Which Baseball Statistic Is the Most Important When Determining Team Success?

Adam Houser

I. INTRODUCTION

The way to measure individual productivity in any working environment is to track individual performance in a working atmosphere. Individuals who work on a commission-based salary have more incentive to work harder, thus more efficiently. The best industry which monitors productivity is clearly the game of baseball, due to its uncanny ability to measure productivity through the countless statistics that are available. Using these statistics, owners can use information to determine what each player should be paid for his services. With the advent of free agency in 1974, statistics became vital information in determining individual salary, with the free agent market clearly setting what each player should be paid for his given statistics.

In recent history however, a disturbing trend has been established within the confines of Major League Baseball (MLB), which states that the rich get richer and the poor get poorer. This is important because there are only a handful of teams which can financially support the best players. If the majority of teams in MLB cannot afford the best players, they cannot expect to contend for a World Series.

Within the past 5 years, there have been some encouraging developments to renounce the trend. Oakland, Anaheim, Minnesota, San Francisco, and Florida have advanced into post-season play despite not being big spenders (Lewis, 2003). Most notably, these teams have not acquired talent from the categories of what history says is statistically significant. Instead of paying top-dollar for home runs and strikeouts, they have found individual bargains who maintain high on-base-percentage (OBP) who have slipped though the cracks of the market. With so many different theories amongst player personnel, it is easy to get confused about which statistics matter the most. This study will attempt to settle the classic argument of which baseball statistic is the best measure of team success.

"With the advent of free agency in 1974, statistics became vital information in determining individual salary."

II. THEORY

The underlying economic theory within the game of baseball is undoubtedly the human capital theory, which states that players will be paid for their productivity; fortunately for baseball productivity can be readily measured by the plethora of performance statistics available on the internet. According to common baseball knowledge, higher than average statistics at both the individual and the team level should bolster the amount of wins each team has annually. According to Gary Becker (1975), the idea of the human capital model of investment is that increases in human capital, which are measured by statistics, lead to an increase in productivity of the team, which leads to an increase in wages.

Theoretically, the team that wins the most games should have the best players. Since they have the best players, they should have the best statistics. Better productivity measures (statistics) should lead to an increase in player wages. Hence, the best team should have the highest payroll amount. However, previous literature (Blass, 1992) has shown that theoretically the human capital model of investment does not hold in baseball. This is due to the nature of baseball's collective bargaining agreement and arbitration process. Generally speaking, Blass (1992) found that players are underpaid when they are younger and overpaid when they are older, which is a violation of the human capital theory.

The profit-maximizing firm will attempt to field a budget in an effort that receipts are equivalent to expenditures. However, this notion is skewed when taking into account the effects of on-the-job-training. Becker (1975) has successfully claimed that firms will increase their on-the-job-training in an effort to either lower future expenditures or increase future receipts of the firm. The type of training offered by firms can either be specific or general. General training increases marginal productivity of trainees by exactly the same amount in the firms providing the training as in other firms. In specific training, the marginal productivity of the individual is increased specifically for that firm. The type of training offered in baseball is general training. Baseball skills are not team specific; what works for one team will most likely work for another. Firms invest in training in an effort to increase human capital, which could potentially lead to an increase in productivity in the worker, which would ultimately bring the firm more revenue.

"Players are underpaid when they are younger and overpaid when they are older."

In relation to the topic at hand, teams in Major League Baseball gladly spend top-dollar for the best players in an effort to increase their ability to win baseball games. It should also be noted that this increase in firm expenditures can be offset by an increase in receipts. If the player is successful and helps the team win more games, attendance at home games will increase and thus, profitability of the firm increases and the increase in firm expenditures is offset. Thus, in order to field the best team possible, teams should invest heavily in training; which in this case is player productivity, if they want to win as many games as possible.

This theory sounds very elitist, and it suggests that the teams from smaller market cities cannot compete with the likes of the Yankees, Red Sox, Dodgers, and Cubs among others. In fact, Zimbalist (2003) did research on this topic and concluded that payroll size is the biggest factor of teams in the playoffs. As it was previously noted however, there are several examples in the past few years in which teams made it to the post-season without fielding a huge payroll (Lewis, 2003).

To complicate matters even more, the vast majority of teams in major league baseball cannot afford the best free-agents because of the fact that their fan base is limited; hence their ability to increase attendance is marginal at best. These teams are un-willing to increase their expenditures in an effort to increase the success of their teams. Rather they have found loop-holes in the economic set-up of major league baseball, in which several sets of skills are under-valued in the market for Major League Baseball players. The Oakland A's and several other teams have found great success by fielding a competitive team by stressing on-base-percentage and keeping payroll relatively constant. Later on in this research paper, I plan to determine if Michael Lewis' (2003) claim that n-base-percentage is indeed the best statistic in measuring the team success, in this case number of wins.

III. REVIEW OF LITERATURE

The inspiration I had for under-taking this research project occurred this past summer, when I read Lewis's (2003) document entitled *Moneyball*. This book laid the groundwork for which I plan to proceed with my investigations. This book followed the Oakland A's, a team able to stay extremely competitive while having a payroll a fraction of their competitors. The way they remained competitive was searching for players with high on-base-percentage (OBP), which the market for ballplayers undervalued. In fact, the book suggested that "the market for baseball players was so inefficient, and the general grasp of sound baseball strategy so weak, that superior management could still run circles around taller piles of cash." (Lewis, 2003) This quote was not from a baseball front office executive, rather from the mouth of Paul Volcker, who led an investigation into the economic situation of professional sports for the US government in 1999.

Scully's (1989) results stated that OPS is the best statistic in determining winning percentage. However, since 1989 many aspects of the game have changed, most notably the expansionary measures taken within the game of baseball. For example, homeruns have increased at an exponential rate, payrolls have increased massively and the disparity between payrolls has been ever increasing. Many aspects have changed, but has theory tested 15 years ago remained the same until present day?

Cook's *Percentage Baseball* (1964) was far ahead of it's time when it was published. The author used economic regressions to question several strategies: such as the sacrifice bunt, attempt to steal a base, when to change pitchers, etc. The results were in a direct contrast with those mentioned in Lewis' (2003) *Moneyball.* Cook concluded that the sacrifice bunt and stolen base were strategic necessities for managers in ballgames. Lewis (2003) argued 40 years later that the sacrifice bunt and attempt to steal a base significantly lower the amount of runs teams are predicted to score.

This research project will attempt to answer the age-old debate and hopefully shed new knowledge as to which single statistic is the best when determining team wins.

IV. DATA

This study will test the following hypotheses: Teams with higher win totals will have better offensive statistics. Teams with better offensive statistics will have higher payroll figures. OBP will be the best statistic in determining team wins.

The reason why Lewis (2003) and the Oakland A's like OBP so much is because of the fact that OBP includes walks, whereas the more common batting average does not. The ability to get on base significantly increases the projected number of runs the team will score per game (www.baseballprospectus.com). Since batting average overlooks this simple fact, many have argued successfully that batting average is the most over-used statistic in all of major league baseball

To successfully test my hypotheses, data will be compiled of team statistics for every team in baseball from the years 2000-04. This data is readily available at www.mlb.com. I am testing the on-base-percentage in an effort to determine whether or not Lewis' claim is valid. The dependent variable in this equation will be the number of wins that the team has in the corresponding year.

I have excluded earned-run-average, a very common pitching statistic, from my regression because of the fact that it accounts for errors in the equation. The ruling for whether or not a ball in play is an error or a base hit is subject to one person's opinion, and can vary heavily from team-to-team. Also, faster players may reach a ball that a normal player would not make, and still make an error. Whereas the slower player would not have come close to the ball, so it would have been a base hit. An example that ERA is not a very good statistic comes this past baseball season by the Boston Red Sox. Between the months of May and July, Boston's record was merely 41-40 despite leading the major leagues in ERA and most offensive statistics. The reason that Boston did not have a very good record was because of their atrocious defense they displayed every night that usually led to a few unearned runs per game.

TABLE 1
Empirical Model 1

Dependent Variable		
Wins		Number of team wins annually
Explanatory Variables		
BA	(+)	Batting average (hits/at bats)
OBP	(+)	On-Base-Percentage [(hits+ walks) / at bats]
SLG	(+)	Slugging Percentage [(hits + 2B + {3B*2} + {Home Runs*3}) / AB]
STL	(+)	Percentage of successful stolen base attempts per team
WHIP	(-)	(Walks + Hits) / Innings Pitched
K per 9	(+)	Strikeouts per 9 Innings Pitched (Total strikeouts / innings pitched)
Field	(+)	Fielding Perc entage
Payroll	(+)	Team budgeted payroll amount

V. EMPIRICAL MODEL

Within the design of my research, I plan to use a simple OLS regression to find the results of my study. I will use multiple regressions to test my multiple hypotheses because of the fact of the very likely probability I will incur omitted variable bias issues if I do not run multiple regressions. The data was taken from a time-series panel over the last 5 years. Tables 1 and 2 will be the format for my OLS regressions. The parentheses indicate the predicted signs for each

TABLE 2
Empirical Model 2

Dependent Variable		
Payroll		Team budged payroll amount
Explanatory Variables		
BA	(+)	Batting average (hits/ at bats)
OBP	(+)	On-Base-Percentage [(hits + walks) / at bats]
SLG	(+)	Slugging Percentage [(hits + 2B + {3B*2} + {Home Runs*3}) / AB]
STL	(+)	Percentage of successful stolen base attempts per team
WHIP	(-)	Walks + Hits / Innings Pitched
K per 9	(+)	Strikeouts per 9 Innings Pitched (T otal strikeouts / innings pitched)
Field	(+)	Fielding Percentage

explanatory variable.

I will also run a 2nd regression to test the 2nd piece of my hypothesis. All statistics used in this regression are exactly the same as the above regression. This model will give accurate figures for what high revenue teams pay for.

I will also run a final test to see if I can determine which single team statistic is the best when attempting to explain team wins. To accomplish this, I plan on testing every explanatory variable individually against team wins. The baseball statistic with the highest concentration of coefficients, T-statistics, and significance values will be the variable that I will conclude to be the best when determining team wins. The process had in determining which statistics I would test was rather simple. I came across an ESPN.com article which a writer determined to be the most important statistics in baseball. For my research, I used some, but not all of the listed stats.

A. Offensive Statistics

Previous research indicates that offensive statistics are by far the most effective way of empirically testing baseball statistics. Hence, that is why this regression includes mostly offensive stats. I chose the listed stats because of the fact that they are all extremely measurable and because of the fact that previous literature has told me that they are important in not only team success, but when testing empirical research as well.

B. Pitching Statistics

In order to measure pitching, the statistics of WHIP and Strikeouts per nine innings have been chosen. I wanted to include stats such as ERA and Saves, but I felt that they would not be good fits in the regression. As was previously mentioned, ERA is not a good measure because of the fact it is so subject to errors and WHIP is not. And since the process of ruling a play a hit or an error is very arbitrary, I did not include it in the regression. Although I anticipate that ERA would be statistically significant at the .01 level, it would be biased and therefore, not a good fit. I could have included Saves, but they are subject to amount of wins. A team cannot get a Save if it does not win the game. Again, Saves would in all likelihood be statistically significant at the .01 level of output, but they too would be biased.

C. Defensive Statistics

In this regression I included fielding percentage in an effort to test if fielding percentage is important in determining number of wins. Fielding percentage takes some measure of errors into the equation, so hopefully there will be some significance. However, as was previously stated the process is completely arbitrary. A faster player may reach a fly ball that a slower player would not, and get penalized with an error whereas a slower player would not reach the ball and it would be ruled a base hit. Because of this fact, defensive statistics are not measured with as much certainty as offensive statistics are. There is no concrete, undisputable way to measure defensive productivity as there exists with offense. Even the players judged to be the best defensive players annually are subject to sports writer's opinions, which can vary from person to person. Although I do not expect fielding percentage to be statistically significant, perhaps this test will at least show if it is considered important at all.

IV. RESULTS

The first regression will test significance for statistics in determining number of team wins annually. The results are a direct result from the model in Table 1 above. For independent variables, the unstandardized coefficients, t-statistics, and signifi-

TABLE 3
Regression Results: Predicting Wins

Explanatory Variable	â	Std. Error	*t*	Sig.
(Constant)	-174.417	167.798	-1.039	0.3
BA	-94.887	76.421	-1.242	0.216
OBP	403.729	62.577	6.452	0
SLG	139.462	30.517	4.57	0
STL	0.062	0.076	0.808	0.421
WHIP	-92.798	5.87	-15.808	0
K per 9	-0.908	0.731	-1.241	0.217
Field	221.541	171.919	1.289	0.2
Payroll	n/a	n/a	0.054	0.957

R^2 = .849 n = 150

cance values are presented for the first regression below in Table 3.

Within this regression, there are three significant variables at the .01 level; on-base-percentage, slugging percentage, and WHIP. I am not surprised at all that these three are considered to be statistically significant because of the fact that they are all on the Gammons' list on ESPN.com of the most important statistics (2004). Also, previous literature told me that these three would be among the most statistically significant (Sommers, 1992).

It came as a surprise to me that batting average, the most common statistic in all of baseball, had a negative relationship with wins when I predicted it would have a positive one. Theoretically, it makes sense that teams with higher batting averages should have higher win totals, but this regression points to other conclusions. Also, strikeouts per nine innings had a different sign than I predicted. The more strikeouts a team's pitching staff has, the fewer amount of balls that are put into play. If fewer balls are put into play, then it should lower the expected run total for the batting team. If expected run totals for the opposing team are lowered, then one can assume that strikeouts should help teams win more games, as opposed to losing more as my results suggest.

The most interesting finding of my first regression however, is the fact that payroll has absolutely no relationship at all to wins, which contradicts my earlier research findings and it also contradicts the recent history of teams making the playoffs in major league baseball. One possible explanation is the fact that other variables are simply more important, therefore a statistic like payroll could potentially be pushed to the back in terms of significance. However, later on when I test each statistic individually against wins I think I will get a better gauge of just how significant payroll is.

The second regression test will test my wsecond hypothesis, which stated that teams with better offensive statistics will have higher payroll figures. The model for this is listed in Table 2. The results are in the same format as Table 3 and are below.

This regression only has two significant values: slugging percentage at the .05 level and strikeouts per nine innings at the .01 level. It does not surprise me that those two statistics are the only statistically significant variables because variables can be described as "sexy", and appeal to fans very much. Teams with high slugging percentages have extraordinary amounts of extra-base hits, which include doubles, triples, and home runs. Because the average baseball fan is fascinated by home runs in this current era of baseball, the demand for players who can hit home runs has increased exponentially within the last 15 years. Because the demand for these players has increased, it makes sense that their salaries have increased as well. If those players who have the ability to hit homeruns and extra base hits are grouped together on one team, theory states that the payroll of that team must be high; therefore one can conclude that money buys extra base hits and home runs.

TABLE 4
Regression Results: Pre dicting Payroll Factor

Explanatory Variables	â	Std. Error	*t*	Sig.
(Constant)	500268775.63	787317572.62	0.635	0.526
BA	-278249287.9	358323736.35	-0.777	0.439
OBP	15437463.69	294027844.27	0.053	0.958
SLG	353576689.88	140287961.61	2.52	0.013
STL	215789.271	357458.957	0.604	0.547
WHIP	-88916811.67	26554794.19	1.941	0.054
K per 9	6581810.157	3391805.723	-3.348	0.001
Field	-458054278.4	806884175.00	-0.568	0.571

R2 = .241 n = 150

Strikeouts per nine innings were also shown to be even more statistically significant than slugging percentage, which comes as a small surprise to me. Although strikeouts are certainly entertaining and to the author, there is nothing better than watching a classic pitcher's duel, I would have expected the owners of teams to place a higher premium on home runs than on strikeouts.

Like power hitters, power pitchers have seen their salaries increase at a faster than normal rate in recent history. Players entering free agency with the ability to strikeout numerous hitters are usually rewarded more handsomely than a pitcher that relies on change of speeds and deception. For example, last February when Greg Maddux was a free agent, he signed with the Cubs for an average of $8 million per season. However, Kerry Wood was also entering free agency at the same time and the Cubs resigned him for an average of $10.5 million per season. There are many reasons for this disparity between the two, but when comparing strikeouts among the two, Wood averages far more strikeouts than Maddux. However, Maddux has won a Major League Baseball record at least 15 baseball games as a pitcher for the last 16 years. Wood has failed to win 15 games once in his career. It seems as money was not spent wisely in this case.

An interesting result of this study so far finds that Tables 3 and 4 do not match: i.e. the owners are overpaying for the wrong players for the wrong statistics. A possible explanation for this come from simple economics: if the owners are profit-maximizers, they will field a team that can sell out the stadium everyday. Table 4 clearly states that owners should maximize home runs and strikeouts. Fans will enjoy the excitement that this team brings, and will ultimately bring the owner a greater source of revenue.

However, if owners are profit-maximizers, they are not necessarily going to field a team that has the best chance of winning a championship, which should be the goal of every owner in every sport. If the goal of an owner isn't to win a championship, then they are in the wrong business. Most likely, the team that doesn't win the maximum amount of games possible will eventually suffer a decline in attendance. Previous research the author has done has revealed that team success is correlated with attendance at the highest level of output (Houser, 2004). According to that theory, owners should field a team that wins the more games and the fans will come through the gates. It is most likely going to be more cost-effective to field a team with a lower payroll that wins more games in the long run.

For the following regression each explanatory variable was tested independently against team wins, in an effort to determine which statistic is the most statistically significant when determining the amount of wins per team.

The above table was not a complete regression, hence excluded is any values for the constant because each of the explanatory variables was tested independently. The variables with the highest absolute value t-statistics are deemed to be the most statistically significant. Under the above table, WHIP, OBP, and SLG are the most important variables. These same variables also have the highest listed R^2 values in the regressions. I do not think that it is a coincidence that these are the same exact variables that Table 3 listed as statistically significant. Rather, they are the most important statistic when determining team wins.

TABLE 5
Regression Results: Best Statistic

Explanatory Variables	**â**	**Std. Error**	***t***	**Sig.**	**R^2**
BA	591.768	82.312	7.189	0	0.259
OBP	605.572	59.435	10.189	0	0.412
SLG	285.704	33.741	8.468	0	0.326
STL	0.466	0.181	2.572	0.011	0.043
WHIP	-97.222	8.49	-11.451	0	0.47
K per 9	7.092	1.417	5.004	0	0.145
Field	1931	344.736	5.601	0	0.175
Payroll	n/a	n/a	5.285	0	0.159

In relation to my hypothesis, as was stated earlier, I stated that I thought that OBP would be the most important of all statistics tested. However, Table 5 clearly states that WHIP is more important than on-base percentage. I think it is extremely interesting that the two most statistically significant variables are both statistics that record number of base runners per game and number of base runners allowed per game. With the vast amount of statistics available and tested, I can conclude that in order to win more games, all you have to do is have more base runners than your opponent. This statement may seem extremely obvious, but with the recent fascination with home runs, it appears that home runs are not as important as current markets say they are. It is satisfying to the author that at its core, the game of baseball remains very simple: get base runners on base and you stand a great chance of winning the game. Since slugging percentage is not as statistically significant, teams that do not have the capability to spend large amounts of money can still remain competitive if they invest heavily in OBP, which fortunately for them is still extremely undervalued in the market for baseball players.

Even Paul Volcker's statement that the market for baseball players is so weak and incorrect is proven to be correct by this theory. Empirical tests have shown that teams with high payroll's continually overpay for slugging percentage when in fact Lewis' claim that on base percentage is a better way to win baseball games. Hence, the theory that the Oakland A's have taken in the last decade or so is the most efficient theory to date. According to Table 4, which measured payroll against statistics, OBP has a significance value of only .958, which suggests that teams that spend heavily spend their money in all the wrong ways.

VII. CONCLUSIONS

My findings show that the best way to field a team is by investing heavily in WHIP and on base percentage. Although Major League Baseball is not likely to expand to any new markets anytime soon, when they do it will be interesting to see how decision makers build their personnel on those teams. Recently, Tampa Bay was awarded a franchise and they invested extremely heavily in home runs and slugging percentage as opposed to pitching and on base percentage. As a result, their record plummeted and their fan base dwindled. Within the past couple of years, their payrolls have been the lowest in baseball by far (www.usatoday.com/mlb/payroll).

In 2003, Alex Rodriguez made more money annually than every player of the Tampa Bay Devil Rays combined (Lewis, 2003). This philosophy error cost Tampa Bay a chance at being competitive for at least 10 years. Now Tampa Bay is focusing their drafting efforts and free agent signings through pitching, defense, and on base percentage. It may take several years, but eventually I would expect them to be able to field a competitive team, something they have not been able to do in the history of their franchise. Although they still cannot afford marquee players, if they win enough games with their payroll at a low level, eventually they will start to increase revenues. Thus they would then be able to afford an impact player that can take their team to the next level.

For teams with fixed budgets, which are the vast majority of Major League Baseball, the ability to pay players their MRP is extremely important. The Oakland A's have made a habit of fielding personnel in which the players MRP exceeded their given salary. To contrast, the Mets always have a payroll in the top 5 and always seem to lose way too many games by investing in players whose MRP is far below what their contract pays them. With this information, teams have a sound theory that should allow them to maximize the amount of wins out of their given budget.

REFERENCES

Blass, Asher A. "Does the baseball labor market con tradict the human capital model of Investment?" *Review of Economics and Statistics* 74 (1992) 261-68.

Becker, Gary Stanley. "Human capital: a theoretical and empirical analysis, with special reference to education" *Human Behavior and Social Institutions*. New York: National Bureau of Economic Research. 1975

Burk, Robert F. *Much More Than a Game*. Chapel Hill, North Carolina: University of North Carolina Press, 2001.

Cook, Earnshaw. *Percentage Baseball*. Boston: MIT Press, 1964.

Fizel, John and Elizabeth Gustafson and Lawrence Hadley. *Baseball Economics*. Westport, CT: Praeger Publishers, 1996.

Houser, Adam. "The Effect of Outside Influences on Attendance Levels of Teams in Major

League Baseball". Applied Ecconometrics Research Project. Illinois Wesleyan University, 2004.

Irwin, Will. "Racial Discrimination in Major League Baseball, can it exist when Productivity is crystal clear?" Research Honors Project. Illinois Wesleyan University, 2004.

Lewis, Michael. *Moneyball*. New York: Bantam, 2003.

Major League Baseball. 2004. 8 February 2004. Available <http://www.mlb.com>

Markham, Jesse W. and Paul Teplitz. *Baseball Economics and Public Policy*. Toronto: Lexington, 1981.

Sommers, Paul M. *Diamonds are Forever*. Washington: Brookings, 1992.

Weiner, Ross D. "Power Hitters Strike Out: New Perspectives on Baseball and Slavery" *Rethinking Marxism.* 15 (2003) 33-48.

Zimbalist, Andrew. *May the Best Team Win: Baseball Economics and Public Policy*. Washington: Brookings, 2003.