

Ouachita Baptist University

Scholarly Commons @ Ouachita

Scholars Day Conference

Scholars Day 2022

Apr 27th, 1:30 PM - 1:45 PM

The Correlation of Winning and Money-Baseball

Jacob Bowman

Ouachita Baptist University

Follow this and additional works at: https://scholarlycommons.obu.edu/scholars_day_conference



Part of the [Finance and Financial Management Commons](#), and the [Statistics and Probability Commons](#)

Bowman, Jacob, "The Correlation of Winning and Money-Baseball" (2022). *Scholars Day Conference*. 1. https://scholarlycommons.obu.edu/scholars_day_conference/2022/oral_presentations_b/1

This Thesis is brought to you for free and open access by the Carl Goodson Honors Program at Scholarly Commons @ Ouachita. It has been accepted for inclusion in Scholars Day Conference by an authorized administrator of Scholarly Commons @ Ouachita. For more information, please contact mortensona@obu.edu.

An abstract graphic consisting of several thin, black, overlapping lines that form a complex, geometric pattern in the upper left quadrant of the page. The lines intersect to create various shapes, including triangles and polygons, some of which are nested within others.

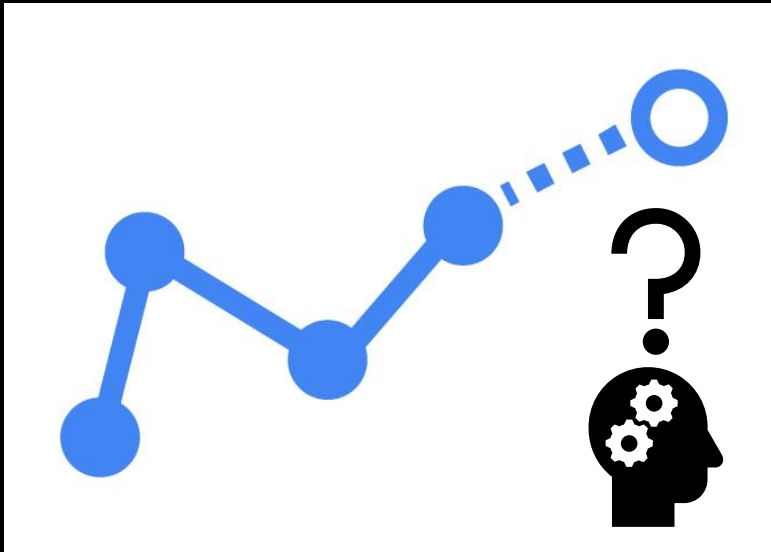
MONEY AND THE DIAMOND

An Honors Thesis Presentation

By: Jacob Bowman

OBJECTIVE OF THIS THESIS

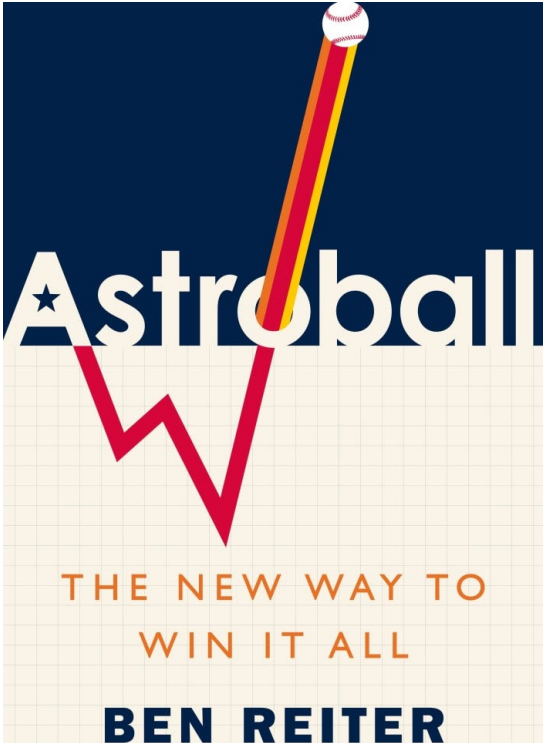
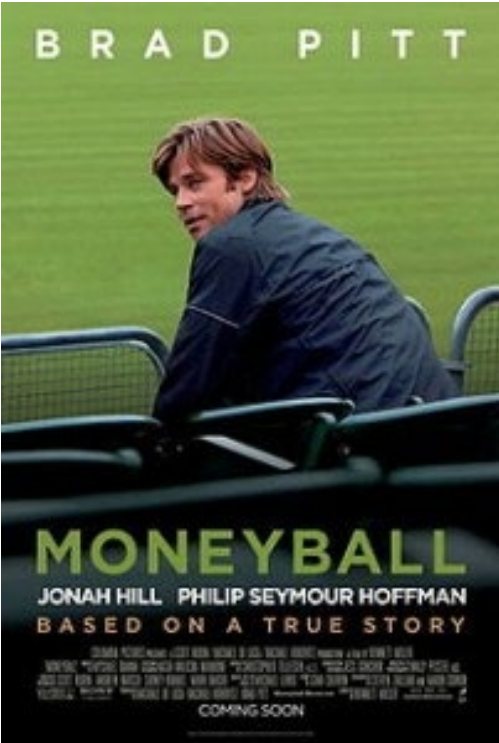
- Determine if financial metrics can be used to forecast performance for Major League Baseball (MLB) Teams.



BACKGROUND OF OBJECTIVE

- *Moneyball*
- *Astrobball*

Two real-world occurrences of organizations taking a financial and data-driven approach to baseball.



THE “WHY” AND “HOW”



Why:

- Baseball is notoriously unpredictable.
- 2005 Yankees: 10% of league-wide salary cap. Worst defense in the league measured by Defensive Runs Saved (DRS) and Ultimate Zone Rating (UZR). Barely made it to the playoffs, and lost their first series.
- If one realized the financial actions undertaken by the Yankees, they would have realized their exposure to suboptimal results.

How:

1. Gathering of relevant data.
2. Data analysis on relationships between financial and on-field data.
3. Take correlations discovered and impose on data from outside the sample.

DATA GATHERING





- Current literature on the subject suggests two main areas of data are needed. Data concerning Liabilities (Ecer, 2014) and data concerning previous winning trends (Pinnuk, 2006)
- Unfortunately, MLB financial data is private due primarily to leverage between the MLBPA and MLB.
- However, close estimations of the financial data is available from *Forbes'* "Business of Baseball" Valuations.
- Data on winning trends was obtained from *BaseballReference.com*, an online encyclopedia of historical baseball statistics.

The Business Of Baseball

Ads by Google
[Send feedback](#) [Why this ad? ▾](#)

The List [Spreadsheet](#) [Reprints](#)

Search by team name

Rank	Team	Current Value	1-Yr Value Change	Debt/Value	Revenue	Operating Income
#1	 New York Yankees	\$7.1 B	18%	0%	\$657 M	\$16 M
#2	 Los Angeles Dodgers	\$4.8 B	18%	9%	\$581 M	\$14 M
#3	 Boston Red Sox	\$4.5 B	15%	0%	\$513 M	\$72 M
#4	 Chicago Cubs	\$4.1 B	8%	10%	\$451 M	\$57 M


Sports Reference | Baseball | Football (college) | Basketball (college) | Hockey | Calcio | Blog | Stathead | Questions or Comments? [Create Account](#) | [Login](#)

BASEBALL REFERENCE [Search](#)

Players Teams Seasons Leaders Scores¹⁶ Playoffs Stathead Newsletter Full Site Menu Below ▾

Baseball Stats and History The complete source for current and historical baseball players, teams, scores and leaders.

22,899 Major League Players



View any Active Player:
 Choose a Team ▾
 Then a player ▾
[Go!](#)

Select a Hall of Famer:
 Select a player ▾

Trending Player Pages
[Shohei Ohtani](#), [Pete Alonso](#), [Aaron Judge](#), [Mike Trout](#), [Gerrit Cole](#),
[Maddux Bumgarner](#), [Cody Bellinger](#), [Jose Lucchesi](#), [Juan Soto](#),
[Alejandro Kirk](#)

Recent Debuts
[Tristan Beck](#) (SFG), [Levi Stoudt](#) (CIN), [Mason Miller](#) (OAK), [Brent Headrick](#) (MIN), [Emmanuel Valdez](#) (BOS) and [Blake Perkins](#) (MIL)

Every Major League Team

2023 MLB Standings
[MLB Summary](#) | [Scores](#) | [Schedule](#) | [Leaders](#) | [Standings](#) | [Transactions](#)
[AL Summary](#) | [Leaders](#) | [Batting](#) | [Pitching](#) | [Standings](#) | [Attendance](#)
[NL Summary](#) | [Leaders](#) | [Batting](#) | [Pitching](#) | [Standings](#) | [Attendance](#)

AL					NL				
W	L	GB	SRS		W	L	GB	SRS	
TBB	18	3	3.1		ATL	14	7	0.9	
BAL	13	7	4.5	0.3	NYM	14	8	0.5	0.3
NYG	13	8	5.0	0.7	MIA	12	9	2.0	-1.2
TOB	12	9	6.0	-0.3	PHI	10	12	4.5	-1.0
BOS	11	11	7.5	0.8	WSN	7	13	6.5	-1.0

Central

W	L	GB	SRS		W	L	GB	SRS	
MIN	11	10	0.3		MIL	15	6	2.1	
CLE	10	11	1.0	-1.3	PIT	15	7	0.5	0.8
DET	7	12	3.0	-1.7	CHC	12	8	2.5	2.6
CHW	7	14	4.0	-0.9	STL	8	13	7.0	0.0
KCB	5	16	6.0	-1.8	CIN	7	14	8.0	-0.2

West

W	L	GB	SRS		W	L	GB	SRS	
TEX	13	7	2.4		ARI	12	10	0.4	
HOU	11	10	2.5	1.3	LAD	11	11	1.0	1.2
LAA	10	11	3.5	-0.3	SDP	11	12	1.5	-0.3
SEA	10	11	3.5	0.1	SEG	7	13	4.0	-1.2
OAK	4	17	9.5	-3.9	COL	6	16	6.0	-2.3

2023 MLB Rules Changes Effect on Stats and Pace of Play
 Game time, batting average, stolen bases, clock violations

Most Likely Playoff Scenario
 NL: Pves, I AD, ATL Wild Card Series: MIL vs STL and NYM vs CHC

The sports search engine that was made for fans like you

STATHEAD POWERED BY **BASEBALL REFERENCE**

Exploring the Baseball Reference database just got easier

[Start your FREE trial](#)

Stathead Commercial | Subscribe to The... **STATHEAD**

THE MOST POWERFUL SPORTS DATABASE ON THE INTERNET

[Learn More](#)

Benefits of Stathead:

- Access to the most complete sports database on the internet
- Affordable (as low as \$8/month)
- Discovery tools to search the Baseball Reference database
- Customized leaderboards
- Ad free viewing on Baseball Reference

DATA GATHERING PT.2

- The next step was to put the data in one place and establish relationships with the data. As well as the relevant time the data represents.

	2022	2021	2020	2019	2018	2017	2016	2015	Current Year	
	Current Value (Billions)	1-Yr Value Change	Debt/Value Ratio	Revenue (Millions)	Operating Income (Millions)	Playoffs (Y/N)	Win % (Previous Year)	Playoffs	Win % Ratio	
New York Yankees	6.000	14%	0%	482	-40	1	0.568	1	0.615	
Los Angeles Dodgers	4.075	14%	11%	565	-7.9	1	0.654	1	0.683	
Boston Red Sox	3.900	13%	0%	479	69	1	0.568	0	0.478	
Chicago Cubs	3.800	13%	11%	425	68	0	0.438	0	0.453	
San Francisco Giants	3.500	10%	4%	384	32	1	0.660	0	0.497	
New York Mets	2.650	8%	17%	302	-96	0	0.475	1	0.621	
St. Louis Cardinals	2.450	9%	9%	287	-34	1	0.556	1	0.578	
Philadelphia Phillies	2.300	12%	6%	323	-17	0	0.506	1	0.540	
Los Angeles Angels	2.200	9%	5%	331	-2.4	0	0.475	0	0.453	
Atlanta Braves	2.100	12%	23%	443	83	1	0.584	1	0.627	
Texas Rangers	2.050	15%	37%	387	97	0	0.370	0	0.416	
Washington Nationals	2.000	4%	25%	322	36	0	0.401	0	0.342	
Houston Astros	1.980	6%	15%	388	29	1	0.586	1	0.652	
Toronto Blue Jays	1.780	6%	0%	238	-52	0	0.562	1	0.589	
Chicago White Sox	1.760	4%	9%	258	-9.5	1	0.574	0	0.503	
Seattle Mariners	1.700	4%	14%	313	71	0	0.566	1	0.553	
San Diego Padres	1.575	5%	19%	282	-32	0	0.488	1	0.553	
Detroit Tigers	1.400	11%	13%	268	31	0	0.475	0	0.410	
Minnesota Twins	1.390	5%	20%	268	10	0	0.451	0	0.478	
Colorado Rockies	1.385	7%	9%	270	14	0	0.460	0	0.422	
Arizona Diamondbacks	1.380	5%	9%	267	40	0	0.321	0	0.453	
Baltimore Orioles	1.375	-4%	16%	251	83	0	0.321	0	0.513	
Pittsburgh Pirates	1.320	3%	11%	258	64	0	0.377	0	0.379	
Cleveland Indians	1.300	12%	10%	267	71	0	0.494	1	0.565	
Milwaukee Brewers	1.280	5%	12%	269	29	1	0.586	0	0.534	
Cincinnati Reds	1.190	10%	13%	266	0.4	0	0.512	0	0.385	

- Items in yellow were taken from *BaseballReference.com*

DATA ANALYSIS

- The primary tool used was multiple regression analysis.
- Reason is due to the ability of multiple regression analysis to provide a linear equation that can be used to predict the dependent variable.
- The dependent variable (what we want to eventually predict) was the win ratio for that year.

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.821053401
R Square	0.674128688
Adjusted R Square	0.661158685
Standard Error	0.049648181
Observations	210

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	8	1.024942307	0.128117788	51.97598763	6.42041E-45
Residual	201	0.495453317	0.002464942		
Total	209	1.520395624			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.344985268	0.030750994	11.21867043	5.39168E-23
Current Value (Billions)	0.01169493	0.006798937	1.720111436	0.08695131
1-Yr Value Change	0.027685501	0.020282609	1.36498719	0.173783004
Debt/Value Ratio	-0.067600539	0.037492156	-1.803058216	0.072876812
Revenue	-8.54916E-05	6.7992E-05	-1.257377009	0.210076133
Operating Income (Millions)	0.00011894	0.000120363	0.988184116	0.324250544
Playoffs (Y/N) Previous Year	0.007929519	0.010564579	0.750575977	0.453785637
Win % (Previous Year)	0.232303785	0.066324116	3.502553824	0.000567816
Playoffs	0.113848751	0.00796451	14.29450836	1.92E-32

DATA ANALYSIS PT. 2

- The initial results were promising. However, another multiple regression analysis was needed, only this time it would only include independent variables whose p value < 0.15 in the initial multiple regression analysis. One variable, Playoff Appearance, was removed due to it being essentially the equivalent of Win Ratio.

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.577749522
R Square	0.33379451
Adjusted R Square	0.324092488
Standard Error	0.070121113
Observations	210

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	3	0.507499712	0.169166571	34.40463445	4.54929E-18
Residual	206	1.012895912	0.00491697		
Total	209	1.520395624			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.213299335	0.031939093	6.678315257	2.20512E-10
Current Value (Billions)	0.014435227	0.005923963	2.43675183	0.015669062
Debt/Value Ratio	-0.023185724	0.052541475	-0.441284233	0.659470051
Win % (Previous Year)	0.529791768	0.062591517	8.464274327	4.80687E-15

RESIDUAL OUTPUT OF THE SECOND MULTIPLE REGRESSION ANALYSIS

RESIDUAL OUTPUT

<i>Observation</i>	<i>Predicted Win % Ratio</i>	<i>Residuals</i>
1	0.600832421	0.014167579
2	0.616056271	0.066943729
3	0.570518444	-0.092518444
4	0.497651562	-0.044651562
5	0.612557767	-0.115557767
6	0.499262203	0.121737797
7	0.541143149	0.036856851
8	0.513183848	0.026816152
9	0.495548638	-0.042548638
10	0.547678987	0.079321013
11	0.430335786	-0.014335786
12	0.448819857	-0.106819857
13	0.548861202	0.103138798
14	0.536737012	0.052262988
15	0.540719094	-0.037719094
16	0.53445536	0.01854464
17	0.490167912	0.062832088
18	0.482145598	-0.072145598
19	0.467663243	0.010336757
20	0.474909622	-0.052909622

- Residual Output is the Regression Analysis testing the validity of it's formula.
- Over 210 separate observations were included in the sample data.
- The closer a residual is to 0, closer a model's prediction was to reality.

<i>Residual Output Statistics</i>	
Mean	1.16309E-17
Median	0.003739692
Standard Deviation	0.069616032
Minimum	-0.182437337
Maximum	0.1557544
Count	210

THE FINAL MODEL

Future Win Ratio

$$= \alpha + \beta_1 \text{Current Value} + \beta_2 \text{Debt to Value Ratio} \\ + \beta_3 \text{Win Ratio Previous Year}$$

Where:

α is the coefficient of the intercept

β_1 is the coefficient of Current Value

β_2 is the coefficient of Debt to Value Ratio

β_3 is the coefficient of the Previous Year's Win Ratio.

TESTING THE MODEL

Two things must be understood prior to testing:

- Regression Statistics – tell how well the equation fits with sample data
- Margin of Error for the Model

REGRESSION STATISTICS

Significance *F* Statistic

- Determines whether or not the model with its independent variables explains the variability of the dependent variable better than a model with no independent variables.
- The Significance *F* Statistic for the model is 4.5428 E-12.
- Therefore the model is sufficient at explaining the variability of an organizations Win Ratio.

Standard Error Statistic

- The Standard Error of the Model was .070.
- .070 is the equivalent of roughly 11 games.
- Without the model, the chances of accurately predicting a 162 game win ratio is .006%, as there are 162 different individual regular season ratios one can choose.
- The model can now “shrink” the range of acceptable guesses using its Standard Error. If one takes the predicted value provided by the model, and applies the standard of error, there is roughly a 20 game stretch from which one can guess.
- This means that one has a 5% chance of correctly predicting a team’s record.
- Additionally, the Standard of Error will continue to shrink as more sample data is added to the model, creating even better chances of predicting the win ratio.

TESTING THE MODEL

- Using data originating from outside the sample data, one can see if the model actually works.
- Three tests were performed. The 2012 Mariners, 2014 Pirates, and 2014 Cardinals.

2012 Mariners

Current Value: .585 Billion
 Debt/Value Ratio: 30%
 Previous Year Win Ratio: .414

Model Prediction: .434
 Actual: .460



2014 Cardinals

Current Value: .820 Billion
 Debt/Value Ratio: 35%
 Previous Year Win Ratio: .598

Model Prediction: .533
 Actual: .598

2014 Pirates

Current Value: .572 Billion
 Debt/Value Ratio: 16%
 Previous Year Win Ratio: .580

Model Prediction: .525
 Actual: .540



CONCLUSION

- The Model will continue to improve as more data is added into it.
- There are potential repercussions in areas such as sports betting.
- Could be used as a new tool for evaluating coaching and box-office management.
- The Money can now tell you about The Diamond

Questions??