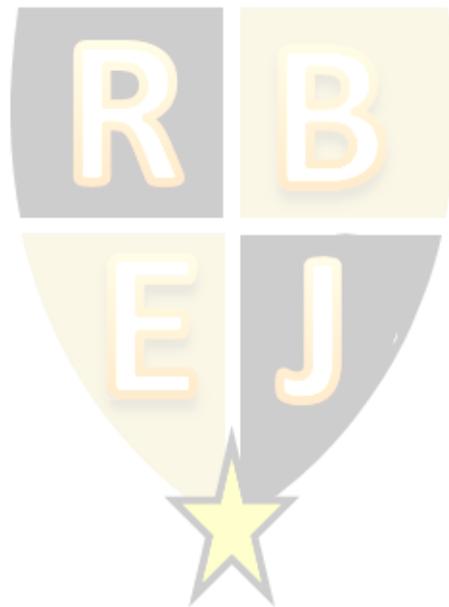


The econometrics of baseball: A statistical investigation

Mary Hilston Keener
The University of Tampa

The purpose of this paper is to use various baseball statistics available at the beginning of each baseball season to create a model that predicts the number of wins that each Major League Baseball team will have during the upcoming season. Only statistics available right before the season starts will be used in this model because that improves the value of the prediction model. If statistics that are not known until the season begins were included in the model, the model would not be as useful because many games would have already been played. The teams that are predicted to be the most successful typically have experienced pitching staffs, have been to the playoffs recently, score a lot of runs, and possess higher payrolls.

Keywords: econometrics, model specification tests, baseball statistics



Copyright statement: Authors retain the copyright to the manuscripts published in AABRI journals. Please see the AABRI Copyright Policy at <http://www.aabri.com/copyright.html>.

INTRODUCTION

“The sheer quantity of brain power that hurled itself voluntarily and quixotically into the search for new baseball knowledge was either exhilarating or depressing, depending on how you felt about baseball. The same intellectual resources might have cured the common cold, or put a man on Pluto” (Lewis, 2003).

This quote by Michael Lewis, the author of the popular 2003 book, Moneyball, demonstrates the extent of the time and effort that has been put into research on baseball. This book about the 2002 Oakland A's was so popular that it inspired a 2011 biographical sports drama film starring Brad Pitt. Lewis (2003) documents how the A's' front office took advantage of many analytical measures of player performance to construct a team that could compete successfully against Major League Baseball (MLB) teams with much larger payrolls. Statistics are one of the most important features of the game of baseball. Almost every baseball fan wakes up in the morning during baseball season and begins the day by looking through the box scores which full of statistics for their favorite baseball team. A daily baseball box score lists everything from batting average to the number of walks that occurred in the previous day's game. Sometimes even the most knowledgeable baseball fans have some difficulty interpreting all of the statistics that are included in a box score.

The MLB season begins with Spring Training in early February and doesn't end until the World Series is completed in October. Each baseball team plays exactly 162 games per year, 81 at home and the other 81 on the road. There are 30 Major League Baseball teams, with 15 being members of the National League and 15 being members of the American League. Because of how long the season is, it seems as if anything can happen throughout the season to help in deciding the number of wins and losses each team has. In some cases, a team that has been very good for several years will have injuries to one or more key players during the season and end up having a very poor season. It is also possible for a team to be in last place one year and in first place the next. The purpose of this paper is to use various baseball statistics available at the beginning of the baseball season to create a model that predicts the number of wins that each MLB team will have during the upcoming season. Only statistics available right before the season starts will be used in this model because that improves the value of the prediction model. If statistics that are not known until the season begins were included in the model, the model would not be as useful because many games would have already been played.

LITERATURE REVIEW

Many previous studies have examined the use of baseball statistics, but very few studies have focused on examining the econometrics of baseball. Several studies have examined the economics elements of baseball. Slottje et al. (1994) examine the pay and performance levels of MLB players using a new econometric technique called frontier estimation. Skelly (2004) examines the “economics of baseball” to identify the economic problems with Major League Baseball. Rushen (1999) examine the economic impact of the Pittsburgh Pirates on the Pittsburgh region using various economic models. Hakes and Sauer (2006) conduct an economic evaluation of the hypotheses raised in Michael Lewis' book, Moneyball, and they confirm, using econometric tools, that there is indeed very little correlation between pay and productivity in Major League Baseball. Regan (2012) find that from 1998 to 2008 the Oakland Athletics were the most payroll efficient team in Major League Baseball, which further demonstrates the

success of the A's general manager, Billy Beane, during the era discussed in Moneyball. Interestingly, the author finds that these efficient payroll strategies tend to have a diminishing impact on fan interest. Hayward and Patrick (2008) attempt to determine whether a player's entire career or only the most recent year of his contract will have more of an impact on the player's new salary that will result from salary negotiations.

Various studies have examined the factors that lead to team success in Major League Baseball. Martin and Troendle (1999) examine the effect of MLB play-off configurations on the likelihood of a team making it to the World Series. The authors find that the extra division series round of the playoffs added in 1994 lowers the likelihood that the team with the best record will make it to the World Series while home field advantage does not appear to have a significant impact. Lewis et al. (2009) examine organizational capability, efficiency and effectiveness in MLB, and they find that organizational capability and, to a lesser extent, efficiency lead to regular season success. However, the authors find that post-season success is not related to capability and managerial performance. Finally, Somberg and Sommers (2012) find that teams with higher payrolls are more likely to make the playoffs.

Several studies also examine factors related to the competitive balance between MLB teams. Schmidt and Berri (2004) use convergence clusters to determine whether the overall competitiveness of Major League Baseball teams has improved or declined over time, and the authors find that competitive balance has continued to improve. Koop (2004) also examines the changes in the competitive balance of MLB teams over time, and he finds that competitive balance has remained constant over time and across teams. However, the author also finds that one team, the New York Yankees, have outperformed all other teams. Gustafson and Hadley (2007) evaluate the impact of hometown market size on competitive balance for MLB teams. The authors find that larger consolidated metropolitan population does lead to a significant increase in local revenues, payrolls and win percentages. Bristow et al. (2010) examine the fan loyalty for two MLB teams, the Chicago Cubs and the Arizona Diamondbacks. The authors find that Cubs fans demonstrate significantly higher levels of team and fan loyalty than Diamondbacks fans. A study by Denaux et al. (2011) shows that many variables including time factors, fan interest, city characteristics, team performance and fan's attendance behavior have a strong influence on attendance.

Many others studies have examined statistics that can be used for evaluating individual Major League Baseball players. Pankin (1978) introduces a new statistic for evaluating offensive performance in baseball called the offensive performance average (OPA). This statistic indicates the increase in expected runs produced by the batter and includes his stolen bases. Rosner and Mosteller (1996) develop a model for the number of batters faced and the number of runs scored against starting pitchers. Anderson and Sharp (1996) develop another new statistic for evaluating MLB players called the Composite Batter Index (CBI). Koop (2002) creates a model for comparing the performance of MLB players that includes many dimensions of batting performance.

Studies on MLB players have also examined many unusual factors that may lead to success during a season or over a player's entire career. McCullough and McWilliams (2010) attempt to determine whether players whose first or last name begin with the letter "K" are more likely to strike out more frequently than players without this initial, and they find that these players do not strike out more. Several studies also attempt to determine the impact of performance-enhancing drugs (PEDs) on Major League Baseball. De Vany (2011) find that there has been no change in MLB home run hitting in the last 45 years, despite the new records

that have been set in recent years. The author states that “the greatest home run hitters are as rare as great scientists, artists, or composers.” Pantuosco (2011) attempts to determine whether it pays for MLB players to be unethical by using PEDs, and he finds that players benefit financially from the use of illegal substances. Longley and Wong (2011) examine the speed of human capital formation in MLB, and they find that minor league player statistics are of limited value for projecting player success in MLB. Miceli and Volz (2012) use data envelopment analysis to examine the voting process for the Baseball Hall of Fame, and they find that about a third of the current members of the Hall of Fame should be replaced by more deserving players. Several research studies have attempted to determine whether racial discrimination is present in MLB (Slotje et al. 1994; Groothuis and Hill 2008).

Finally, several papers indicate the importance of using baseball statistics to bring realistic examples into the statistics classroom. Wiseman and Chatterjee (1997) explain how a set of data containing the salaries of MLB players can be used to teach data analysis. Horowitz and Lee (2002) demonstrate how to use semiparametric statistical models in applied economics. Strow and Strow (2005) describe how to teach cliometrics using baseball statistics.

METHODOLOGY

Many factors need to be considered when creating a model that will predict the number of wins that each Major League baseball team will have. It should also be noted that it is not possible to create a perfect prediction model because of the unexplainable success or failure that some teams end up having. For example, in some cases teams with extremely low payrolls that have been doing very poorly in recent years will all of a sudden go from “worst to first” in their divisions as their talented young players gain more experience. It is possible, though, to come up with a fairly accurate model that will determine which teams will have many wins and which teams will have very few wins.

A significant amount of baseball expertise needs to be used in coming up with the factors that are most likely to predict the number of wins that a particular team will have. The dependent variable in this paper is the number of wins each team will have in 2001. For the purposes of this paper, the predictor variables indicated in Table 1 (Appendix) are initially considered.

The LGYEARS variable represents the total number of years that the team has existed, and teams that have been around longer are predicted to have better records. The next variable considered for the model is DUMMYWINS. This is a dummy variable for the number of wins each team had in the previous year. DUMMYWINS is assigned a value of one if the team has 80 or more wins and is assigned a value of zero otherwise. The third variable considered is NEWMGR, a dummy variable that equals one if the team has a new manager for the beginning of the current season. PITCAGE is a dummy variable that equals 1 if the average age of the team’s pitching staff for the current year is greater than or equal to 28.7 years, the average age of all the pitchers from all the teams. The average age of the batters on each team for the current year is also considered, but it is not ultimately included because it is highly correlated with PITCAGE. The next variable considered is ALLSTR, the number of all-stars that a team had in the previous year. PLAYOFF is a dummy variable that equals 1 if the team made the playoffs in the previous season. NEWRUNS is a dummy variable that equals 1 if the number of runs scored in the previous year is greater than or equal to 900. ERA is the earned run average of each team’s pitchers for the previous season. Earned run average is the average number of runs that

would have been scored off a pitcher in a nine inning game, and teams would like this variable to be as low as possible. TOTPAY is a continuous variable that discloses the salary that each team will be paying their players for the current season. Finally, ATTENDAN is the total attendance for each team for the previous season.

Preliminary descriptive statistics were obtained for the data set to see the average values for each the variables. Most of the dummy variables just look at whether each individual team is above or below the average value for that variable. Although many independent variables are considered initially, it is determined that there is significant correlation between many of the variables and that some of the variables are not likely to add additional predictive power to the model. The number of variables to be included in the model is therefore narrowed systematically down to four. It is determined that LGYEARS, NEWMGR, ALLSTR, and ATTENDAN do not offer very much additional predictive power, so these variables are not included in the final model. Also, it is clear that the variables DUMYWINS and ERA both correlate highly with PLAYOFF, so DUMYWINS and ERA are both left out of the final model. The variable BATAGE is not included because of its high correlation with PITCAGE. The final model is therefore:

$$WINS_t = \alpha + \beta_1 NEWRUNS_{t-1} + \beta_2 PITCAGE_t + \beta_3 PLAYOFF_{t-1} + \beta_4 TOTPAY_t$$

The purpose of the model is to predict the number of wins that a baseball team will have during a season before the season even begins. The hypothesized sign of the coefficient on NEWRUNS is positive, because teams that scored more runs in the previous year should theoretically have more wins in the coming year. The coefficient of the variable PITCAGE is expected to have a positive sign because teams with older, more experienced pitchers are generally expected to have more wins. The variable PLAYOFF should also have a positive coefficient because teams that made the playoffs last year are more likely to have a higher number of wins this year. Finally, teams that have higher payrolls are expected to win more games, so the coefficient on TOTPAY is also expected to be positive.

RESULTS

The results of an ordinary least squares regression on the model are as indicated in Table 2 (Appendix). The F-value for the model of 16.691 is significant at a p-value of less than .001. Also, the R-squared of .728 and the adjusted R-squared value of .684 both show that a significant amount of the variance in the number of wins is explained by the four predictor variables included in this model. The coefficients have the expected signs and are significant predictors in all cases except for the TOTPAY variable.

It is clear from the casewise diagnostics shown in Table 3 (Appendix) that the prediction model does a fairly good job of predicting the number of wins that each MLB team will have during 2012. Each case above represents one of the thirty Major League baseball teams. Comparisons can easily be made between the actual and predicted number of wins for 2012. One thing to note is that the model was not able to correctly identify cases of teams like the Washington Nationals and Baltimore Orioles who improved significantly during and somewhat unexpectedly during 2012 from the previous year. A few other comparisons show how valuable this model would really be to a baseball analyst. A comparison of the results of the model created for this study with the actual results can be made to test the value of this model for predicting the number of wins that each team will have and where each team might rank in their division.

As shown in Table 4 (Appendix), the predicted value for the number of wins of each time is the number immediately following the team's name, and the actual number of wins for each team is listed in parentheses. It is clear that the prediction model discussed in this paper is a fairly successful prediction model. The teams that are predicted to be the most successful typically have experienced pitching staffs, have been to the playoffs recently, score a lot of runs, and possess higher payrolls.

MODEL SPECIFICATION TESTS

Although the model has significant predictive power, it is necessary to run specification tests to assure that no problems exist. The first tests run examine whether or not there is heteroskedasticity in the model. First, the squared residuals are graphed against the predicted values to see if any heteroskedasticity is evident in the model. The resulting graph in Figure 1 (Appendix) shows that there is no pattern to the data points and thus no obvious heteroskedasticity.

The Breusch-Pagan test is used to formally rule out a heteroskedasticity problem. After regressing P_i on the model, the results show that one half of the explained sum of squares (ESS = 4.814) is equal to 2.407, and thus the null hypothesis of homoskedasticity cannot be rejected. Therefore, the model does not appear to have a heteroskedasticity problem.

Next, tests are run to see if there is an autocorrelation problem in the model. The Durbin-Watson statistic for the model is 2.257. After finding the correct Durbin-Watson statistic from the table, it is clear that 2.257 is between D_u (1.57) and $4 - D_u$ (2.43), so we can conclude that there is not an autocorrelation problem in this model.

Further tests are conducted to see if there is a multicollinearity problem in the model. This model does have a fairly high R-squared value and several significant t-ratios, so the model should be looked at more closely to test for check for multicollinearity. All of the Pearson correlations are less than or equal to .722 (as shown below), and the fact that the variance inflation factors are all pretty low (between one and two) indicates that there is not a problem with multicollinearity in the model.

Tests are also conducted to see if there is an underfitting or overfitting problem in the model. To test for underfitting, Ramsey's RESET test is run on the data, and the significance of the resulting F value (as seen below) indicates that there are in fact some omitted variables. Although there are missing variables, no changes will be made to the model for the purposes of this study. Next, the model is examined to see if there is a problem with overfitting. Only one of the t-ratios is not significant at the .05 level, so the regression is run again without TOTPAY. The resulting R-squared and adjusted R-squared values (as seen below) are lower than they were before taking out TOTPAY, so it appears that the model is better when it includes this variable.

Tests of functional form would normally be the next tests run on this model, but the fact that dummy variables are used makes it impossible to use the log form of the model. It is impossible because the log of one is equal to zero, and the log of zero cannot be calculated. Therefore, the final specification tests conducted on the model are tests of model stability. When the model is broken into two independent groups so that RSS1 and RSS2 can be used to calculate URSS, the resulting insignificant F-value of -1.115 indicates that there is not a problem with stability for this model.

CONCLUSION

It appears from all the model specification tests that the model created for the purposes of this paper is a fairly accurate prediction model. An interesting way to further test the results of this model would be to collect of the data for the model for several other baseball seasons and to see if the model can as accurately predict the number of wins that each team will have in other seasons. Until these tests are done, it is not clear if this model will be generalizable to other baseball seasons.

This model provides many interesting implications for sports analysts. If this model can indeed give a more clear idea which teams are likely to have the best records for a particular baseball season, it could be a very handy tool. Baseball analysts could use the model created in this study as a more accurate model than just using their own best intuitions about baseball. Unfortunately it is probably not possible to create a more accurate model because many of the unexplainable things that happen each year in a baseball season. For example, the model in this paper predicts that both the Colorado Rockies and the Chicago White Sox will have more wins than the number they ended up having. All factors would have lead anyone to believe that Chicago and Colorado would have ended up having better years than they ended up having. Analysts at ESPN (a sports news company) frequently make predictions about the coming season before the baseball season begins, so a model like this could assist them as they make their predictions.

Further studies could be done to test the model discussed in this paper using different seasons. If the model is a significant predictor of the number of wins for seasons other than 2012, this model may help analysts in the future to make significantly more accurate predictions than those listed in the preceding table. Also, future research could attempt to determine what additional variables could be added to the model to increase its predictive ability. Future studies could also apply a model of this type to other professional sports to determine the variables that are most useful for projecting success in those sports. Several limitations on the results of this study should also be noted. First, the study was only conducted for one season. Next, this paper chose a small number of variables that are representative of team performance in certain areas, but other variables may improve this model. Finally, the model is run using an ordinary least squares regression, and this may have some impact on the results. Future studies could instead use a binary logistic regression model by incorporating a dummy dependent variable such as whether or not a team has more than a certain target level of wins

REFERENCES

- Anderson, T. R., & Sharp, G. P. (1996). A new measure of baseball batters using DEA. *Annals of Operations Research*, 66 (1-4), 141-155.
- Baseball salary information. (n.d.). Retrieved June 12, 2013 from <http://content.usatoday.com/sportsdata/baseball/mlb/salaries/team>
- Baseball statistical data. (n.d.). Retrieved June 12, 2013 from <http://www.baseball-reference.com>
- Bristow, D., Schneider, K., & Sebastian, R. (2010). Thirty games out and sold out for months! An empirical examination of fan loyalty to two Major League Baseball teams. *Journal of Management Research*, 2 (1), 1-14.
- De Vany, A. (2011). Steroids and home runs. *Economic Inquiry*, 49 (2), 489-511.

- Denaux, Z. S., Denaux, D. A., & Yalcin, Y. (2011). Factors affecting attendance of Major League Baseball: Revisited. *Atlantic Economic Journal*, 39 (2), 117-127.
- Groothuis, P. A., & Hill, J. R. (2008). Exit discrimination in Major League Baseball: 1990-2004. *Southern Economic Journal*, 75 (2), 574-590.
- Gustafson, E., & Hadley, L. (2007). Revenue, population, and competitive balance in Major League Baseball. *Contemporary Economic Policy*, 25 (2), 250-261.
- Hakes, J. K., & Sauer, R. D. (2006). An economic evaluation of the Moneyball hypothesis. *Journal of Economic Perspectives*, 20 (3), 173-185.
- Hayward, P., & Patrick, T. (2008). How good is a baseball owner's memory? The importance of career statistics vs. recent performance in salary negotiations. *Proceedings of the Northeast Business & Economics Association*, 5-13.
- Horowitz, J. L., & Lee, S. (2002). Semiparametric methods in applied econometrics: Do the models fit the data? *Statistical Modelling: An International Journal*, 2 (1), 3-22.
- Koop, G. (2002). Comparing the performance of baseball players: A multiple-output approach. *Journal of the American Statistical Association* 97 (459), 710-720.
- Koop, Gary. (2004). Modelling the evolution of distributions: An application to Major League Baseball. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 167 (4), 639-655.
- Lewis, H. F., Lock, K. A., & Sexton, T. R. (2009). Organizational capability, efficiency, and effectiveness in Major League Baseball: 1901–2002. *European Journal of Operational Research* 197 (2), 731-740.
- Lewis, M. (2003). *Moneyball: The art of winning an unfair game*. New York: W.W. Norton.
- Longley, N., & Wong, G. (2011). 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*, 32 (3), 193-204.
- Martin, D. E. K., & Troendle, J. F. (1999). Paired comparison models applied to the design of the Major League Baseball play-offs. *Journal of Applied Statistics*, 26 (1), 69-80.
- McCullough, B. D., & McWilliams, T. P. (2010). Baseball players with the initial “K” do not strike out more often. *Journal of Applied Statistics*, 37 (6), 881-891.
- Miceli, T. J., & Volz, B. D. (2012). Debating immortality: Application of data envelopment analysis to voting for the Baseball Hall of Fame. *Managerial & Decision Economics*, 33 (3), 177-188.
- Pankin, M. D. (1978). Evaluating offensive performance in baseball. *Operations Research* 26 (4), 610-619.
- Pantuosco, L. J. (2011). Does it pay to be unethical? The case of performance enhancing drugs in MLB. *American Economist*, 56 (2), 58-68.
- Regan, C. S. (2012). The price of efficiency: Examining the effects of payroll efficiency on Major League Baseball attendance. *Applied Economics Letters*, 19 (11), 1007-1015.
- Rosner, B., & Mosteller, F. (1996). Modeling pitcher performance and the distribution of runs per inning in Major League Baseball. *American Statistician*, 50 (4), 352-360.
- Rushen, S. (1999). Economic impact of the Pirates on the Pittsburgh region. *Public Administration Quarterly*, 23 (3), 354-367.
- Schmidt, B., & Berri, D. J. (2004). Convergence and clustering in Major League Baseball: The haves and have nots? *Applied Economics*, 36 (18), 2007-2014.
- Skelly, K. (2004). Economics of baseball. *Monthly Labor Review*, 127 (1), p54-54.

Slottje, D. J., Hirschberg, J. G., Hayes, K. J., & Scully, G. W. (1994). A new method for detecting individual and group labor market discrimination. *Journal of Econometrics*, 61 (1), 43-64.

Somberg, A., & Sommers, P. (2012). Payrolls and playoff probabilities in Major League Baseball. *Atlantic Economic Journal* 40 (3), 347-348.

Strow, B. K., & Strow, C. W. (2005). How to pass down ideas via the national pastime or teaching cliometrics using baseball statistics. *Journal of Applied Economics & Policy*, 24 (2), 40-51.

Wiseman, F., & Chatterjee, S. (1997). Major League Baseball salaries: Bringing realism into introductory statistics courses. *American Statistician*, 51 (4), 350-352.

Table 1. Independent Variable Descriptions

Variable	Description
LGYEARS	Total # of years the team has existed
DUMYWINS	A dummy for the # of wins in 2011
NEWMGR	New manager or not for 2012?
BATAGE	Average age of batters for 2012
PITCAGE	Average age of pitchers for 2012
ALLSTR	# of All-Stars from 2011
PLAYOFF	Did the team make the playoffs in 2011?
NEWRUNS	Total # of runs scored in 2011
ERA	Average ERA from 2011
TOTPAY	Total Team Payroll for 2012
ATTENDAN	Attendance from 2011

Table 2: OLS Regression Predicting the Number of Wins in 2012

$$WINS_t = \alpha + \beta_1 NEWRUNS_{t-1} + \beta_2 PITCAGE_t + \beta_3 PLAYOFF_{t-1} + \beta_4 TOTPAY_t$$

Independent Variables	Expected Sign	Coefficients (t-Statistics)
Constant		75.611 (17.552)*
NEWRUNS	+	7.541 (2.307)**
(PITCAGE	+	14.349 (3.807)*
TOTPAY	+	-0.001 (-1.262)
PLAYOFF	+	13.305 (3.677)*

Adjusted R² = .684

*Indicates significance at the .01 level.

**Indicates significance at the .05 level

Table 3: Predicted vs. Actual Wins for the 2012 MLB Season

Team	Predicted Wins	Actual Wins	Residual
Arizona	87.00	81	-6.01
Atlanta	85.06	94	8.94
Baltimore	73.65	93	19.35
Boston	82.12	69	-13.12
Chicago Cubs	75.45	61	-14.45
Chicago White Sox	78.84	85	6.16
Cincinnati	78.02	97	18.98
Cleveland	79.79	68	-11.79
Colorado	76.70	64	-12.70
Detroit	86.27	88	1.73
Houston	76.10	55	-21.10
Kansas City	67.62	72	4.38
Los Angeles Angels of Anaheim	75.99	89	13.01
Los Angeles Dodgers	85.06	86	0.94
Miami	81.93	69	-12.93
Milwaukee	89.95	83	-6.95
Minnesota	71.02	66	-5.02
New York Mets	81.91	74	-7.91
New York Yankees	87.73	95	7.27
Oakland	78.35	94	15.65
Philadelphia	90.68	81	-9.68
Pittsburgh	81.04	79	-2.04
San Diego	77.34	76	-1.34
San Francisco	83.42	94	10.58
Seattle	77.66	75	-2.66
St. Louis	85.21	88	2.79
Tampa Bay	87.09	90	2.91
Texas	87.63	93	5.37
Toronto	80.43	73	-7.43
Washington	79.40	98	18.60

Table 4: 2012 MLB Standings with Predicted and Actual Runs Listed

AL East Predicted Wins (Actual):

1. New York Yankees 88 (95)
2. Baltimore Orioles 74 (93)
3. Tampa Bay Rays 87 (90)
4. Toronto Blue Jays 80 (73)

NL East Predicted Wins (Actual):

1. Washington Nationals 79 (98)
2. Atlanta Braves 85 (94)
3. Philadelphia Phillies 91 (81)
4. New York Mets 82 (74)

5. Boston Red Sox 82 (69)

5. Miami Marlins 82 (69)

AL Central Predicted Wins (Actual):

1. Detroit Tigers 86 (88)
2. Chicago White Sox 79 (85)
3. Kansas City Royals 68 (72)
4. Cleveland Indians 80 (68)
5. Minnesota Twins 71 (66)

NL Central Predicted Wins (Actual):

1. Cincinnati Reds 78 (97)
2. St. Louis Cardinals 85 (88)
3. Milwaukee Brewers 90 (83)
4. Pittsburgh Pirates 81 (79)
5. Chicago Cubs 75 (61)

AL West Predicted Wins (Actual):

1. Oakland A's 78 (94)
2. Texas Rangers 88 (93)
3. Los Angeles Angels of Anaheim 76 (89)
4. Seattle Mariners 78 (75)
5. Houston Astros 76 (55)

NL West Predicted Wins (Actual):

1. San Francisco Giants 83 (94)
2. Los Angeles Dodgers 85 (86)
3. Arizona Diamondbacks 87 (81)
4. San Diego Padres 77 (76)
5. Colorado Rockies 77 (64)



Figure 1:

Residuals squared vs. predicted wins for 2012

