Beyond Moneyball

Logan Munson and Taylor Klopp

Linfield Department of Economics • Spring 2015

I. ABSTRACT

This study provides an updated test of Billy Beane's Moneyball hypothesis using a panel model over the years 1999-2013. We regressed winning percentage as a function of the original Moneyball variables, which included on-base percentage, slugging percentage, on-base percentage against and slugging percentage against. In turn we created our own model which replaced the "against" statistics with earned run average and fielding percentage. Within both models, we concluded that the coefficient of on-base percentage was significantly greater than slugging percentage, which supports Beane's theory that in today's game on-base percentage is more important than slugging in determining winning percentage. These conclusions can be used by major league managers and owners to decide which players to trade for or to pick up in free agency.

II. Empirical Model and Variables

(Beane's Model) WIN_{it}=f(OBP_{it} $SLUG_{it}$ $OBP\_AGAINST_{it}$ $SLUG\_AGAINST_{it}$)

(Klopp/Munson Model) WIN_{it}=f(OBP_{it} $SLUG_{it}$ ERA_{it} $FIELD_{it}$)

WIN_{it} = Winning Percentage - Percentage of wins vs losses

OBP_{it} = On-Base Percentage - Percentage of times a hitter gets on base per plate appearance

$SLUG_{it}$ = Slugging Percentage - Number of total bases (single=1 double=2 triple=3 home run=4) divided by the total number of at bats

$OBP\_AGAINST_{it}$ = On-Base Percentage Against - Percentage of times the opposing team's batters get on base per plate appearance

$SLUG\_AGAINST_{it}$ = Slugging Percentage Against - The total bases your opponent reaches divided by their total number of at bats

ERA_{it} = Earned Run Average - The total amount of Earned Runs given up by a team per 9 innings.

$FIELD_{it}$ = Fielding Percentage - The percentage of times players in the field properly field a batted ball

*i denotes teams, t denotes years

III. Theory and Hypothesis

OBP_{it} is hypothesized to have a positive relationship with WIN_{it}. As more players get on base, more runs will be scored, resulting in a greater chance of winning.

$SLUG_{it}$ is hypothesized to have a positive relationship with WIN_{it}. The more bases you get to with each hit, the more likely you are to score, resulting in a greater chance of winning.

$OBP\_AGAINST_{it}$ is hypothesized to have a negative relationship with WIN_{it}. The more your opponent gets on base, the more likely they are to score, reducing your chances of winning.

$SLUG\_AGAINST_{it}$ is hypothesized to have a negative relationship with WIN_{it}. If your opponent gets to more bases with each hit, the more likely they are to score, reducing your chances of winning.

ERA_{it} is hypothesized to have a negative relationship with WIN_{it}. The lower this number is, the fewer runs a team allows on average. The fewer the number of runs allowed, the greater the chance a team has at winning the game.

$FIELD_{it}$ is hypothesized to have a positive relationship with WIN_{it}. The more often a ball is fielded without making an error, the fewer runs a team will allow, resulting in a greater chance of winning.

IV. Data

Panel model data set of all 30 MLB teams over 15 years (1999-2013)
Sample size: 450

Data Limitations:

- Due to the highly statistical nature of baseball we had no limitations in finding sufficient data for our project.

Data Sources:

- ERA, OBP, SLUG, and FIELD data all came from Baseballreference.com
- OBP_AGAINST and SLUG_AGAINST data came from both ESPN.com and MLB.com

V. Empirical Results

Beane's Model

Dependent Variable: WIN
Method: Panel Least Squares
Date: 03/20/15 Time: 10:01
Sample: 1999 2013
Periods included: 15
Cross-sections included: 30
Total panel (balanced) observations: 450

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.473365	0.044096	10.73486	0.0000
OBP	1.790287	0.170300	10.51254	0.0000
SLUG	1.006864	0.093878	10.72521	0.0000
OBP_AGAINST	-1.512180	0.151049	-10.01119	0.0000
SLUG_AGAINST	-1.163417	0.092161	-12.62373	0.0000

R-squared	0.803118	Mean dependent var	0.500000
Adjusted R-squared	0.801348	S.D. dependent var	0.071943
S.E. of regression	0.032065	Akaike info criterion	-4.031047
Sum squared resid	0.457536	Schwarz criterion	-3.985389
Log likelihood	911.9856	Hannan-Quinn criter.	-4.013052
F-statistic	453.8079	Durbin-Watson stat	1.600689
Prob(F-statistic)	0.000000		

Klopp/Munson Model

Dependent Variable: WIN
Method: Panel Least Squares
Date: 03/20/15 Time: 09:59
Sample: 1999 2013
Periods included: 15
Cross-sections included: 30
Total panel (balanced) observations: 450

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.260845	0.490677	-4.607604	0.0000
OBP	1.602465	0.144129	11.11830	0.0000
SLUG	1.116443	0.079273	14.08359	0.0000
ERA	-0.099842	0.002678	-37.28585	0.0000
FIELD	2.231919	0.497878	4.482867	0.0000

R-squared	0.859185	Mean dependent var	0.500000
Adjusted R-squared	0.857919	S.D. dependent var	0.071943
S.E. of regression	0.027118	Akaike info criterion	-4.366206
Sum squared resid	0.327241	Schwarz criterion	-4.320548
Log likelihood	987.3964	Hannan-Quinn criter.	-4.348211
F-statistic	678.7929	Durbin-Watson stat	1.740477
Prob(F-statistic)	0.000000		

VI. Conclusions

- As indicated by the adjusted R^2, 85.8% of the variation in WIN_{it} is explained by the Klopp/Munson model. The Beane model's adjusted R^2 is less at 80.3%.
- On-Base Percentage is significantly more important in determining winning percentage than is slugging percentage in both Beane's model and the Klopp/Munson model.
- We tested for the effects of the 2005 steroid ban on the importance of OBP and SLUG in determining winning percentage and found no statistically significant implications of the ban.
- We determined the Klopp/Munson Model predicts winning percentage better than Beane's does.