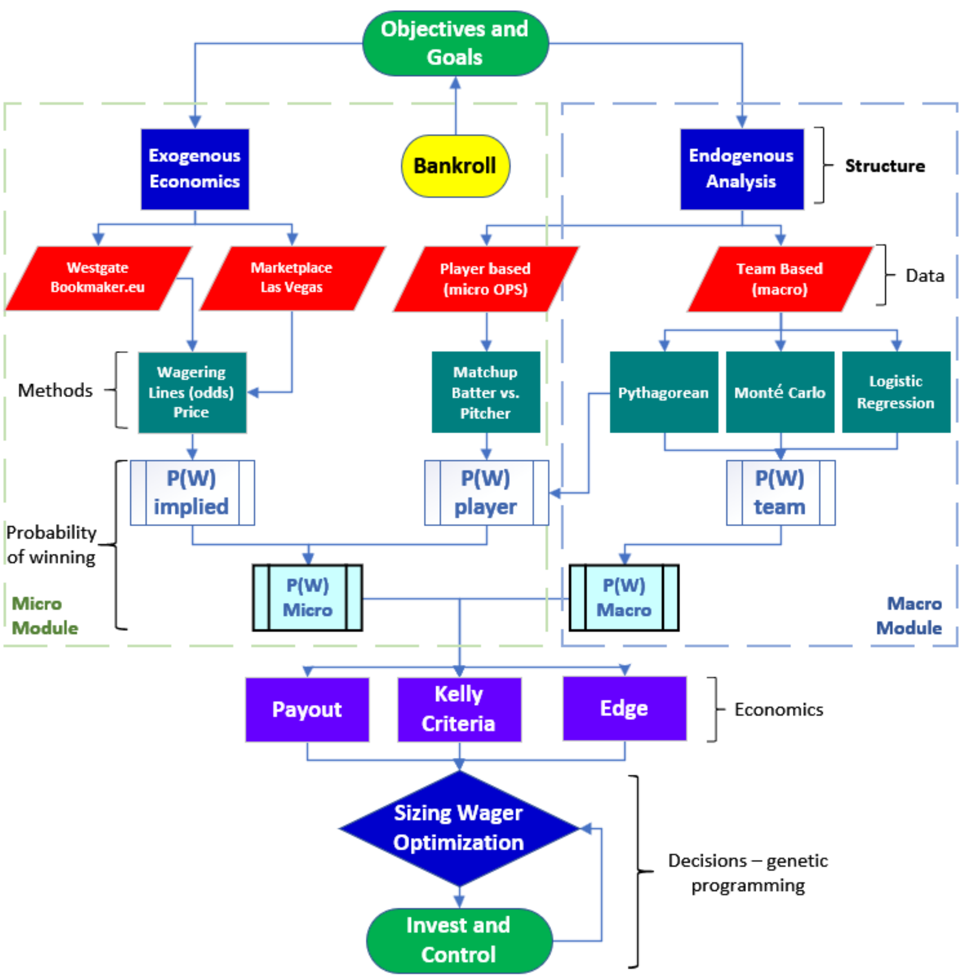
Bettor Up: Assess, Analyze, and Achieve

Tracks – Business of Sports and Open Source
ID 1548878

1. Introduction

With the repeal of the "Professional and Amateur Sports Protection Act" (PASPA) betting on athletic competition in the United States is now taking on a modicum of respectability. Our focus will be investing on baseball games. At the heart of sports betting is calculating the probability of winning the game in general and the wager in particular. Accurately quantifying matchups is essential for prediction consistency and reproducibility. Concurrently sizing the stake (bet) is the foundation to enhancing profitability. Figure 1 profiles the process to be undertaken.

Figure 1- Pathway to Economic Achievement



Baseball is unique in sports from a ***micro*** perspective with the matchup between batters and pitchers and from a ***macro*** view as a team. Several research questions present themselves, can we:

1. Identify inequities between the wagering line(s) and game's expected production,
2. Capitalize on such inequities so as to profit from the investment,
3. Build an investment model to optimally select and size each wager in order to maximize profits.

To answer these questions, it is essential to measure the match-ups between individual players (***micro***) and the competing teams (***macro***). By quantifying match-ups the probability of winning the game may be ascertained. Combining the probability of winning with the betting lines the economics of the wager will be measured. Sounds simple enough, but just how?

The schematic in Figure 1 shows the process that we'll be undertaking and will be addressed in the following sections:

- Exogenous Economics,
- Endogenous Analysis,
- Sizing Wager Optimization.

As a reference terms and nomenclature are provided in Appendix A. Let's proceed into the hinter land of sports gambling!

2. Exogenous Economics

At the foundation of economics are those who provide the service or product and those who consume them. The meeting point is the price. The ***moneyline*** is the price of the wager and starts when published. First notice is usually provided by the Westgate Casino in Las Vegas or on the Internet by Bookmaker.eu.

Stated simply the odds or line may be quoted as:

- Houston Astros -150 (favorite), Washington Nationals +140 (underdog)

This means a winning $100 bet on Houston would net a $66[1] profit. In the case of the Nationals, a winning $100 wager would yield $140[2]. From the house's perspective:

- Collect $100 for the Houston bet and pay out $166 (return of $100 bet plus $66)

Should the Nationals win:

- Collect $100 on the Nationals and pay out $240 (return of original bet plus $140)

---

[1] Astros profit calculations with a line of -150: $100*(100/|150|) = $66
[2] Nationals profit calculations with a line of 140: $100*(140/100) = $140

The House's objective is to balance the lines so as to make a profit. The difference between moneylines of -150 and +140 is -10, know is a "dime line". The line can be converted to an implied probability of winning[3].

2.1. Implied probability of winning[4]

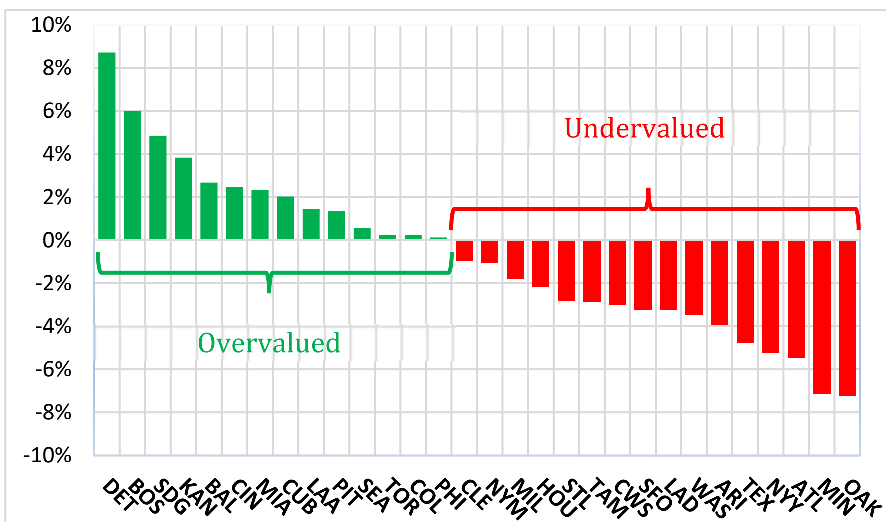
In this case the implied probability of winning for the Houston and Nationals respectively are:

- |150| / (|150| + 100) = 60% ⇒ 60%/ (60%+42%) = 58.8%
- 100 / (|140| + 100) = 42% ⇒ 42% / (60% + 42%) = 41.2%

Normalizing for the house's profit (the juice or vig[5]) generates P(W) for Houston of 58.8% and Nationals of 41.2%.

A question arises, are all things fair in Mudville[6]? Specifically, over a season, is a team's winning percentage about the same as its implied probability of winning (value[7])?

Figure 2 - Team Relative Wager Value 2019 Season



Source: https://sportsbookreviewsonline.com/scoresoddsarchives/mlb/mlboddsarchives.htm
Tabulated: by authors

---

3 Probability of "i" winning will be abbreviated as: $P(W)_i$
4 The implied probability of winning is the conversion of odds (line) into a percentage - while taking into account the bookmaker's profit. https://www.pinnacle.com/en/esports-hub/betting-articles/educational/implied-probability-odds-conversion/72m2z3g3g22m5tps
5 Vig is short for vigorish.
6 https://www.baseball-reference.com/bullpen/Mudville_Nine
7 Value = (Actual Winning % - Implied probability of winning P(W))

Figure 2 demonstrates that over the 2019 season a team's value varied quite widely. The Tigers on average were overvalued by 8% while Oakland was undervalued by the same amount. Hence, we now know that if we can estimate the actual probability and compare it to the implied probability of winning, we may then measure the economic ***edge***[8]. Let's move on to measuring team and player performances and resultant probabilities of winning.

3. Endogenous Analysis (probability of winning)

Considering that baseball has more data available than a Microsoft Cloud Server, we must narrow the scope of our inquiry. A simple question, what baseball statistics from an offensive and defensive perspective correlate well with a team's winning percentage? Ranking a team's batting and pitching production variables and relating them with winning percentage generates a correlation matrix and is shown in its entirety in Appendix B, partially summarized in Table 1.

Table 1- Coefficients of Correlation Selected Variables with Winning

Batting			Pitching		
Variable	|%Win|	Rank	Variable	|%Win|	Rank
Δ(OPS)	0.933	1	Δ(OPS)	0.933	1
OBP	0.811	2	Wins	0.851	2
R	0.776	3	ERA	0.793	3
RBI	0.763	4	ER	0.782	4
OPS	0.758	5	OBP	0.764	5
Scr-Al	0.706	6	OPS	0.753	6
SLG	0.692	7	BAA	0.736	7
BB	0.684	8	Scr-Alw	0.706	8
TB	0.670	9	Losses	0.704	9
TPA	0.666	10	SLG	0.695	10
XBH	0.631	11	K/BB	0.656	11
AVG	0.621	12	SHO	0.641	12
HR	0.571	13	SV	0.640	13
H	0.569	14	IP	0.577	14
AB/HR	0.542	15	DIP%	0.568	15

Source: Appendix B

It is significant that the difference between a team's batters' and pitchers' on base percentages (ΔOPS) is actually more highly correlated with winning than runs scored and runs allowed (SCR-AL) e.g., .933 vs. .706. This result will be useful in building a more accurate game prediction model.

[8] The ***edge*** is the: (calculated probability of winning – implied probability of winning)

3.1 Baseball's Pythagorean Theorem

Deep in the lore of baseball analytics is the work of Bill James[9]. One of his many contribution to the literature was identifying a nonlinear relationship between runs scored and runs allowed. It was specifically postulating that:

$$\text{Probability of winning} = \text{Runs Scored}^2 / (\text{Runs scored}^2 + \text{Runs allowed}^2) \qquad (1)$$

Because of equation (1)'s format, it was dubbed Baseball's Pythagorean Theorem. Mathematically derived, the exponent in actuality is about 1.77 for Major League Baseball (seasons 2015-19). Since OPS is a more accurate variable for measuring the winning percentage the following defines ΔOPS as:

$$\Delta OPS_{home} = (OPS_{batting} - OPS_{pitching})_{home} \qquad (2)$$

$$\Delta OPS_{road} = (OPS_{batting} - OPS_{pitching})_{road} \qquad (3)$$

Hence, a new equation evolves:

$$\text{Probability of Winning}_{home} = (\Delta OPS_{home})^{\alpha} / ((\Delta OPS_{home})^{\alpha} + (\Delta OPS_{road})^{\alpha}) \qquad (4)$$

Where $\alpha = 3.374$.[10]

Equation (4) is the predictor and provides a more accurate mechanism to calculate the probability of winning *a priori* in contrast to having to use estimators for runs scored and allowed. It will next be incorporated into a player-based matchup model.

3.2 Matchups

Competitive matchups can be measured from two major perspectives. First is a ***micro*** vantage point of player vs. player and second is a ***macro*** or team view. Each will be examined.

3.2.1 Matchups: Player-based OPS

The matchup between batter and pitcher is one of the most unique characteristics of baseball as a team sport. Player performance is affected by whether they are playing at home or on the road. In a similar fashion, pitchers and batters facing their left and right counterparts yield different results. On October 25th the Houston Astros traveled to Washington D.C to meet the Nationals in the third game of the 2019 World Series. Table 2 shows the player by player matchups between the Astros and Nationals. Incorporating the resultant OPS team calculations in Table 2 with Equation (4) yields the following:

$$\text{Probability Houston winning} = .7988^{3.374} / (.7988^{3.374} + .7279^{3.374}) = 58\% \qquad (5)$$

This equation represents a player-based estimation of the probability of winning. Viable data needs to cure before it decays. The most reliable time period for a batters' data was a moving 12-month aggregation while a pitchers' statistics were most representative over a 24-month horizon. Moving from a player-based example let's proceed to look at team-based profiles.

---

[9] Bill James ushered in the Sabermetric era by publishing the book *The Bill James Baseball Abstract*.

[10] Calculation in Appendix C.

Table 2 - OPS Player Match-up Houston at Nationals

		-150	140	-4%	1%
20192723	HOU at WSH	OPS			P(Wh)
		HOU	WSH	Adv	IPW
	Friday, October 25, 2019	0.799	0.728	-9%	42%
		4	1	1%	42%

HOU Best Bet

Pyth Exponent 3.374

PA>= 50

Opposing Pitcher

45578 **Anibal Sanchez** R

Road	Houston Astros					Pitching Away		Pitching Home		Batting Away		Batting Home		Toal PA	Home Pitcher	
															vs LHB	vs. RHB
															0.7172	0.7872
11		Club	Pos	Bats	Throws	vs LHB	vs. RHB	vs LHB	vs. RHB	vs LHP	vs. RHP	vs LHP	vs. RHP		0.8052	0.7970
31734	Zack Greinke	HOU	P	R	R	0.6477	0.6759	0.6178	0.6446	1.0380	0.5904	0.9374	0.5331	127		0.6888
55877	Jose Altuve	HOU	2B	R	R					0.8680	0.8153	1.0057	0.9445	1,147		0.8012
49264	Michael Brantley	HOU	LF	L	L					0.7682	0.9803	0.7913	1.0097	1,268	0.8837	
70607	Alex Bregman	HOU	3B	R	R					1.1750	1.0102	0.9881	0.8495	1,395		0.8987
33829	Robinson Chirinos	HOU	C	R	R					0.7121	0.6392	0.9317	0.8362	863		0.7132
100502	Carlos Correa	HOU	SS	R	R					0.8712	0.8015	1.0668	0.9815	789		0.7943
51408	Yuli Gurriel	HOU	1B	R	R					0.8084	0.7683	0.9280	0.8819	1,185		0.7777
56609	Josh Reddick	HOU	RF	L	R					0.8144	0.6663	0.7635	0.6247	1,037	0.7268	
65992	George Springer	HOU	RF	R	R					1.0115	1.0229	0.9069	0.9171	1,176		0.9050
														9	HOU OPS	0.7988
															P(W) HOU	58%
															OPS Adv	5%

Team's OPS

Opposing Pitcher

31734 **Zack Greinke** R

Home	Washington Nationals					Pitching Away		Pitching Home		Batting Away		Batting Home		Toal PA	ROAD Pitcher	
															vs LHB	vs. RHB
															0.6477	0.6759
30			Pos	Bats	Throws	vs LHB	vs. RHB	vs LHB	vs. RHB	vs LHP	vs. RHP	vs LHP	vs. RHP		0.8086	0.6875
45578	Anibal Sanchez	WSH	P	R	R	0.7529	0.8263	0.7172	0.7872	0.2098	0.4065	0.0885	0.1715	100		0.4237
45398	Asdrubal Cabrera	WSH	3B	S	R					0.6668	0.7524	0.7885	0.8896	1,106	0.7687	
67746	Adam Eaton	WSH	RF	L	L					0.6944	0.8497	0.7454	0.9121	1,026	0.7799	
70755	Anthony Rendon	WSH	3B	R	R					1.0112	0.9684	1.0851	1.0392	1,243		0.8575
104023	Victor Robles	WSH	CF	R	R					0.7261	0.6581	0.8457	0.7665	683		0.7212
107182	Juan Soto	WSH	LF	L	L					0.8283	0.9519	0.9631	1.1068	1,153	0.8773	
49076	Kurt Suzuki	WSH	C	R	R					0.7989	0.6871	0.9263	0.7967	697		0.7363
70917	Trea Turner	WSH	SS	R	R					0.8278	0.8288	0.8601	0.8612	1,309		0.7685
45623	Ryan Zimmerman	WSH	1B	R	R					0.9129	0.5901	0.8665	0.5601	513		0.6180
														9	WSH OPS	0.7279
															P(W) WSH	42%
															OPS Adv.	-9.3%

Source: https://legacy.baseballprospectus.com/sortable/
Tabulated by authors

3.2.2 Matchups: Team-Based Averages

To place a perspective on performance, looking at a team's ability to score and allow runs is fundamental. Using the average runs per game in Table 3 provides details from the 2019 season.

Matching up the Astros and Nationals generates a probability of winning (Astros) by combining estimates of run production from Table 3, relative park factor[11], Equation (1), and derived exponent 1.77 into Table 4.

Table 4 - Estimated Astros Probability of Winning Team Based Data

Team based Probability of Winning Astros vs Nationals October 25, 2019				
	Scored	Allowed by Competing Team	Net EV runs	Adj Park Factor
Astros (road)	5.272	4.741	5.006	4.924
Nationals (home)	5.593	3.926	4.759	4.759
Pythagorean P(W) Astros	$4.924^{1.77} / (4.924^{1.77} + 4.759^{1.77})$ = **51.5%**			

Table 4 demonstrates how the net expected run production is used to calculate the probability of winning.

3.2.3 Matchups: Team-Based Rankings

Identifying the overall comparative advantage in a simple to understand calculation would be helpful in the decision-making process. In Table 3, each team has four rankings, specifically:

- a = Scoring by road team,
- b = Runs allowed by home team,
- c = Scoring by home team,
- d = Runs allowed by road team.

The ranking value will range between 1 and 30 (30 is best) for each team category. From a home team perspective, the measure of competitive advantage becomes:

$$\text{Competitive advantage} = (c - d) - (a - b) \quad (6)$$

In our case:

$$\text{Nationals competitive advantage} = (27 - 30) - (25 - 15) = -13 \quad (7)$$

A negative number means the Nationals are at a disadvantage. Conversely, the Astros competitive advantage is +13. The competitive advantage rank is highly correlated with the winning percentage. A Spearman Coefficient of .96 adds confidence for using this statistic and its derivation is detailed in Appendix D.

[11] Relative park factor levels the p[laying field by scaling the road teams park influence with that of the home team: Park factor Astros / Park factor Nationals = (1.0083 / 1.101) = .9158 http://proxy.espn.com/mlb/stats/parkfactor?order=false

Table 3 - MLB 2019 Scoring Profile by Team

YearMo	(Multiple Items)	ROAD				
	Road Scoring		Home Allowed		Scoring Rank	
Row Labels	Avg RdScr	StdDev RdScr	Avg HmScr	StdDev HmScr	Road Scoring	Home Allowe
Angels	4.667	3.202	5.346	3.046	13	4
Astros	5.272	3.798	3.926	3.126	25	30
Athletics	5.444	4.135	4.506	3.009	26	23
Blue Jays	4.531	3.139	4.914	3.079	8	14
Braves	5.210	3.255	4.642	2.865	24	20
Brewers	4.704	3.136	4.889	3.457	15	16
Cardinals	4.605	3.408	4.543	2.971	12	22
Cubs	4.840	3.494	4.951	3.424	18	11
Diamondbacks	5.111	3.788	4.605	3.559	22	21
Dodgers	5.494	3.568	4.222	3.098	27	27
Giants	5.025	4.071	4.938	3.080	20	12
Indians	4.778	3.708	4.148	3.233	16	29
Mariners	4.827	3.293	5.617	3.659	17	1
Marlins	3.568	3.146	4.852	3.017	1	18
Mets	5.074	3.181	4.901	3.068	21	15
Nationals	5.185	3.432	4.198	2.939	23	28
Orioles	4.568	3.178	5.543	3.248	9	3
Padres	4.568	3.616	5.198	3.393	9	6
Phillies	4.494	3.062	4.926	3.474	7	13
Pirates	4.691	3.145	5.593	3.898	14	2
Rangers	4.395	3.434	4.889	3.785	6	16
Rays	5.013	3.196	4.288	3.238	19	25
Red Sox	5.543	3.539	4.802	2.861	28	19
Reds	4.321	3.247	4.235	3.071	5	26
Rockies	4.136	2.919	5.111	3.402	3	9
Royals	4.288	3.191	4.988	2.893	4	10
Tigers	3.713	2.620	5.113	2.792	2	8
Twins	6.136	3.754	4.444	2.725	29	24
White Sox	4.580	2.743	5.123	3.276	11	7
Yankees	6.210	3.797	5.222	3.237	30	5
Grand Total	**4.834**	**3.422**	**4.822**	**3.221**		

YearMo	(Multiple Items)	HOME				
	Road Allowed		Home Scoring		Scoring Rank	
	Avg RdScr	StdDev RdScr	Avg HmScr	StdDev HmScr	Road Allowe	Home Scoring
Angels	5.370	3.348	4.827	2.957	8	17
Astros	3.975	3.661	6.037	4.143	23	29
Athletics	3.889	2.725	4.963	3.124	27	19
Blue Jays	5.309	3.635	4.432	3.346	10	8
Braves	4.531	3.190	5.346	3.171	21	22
Brewers	4.568	3.691	4.790	2.714	17	15
Cardinals	3.630	2.857	4.790	3.338	29	15
Cubs	3.901	2.764	5.210	3.545	25	21
Diamondbacks	4.568	2.824	4.926	3.442	17	18
Dodgers	3.346	2.486	5.444	3.630	30	24
Giants	4.605	3.145	3.346	2.276	16	1
Indians	3.963	2.799	4.716	2.976	24	14
Mariners	5.358	3.665	4.531	3.283	9	11
Marlins	5.123	3.059	4.025	3.539	12	4
Mets	4.160	3.160	4.691	2.677	22	13
Nationals	4.741	3.420	5.593	3.653	15	27
Orioles	6.568	4.585	4.432	2.898	2	8
Padres	4.543	3.170	3.815	2.330	19	3
Phillies	4.840	3.341	5.062	3.257	14	20
Pirates	5.654	3.665	4.667	3.557	6	12
Rangers	5.951	3.860	5.605	2.836	4	28
Rays	3.790	2.553	4.519	2.613	28	10
Red Sox	5.438	3.337	5.588	3.676	7	26
Reds	4.543	3.241	4.333	3.182	19	7
Rockies	6.716	3.982	6.173	3.438	1	30
Royals	5.728	3.511	4.247	2.777	5	6
Tigers	6.222	3.406	3.519	2.784	3	2
Twins	4.864	3.327	5.457	3.182	13	25
White Sox	5.213	3.542	4.175	2.967	11	5
Yankees	3.900	3.075	5.425	2.889	26	23
Grand Total	**4.834**	**3.422**	**4.822**	**3.221**		

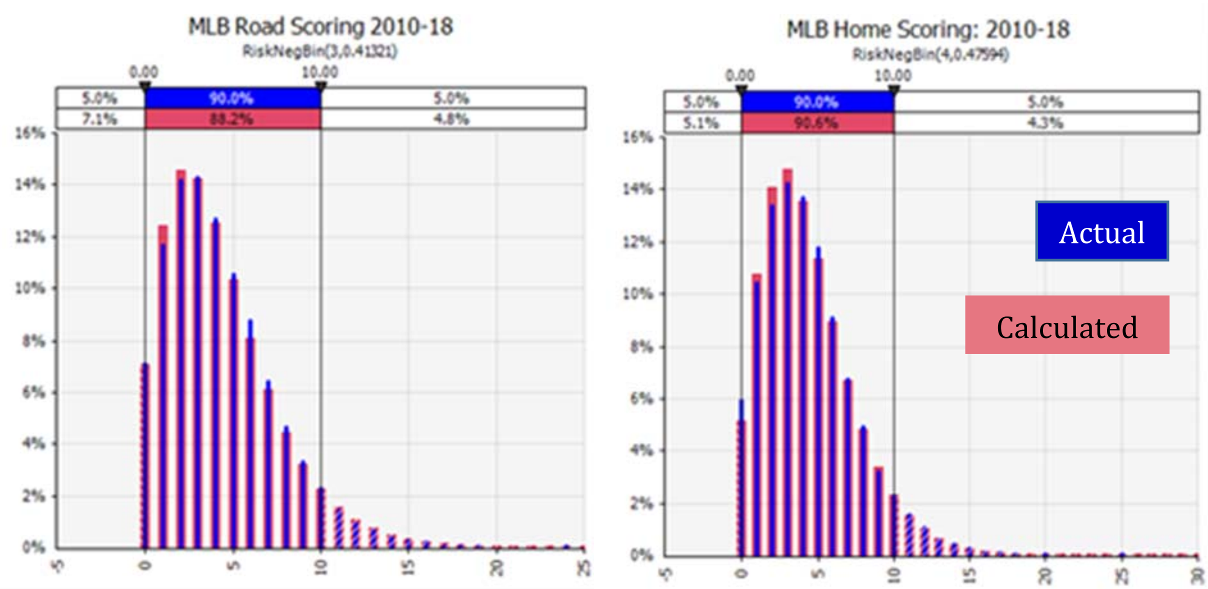
Source: http://www.seanlahman.com/baseball-archive/statistics/
Tabulated by authors

3.2.4 Matchups: Team-Based Monté Carlo

Team Scoring can be represented by an average as in Table 3 but more accurately by density functions. Figure 3 shows the league average runs/game for road and home teams between 2010 and 2018. A negative binomial discrete distribution generates the best fit. As a results, incorporating each team's data requires four distributions:

1. Scoring at home,
2. Runs allowed at home by road team,
3. Scoring on the road,
4. Runs allowed on the road by home team.

Figure 3 - MLB Home and Road Team Scoring Distributions 2010 - 2018



Source: MLB.com
Tabulated by authors

The negative binomial yielded the best fits at the individual team level. Figure 4 identifies examples of the four distributions generated for each team necessary to build the required Monté Carlo matchups. A total of 120 unique distributions are used to matchup each day's competitors. The distributions are updated daily with a rolling six-month time horizon. Again, time-based data is critical enhanced accuracy.

Figure 4 - Negative Binomial Definition Matrix by Team, Road, Home, Runs Scored, and Allowed

Each Club 4 Negative Binomial Distributions

3/21/2019	10/21/2019	
Team & Status	Param1	Param2
Angels-HmAlw	5	0.48406
Angels-HmScr	6	0.55281
Angels-RdAlw	8	0.59636
Angels-RdScr	3	0.38818
Astros-HmAlw	[illegible]	0.3308
Astros-HmScr	3	0.33376
Astros-RdAlw	3	0.43[illegible]
Astros-RdScr	3	0.36019
Athletics-HmAlw	4	0.51104
Athletics-HmScr	6	0.54484
Athletics-RdAlw	5	0.52109
Athletics-RdScr	4	0.40945
Blue Jays-HmAlw	3	0.36192
Blue Jays-HmScr	3	0.40097

Angels at home runs allowed
Angels at home runs scored
Angels on road Runs allowed
Angels on road runs scored

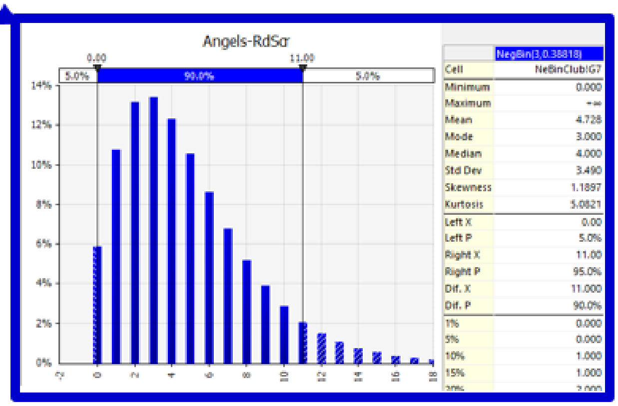
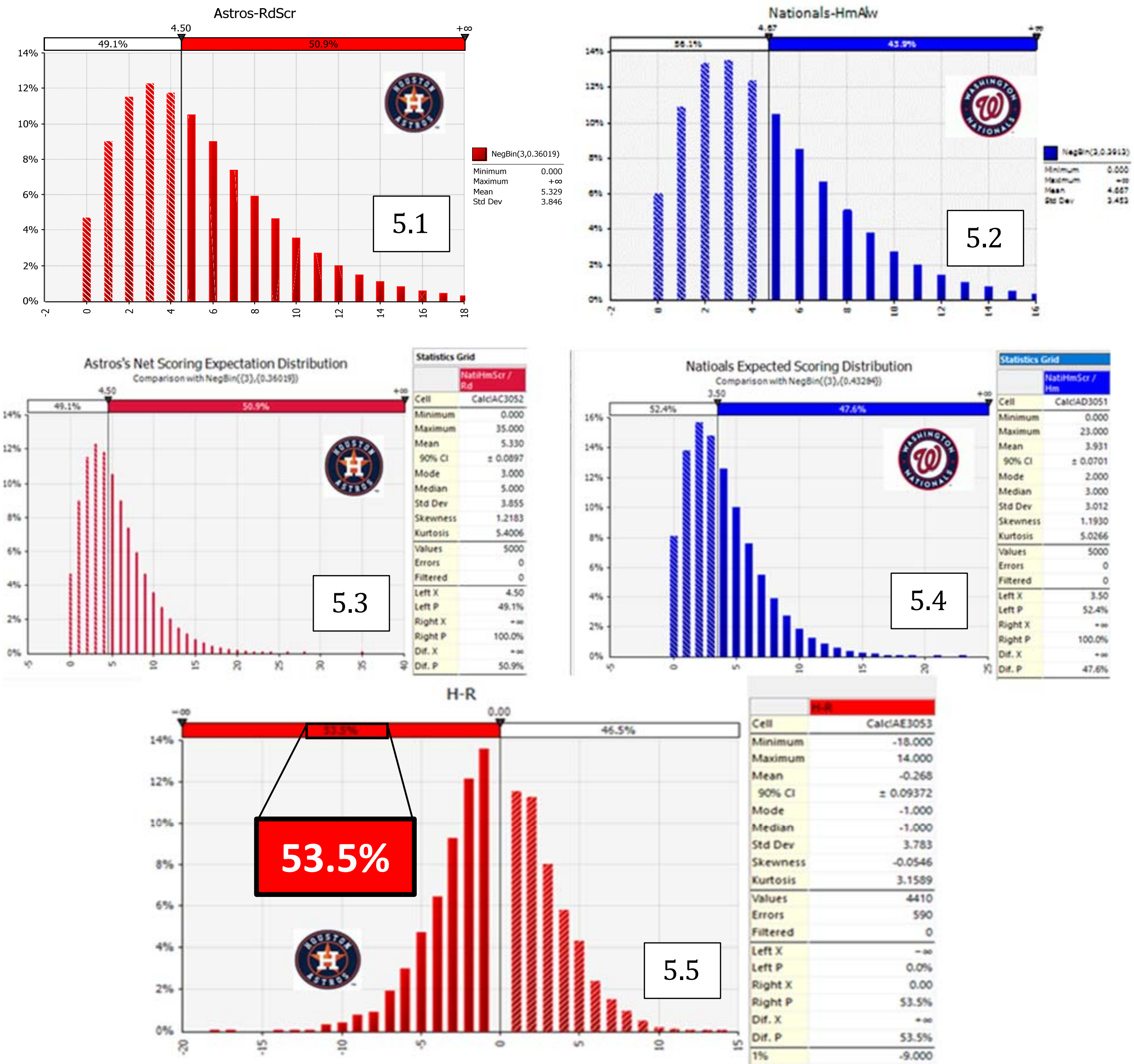


Figure 5 has five density functions representing the game's Monté Carlo matchups including:

5.1 Houston runs scored on the road
5.2 Nationals runs allowed on road
5.3 Combining 5.1 and 5.2 to generate Astros resultant net scoring
5.4 Same as 5.3 but for Nationals net scoring
5.5 Aggregate scoring distribution for each team
Probability of Astros winning 53.5 %

In Appendix E, the 120 negative binomial distributions are specified for each club as in Figure 4.

Figure 5 (5.1-5.5) - Monté Carlo Density Function Matchups: Astros vs. Nationals



The Monté Carlo simulation places a perspective on the scoring characteristics by each team as well as the likely outcome of the game itself. Figure 5.1 matchup with Figure 5.2 to generate Figure 5.3, the resultant scoring of Houston on the road. In a similar fashion Figure 5.4 (resultant Washington scoring distribution) is the consequence of combining the runs scored distribution by Washington and runs allowed distribution by Houston.

Figure 5.5 shows the scoring distributions resulting in probabilities of winning for Houston of 53.5% and Washington 46.5%.

3.2.5 Matchups: Team-Based Logistic Regression

Logistic regression is used when the dependent variable is binary (win-1, loss-0). It is utilized to build the relationship between one dependent binary variable and one or more independent variables.

Four variables functionally represent a games performance from an offense and defensive posture and include:

- (X_1) = Strikeouts / Base on Balls Road,
- (X_2) = Strikeouts / Base on Balls Home,
- (X_3) = OBP + SLG (OPS) Road,
- (X_4) = OBP + SLG (OPS) Home.

Appendix F has the statistical tabulation for the model and its parameters. These are the tools to calculate the probability of winning.

The probability of the home team winning is:

$$P(W)_H = \exp(.224+.066*X_1-.100*X_2-10.984*X_3+10.741*X_4) / (1+\exp(.224+.066*X_1-.100*X_2-10.984*X_3+10.741*X_4)) \quad (8)$$

For the Houston vs. Astros game:

$$X_1 = 2.94 \quad (9)$$
$$X_2 = 2.96 \quad (10)$$
$$X_3 = .778 \quad (11)$$
$$X_4 = .737 \quad (12)$$

$$P(W)_{Nationals} = \exp(.224 + .066*2.94 - .100*2.96 - 10.984*.778 + 10.7841*.737) / (1 + \exp(.224 + .066*2.94 - .100*2.96 - 10.984*.778 + 10.7841*.737)) = 36.6\% \quad (13)$$

$$P(W)_{Astros} = 1- P(W)_{Nationals} = 53.4\% \quad (14)$$

Logistic regression is a powerful methodology that provides a direct calculation of the probability of winning using key operational variables.

3.3 Probability of Winning Summary

A multitude of probabilities of winning for our sample game have been calculated and are shown in Table 5:

Table 5. Probabilities of Winning: Astros vs Nationals October 25, 2019		
Type of Measurement	P(W) Astros	Aggregate P(W) Astros
MICRO		**58.9%**[12]
a. Implied P(W) Astros (-150)	**60.0%**	
b. Batter Pitcher OPS matchups	**57.8%**	
MACRO		**52.8%**[13]
c. Pythagorean – team-based	**51.5%**	
d. Monté Carlo distributions	**53.5%**	
e. Logistic regression	**53.4%**	

The implied probability of winning (a) and batter pitcher OPS matchups (b) are responsive to daily changes. The implied probability of winning is a function of the money line. That line will respond to the market place with change in:

- Pitcher designation,
- Injuries,
- Weather,
- Match up characteristics,
- Batting order,
- Other elements.

Likewise, Batter Pitcher OPS matchups will vary with line-ups and recent player performance history. Aggregating these two measures will provide a ***micro*** or highly responsive estimation of the probability of winning.

In contrast, the Pythagorean (c), Monté Carlo (d), and Logistic regression (e) are all team-based calculations and have a ***macro*** perspective. It is essential that a viable decision model incorporate both ***micro*** and ***macro*** components. Just how that is done will be covered in Section 4.

[12] Simple average of a. and b.
[13] Simple average of c., d., and e.

4. Sizing Wager Optimization

The most critical decision to be made is the level of investment for each wager. Sizing of the economic commitment is determined by a combination of the following factors:

- Probabilities of winning,
- Economics
    - Edge(s),
    - Payoff(s),
    - Return(s) on investment,
    - Risk,
    - Bankroll,
- Hot tips from the Gnome of Zurich.

Just how can these elements be brought together to maximize profits? By incorporating genetic programming, it will be possible handle all non-linear and mixed integer variables, impose filtering constraints, and specify secondary objectives.

The objective function will be to maximized profits with the critical measure to be derived is the percentage of bankroll to be invested.

4.1 Probability of Winning

The probability of winning lays at the heart of any wagering decision. In section 3, five methods for calculation the probability of winning were discussed. Table 5 summarizes the probabilities for a single game and high lights the Micro and Macro characteristics. A primary constraint of the model will be that the probability of winning will be a convex combination of ***micro*** and ***macro*** probabilities (Equation 15) and determined within the model itself.

$$1 = P(W)_{micro} + P(W)_{macro}\ ,\quad \text{where } P(W)_{micro} \text{ and } P(W)_{micro} >= 0 \tag{15}$$

The coefficients of those variables will show the relative impact of a team's aggregate performance versus the market place and direct batter matchups. At this juncture, the introduction of economics becomes the task.

4.2 Edge

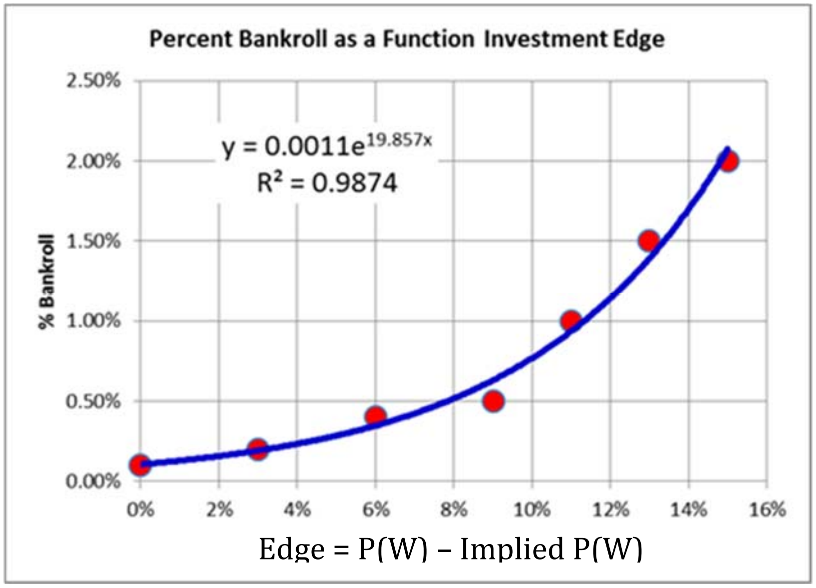
The Edge is defined as:

$$\text{Edge} = \text{Calculated probability of winning} - \text{Implied probability of winning} \tag{16}$$

Joe Peta built a wagering model[14] that calculated percent bankroll to invest with the ***edge*** as the governing variable and is shown in Figure 6.

[14] Peta, Joe, (2013), *Trading Bases,* 1st Edition, Penguin Group, New York, New York

Figure 6 – Percent Bankroll to Invest as a Function of Competitive "Edge"



Source: Joe Peta, Trading Bases, 1st Edition, New York, New York, 2013. Tabulated by authors

Having an ***edge*** is not only intuitively logical, but a real economic mandate. The formulation in Figure 6 will become part of our decision model.

4.3 Payoffs

Knowing the consequence of an investment is a basic requirement and essential in the decision-making process. The games moneyline defines the price of the wager as well as its payoff. From footnotes 1 and 2 (page 2) details of the calculations are provided. The lines (odds) usually change after they are first published. Since the US moneyline is discontinuous, tabulations will first be converted to the decimal equivalents as demonstrated in Appendix G. These changes impact the results of the wager and must be an integral part of the decision model.

4.4 Kelly Criteria (Return on investment based)

In 1956 John Larry Kelly published a seminal work on sizing the bet based upon maximizing the expected growth rate. The original formula was postulated as:

$$f = (bp - q) / b \tag{17}$$

where:

f = fraction of the bankroll to bet
b = net odds received on the wager ("b to 1")
p = probability of winning
q = probability of losing (same as 1 – p)

The b term is a bit cumbersome for those familiar with US nomenclature. Equation (18) can be incorporated as a substitute for the b term.

$$b = (100/ (|\text{US Moneyline}|)^{\text{if(USMoneyline <0 , 1,-1)}} \tag{18}$$

The Kelly Criteria averaged about 13% of bankroll per bet over the 2018 and 2019 seasons. This level of financial commitment is just a bit too risky for most investors. However, the formulation is viable for inclusion in our decision model as long as the influence of the Kelly Criteria is scaled.

4.5 Decision Model

It's time to make sausage.

The model itself is comprised of two parts, Phase I calculates the optimal size of the wager for every game. Phase II selects which games in which to actually invest.

4.5.1 Phase I Decision Model – Sizing Investment

Now the components of our analysis are ready to be integrated and include the following:

Objective function:
Maximize Profits

Subject to:
micro P(W) + ***macro*** P(W) = 1
weighted Kelly criteria
weighted Edge function
weighted Payoff function
management guidelines on bankroll and maximum exposure to risk

Table 6 is an annotated version of the model worksheet. showing the salient variables incorporated into the decision-making process. The critical level of a wager (expressed as a % bankroll) and selected team are noted for each game. From this juncture, each game will be filtered according to our risk sensitive criteria.

Table 6 - Decision Model, League Championship Series 2019

17%	83%												-0.060	-0.080		Filters / Constraints			Max Bet 25,000	Min Line -138
P(W) micro	P(W) macro	Date	Score Rd	Score Hm	Club Select			IP(W)	MoneyLine Rd	MoneyLine Hm	Edge Rd	Edge Hm	Kelly Rd	Kelly Hm	Payout	% Bankroll	Calc Inv	Target	Select	P&L
53%	46%	10/11/19	2	0	WAS	WAS	STL	44%	129	-139	9%	-11%	0.157	-0.255	129%	3.41%	17,067	WAS	17,000	21,930
66%	63%	10/12/19	7	0	HOU	NYY	HOU	61%	144	-154	-5%	3%	-0.080	0.079	65%	2.35%	11,731			
50%	45%	10/12/19	3	1	STL	WAS	STL	44%	-138	128	-4%	2%	-0.097	0.040	128%	3.77%	18,831			
67%	62%	10/13/19	2	3	HOU	NYY	HOU	64%	167	-181	-1%	-1%	-0.010	-0.035	55%	2.58%	12,906			
62%	58%	10/14/19	1	8	WAS	STL	WAS	57%	122	-132	-3%	2%	-0.064	0.036	76%	2.73%	13,661	WAS	13,600	10,303
61%	40%	10/15/19	4	1	NYY	HOU	NYY	40%	-158	148	-5%	3%	-0.120	0.052	148%	4.12%	20,604			
71%	59%	10/15/19	4	7	WAS	STL	WAS	61%	147	-157	-1%	0%	-0.021	-0.009	64%	2.65%	13,257			
48%	40%	10/17/19	8	3	HOU	HOU	NYY	45%	120	-130	13%	-15%	0.244	-0.351	120%	2.95%	14,745	HOU	14,700	17,640
48%	40%	10/18/19	1	4	HOU	HOU	NYY	59%	-142	132	0%	-2%	0.002	-0.033	70%	2.75%	13,740			
42%	62%	10/19/19	4	6	HOU	NYY	HOU	56%	120	-129	-4%	2%	-0.076	0.053	78%	2.71%	13,528	HOU	13,500	10,465
53%	64%	10/22/19	5	4	HOU	WAS	HOU	66%	175	-195	2%	-4%	0.024	-0.118	51%	2.83%	14,138			
60%	64%	10/23/19	12	3	HOU	WAS	HOU	64%	166	-179	-1%	-1%	-0.016	-0.021	56%	2.54%	12,705			
41%	42%	10/25/19	4	1	HOU	HOU	WAS	60%	-150	140	-2%	0%	-0.046	0.003	67%	2.85%	14,256			
57%	39%	10/26/19	8	1	HOU	HOU	WAS	49%	106	-114	9%	-11%	0.184	-0.241	106%	2.85%	14,247	HOU	14,200	15,052
34%	39%	10/27/19	7	1	HOU	HOU	WAS	60%	-149	139	2%	-4%	0.049	-0.063	67%	2.51%	12,530			
59%	64%	10/29/19	7	2	HOU	WAS	HOU	63%	159	-172	-2%	0%	-0.028	-0.003	58%	2.52%	12,582			
54%	64%	10/30/19	6	2	HOU	WAS	HOU	58%	127	-137	-6%	4%	-0.114	0.107	73%	2.42%	12,085	HOU	12,000	-12,000
																			85,000	63,390

Implied Probabil;ity of winning
Moneyline road team
Moneyline home team
Edge road
Edge home
Kelly criteria road
Kelly criteria home
Payout selected team
% Bankroll

Tabulated by Authors

4.5.2 Phase II Decision Model – Risk Filtering

The specific team and bet size will now be narrowed based upon the calculations from Section 4.5.2.

First, the minimum odds line was derived not to fall lower than -138 (shown in Table 6). This is where the level of risk is not economically justified.

Second, the relative competitive advantage (Section 3.2.3) must be below -13 for the road team or above +15 for the home team to qualify for entry into the selection.

Third, maximum size bet is set by management based upon the size of the bankroll and risk tolerances, and exposure to casino warning flags (see Table 6).

Once the afore referenced constraints are satisfied, the investment is calculated and rounded to the nearest $100.

5. Results

Outlined in Section 1 were three objectives for the study:

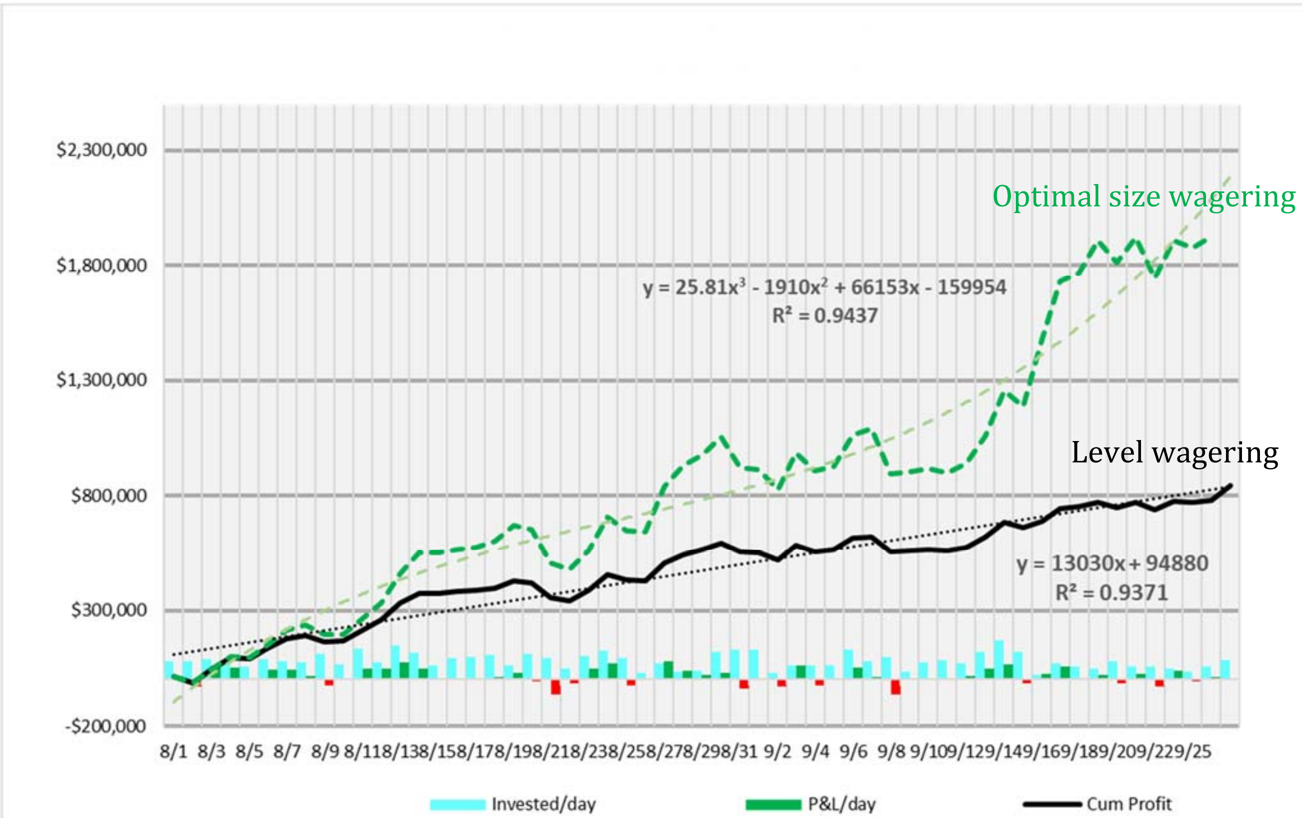
1. **"Identify inequities between the wagering line(s) and each game's expected production,"**
    - Section 2 quantified the over and under valuation for each team (inconsistency between betting odds and actual results).
    - Section 4.2 measured the ***edge*** for each team identifying economic expectations.
    - Section 3.2.1 introduced using OPS to accurately compare both batter-pitcher matchups and team level probability of winning.
    - Quantifying the probability of winning:
        - Five measures of P(W) were used
        - P(W) was aggregated into ***micro*** and ***macro*** components

2. **"Capitalize on such inequities so as to profit from the investment."**
    - Inequities between the market place (odds) and performance (P(W)) were exploited using:
        - Kelly Criteria in Section 4.4,
        - ***Edge*** valuation in Section 4.2
        - Optimizing algorithm of genetic programming Section 4.5.

    - Competitive advantage rankings 3.2.3

3. **"Build an investment model to optimally select and size each wager in order to maximize profits."**
    - Section 4 brings together the elements of:
        - Probabilities of winning
        - Economic consequences
        - Methods of selection and sizing
            - Kelly criteria
            - ***Edge*** based identification
    - Figure 7 shows actual results between 8/1/2019 and 9/25/2019
        - Growing $500K bankroll to over $800K with level betting
        - Demonstrating how applying sizing investment function would improve results through compounding.
    - Table 6 demonstrates how applying investment model to the League Championship series generated a profit $63K while investing $85K while risking just $17!
    - Time dynamics played a roll throughout all aspects of the study.

Figure 7 - MLB 2019 Season: Investment and P&L 8/1-9/25 Level Wagering vs. Applying Wagering Size Model



Tabulated by authors

6. Conclusions

"At gambling, the deadly sin is to mistake bad play for bad luck."

Ian Fleming

Sports gambling's focus is on making a profit, not just picking winners. The importance of accurately accessing the probability of winning is essential for wagering success. Using ΔOPS, dynamic time lines, ***micro / macro*** perspectives, multiple investment tactics honed the source data. Employing multiple analytical tools concurrently with enhanced data generated an investors competitive edge against the "house."

Wagering success followed in the wake of applying the refined data base and complementary analytical methodologies. Coupling responsive data and innovative analytics with sound management constraints (filters) created one of the most profitable baseball models.

The overall approach presented can be applied to baseball management in order to determine:

- Game day player selection
- Player expected performance
- Batting order
- Over and under-valued batters and pitchers
- Matchups to capitalize on players' strength and weaknesses

Appendix H contains may mathematical functions (VBA for EXCEL) that can be helpful in the modeling process.

With the tools presented, it's time: "Bettor Up"!

References

[1] Kelly, J. L. (1956). "A New Interpretation of Information Rate" (PDF). *Bell System Technical Journal*. **35** (4): 917–926.

[2] Lewis, Michael, (2003), *Moneyball: The Art of Winning an Unfair Game*, W.W. Norton & Co., New York

[3] Peta, Joe, (2003), *Trading Bases*, Penguin Group, 1st Edition, New York, New York.

[4] Thorp, Edward, (1962), *Beat the Dealer*, Vintage Books Edition, New York

[5] Mezrich, Ben, (2002), *Bringing down the House*, Atria Paperback, New York

Appendix A

Terms and Nomenclature

Abbreviation	Description
@RISK	bet investment at risk
α	Gamma function shape parameter
β	Gamma function scale parameter
B	Batter as a prefix
color-up	Gambling slang for exchanging for higher valued chips
Δ	Delta variation or change
dawg	southern for underdog
EDGE	betting edge: $P(W_i)$ - IP(Wi)
EV	Expected value as a prefix
EVR_i	expected value runs event i
$EVRA_i$	expected value runs allowed event i
$EVRS_i$	expected value runs scored event i
EVROI	expected value return on investment
EW%	Expected winning percentage
Favorite	Money Line < -100
γ	parameter
Hm	Home team
$\Gamma(\alpha, \beta)$	Gamma distribution
$IP(W_i)$	implied probability of winning for variable i
K	strike outs
K/BB	ratio: strikeouts/base on balls (walks)
MAD	mean absolute deviation
ML	money line
MLB	major league baseball
μ	population mean
N	size of sample data set
n	size of sample subset
MLB	Major League Baseball
$NIP(W_i)$	normalized implied probability of winning for variable i
OBP	on-base percentage
OPS	on-base percentage plus slugging percentage
P	Pitcher as a prefix
O/U	over under
$P(W_i)$	probability of winning for team i
PF_i	park factor field i
r	coefficient of correlation
r^2	r-squared value coefficient of determination
RA_i	Runs allowed team i
Rd	road team
Rs_i	runs scored
s	sample standard deviation
s^2	sample variance
δ	population standard deviation
δ^2	population variance
SLG	slugging percentage (total bases/AB)
Underdog	Money Line > 100
wager	payout for a bet @risk
$\bar{x}$	sample mean
$\mathbf{X}_i$	X variable team i

Appendix B

Coefficient of Correlation: %Win with Selected Variables

Batting			Pitching		
Variable	|%Win|	Rank	Variable	|%Win|	Rank
Δ(OPS)	0.933	1	Δ(OPS)	0.933	1
OBP	0.811	2	Wins	0.851	2
R	0.776	3	ERA	0.793	3
RBI	0.763	4	ER	0.782	4
OPS	0.758	5	OBP	0.764	5
Scr-Al	0.706	6	OPS	0.753	6
SLG	0.692	7	BAA	0.736	7
BB	0.684	8	Scr-Alw	0.706	8
TB	0.670	9	Losses	0.704	9
TPA	0.666	10	SLG	0.695	10
XBH	0.631	11	K/BB	0.656	11
AVG	0.621	12	SHO	0.641	12
HR	0.571	13	SV	0.640	13
H	0.569	14	IP	0.577	14
AB/HR	0.542	15	DIP%	0.568	15
2B	0.427	16	SVO	0.555	16
SO	0.409	17	QS	0.537	17
IBB	0.393	18	RS	0.528	18
SF	0.386	19	SO	0.474	19
FB	0.274	20	K/9	0.450	20
AB	0.256	21	R	0.417	21
HBP	0.222	22	SV%	0.352	22
PH-BA	0.218	23	GP	0.321	23
CS	0.209	24	3B	0.271	24
GDP	0.179	25	P/PA	0.260	25
PH-H	0.177	26	WHIP	0.231	26
SB%	0.161	27	2B	0.226	27
CI	0.148	28	TB	0.209	28
3B	0.137	29	IBB	0.206	29
GB	0.136	30	H	0.190	30
PH-AB	0.125	31	CG	0.190	31
SB	0.029	32	BB	0.189	32
SH	0.014	33	BK	0.182	33
G/F	0.010	34	BLSV	0.174	34
			WP	0.149	35
			CS	0.136	36
			HBP	0.129	37
			CS%	0.083	38
			HR	0.077	39
			ERC%	0.071	40
			SB	0.070	41

Source: http://www.espn.com/mlb/stats/team/_/stat/
Tabulated by authors

Appendix C

Derivation of Exponent for Baseball's Pythagorean OPS Based Theorem

									3.3740	2.6%
Club Year	**TEAM**	**YEAR**	**Win%**	**OPS_s**	**OPS_a**	**$OPS_{(s\text{-}a)}$**	**$Runs_s$**	**$Runs_a$**	**PythCalc**	**ABS(Δ)**
ANA 2016	ANA	2016	0.457	0.726	0.772	-0.046	717	727	0.4484	0.0086
ANA 2017	ANA	2017	0.494	0.712	0.742	-0.030	710	709	0.4652	0.0288
ANA 2018	ANA	2018	0.494	0.726	0.737	-0.011	721	722	0.4873	0.0067
ANA 2019	ANA	2019	0.495	0.775	0.776	-0.001	457	460	0.4989	0.0039
ARI 2016	ARI	2016	0.426	0.752	0.799	-0.047	752	890	0.4490	0.0230
ARI 2017	ARI	2017	0.574	0.774	0.705	0.069	812	659	0.5781	0.0041
ARI 2018	ARI	2018	0.506	0.707	0.696	0.011	693	644	0.5132	0.0072
ARI 2019	ARI	2019	0.505	0.770	0.739	0.031	464	411	0.5346	0.0296
ATL 2016	ATL	2016	0.422	0.705	0.741	-0.036	649	779	0.4581	0.0361
ATL 2017	ATL	2017	0.444	0.738	0.774	-0.036	732	821	0.4599	0.0159
ATL 2018	ATL	2018	0.556	0.742	0.682	0.060	759	657	0.5706	0.0146
ATL 2019	ATL	2019	0.593	0.800	0.746	0.054	491	432	0.5587	0.0343
BAL 2016	BAL	2016	0.549	0.760	0.748	0.012	744	715	0.5134	0.0356
BAL 2017	BAL	2017	0.463	0.747	0.799	-0.052	743	841	0.4435	0.0195
BAL 2018	BAL	2018	0.290	0.689	0.819	-0.130	622	892	0.3582	0.0682
BAL 2019	BAL	2019	0.303	0.704	0.838	-0.134	375	540	0.3571	0.0541
BOS 2016	BOS	2016	0.574	0.810	0.709	0.101	878	694	0.6105	0.0365
BOS 2017	BOS	2017	0.574	0.736	0.711	0.025	785	668	0.5291	0.0449
BOS 2018	BOS	2018	0.667	0.792	0.698	0.094	876	647	0.6050	0.0620
BOS 2019	BOS	2019	0.544	0.807	0.749	0.058	509	451	0.5626	0.0186
CHA 2016	CHA	2016	0.481	0.727	0.744	-0.017	686	715	0.4805	0.0005
CHA 2017	CHA	2017	0.414	0.731	0.786	-0.055	706	820	0.4391	0.0251
CHA 2018	CHA	2018	0.383	0.703	0.761	-0.058	656	848	0.4335	0.0505
CHA 2019	CHA	2019	0.488	0.726	0.798	-0.072	378	449	0.4209	0.0671
CHN 2016	CHN	2016	0.640	0.772	0.632	0.140	808	556	0.6626	0.0226
CHN 2017	CHN	2017	0.568	0.775	0.711	0.064	822	695	0.5722	0.0042
CHN 2018	CHN	2018	0.583	0.744	0.696	0.048	761	645	0.5560	0.0270
CHN 2019	CHN	2019	0.522	0.788	0.732	0.056	455	400	0.5619	0.0399
CIN 2016	CIN	2016	0.420	0.724	0.798	-0.074	716	854	0.4186	0.0014
CIN 2017	CIN	2017	0.420	0.761	0.807	-0.046	753	869	0.4507	0.0307
CIN 2018	CIN	2018	0.414	0.729	0.780	-0.051	696	819	0.4432	0.0292
CIN 2019	CIN	2019	0.471	0.712	0.700	0.012	368	341	0.5143	0.0433
CLE 2016	CLE	2016	0.584	0.759	0.710	0.049	777	676	0.5561	0.0279
CLE 2017	CLE	2017	0.630	0.788	0.673	0.115	818	564	0.6300	0.0000
CLE 2018	CLE	2018	0.562	0.766	0.713	0.053	818	648	0.5602	0.0018
CLE 2019	CLE	2019	0.568	0.739	0.726	0.013	396	369	0.5150	0.0530
COL 2016	COL	2016	0.463	0.794	0.788	0.006	845	860	0.5064	0.0434
COL 2017	COL	2017	0.537	0.781	0.768	0.013	824	757	0.5142	0.0228
COL 2018	COL	2018	0.558	0.757	0.735	0.022	780	745	0.5249	0.0331
COL 2019	COL	2019	0.494	0.779	0.802	-0.023	490	488	0.4755	0.0185
DET 2016	DET	2016	0.534	0.769	0.740	0.029	750	721	0.5324	0.0016
DET 2017	DET	2017	0.395	0.748	0.810	-0.062	735	894	0.4332	0.0382
DET 2018	DET	2018	0.395	0.680	0.761	-0.081	630	796	0.4062	0.0112
DET 2019	DET	2019	0.329	0.675	0.803	-0.128	311	470	0.3576	0.0286
HOU 2016	HOU	2016	0.519	0.735	0.737	-0.002	724	701	0.4977	0.0213
HOU 2017	HOU	2017	0.623	0.823	0.719	0.104	896	700	0.6120	0.0110
HOU 2018	HOU	2018	0.636	0.754	0.640	0.114	797	534	0.6348	0.0012
HOU 2019	HOU	2019	0.633	0.816	0.692	0.124	458	367	0.6356	0.0026
KCA 2016	KCA	2016	0.500	0.712	0.748	-0.036	675	712	0.4585	0.0415
KCA 2017	KCA	2017	0.494	0.731	0.764	-0.033	702	791	0.4628	0.0312
KCA 2018	KCA	2018	0.358	0.697	0.787	-0.090	638	833	0.3990	0.0410

Source: http://www.espn.com/mlb/stats/team/_/stat/
Tabulated by authors

Appendix D

Rank Correlation and Competitive Edge

Spearman Rank Correlation: Ranks of Competitive Advantage vs. Winning Percentage
2018 Regular Season

Club	Scoring Runs		Runs Allowed		Mean Rank	Rank of Means	Win %	Ranks of Win%	$(\Delta \text{Ranks})^2 = d_i^2$
	Hm	Rd	Hm	Rd					
Angels	13	17	15	14	14.75	17	49.4%	17	0
Astros	17	28	30	28	25.75	1	63.6%	2	1
Athletics	16	30	17	26	22.25	6	59.9%	4	4
Blue Jays	15	15	4	7	10.25	23	45.1%	21	4
Braves	23	20	28	13	21	9	55.6%	10	1
Brewers	21	19	20	25	21.25	8	58.9%	5	9
Cardinals	11	26	19	16	18	12	54.3%	13	1
Cubs	20	22	29	17	22	7	58.3%	6	1
Diamondbacks	14	11	24	21	17.5	13	50.6%	15	4
Dodgers	12	29	27	29	24.25	3	56.4%	7	16
Giants	5	1	16	20	10.5	22	45.1%	21	1
Indians	28	21	26	18	23.25	4	56.2%	8	16
Mariners	3	23	13	19	14.5	18	54.9%	12	36
Marlins	2	9	1	23	8.75	25	39.1%	27	4
Mets	1	25	9	26	15.25	15	47.5%	20	25
Nationals	25	16	23	9	18.25	11	50.6%	15	16
Orioles	10	2	2	4	4.5	30	29.0%	30	0
Padres	4	4	11	8	6.75	26	40.7%	25	1
Phillies	18	7	10	15	12.5	20	49.4%	17	9
Pirates	8	17	12	24	15.25	15	50.9%	14	1
Rangers	26	5	14	1	11.5	21	41.4%	23	4
Rays	19	14	18	30	20.25	10	55.6%	10	0
Red Sox	30	26	22	22	25	2	66.7%	1	1
Reds	22	8	8	3	10.25	23	41.4%	23	0
Rockies	27	10	20	6	15.75	14	55.8%	9	25
Royals	9	5	7	2	5.75	29	35.8%	29	0
Tigers	7	3	5	9	6	28	39.5%	26	4
Twins	24	13	6	11	13.5	19	48.1%	19	0
White Sox	5	12	3	5	6.25	27	38.3%	28	1
Yankees	29	24	25	12	22.5	5	61.7%	3	4
								Σd^2	189
								$6 * \Sigma d^2$	1134
								$n/(n^2 - 1)$	26970
									0.958

Sperman Rank Correlation: $r_R = 1 - \left\{ \left(6 * \Sigma_{i=1}^{n} d_i^2 \right) / \left(n / (n^2 - 1) \right) \right\}$

Source: mlb.com
Tabulated by authors

Appendix E

Negative Binomial Parameters by Team, Road, Home, Runs Scored, Runs Allowed

3/21/2019	10/21/2019										
Team & Status	**Param1**	**Param2**	**Team & Status**	**Param1**	**Param2**	**Team & Status**	**Param1**	**Param2**	**Team & Status**	**Param1**	**Param2**
Angels-HmAlw	5	0.48406	Cubs-RdAlw	3	0.37900	Nationals-HmAlw	3	0.39130	Red Sox-RdAlw	6	0.55955
Angels-HmScr	6	0.55281	Cubs-RdScr	3	0.38436	Nationals-HmScr	4	0.41727	Red Sox-RdScr	4	0.42740
Angels-RdAlw	8	0.59636	Diamondbacks-Hm	6	0.57111	Nationals-RdAlw	3	0.42359	Reds-HmAlw	3	0.39089
Angels-RdScr	3	0.38818	Diamondbacks-Hm	4	0.44796	Nationals-RdScr	4	0.43085	Reds-HmScr	3	0.41089
Astros-HmAlw	2	0.33080	Diamondbacks-Rd	3	0.39320	Orioles-HmAlw	3	0.31021	Reds-RdAlw	4	0.48186
Astros-HmScr	3	0.33376	Diamondbacks-Rd	3	0.36641	Orioles-HmScr	5	0.53109	Reds-RdScr	3	0.41026
Astros-RdAlw	3	0.43284	Dodgers-HmAlw	3	0.46667	Orioles-RdAlw	7	0.55894	Rockies-HmAlw	4	0.37178
Astros-RdScr	3	0.36019	Dodgers-HmScr	4	0.42424	Orioles-RdScr	3	0.41197	Rockies-HmScr	7	0.53059
Athletics-HmAlw	4	0.51104	Dodgers-RdAlw	3	0.41311	Padres-HmAlw	3	0.39901	Rockies-RdAlw	4	0.44208
Athletics-HmScr	6	0.54484	Dodgers-RdScr	3	0.34969	Padres-HmScr	9	0.70164	Rockies-RdScr	4	0.49608
Athletics-RdAlw	5	0.52109	Giants-HmAlw	4	0.46927	Padres-RdAlw	5	0.49076	Royals-HmAlw	5	0.46485
Athletics-RdScr	4	0.40945	Giants-HmScr	5	0.60694	Padres-RdScr	3	0.39806	Royals-HmScr	4	0.48164
Blue Jays-HmAlw	3	0.36192	Giants-RdAlw	5	0.50124	Phillies-HmAlw	4	0.45378	Royals-RdAlw	7	0.58507
Blue Jays-HmScr	3	0.40097	Giants-RdScr	2	0.28419	Phillies-HmScr	4	0.44262	Royals-RdScr	3	0.41709
Blue Jays-RdAlw	6	0.55219	Indians-HmAlw	3	0.43310	Phillies-RdAlw	3	0.38015	Tigers-HmAlw	7	0.53406
Blue Jays-RdScr	4	0.47230	Indians-HmScr	5	0.51899	Phillies-RdScr	4	0.46039	Tigers-HmScr	3	0.45570
Braves-HmAlw	4	0.46639	Indians-RdAlw	2	0.32046	Pirates-HmAlw	4	0.41500	Tigers-RdAlw	8	0.60845
Braves-HmScr	6	0.53180	Indians-RdScr	3	0.37500	Pirates-HmScr	3	0.38906	Tigers-RdScr	5	0.57269
Braves-RdAlw	6	0.56335	Mariners-HmAlw	4	0.42564	Pirates-RdAlw	4	0.41837	Twins-HmAlw	4	0.45455
Braves-RdScr	5	0.48969	Mariners-HmScr	3	0.40226	Pirates-RdScr	4	0.45957	Twins-HmScr	6	0.52850
Brewers-HmAlw	2	0.31086	Mariners-RdAlw	4	0.41969	Rangers-HmAlw	4	0.40147	Twins-RdAlw	6	0.57679
Brewers-HmScr	10	0.67645	Mariners-RdScr	3	0.37795	Rangers-HmScr	12	0.68050	Twins-RdScr	4	0.39532
Brewers-RdAlw	4	0.44804	Marlins-HmAlw	6	0.54066	Rangers-RdAlw	3	0.38132	White Sox-HmAlw	3	0.36176
Brewers-RdScr	4	0.46316	Marlins-HmScr	2	0.33401	Rangers-RdScr	3	0.41000	White Sox-HmScr	4	0.49175
Cardinals-HmAlw	3	0.45470	Marlins-RdAlw	6	0.55324	Rays-HmAlw	5	0.57377	White Sox-RdAlw	4	0.44326
Cardinals-HmScr	3	0.39426	Marlins-RdScr	2	0.36321	Rays-HmScr	9	0.66432	White Sox-RdScr	6	0.56552
Cardinals-RdAlw	5	0.52503	Mets-HmAlw	3	0.41157	Rays-RdAlw	3	0.41447	Yankees-HmAlw	3	0.43294
Cardinals-RdScr	3	0.39698	Mets-HmScr	8	0.62938	Rays-RdScr	5	0.49751	Yankees-HmScr	9	0.62963
Cubs-HmAlw	4	0.50774	Mets-RdAlw	6	0.54665	Red Sox-HmAlw	4	0.42932	Yankees-RdAlw	5	0.49771
Cubs-HmScr	4	0.43272	Mets-RdScr	5	0.49820	Red Sox-HmScr	4	0.41997	Yankees-RdScr	5	0.43870

Tabulated by authors

Appendix F

Probability of Winning Calculated through Logistic Regression

Analysis: **Logistic Regression**

Updating: Static

Variable: HwinFlag

Logistic Regression for HwinFlag

Summary Measures	
Null Deviance	14454.87
Model Deviance	6481.25
Improvement	7973.62
p-Value	< 0.0001

Regression Coefficients	Coefficient	Standard Error	Wald Value	p-Value	Lower Limit	Upper Limit
Constant	0.224	0.177	1.261	0.2073	-0.12	0.57
K/BBRd	0.066	0.012	5.357	< 0.0001	0.04	0.09
K/BBHm	-0.100	0.013	-7.455	< 0.0001	-0.13	-0.07
OBP+SLGRd	-10.984	0.249	-44.051	< 0.0001	-11.47	-10.50
OBP+SLGHm	10.741	0.240	44.837	< 0.0001	10.27	11.21

Reliable results

Classification Matrix	1	0	Percent Correct
1	4828	718	87.05%
0	780	4129	84.11%

Consistency between road and home teams

Summary Classification	Percent
Correct	85.67%
Base	53.05%
Improvement	69.48%

Powerful improvement through model

Source: MLB.com

Appendix G

Implied Probabilities of Winning (ImP(W)) Conversions: US Lines, Odds, and Decimal Lines

Gaming Conversions

Favorite					Underdog			
US	Decimal	Fractional	ImpP(W)		US	Decimal	Fractional	ImpP(W)
-100	2.000	1/1	50.0%		100	2.000	1/1	50.0%
-105	1.952	20/21	51.2%		105	2.050	21/20	48.8%
-110	1.909	10/11	52.4%		110	2.100	11/10	47.6%
-115	1.870	87/100	53.5%		115	2.150	100/87	46.5%
-120	1.833	5/6	54.5%		120	2.200	6/5	45.5%
-125	1.800	4/5	55.6%		125	2.250	5/4	44.4%
-130	1.769	77/100	56.5%		130	2.300	100/77	43.5%
-135	1.741	37/50	57.4%		135	2.350	50/37	42.6%
-140	1.714	71/100	58.3%		140	2.400	100/71	41.7%
-145	1.690	69/100	59.2%		145	2.450	100/69	40.8%
-150	1.667	4/6	60.0%		150	2.500	6/4	40.0%
-155	1.645	13/20	60.8%		155	2.550	20/13	39.2%
-160	1.625	5/8	61.5%		160	2.600	8/5	38.5%
-165	1.606	61/10	62.3%		165	2.650	10/61	37.7%
-170	1.588	59/10	63.0%		170	2.700	10/59	37.0%
-175	1.571	4/7	63.6%		175	2.750	7/4	36.4%
-180	1.556	14/25	64.3%		180	2.800	25/14	35.7%
-185	1.541	27/50	64.9%		185	2.850	50/27	35.1%
-190	1.526	53/100	65.5%		190	2.900	100/53	34.5%
-195	1.513	51/100	66.1%		195	2.950	100/51	33.9%
-200	1.500	1/2	66.7%		200	3.000	2/1	33.3%
-210	1.476	12/25	67.7%		210	3.100	25/12	32.3%
-220	1.455	9/20	68.8%		220	3.200	20/9	31.3%
-230	1.435	43/100	69.7%		230	3.300	100/43	30.3%
-240	1.417	21/50	70.6%		240	3.400	50/21	29.4%
-250	1.400	2/5	71.4%		250	3.500	5/2	28.6%
-260	1.385	19/50	72.2%		260	3.600	50/19	27.8%
-270	1.370	37/100	73.0%		270	3.700	100/37	27.0%
-280	1.357	9/25	73.7%		280	3.800	25/9	26.3%
-290	1.345	17/50	74.4%		290	3.900	50/17	25.6%
-300	1.333	1/3	75.0%		300	4.000	3/1	25.0%
-325	1.308	31/100	76.5%		325	4.250	100/31	23.5%
-350	1.286	2/7	77.8%		350	4.500	7/2	22.2%
-375	1.267	27/100	78.9%		375	4.750	100/27	21.1%
-400	1.250	1/4	80.0%		400	5.000	4/1	20.0%
-425	1.235	6/25	81.0%		425	5.250	25/6	19.0%
-450	1.222	2/9	81.8%		450	5.500	9/2	18.2%
-475	1.211	21/100	82.6%		475	5.750	100/21	17.4%
-500	1.200	1/5	83.3%		500	6.000	5/1	16.7%
-550	1.182	9/50	84.6%		550	6.500	50/9	15.4%
-600	1.167	17/100	85.7%		600	7.000	100/17	14.3%
-700	1.143	7/50	87.5%		700	8.000	50/7	12.5%
-800	1.125	12/100	88.9%		800	9.000	100/12	11.1%
-900	1.111	11/100	90.0%		900	10.000	100/11	10.0%
-1000	1.100	1/10	90.9%		1000	11.000	10/1	9.1%

Tabulated by authors

Appendix H (1 / 5)

Useful VBA Functions Applicable for Insertion into an EXCEL Module

```
'Version 19.4

Function EVROI(probwin As Double, Line As Double)

'Keyword - EVROI - Expected Value of Return on Investment
'  probwin - probability of winning, 55
'  line - US line, -150

If Line < 0 Then
   EVROI = (probwin * 100 / Abs(Line)) - (1 - probwin)
   Else
      EVROI = (probwin * Line / 100) - (1 - probwin)
      End If

End Function

Function ALPHA(xbar As Double, sigma As Double)

'Keyword - ALPHA -  first parameter for Gamma distribution
'  xbar - average value from data
'  sigma - standard deviation from data

   ALPHA = (xbar ^ 2) / (sigma ^ 2)

End Function

Function BETA(xbar As Double, sigma As Double)

'Keyword - BETA -  second parameter for Gamma distribution
'  xbar - average value from data
'  sigma - standard deviation from data

   BETA = (sigma ^ 2) / xbar

End Function

Function ROILINE(Line As Double)

'Keyword - ROI return on investment calculated from betting line
'  line - US line, -150

If Line < 0 Then
   ROILINE = 100 / -(Line)
   Else
```

Appendix H (2 / 5)

```
    ROILINE = Line / 100

End If
End Function

Function Dec2US(DecLine As Double)

'Keyword - Dec2US conversion decimal line to US line
'   DecLine - decimal line, 1.65, 2.22

If DecLine < 2 Then
   Dec2US = -100 / (DecLine - 1)
   Else
      Dec2US = (DecLine * 100) - 100
End If
End Function

Function US2Dec(USLine As Double)

'Keyword - US2Dec conversion US line to decimal line
'   USLine - USline, -155, 165

If USLine >= 100 Then
   US2Dec = (100 + USLine) / 100
   Else
      US2Dec = (100 - USLine) / -USLine

End If
End Function

Function Prob2USLine(Prob As Double)

'Keyword - Prob2USLine calculation probability of winning to US line
'   Prob - probability of winning, .56, .44

If Prob >= 0.5 Then
   Prob2USLine = 100 * Prob / (Prob - 1)
   Else
      Prob2USLine = (100 * (1 - Prob)) / Prob
End If
End Function
```

Appendix H (3 / 5)

```
Function Stake(Ab As Double, At As Double, W As Double, X0 As Double, EVROI As Double)

'Keyword - Stake % bankroll, multiplier from EVROI and selected bounder parameters
'  Ab - Lower limit boundry % bankroll .05
'  At - Upper limit boundry % bankroll .20
'  W - width of edge values, .005 - .15
'  X0 - Variable selection point .06
'  EVROI - Expected value return on investment (scaleable)

Stake = Ab + ((At - Ab) / (1 + Exp(-(EVROI - X0) / W)))

End Function

Function LineProbwin(Line As Double)

'Keyword - LineProbWin - Calculate Implied probability of winning from US line
'  Line - US line, -150, 135

If Line < 0 Then
  LineProbwin = Abs(Line) / (Abs(Line) + 100)
  Else
    LineProbwin = 100 / (Line + 100)

End If
End Function

Function Payout(Inv As Double, Line As Double)

'Keyword - Payout - calculate payout given investment and US line
'  Inv - Investment 10000, 500

'  Line - US line, -150, 125

If Line < 0 Then
  Payout = Inv * (100 / -Line)
  Else
    Payout = Inv * (Line / 100)
End If
End Function
```

Appendix H (4 / 5)

```
Function NImProb(Line1 As Double, Line2 As Double)

'Keyword - NImProb, Normalized implied probability of winning, both moneyline values
'  Line1 - moneyline value 1 -150
'  Line2 - moneyline value 2  140


Prob1 = Prob2 = 0

If Line1 < 0 Then
   Prob1 = -Line1 / (-Line1 + 110)
   Else
      Prob1 = 100 / (Line1 + 110)
End If

If Line2 < 0 Then
   Prob2 = -Line2 / (-Line2 + 110)
   Else
      Prob2 = 100 / (Line2 + 110)
End If

NImProb = Prob2 / (Prob1 + Prob2)

End Function

Function Kelly(PW As Double, Line As Double)

'Keyword - Kelly, % bankroll to invest
'  PW - probability of winning .56
'  Line - US line, -150

Kelly = ((PW * US2Dec(Line)) - 1) / (US2Dec(Line) - 1)

End Function

Function IPW(Line As Double)

'Keyword - IPW, Implied probability of winning f(Moneyline)
'  Line, US Moneyline, -150,

If Line < 0 Then
   IPW = Abs(Line) / (Abs(Line) + 100)
   Else
   IPW = 100 / (Abs(Line) + 100)
```

```
End If

End Function

Function LRPW(p0, p1, p2, p3, P4, x1, x2, x3, x4)

'Keyword - LRPW - Logistic regression probability of winning up to 4 variables
'   po,p1... - coefficients from logistic regression
'   x1,x2,...- variable value in location x1,

y1 = Exp(p0 + p1 * x1 + p2 * x2 + p3 * x3 + P4 * x4)

LRPW = y1 / (1 + y1)

End Function
```

Coded by authors