0.315	0.398	0.989	7.3
13	TOR	Bautista, J	8000000	31	81	7	105	155	24	2	43	103	132	111	9	0.447	0.608	0.976	2.4
14	TOR	Escobar, Y	2900000	28	81	4	77	149	24	3	11	48	61	70	3	0.369	0.413	0.974	4.21
15	TOR	Encarnacion, E	2500000	28	81	6	70	131	36	0	17	55	43	77	8	0.334	0.453	0.957	5.17
16	TOR	Lind, A	5150000	28	81	5	56	125	16	0	26	87	32	107	1	0.295	0.439	0.996	9.47
17	TOR	Hill, A	5000000	29	81	6	38	89	15	1	6	45	23	53	16	0.27	0.313	0.985	4.65
18	TOR	Arencibia, J	417400	25	81	1	47	97	20	4	23	78	36	133	1	0.282	0.438	0.993	7.59
19	TAM	Kotchman, C		28	91	7	44	153	24	2	10	48	48	66	2	0.378	0.422	0.998	8.83
20	TAM	Joyce, M	426500	27	91	3	69	128	32	2	19	75	49	106	13	0.347	0.478	0.988	1.97
21	TAM	Zobrist, B	4687300	30	91	5	99	158	46	6	20	91	77	128	19	0.353	0.469	0.99	3.97
22	TAM	Damon, J	5250000	38	91	17	79	152	29	7	16	73	51	92	19	0.326	0.418	-	-
23	TAM	Longoria, E	2000000	26	91	3	78	118	26	1	31	99	80	93	3	0.355	0.495	0.962	2.82
24	TAM	Upton, B	4825000	27	91	6	82	136	27	4	23	81	71	161	36	0.331	0.429	0.992	2.63
25	TAM	Fuld, S	418300	29	91	3	41	74	18	5	3	27	32	49	20	0.313	0.36	0.985	2.57
26	TAM	Rodriguez, S	428600	26	91	3	45	83	20	3	8	36	38	87	11	0.323	0.357	0.975	3.86
27	BOS	Gonzalez, A	6300000	29	90	7	108	213	45	3	27	117	74	119	1	0.41	0.548	0.997	8.88
28	BOS	Ellsbury, J	2400000	28	90	4	119	212	46	5	32	105	52	98	39	0.376	0.552	1	2.61
29	BOS	Ortiz, D	12500000	35	90	14	84	162	40	1	29	96	78	83	1	0.398	0.554	-	-
30	BOS	Pedroia, D	5750000	28	90	5	102	195	37	3	21	91	86	85	26	0.387	0.474	0.99	4.62
31	BOS	Scutaro, M	5500000	36	90	9	59	118	26	1	7	54	38	36	4	0.358	0.423	0.97	4.04
32	BOS	Youkilis, K	12250000	32	90	7	68	111	32	2	17	80	68	100	3	0.373	0.459	0.971	2.77
33	BOS	Crawford, C	14857142	30	90	9	65	129	29	7	11	56	23	104	18	0.289	0.405	0.987	1.93
34	BOS	Saltalamacchia, J	750000	26	90	4	52	84	23	3	16	56	24	119	1	0.288	0.45	0.992	7.72
35	BAL	Guerrero, V	7611455	36	69	16	60	163	30	1	13	63	17	56	2	0.317	0.416	-	-
36	BAL	Markakis, N	10600000	27	69	5	72	182	31	1	15	73	62	75	12	0.351	0.406	1	2.18
37	BAL	Jones, A	3250000	26	69	5	68	159	26	2	25	83	29	113	12	0.319	0.466	0.98	2.78
38	BAL	Hardy, J	5850000	29	69	6	76	142	27	0	30	80	31	92	0	0.31	0.491	0.99	4.88

APPENDIX F – Screenshot of Pitchers Statistics Table

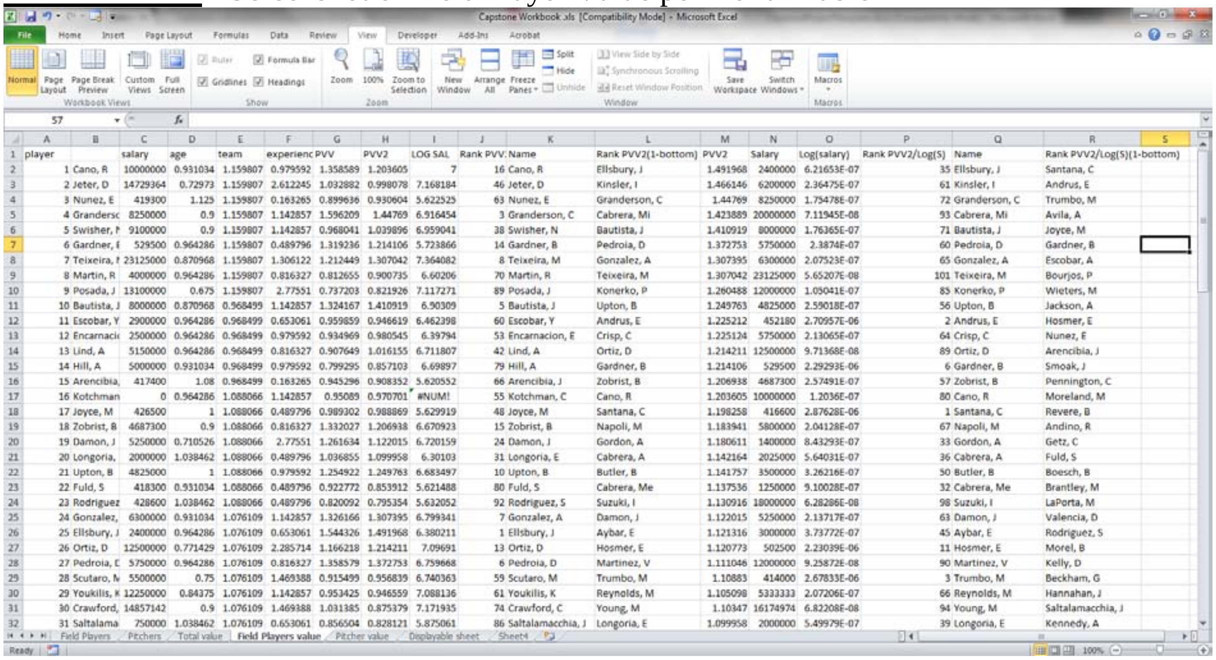
Team	Player	salary	age	team wins	years in league	innings	stike outs per inning	walks per inning	ERA	WHIP	WAR	saves
BAL	Guthrie, J	5750000	32	69	7	208	0.625	0.317307692	4.33	1.34	1.4	
BAL	Britton, Z	486000	23	69	0	154.1	0.629461389	0.402336145	4.61	1.45	0.1	
BAL	Uehara K	3000000	36	69	2	47	1.319148936	0.170212766	1.72	0.7	1.9	
BAL	Johnson J	975000	28	69	5	91	0.637362637	0.230769231	2.67	1.11	2.5	
BAL	Gonzalez M	6000000	33	69	8	46.1	0.997830803	0.390455531	4.27	1.38	0	
BAL	Gregg K	4200000	33	69	9	59.2	0.89527027	0.675675676	4.37	1.64	-0.4	22
BAL	Berken J	4265000	27	69	2	47	0.872340426	0.446808511	5.36	1.79	-0.2	
BOS	Aceves A	1200000	28	90	3	114	0.701754386	0.368421053	2.61	1.11	2.6	2
BOS	Beckett J	17000000	31	90	10	193	0.906735751	0.269430052	2.89	1.03	5.5	
BOS	Papelbon J	12000000	30	90	6	64.1	1.35725429	0.15600624	2.94	0.93	1.5	31
BOS	Bard D	505000	26	90	2	73	1.01369863	0.328767123	3.33	0.96	1.1	1
BOS	Lester J	5750000	27	90	5	191.2	0.951882845	0.392259414	3.47	1.26	4.1	
BOS	Wheeler D	3000000	34	90	12	49.1	0.794297352	0.16293279	4.38	1.11	0.2	
BOS	Albers M	875000	28	90	5	64.2	1.059190031	0.482866044	4.73	1.44	-0.1	
BOS	Wakefield T	2000000	45	90	19	154.2	0.60311284	0.304798962	5.12	1.36	-1.9	
BOS	Lackey J	15950000	33	90	9	160	0.675	0.35	6.41	1.62	-2.1	
NYY	Robertson D	460450	26	97	3	66.2	1.510574018	0.528700906	1.08	1.13	3.8	1
NYY	Rivera M	14911700	41	97	16	61.1	0.981996727	0.130932897	1.91	0.9	3.1	44
NYY	Ayala L	650000	33	97	6	56	0.696428571	0.357142857	2.09	1.27	1.4	
NYY	Sabathia C	24285714	31	97	10	237.1	0.970054829	0.257275411	3	1.23	7	
NYY	Logan B	1200000	27	97	5	41.2	1.116504854	0.315533981	3.46	1.34	0.5	
NYY	Garcia F	1500000	35	97	12	146.2	0.656634747	0.307797538	3.62	1.34	3.3	
NYY	Nova I	432900	24	97	1	165.1	0.593579649	0.345245306	3.7	1.33	3	
NYY	Colon B	900000	38	97	13	164.1	0.822669104	0.243753809	4	1.29	1.8	
NYY	Burnett A	16500000	34	97	12	190.1	0.910047344	0.436612309	5.15	1.43	0.4	
TOR	Janssen C	1095000	30	81	4	55.2	0.960144928	0.253623188	2.26	1.1	1.9	2
TOR	Romero R	1000000	27	81	2	225	0.791111111	0.355555556	2.92	1.14	6.2	
TOR	Frasor J	3500000	34	81	7	42.1	0.878859857	0.356294537	2.98	1.25	1.1	
TOR	Francisco F	4000000	32	81	6	50.2	1.055776892	0.358565737	3.55	1.32	1.1	17
TOR	Camp S	2250000	36	81	7	66.1	0.484114977	0.332829047	4.21	1.52	0.1	1
TOR	Morrow B	2300000	27	81	4	179.1	1.133445003	0.385259631	4.72	1.29	1.2	

APPENDIX G – Screenshot of Total Value per Dollar Table

player		salary	age	team	experience	PVV	PVV2	LOG SAL	Rank PVV2	Name	Rank PVV2(1-bottom)	PVV2	Salary	Log(salary)	Rank PVV2/Log(S)	Name	Rank PVV2/Log(S)(1-bottom)
1	Cano, R	10000000	0.931034	1.159807	0.979591837	1.358589	1.203605	7	43	Cano, R	Verlander J	2.112724	12850000	1.64414E-07	167	Verlander J	Robertson D
2	Jeter, D	14729364	0.72973	1.159807	2.612244898	1.032882	0.998078	7.168184	97	Jeter, D	Rivera M	1.970575	14911700	1.3215E-07	175	Rivera M	Gonzalez G
3	Nunez, E	419300	1.125	1.159807	0.163265306	0.899636	0.930604	5.622525	126	Nunez, E	Weaver J	1.836301	7365000	2.49328E-07	135	Weaver J	Tomlin J
4	Grandersc	8250000	0.9	1.159807	1.142857143	1.596209	1.44769	6.916454	15	Granderson, C	Sabathia C	1.785532	24285714	7.35219E-08	198	Sabathia C	Santana, C
5	Swisher, N	9100000	0.9	1.159807	1.142857143	0.968041	1.039896	6.959041	86	Swisher, N	Uehara K	1.660035	3000000	5.53345E-07	96	Uehara K	Ogando A
6	Gardner, E	529500	0.964286	1.159807	0.489795918	1.319236	1.214106	5.723866	40	Gardner, B	Beckett J	1.615034	17000000	9.5002E-08	193	Beckett J	Hellickson J
7	Teixeira, M	23125000	0.870968	1.159807	1.306122449	1.212449	1.307042	7.364082	24	Teixeira, M	Haren D	1.609463	12750000	1.26232E-07	180	Haren D	Santos S
8	Martin, R	4000000	0.964286	1.159807	0.816326531	0.812655	0.900735	6.60206	137	Martin, R	Romero R	1.594869	1000000	1.59487E-06	54	Romero R	Harrison M
9	Posada, J	13100000	0.675	1.159807	2.775510204	0.737203	0.821926	7.117271	165	Posada, J	Papelbon J	1.539384	12000000	1.28282E-07	179	Papelbon J	Walden J
10	Bautista, J	8000000	0.870968	0.968499	1.142857143	1.324167	1.410919	6.90309	17	Bautista, J	Shields J	1.523764	4250000	3.58533E-07	116	Shields J	Andrus, E
11	Escobar, Y	2900000	0.964286	0.968499	0.653061224	0.959859	0.946619	6.462398	120	Escobar, Y	Valverde J	1.513249	7000000	2.16178E-07	147	Valverde J	Trumbo, M
12	Encarnaci	2500000	0.964286	0.968499	0.979591837	0.934969	0.980545	6.39794	108	Encarnacion, E	Robertson D	1.505592	460450	3.26983E-06	1	Robertson D	Pineda M
13	Lind, A	5150000	0.964286	0.968499	0.816326531	0.907649	1.016155	6.711807	92	Lind, A	Ellsbury, J	1.491968	2400000	6.21653E-07	89	Ellsbury, J	Masterson J
14	Hill, A	5000000	0.931034	0.968499	0.979591837	0.799295	0.857103	6.69897	150	Hill, A	Kinsler, I	1.466146	6200000	2.36475E-07	140	Kinsler, I	Feliz N
15	Arencibia,	417400	1.08	0.968499	0.163265306	0.945296	0.908352	5.620552	132	Arencibia, J	Granderson, C	1.44769	8250000	1.75478E-07	160	Granderson, C	Pestano C
16	Kotchman	0	0.964286	1.088066	1.142857143	0.95089	0.970701	#NUM!	113	Kotchman, C	Cabrera, Mi	1.423889	20000000	7.11945E-08	199	Cabrera, Mi	Avila, A
17	Joyce, M	426500	1	1.088066	0.489795918	0.989302	0.988869	5.629919	102	Joyce, M	Bautista, J	1.410919	8000000	1.76365E-07	159	Bautista, J	Holland G
18	Zobrist, B	4687300	0.9	1.088066	0.816326531	1.332027	1.206938	6.670923	42	Zobrist, B	Wilson C	2.112724	7000000	3.01818E-07	126	Wilson C	Sale C
19	Damon, J	5250000	0.710526	1.088066	2.775510204	1.261634	1.122015	6.720159	64	Damon, J	McCarthy B	1.391399	1000000	1.3914E-06	61	McCarthy B	Nova I
20	Longoria,	2000000	1.038462	1.088066	0.489795918	1.036855	1.099958	6.30103	72	Longoria, E	Pedroia, D	1.372753	5750000	2.3874E-07	139	Pedroia, D	Alburquerque A
21	Upton, B	4825000	1	1.088066	0.979591837	1.254922	1.249763	6.683497	29	Upton, B	League B	1.362108	2250000	6.05381E-07	91	League B	Humber P
22	Fuld, S	418300	0.931034	1.088066	0.489795918	0.922772	0.853912	5.621488	152	Fuld, S	Farnworth K	1.313311	2600000	5.0512E-07	100	Farnworth K	Joyce, M
23	Rodriguez	428600	1.038462	1.088066	0.489795918	0.820092	0.795354	5.632052	170	Rodriguez, S	Gonzalez, A	1.307395	6300000	2.07523E-07	152	Gonzalez, A	Holland D
24	Gonzalez,	6300000	0.931034	1.076109	1.142857143	1.326166	1.307395	6.799341	23	Gonzalez, A	Teixeira, M	1.307042	23125000	5.65207E-08	207	Teixeira, M	Bailey A
25	Ellsbury, J	2400000	0.964286	1.076109	0.653061224	1.544326	1.491968	6.380211	13	Ellsbury, J	Lester J	1.289953	5750000	2.2434E-07	146	Lester J	Gardner, B
26	Ortiz, D	12500000	0.771429	1.076109	2.285714286	1.166218	1.214211	7.09691	39	Ortiz, D	Hernandez F	1.267324	11700000	1.08318E-07	186	Hernandez F	Escobar, A
27	Pedroia, D	5750000	0.964286	1.076109	0.816326531	1.358579	1.372753	6.759668	20	Pedroia, D	Gonzalez G	1.261844	420000	3.00439E-06	2	Gonzalez G	Bourjos, P
28	Scutaro, M	5500000	0.75	1.076109	1.469387755	0.915499	0.956839	6.740363	118	Scutaro, M	Konerko, P	1.260488	12000000	1.05041E-07	188	Konerko, P	Wieters, M
29	Youkilis, K	12250000	0.84375	1.076109	1.142857143	0.953425	0.946559	7.088136	121	Youkilis, K	Upton, B	1.249763	4825000	2.59018E-07	132	Upton, B	Jackson, A
30	Crawford,	14857142	0.9	1.076109	1.469387755	1.031385	0.875379	7.171935	143	Crawford, C	Holland G	1.24683	497150	2.50796E-06	17	Holland G	Hosmer, E
31	Saltalama	750000	1.038462	1.076109	0.653061224	0.856504	0.828121	5.875061	161	Saltalamacchia, J	Tomlin J	1.242634	417200	2.97851E-06	3	Tomlin J	Nunez, E

Field Players | Pitchers | Total value | Field Players value | Pitcher value | Displayable sheet | Sheet4

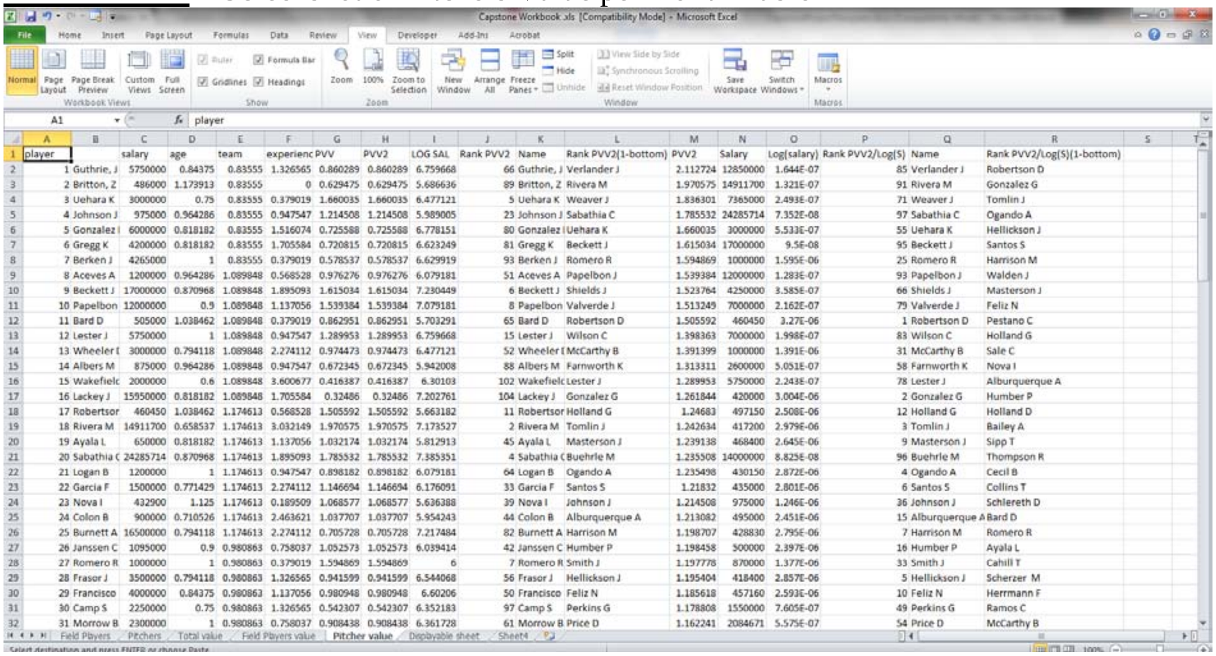
<u>APPENDIX H</u> - Screenshot of Field Player Value per Dollar Table



player		salary	age	team	experienc	PVV	PVV2	LOG SAL	Rank PVV	Name	Rank PVV2(1-bottom)	PVV2	Salary	Log(salary)	Rank PVV2/Log(S)	Name	Rank PVV2/Log(S)(1-bottom)
1	Cano, R	10000000	0.931034	1.159807	0.979592	1.358589	1.203605	7	16	Cano, R	Ellsbury, J	1.491968	2400000	6.21653E-07	35	Ellsbury, J	Santana, C
2	Jeter, D	14729364	0.72973	1.159807	2.612245	1.032882	0.998078	7.168184	46	Jeter, D	Kinsler, I	1.466146	6200000	2.36475E-07	61	Kinsler, I	Andrus, E
3	Nunez, E	419300	1.125	1.159807	0.163265	0.899636	0.930604	5.622525	63	Nunez, E	Granderson, C	1.44769	8250000	1.75478E-07	72	Granderson, C	Trumbo, M
4	Grandersc	8250000	0.9	1.159807	1.142857	1.596209	1.44769	6.916454	3	Granderson, C	Cabrera, Mi	1.423889	20000000	7.11945E-08	93	Cabrera, Mi	Avila, A
5	Swisher, N	9100000	0.9	1.159807	1.142857	0.968041	1.039896	6.959041	38	Swisher, N	Bautista, J	1.410919	8000000	1.76365E-07	71	Bautista, J	Joyce, M
6	Gardner, E	529500	0.964286	1.159807	0.489796	1.319236	1.214106	5.723866	14	Gardner, B	Pedroia, D	1.372753	5750000	2.3874E-07	60	Pedroia, D	Gardner, B
7	Teixeira, M	23125000	0.870968	1.159807	1.306122	1.212449	1.307042	7.364082	8	Teixeira, M	Gonzalez, A	1.307395	6300000	2.07523E-07	65	Gonzalez, A	Escobar, A
8	Martin, R	4000000	0.964286	1.159807	0.816327	0.812655	0.900735	6.60206	70	Martin, R	Teixeira, M	1.307042	23125000	5.65207E-08	101	Teixeira, M	Bourjos, P
9	Posada, J	13100000	0.675	1.159807	2.77551	0.737203	0.821926	7.117271	89	Posada, J	Konerko, P	1.260488	12000000	1.05041E-07	85	Konerko, P	Wieters, M
10	Bautista, J	8000000	0.870968	0.968499	1.142857	1.324167	1.410919	6.90309	5	Bautista, J	Upton, B	1.249763	4825000	2.59018E-07	56	Upton, B	Jackson, A
11	Escobar, Y	2900000	0.964286	0.968499	0.653061	0.959859	0.946619	6.462398	60	Escobar, Y	Andrus, E	1.225212	452180	2.70957E-06	2	Andrus, E	Hosmer, E
12	Encarnacio	2500000	0.964286	0.968499	0.979592	0.934969	0.980545	6.39794	53	Encarnacion, E	Crisp, C	1.225124	5750000	2.13065E-07	64	Crisp, C	Nunez, E
13	Lind, A	5150000	0.964286	0.968499	0.816327	0.907649	1.016155	6.711807	42	Lind, A	Ortiz, D	1.214211	12500000	9.71368E-08	89	Ortiz, D	Arencibia, J
14	Hill, A	5000000	0.931034	0.968499	0.979592	0.799295	0.857103	6.69897	79	Hill, A	Gardner, B	1.214106	529500	2.29293E-06	6	Gardner, B	Smoak, J
15	Arencibia,	417400	1.08	0.968499	0.163265	0.945296	0.908352	5.620552	66	Arencibia, J	Zobrist, B	1.206938	4687300	2.57491E-07	57	Zobrist, B	Pennington, C
16	Kotchman	0	0.964286	1.088066	1.142857	0.95089	0.970701	#NUM!	55	Kotchman, C	Cano, R	1.203605	10000000	1.2036E-07	80	Cano, R	Moreland, M
17	Joyce, M	426500	1	1.088066	0.489796	0.989302	0.988869	5.629919	48	Joyce, M	Santana, C	1.198258	416600	2.87628E-06	1	Santana, C	Revere, B
18	Zobrist, B	4687300	0.9	1.088066	0.816327	1.332027	1.206938	6.670923	15	Zobrist, B	Napoli, M	1.183941	5800000	2.04128E-07	67	Napoli, M	Andino, R
19	Damon, J	5250000	0.710526	1.088066	2.77551	1.261634	1.122015	6.720159	24	Damon, J	Gordon, A	1.180611	1400000	8.43293E-07	33	Gordon, A	Getz, C
20	Longoria,	2000000	1.038462	1.088066	0.489796	1.036855	1.099958	6.30103	31	Longoria, E	Cabrera, A	1.142164	2025000	5.64031E-07	36	Cabrera, A	Fuld, S
21	Upton, B	4825000	1	1.088066	0.979592	1.254922	1.249763	6.683497	10	Upton, B	Butler, B	1.141757	3500000	3.26216E-07	50	Butler, B	Boesch, B
22	Fuld, S	418300	0.931034	1.088066	0.489796	0.922772	0.853912	5.621488	80	Fuld, S	Cabrera, Me	1.137536	1250000	9.10028E-07	32	Cabrera, Me	Brantley, M
23	Rodriguez	428600	1.038462	1.088066	0.489796	0.820092	0.795354	5.632052	92	Rodriguez, S	Suzuki, I	1.130916	18000000	6.28286E-08	98	Suzuki, I	LaPorta, M
24	Gonzalez,	6300000	0.931034	1.076109	1.142857	1.326166	1.307395	6.799341	7	Gonzalez, A	Damon, J	1.122015	5250000	2.13717E-07	63	Damon, J	Valencia, D
25	Ellsbury, J	2400000	0.964286	1.076109	0.653061	1.544326	1.491968	6.380211	1	Ellsbury, J	Aybar, E	1.121316	3000000	3.73772E-07	45	Aybar, E	Rodriguez, S
26	Ortiz, D	12500000	0.771429	1.076109	2.285714	1.166218	1.214211	7.09691	13	Ortiz, D	Hosmer, E	1.120773	502500	2.23039E-06	11	Hosmer, E	Morel, B
27	Pedroia, D	5750000	0.964286	1.076109	0.816327	1.358579	1.372753	6.759668	6	Pedroia, D	Martinez, V	1.111046	12000000	9.25872E-08	90	Martinez, V	Kelly, D
28	Scutaro, M	5500000	0.75	1.076109	1.469388	0.915499	0.956839	6.740363	59	Scutaro, M	Trumbo, M	1.10883	414000	2.67833E-06	3	Trumbo, M	Beckham, G
29	Youkilis, K	12250000	0.84375	1.076109	1.142857	0.953425	0.946559	7.088136	61	Youkilis, K	Reynolds, M	1.105098	5333333	2.07206E-07	66	Reynolds, M	Hannahan, J
30	Crawford,	14857142	0.9	1.076109	1.469388	1.031385	0.875379	7.171935	74	Crawford, C	Young, M	1.10347	16174974	6.82208E-08	94	Young, M	Saltalamacchia, J
31	Saltalama	750000	1.038462	1.076109	0.653061	0.856504	0.828121	5.875061	86	Saltalamacchia, J	Longoria, E	1.099958	2000000	5.49979E-07	39	Longoria, E	Kennedy, A

Who Is The American League MVP On a Per Dollar Basis?
Senior Capstone Project for Patrick Decker

<u>APPENDIX I</u> – Screenshot of Pitchers Value per Dollar Table



player		salary	age	team	experienc	PVV	PVV2	LOG SAL	Rank PVV2	Name	Rank PVV2(1-bottom)	PVV2	Salary	Log(salary)	Rank PVV2/Log(S)	Name	Rank PVV2/Log(S)(1-bottom)
1	Guthrie, J	5750000	0.84375	0.83555	1.326565	0.860289	0.860289	6.759668	66	Guthrie, J	Verlander J	2.112724	12850000	1.644E-07	85	Verlander J	Robertson D
2	Britton, Z	486000	1.173913	0.83555	0	0.629475	0.629475	5.686636	89	Britton, Z	Rivera M	1.970575	14911700	1.321E-07	91	Rivera M	Gonzalez G
3	Uehara K	3000000	0.75	0.83555	0.379019	1.660035	1.660035	6.477121	5	Uehara K	Weaver J	1.836301	7365000	2.493E-07	71	Weaver J	Tomlin J
4	Johnson J	975000	0.964286	0.83555	0.947547	1.214508	1.214508	5.989005	23	Johnson J	Sabathia C	1.785532	24285714	7.352E-08	97	Sabathia C	Ogando A
5	Gonzalez I	6000000	0.818182	0.83555	1.516074	0.725588	0.725588	6.778151	80	Gonzalez I	Uehara K	1.660035	3000000	5.533E-07	55	Uehara K	Hellickson J
6	Gregg K	4200000	0.818182	0.83555	1.705584	0.720815	0.720815	6.623249	81	Gregg K	Beckett J	1.615034	17000000	9.5E-08	95	Beckett J	Santos S
7	Berken J	4265000	1	0.83555	0.379019	0.578537	0.578537	6.629919	93	Berken J	Romero R	1.594869	1000000	1.595E-06	25	Romero R	Harrison M
8	Aceves A	1200000	0.964286	1.089848	0.568528	0.976276	0.976276	6.079181	51	Aceves A	Papelbon J	1.539384	12000000	1.283E-07	93	Papelbon J	Walden J
9	Beckett J	17000000	0.870968	1.089848	1.895093	1.615034	1.615034	7.230449	6	Beckett J	Shields J	1.523764	4250000	3.585E-07	66	Shields J	Masterson J
10	Papelbon	12000000	0.9	1.089848	1.137056	1.539384	1.539384	7.079181	8	Papelbon	Valverde J	1.513249	7000000	2.162E-07	79	Valverde J	Feliz N
11	Bard D	505000	1.038462	1.089848	0.379019	0.862951	0.862951	5.703291	65	Bard D	Robertson D	1.505592	460450	3.27E-06	1	Robertson D	Pestano C
12	Lester J	5750000	1	1.089848	0.947547	1.289953	1.289953	6.759668	15	Lester J	Wilson C	1.398363	7000000	1.998E-07	83	Wilson C	Holland G
13	Wheeler I	3000000	0.794118	1.089848	2.274112	0.974473	0.974473	6.477121	52	Wheeler I	McCarthy B	1.391399	1000000	1.391E-06	31	McCarthy B	Sale C
14	Albers M	875000	0.964286	1.089848	0.947547	0.672345	0.672345	5.942008	88	Albers M	Farnworth K	1.313311	2600000	5.051E-07	58	Farnworth K	Nova I
15	Wakefielc	2000000	0.6	1.089848	3.600677	0.416387	0.416387	6.30103	102	Wakefielc	Lester J	1.289953	5750000	2.243E-07	78	Lester J	Alburquerque A
16	Lackey J	15950000	0.818182	1.089848	1.705584	0.32486	0.32486	7.202761	104	Lackey J	Gonzalez G	1.261844	420000	3.004E-06	2	Gonzalez G	Humber P
17	Robertsor	460450	1.038462	1.174613	0.568528	1.505592	1.505592	5.663182	11	Robertsor	Holland G	1.24683	497150	2.508E-06	12	Holland G	Holland D
18	Rivera M	14911700	0.658537	1.174613	3.032149	1.970575	1.970575	7.173527	2	Rivera M	Tomlin J	1.242634	417200	2.979E-06	3	Tomlin J	Bailey A
19	Ayala L	650000	0.818182	1.174613	1.137056	1.032174	1.032174	5.812913	45	Ayala L	Masterson J	1.239138	468400	2.645E-06	9	Masterson J	Sipp T
20	Sabathia C	24285714	0.870968	1.174613	1.895093	1.785532	1.785532	7.385351	4	Sabathia C	Buehrle M	1.235508	14000000	8.825E-08	96	Buehrle M	Thompson R
21	Logan B	1200000	1	1.174613	0.947547	0.898182	0.898182	6.079181	64	Logan B	Ogando A	1.235498	430150	2.872E-06	4	Ogando A	Cecil B
22	Garcia F	1500000	0.771429	1.174613	2.274112	1.146694	1.146694	6.176091	33	Garcia F	Santos S	1.21832	435000	2.801E-06	6	Santos S	Collins T
23	Nova I	432900	1.125	1.174613	0.189509	1.068577	1.068577	5.636388	39	Nova I	Johnson J	1.214508	975000	1.246E-06	36	Johnson J	Schlereth D
24	Colon B	900000	0.710526	1.174613	2.463621	1.037707	1.037707	5.954243	44	Colon B	Alburquerque A	1.213082	495000	2.451E-06	15	Alburquerque A	Bard D
25	Burnett A	16500000	0.794118	1.174613	2.274112	0.705728	0.705728	7.217484	82	Burnett A	Harrison M	1.198707	428830	2.795E-06	7	Harrison M	Romero R
26	Janssen C	1095000	0.9	0.980863	0.758037	1.052573	1.052573	6.039414	42	Janssen C	Humber P	1.198458	500000	2.397E-06	16	Humber P	Ayala L
27	Romero R	1000000	1	0.980863	0.379019	1.594869	1.594869	6	7	Romero R	Smith J	1.197778	870000	1.377E-06	33	Smith J	Cahill T
28	Frasor J	3500000	0.794118	0.980863	1.326565	0.941599	0.941599	6.544068	56	Frasor J	Hellickson J	1.195404	418400	2.857E-06	5	Hellickson J	Scherzer M
29	Francisco	4000000	0.84375	0.980863	1.137056	0.980948	0.980948	6.60206	50	Francisco	Feliz N	1.185618	457160	2.593E-06	10	Feliz N	Herrmann F
30	Camp S	2250000	0.75	0.980863	1.326565	0.542307	0.542307	6.352183	97	Camp S	Perkins G	1.178808	1550000	7.605E-07	49	Perkins G	Ramos C
31	Morrow B	2300000	1	0.980863	0.758037	0.908438	0.908438	6.361728	61	Morrow B	Price D	1.162241	2084671	5.575E-07	54	Price D	McCarthy B

REFERENCES

Beneventano, Philip, Paul Berger, and Bruce Weinberg. "Predicting Run Production and Run Prevention in Baseball: The Impact of Sabermetrics."*International Journal of Business, Humanities, and Technology*. 2.4 (2012): n. page. Web. 25 Apr. 2013. <http://ijbhtnet.com/journals/Vol_2_No_4_June_2012/7.pdf>.

ESPN (2011): n.pag. *ESPN*. Web. 12 Nov 2012. <espn.go.com>.

Frankel, Matthew. "Secret Sabermetrics: Trade Secret Protection in the Baseball Analytics Field.." *Albany Government Law Review of 2012*. 5.1 (2012): n. page. Web. 25 Apr. 2013. <http://www.albanygovernmentlawreview.org/files/Volume_5_Issue_1/9b_Frankel_website.pdf>.

James, Bill. *The Bill James Historical Baseball Abstract*. 2nd. Villard Books, 1986. Print.
Kaufmann, Eric. "Moneyball: Can Sports Statistics Save Political Studies?." *Political Insight*. N.p., n.d. Web. 12 Dec 2012. <http://www.sneps.net/wp-content/uploads/2011/12/1-moneyball.pdf>.

Lewis, Micheal. *Moneyball: The Art of Winning an Unfair Game*. New York: W.W. Norton &Company Inc., 2003. Print.

Longley, Neil, and Glenn Wong. "The speed of human capital formation in the baseball industry: the information value of minor-league performance in predicting major-league performance." *Managerial & Decision Economics*. 3. (2011): n. page. Web. 25 Apr. 2013. <http://0-web.ebscohost.com.helin.uri.edu/ehost/detail?sid=01f35332-11f9-489e-bd1d-a7519b3b3f36@sessionmgr10&vid=1&hid=10&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ>.

SABR Encyclopedia. 2009. <http://sabrpedia.org/wiki/Main_Page>.

"The SABR story." *Society for American Baseball Research*. SABR, n.d. Web. 10 Dec 2012. <http://sabr.org/about>.

Shaheen, Anna, Brent Shaheen, and Estes Shaheen. "Determinants of Value and Productivity in a Complex Labor Market: How Sabermetrics and Statistical Innovation Changed the Business of Professional Baseball." *Business Studies Journal*. 2.1 (2010): n. page. Web. 25 Apr. 2013. <http://www.alliedacademies.org/Publications/Papers/BSJ Vol 2 SI No 1 2010 p 27-48.pdf>.

Verducci, Tom. "The Art of Winning an (even more) Unfair Game." *Sports Illustrated*. 115.13. 26 sept 2011: 38-49. Print.

Verducci, Tom. "Fall of the Red Sox." *Sports Illustrated*. 117.10 (2012): 38. Web. 12 Dec. 2012. <http://0-web.ebscohost.com.helin.uri.edu/ehost/detail?sid=541b88e1-0a8a-40d8-9811-b72383ca72f1@sessionmgr10&vid=1&hid=10&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ

"2011 MLB Salaries." *USATODAY Salary Database* n.pag.*MLB Salary Database*. Web. 12 Oct 2012. <content.usatoday.com/sportsdata/baseball/mlb/salaries/team>.