Presentation Overview

- History of baseball statistical analysis (sabermetrics)
- Baseball and actuarial science
- Projection models using actuarial techniques
- Baseball insurance

What is Sabermetrics?

- Sabermetrics: empirical analysis of baseball statistics
  - SABR: Society for American Baseball Research

- Bill James: pioneer of baseball analytics
  - Security guard at pork and beans factory
  - *The Bill James Baseball Abstract* (1977-1988)
  - Hired by Boston Red Sox in 2003
    - Won World Series in 2004, 2007, and 2013

History of Sabermetrics

- Henry Chadwick (Mid-19th Century)
  - Invented the box score
  - Derived formulas for batting average and earned run average (ERA)

- Earl Weaver (1960-1980, Manager of Orioles)
  - Made use of platoon advantages: lefty batter vs. righty pitcher
  - "Pitching, defense, and the three-run homer"
    - Didn't believe in giving up "free" outs

History of Sabermetrics (continued)

- *Moneyball: The Art of Winning an Unfair Game*
  - 2002 Oakland Athletics and the analytical approach to player evaluation
  - Identify undervalued indicators of performance (OBP, SLG) to help small-market club compete in league with no salary cap
  - Playoff appearances despite fraction of payroll

Statistical Analysis

- Scouts vs. Stats
  - Traditional scouting relied on counting stats and physical attributes before computers

A: "Artie, who do you like?"
B: **"**I like Perez. He's got a classy swing, it's a real clean stroke."
A: "He can't hit the curve ball."
B: **"**Yeah, there's some work to be done, I'll admit that."

Statistical Analysis

- Traditional vs. Modern Statistical Measures
  - Hitting: Batting Average vs. BABIP
    - Identify slumps vs. player regression
  - Pitching: Wins vs. WHIP
    - Wins – worst statistic in sports?
    - WHIP – number of baserunners allowed per inning
      - Something the pitcher can control

Baseball and Actuarial Science

- What does baseball have to do with actuarial science?
  - Similar analytical techniques applied to baseball statistics or insurance data
  - Projections of future player performance or reserve development
  - Assigning value to player contracts or policy rates
- John Dewan, FSA: actuary turned statistician
  - Consultant at AON in Chicago
  - Founder of STATS, Inc., sold to Fox Sports
  - Founder of Baseball Info Solutions
  - Author of *The Fielding Bible*

Data Mining

- Techniques to extract information from a data set to determine patterns and enhance understanding
  - Factor Analysis: used to reduce the number of parameter estimates by defining a function that approximates a regression line between multiple variables
  - Cluster Analysis: combines small groups of risk into homogeneous groups called “clusters”; minimize differences within category, maximize differences between categories

Data Mining in Baseball Statistics

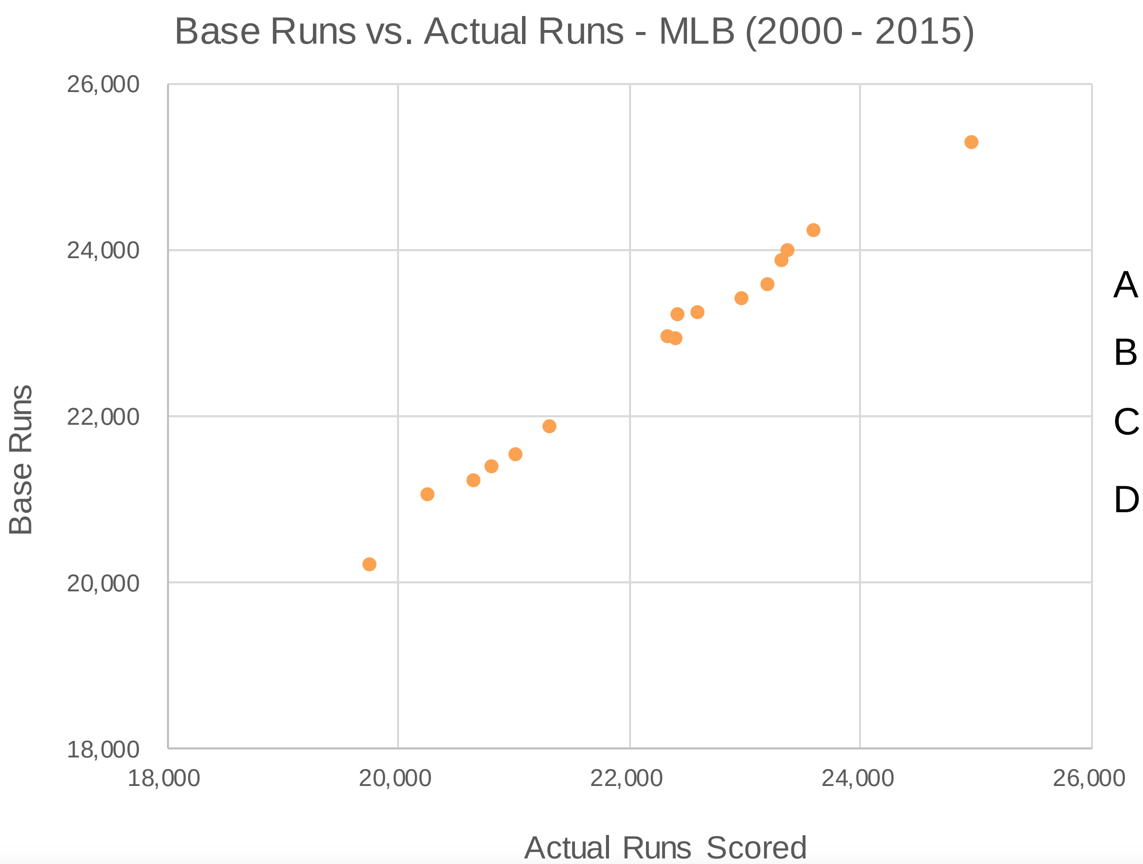
- How does this relate to baseball?
  - Determine which statistics correlate most closely with winning
  - Isolate the most important factors to team performance to more effectively attach values to individual players
  - Want predictive statistics
  - Determining the interdependency of predictor variables in a GLM setting

Generalized Linear Models (GLM)

- GLM: A Method of Multivariate Classification
  - Expresses expected response variable as the sum of linear combination of predictor variables and an error term
  - Considers all predictor variables simultaneously
  - Adjusts for exposure correlations between variables

Generalized Linear Models (GLM) and Baseball

- How does this relate to baseball?
  - Projection models assign weights to specific events



$$\frac{A * B}{B + C} + D$$

A = H + BB - HR

B = (1.4TB - .6H – 3HR + .1BB)*1.02

C = AB - H

D = HR

Credibility

- Predictive or explanatory power assigned to a set of data
  - Credibility (Z)
  - Estimate = Z x Observed Experience + (1-Z) x Related Experience
  - Criteria:
    - $0 \leq Z \leq 1$
    - Z should increase as number of risks increase
    - Z should increase at a non-increasing rate
- Balance credibility vs. responsiveness
  - Too much data: more credible, less responsive
  - Too little data: less credible, more responsive

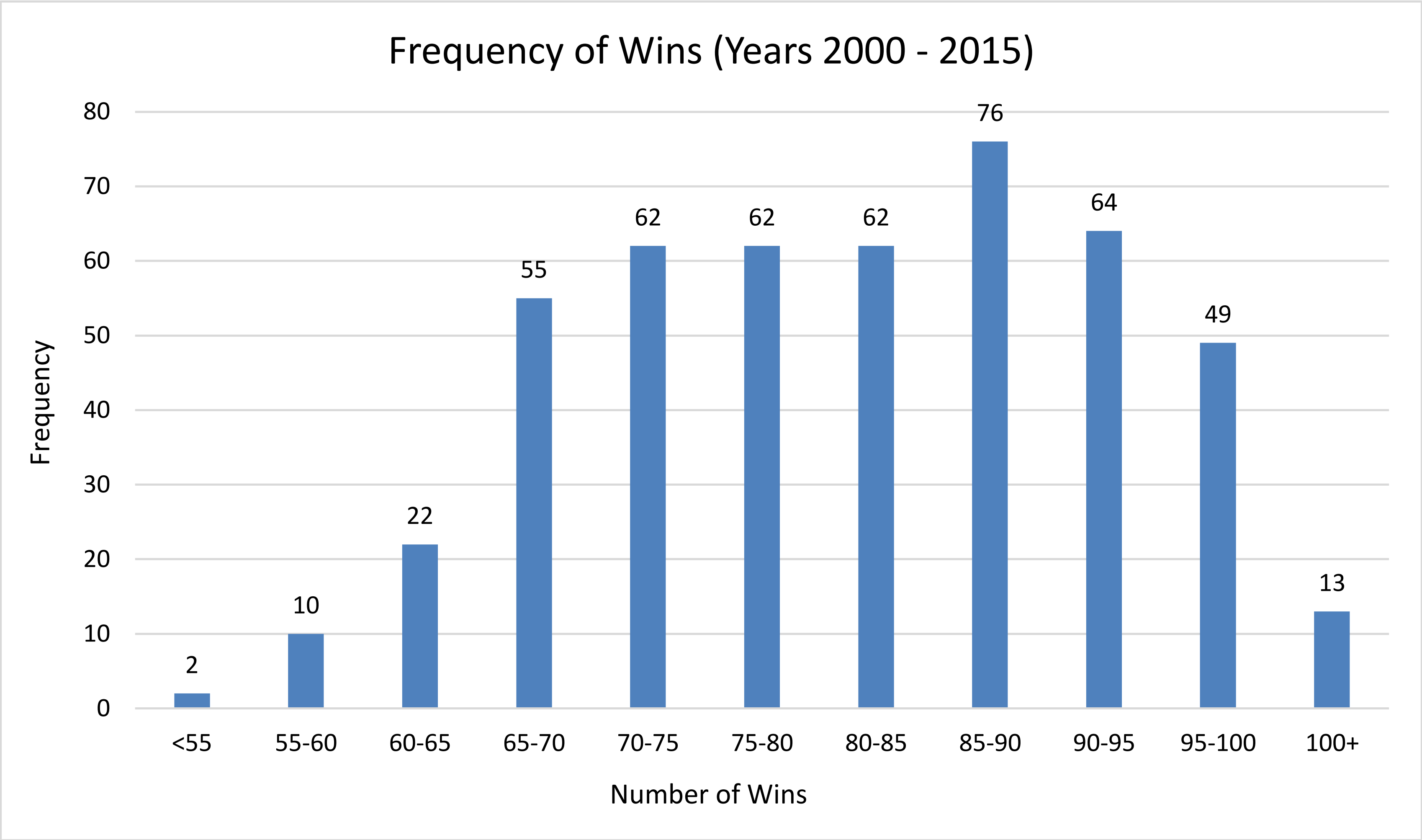
Credibility and Baseball

- How does this relate to baseball?
  - Baseball data is extensive; every pitch generates multiple points of data
  - Managers make decisions on matchups based on recent at-bats which may be limited
  - Consider sample size: clutch hitting

Team Wins Projection Model

- Goal
  - Project 2016 Final Standings
- Projection Methods
  - 3 Types of Win Analyses

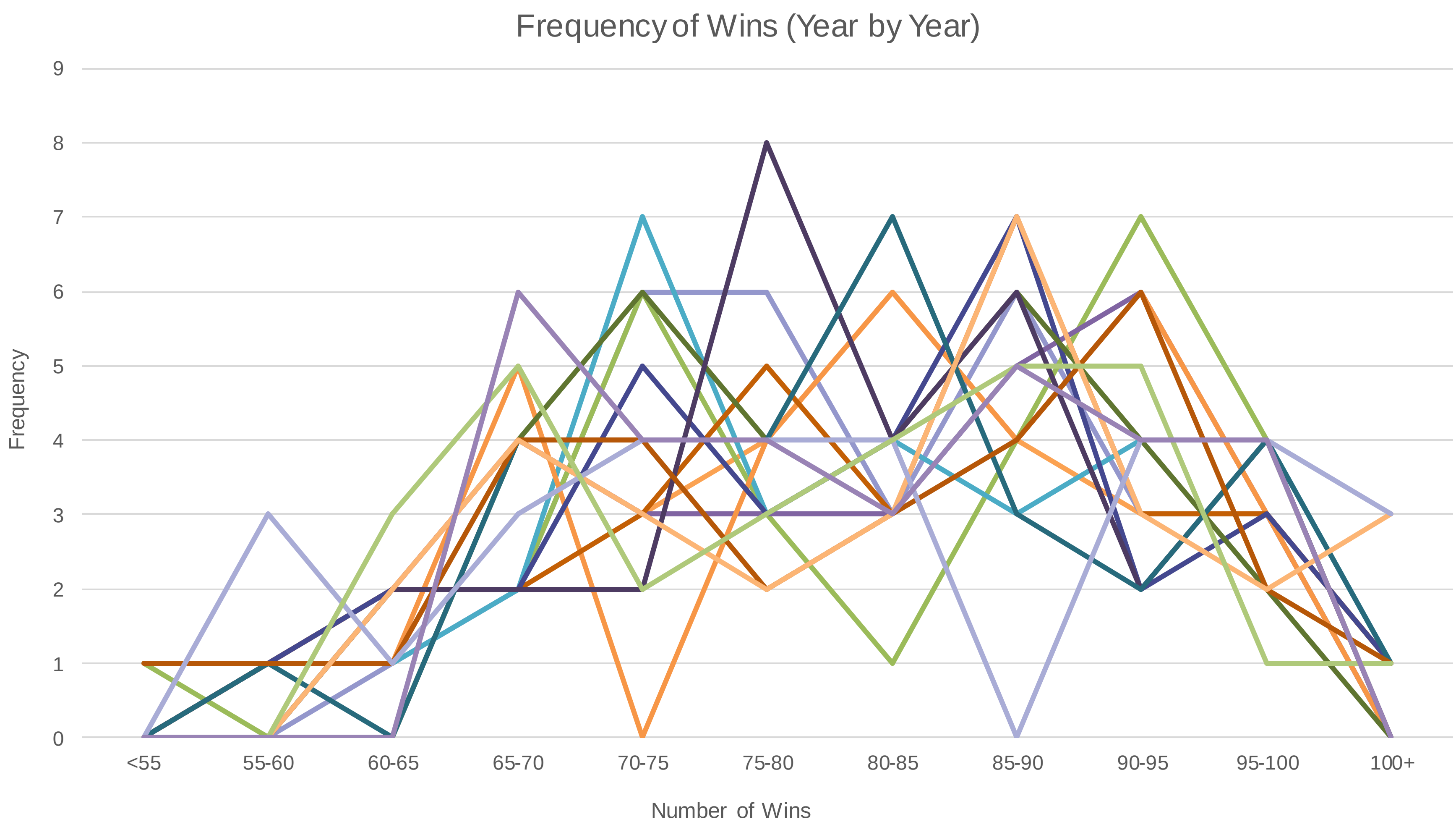
Team Wins Projection Model – Actual Distribution



- Normally distributed, slightly skewed left

PINNACLE ACTUARIAL RESOURCES, INC.

Team Wins Projection Model – Actual Distributions



- Individual years are volatile

Team Wins Projection Model – Choosing Method

- Basic Pythagorean Expectation

$$\text{Exponent} = \left(\frac{R + RA}{G}\right)^{.287}$$

$$\text{Win} = \frac{\text{runs scored}^{1.83}}{\text{runs scored}^{1.83} + \text{runs allowed}^{1.83}} = \frac{1}{1 + (\text{runs allowed}/\text{runs scored})^{1.83}}$$

Peter Brand: "So using this equation on the upper left right here, I'm projecting that we need to **win at least 99 games** in order to make it to the postseason. We need to **score at least 814 runs** in order to win those games and **allow no more than 645**."

$$Win\% = \frac{1}{1 + (\frac{645}{814})^{1.83}} = 60.5\%$$

$$\#\ of\ wins = \ 60.5\% \ \times 162 = 97.99$$

Team Wins Projection Model – Choosing Method

- ELO (Log5) Head to Head Winning Expectation
  - Probability of Team A beating Team B
  - Based on true winning percentages

$$p_{A,B} = \frac{p_A - p_A \times p_B}{p_A + p_B - 2 \times p_A \times p_B}$$

Team Wins Projection Model – Choosing Method

	Basic Pythagorean Expectation				
	Actual - Expected			Win Distribution	
	Sum (Mean)	Average Abs(Error)	Std. Dev Error	Actual Std.Dev	Pythag Std.Dev
2015	(3.313)	4.062	4.731	10.453	10.154
2014	(2.570)	2.883	3.862	9.599	8.876
2013	0.420	3.071	3.713	12.254	12.247
2012	(5.224)	2.877	3.814	11.934	10.667
2011	1.954	3.264	3.973	11.415	10.807
2010	0.016	2.304	2.844	11.005	12.118
Total	(28.425)	3.214			

Over 15 years of results, ELO yielded a MAX error of 4.4 games

	ELO Head to Head - Based on True Winning Percentage				
	Actual - Expected			Win Distribution	
	Sum (Mean)	Average Abs(Error)	Std. Dev Error	Actual Std.Dev	ELO Std.Dev
2015	0.000	1.258	1.594	10.453	11.104
2014	0.000	0.884	1.142	9.599	10.331
2013	0.000	1.018	1.339	12.254	13.074
2012	0.000	1.438	1.790	11.934	12.645
2011	0.000	1.161	1.446	11.415	12.381
2010	(0.000)	1.279	1.610	11.005	11.973
Total		1.267			

Note:

Retroactive

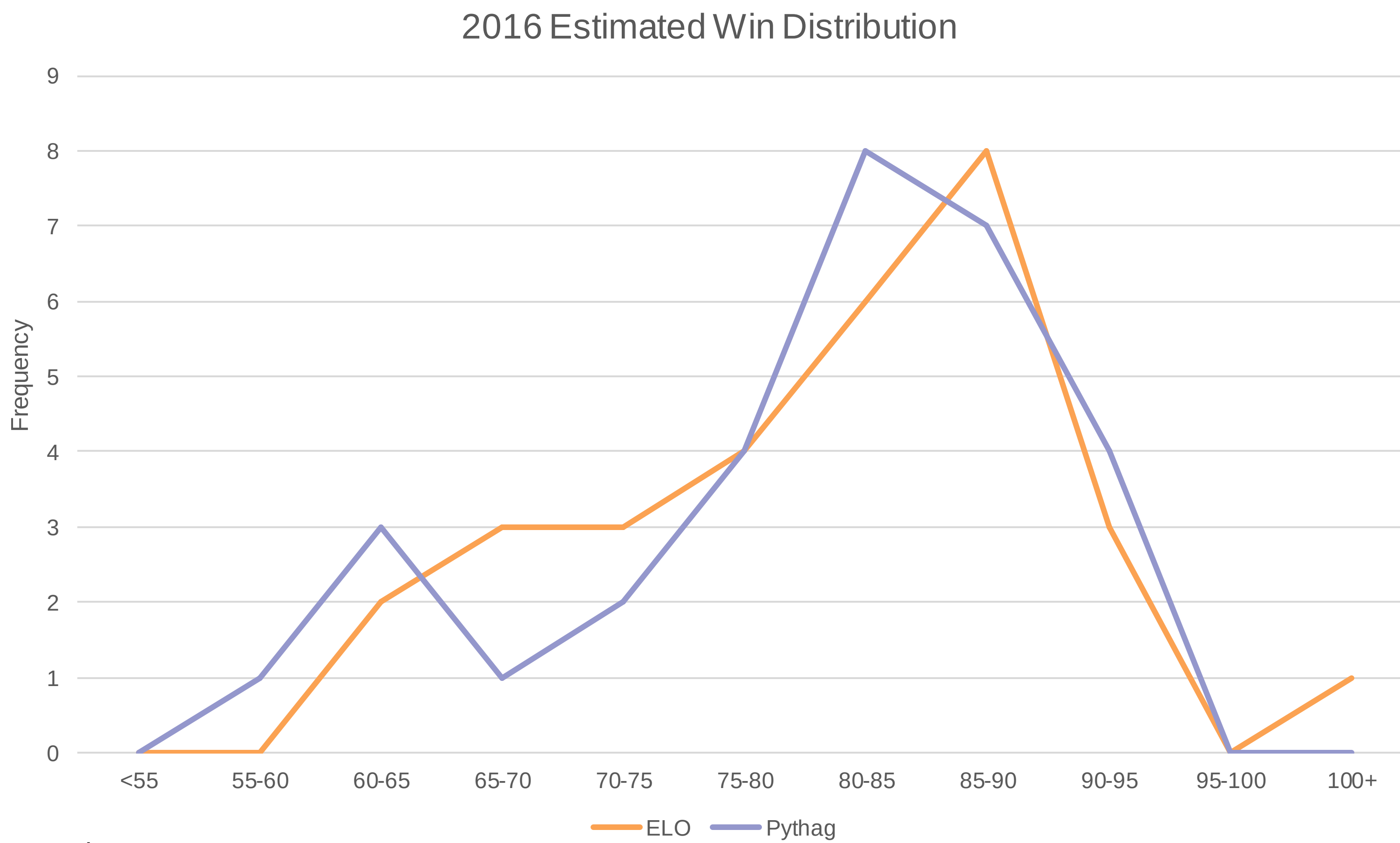
Based on True Win%

Team Wins Projection Model – Choosing Method

- Modified ELO Expectation
  - Use Pythagorean Winning Percentage instead of "True"
  - A "retroactive" estimation

	ELO - Head to Head based on Pythagorean Expectation				
	Actual - Expected			Win Distribution	
	Sum (Mean)	Average Abs(Error)	Std. Dev Error	Actual Std.Dev	ELO Std.Dev
2015	(0.000)	4.148	4.960	10.453	10.784
2014	(0.000)	2.864	3.905	9.599	9.529
2013	0.000	3.246	4.020	12.254	13.060
2012	0.000	3.077	4.177	11.934	11.438
2011	0.000	3.282	3.965	11.415	11.581
2010	0.000	2.826	3.616	11.005	13.110
Total		3.321			

Team Wins Projection Model – Distributions



- Pythagorean Expectation
  - Derived to regress towards mean
  - Prefer ELO for High and Low win totals

Team Projection Model – Calculating Expected Wins

Interleague Schedule Rotation			
	NL Cent	NL East	NL West
2016	AL West	AL Cent	AL East
2015	AL Cent	AL East	AL West
2014	AL East	AL West	AL Cent
2013	AL West	AL Cent	AL East

2016 Games Played Per Division						
	AL Cent	AL East	AL West	NL Cent	NL East	NL West
AL Cent	76	33	33	0	20	0
AL East	33	76	33	0	0	20
AL West	33	33	76	20	0	0
NL Cent	0	0	20	76	33	33
NL East	20	0	0	33	76	33
NL West	0	20	0	33	33	76

Team Projection Model – Calculating Expected Wins

2016 Games Played Per Division						
	AL Cent	AL East	AL West	NL Cent	NL East	NL West
AL Cent	76	33	33	0	20	0
AL East	33	76	33	0	0	20
AL West	33	33	76	20	0	0
NL Cent	0	0	20	76	33	33
NL East	20	0	0	33	76	33
NL West	0	20	0	33	33	76

	Ari	Atl	Bal	Bos	ChC	ChW
Ari	50.00%	60.73%	45.80%	42.34%	46.38%	45.21%
Atl	39.27%	50.00%	35.34%	32.20%	35.87%	34.80%
Bal	54.20%	64.66%	50.00%	46.50%	50.58%	49.41%
Bos	57.66%	67.80%	53.50%	50.00%	54.08%	52.91%
ChC	53.62%	64.13%	49.42%	45.92%	50.00%	48.82%
ChW	54.79%	65.20%	50.59%	47.09%	51.18%	50.00%

Team Projection Model – Reasonability

	R	AB	H	2B	3B	HR
2015	20,647	165,488	42,106	8,242	939	4,909
2014	19,761	165,614	41,595	8,137	849	4,186
2013	20,255	166,070	42,093	8,222	772	4,661
2012	21,017	165,251	42,063	8,261	927	4,934
2011	20,808	165,705	42,267	8,399	898	4,552
2010	21,308	165,353	42,554	8,486	866	4,613
Max	24,971	167,783	45,246	9,197	952	5,693
Min	19,761	165,251	41,595	8,137	772	4,186
Mean	22,186	166,294	43,449	8,678	906	5,000
Last 3	20,221	165,724	41,931	8,200	853	4,585
Last 5	20,498	165,626	42,025	8,252	877	4,648
2016	20,693	162,015	42,304	8,174	943	4,775
Proj Percentile	0.152	0	0.275	0.029	0.826	0.235
Last 3 Percentile	0.062	0.342	0.047	0.049	0.083	0.103
Last 5 Percentile	0.107	0.275	0.061	0.169	0.17	0.182

- Recent years major decline in offense
- Look at ratios of HR/AB, etc.

Vegas Projected Win Totals – Reasonable?

2016 MLB Projected Win Totals

League	Division	Team	Vegas Proj Wins	ELO Proj Wins	Pythag Proj Wins
NL	Central	Chicago Cubs	89.0	87.8	82.9
		St. Louis Cardinals	87.5	92.3	87.0
		Pittsburgh Pirates	87.0	82.6	78.1
		Milwaukee Brewers	71.5	66.1	63.3
		Cincinnati Reds	71.0	67.2	64.2
	East	New York Mets	88.0	100.5	94.2
		Washington Nationals	87.0	85.8	80.7
		Miami Marlins	80.5	79.8	75.2
		Philadelphia Phillies	66.5	64.4	61.4
		Atlanta Braves	65.0	62.7	59.8
	West	San Francisco Giants	90.0	93.3	89.8
		Los Angeles Dodgers	87.0	88.3	85.2
		Arizona Diamondbacks	84.5	79.3	77.0
		San Diego Padres	74.0	72.1	70.5
		Colorado Rockies	68.5	75.0	73.1

Source: VegasInsider as of March 17, 2016

Vegas Projected Win Totals – Reasonable?

2016 MLB Projected Win Totals

League	Division	Team	Vegas Proj Wins	ELO Proj Wins	Pythag Proj Wins
AL	Central	Kansas City Royals	87.0	81.1	84.0
		Detroit Tigers	85.0	85.7	88.2
		Cleveland Indians	84.0	81.9	84.7
		Chicago White Sox	80.5	81.9	84.8
		Minnesota Twins	77.5	73.8	77.4
	East	Toronto Blue Jays	87.0	79.1	84.6
		Boston Red Sox	85.5	84.5	89.5
		New York Yankees	85.0	88.7	93.1
		Baltimore Orioles	80.5	78.2	83.8
		Tampa Bay Rays	78.0	90.3	94.6
	West	Texas Rangers	86.0	86.8	89.6
		Houston Astros	85.5	85.5	88.4
		Seattle Mariners	83.0	65.1	69.9
		Los Angeles Angels	82.5	80.7	84.1
		Oakland Athletics	75.5	89.6	92.1

Source: VegasInsider as of March 17, 2016

Vegas Projected Win Totals – Reasonable?

	Average Wins			Median Wins			Std. Dev. Wins		
	Vegas	Elo	Pythag	Vegas	Elo	Pythag	Vegas	Elo	Pythag
NL Cent	81.2	79.2	75.1	87.0	82.6	78.1	9.1	12.0	10.8
NL East	77.4	78.7	74.3	80.5	79.8	75.2	11.0	15.7	14.3
NL West	80.8	81.6	79.1	84.5	79.3	77.0	9.1	8.9	8.1
AL Cent	82.8	80.9	83.8	84.0	81.9	84.7	3.8	4.4	3.9
AL East	83.2	84.2	89.1	85.0	84.5	89.5	3.8	5.5	4.9
AL West	82.5	81.5	84.8	83.0	85.5	88.4	4.2	9.8	8.8
NL	79.8	79.8	76.2	84.5	79.8	77.0	9.2	11.7	10.7
AL	82.8	82.2	85.9	84.0	81.9	84.8	3.6	6.6	6.3
MLB	81.3	81.0	81.0	84.3	81.9	84.1	7.1	9.4	10.0

	Min Wins			Max Wins		
	Vegas	Elo	Pythag	Vegas	Elo	Pythag
MLB	65.0	62.7	59.8	90.0	100.5	94.6

- Vegas Odds are much more centralized than typical

Conclusions Drawn from Team Projection

- What we know
  - Model is reasonably accurate given the assumption runs and runs against can be predicted
  - If we can predict runs and runs against, we can rely on this projection model

- How to accomplish this
  - Project component stats (runs, hits, etc.)
    - Simulate individual components per player
    - Aggregate components for each team

Player Projection Model - Parameters

- Projected 2016 Home Runs
  - Justin Upton, outfielder for Detroit Tigers
- Adjustment Factors
  - Ballpark Factor
  - Age Factor
  - MLB Trend in Home Runs per At-Bat (HR/AB)
- Credibility-Weighted
  - Full Credibility Standard = 5,000 At-Bats
  - Complement of Credibility – Trended MLB Average HR/AB

Player Projection Model - Data

Exhibit 1

Justin Upton
2016 Home Run Projection Model

Home Runs Per At-Bat

Year	Team	Age	Home Runs (HR)	At-Bats (AB)	Home Runs Per At-Bat (HR/AB)	At-Bats Per Home Run (AB/HR)
	(1)	(2)	(3)	(4)	(5)	(6)
2007	AZ	20	2	140	0.0143	70.0
2008	AZ	21	15	356	0.0421	23.7
2009	AZ	22	26	526	0.0494	20.2
2010	AZ	23	17	495	0.0343	29.1
2011	AZ	24	31	592	0.0524	19.1
2012	AZ	25	17	554	0.0307	32.6
2013	ATL	26	27	558	0.0484	20.7
2014	ATL	27	29	566	0.0512	19.5
2015	SD	28	26	542	0.0480	20.8
Total			190	4,329	0.0439	22.8

Notes:

(1)-(4) Source: baseball-reference.com

(5) = (3) / (4)

(6) = (4) / (3)

Player Projection Model – MLB HR/AB Trend

Exhibit 2

Justin Upton
2016 Home Run Projection Model

MLB Home Runs Per At-Bat Adjustment Factors

Year	Team (1)	Age (2)	MLB Home Runs (HR) (3)	MLB At-Bats (AB) (4)	MLB Home Runs Per At-Bat (HR/AB) (5)	HR/AB Trend (6)	HR/AB Trended to 2016 Level (7)
2007	AZ	20	4,957	167,783	0.0295	0.914	0.0270
2008	AZ	21	4,878	166,714	0.0293	0.923	0.0270
2009	AZ	22	5,042	165,849	0.0304	0.932	0.0283
2010	AZ	23	4,613	165,353	0.0279	0.941	0.0263
2011	AZ	24	4,552	165,705	0.0275	0.951	0.0261
2012	AZ	25	4,934	165,251	0.0299	0.961	0.0287
2013	ATL	26	4,661	166,070	0.0281	0.970	0.0272
2014	ATL	27	4,186	165,614	0.0253	0.980	0.0248
2015	SD	28	4,909	165,488	0.0297	0.990	0.0294
Total			42,732	1,493,827	0.0286		0.0272

Notes:					
		9-Yr Trend	-0.9%	7-Yr Avg	0.0273
(1)-(4)	Source: baseball-reference.com	7-Yr Trend	-0.9%	7-Yr x-H/L	0.0273
(5)	= (3) / (4)	5-Yr Trend	-0.1%	5-Yr Avg	0.0272
(6)	Based on selected trend of -1.0%	4-Yr Trend	-1.2%	4-Yr Avg	0.0275
(7)	= (5) x (6)	3-Yr Trend	2.8%	3-Yr Avg	0.0271
		Selected	-1.0%	Selected	0.0273

Player Projection Model – Adjustment Factors

Exhibit 3

Justin Upton
2016 Home Run Projection Model

Ballpark and Age Adjustment Factors

Year	Team	Age	Ballpark Factor	5.0% Age Factor
	(1)	(2)	(3)	(4)
2007	AZ	20	1.112	0.698
2008	AZ	21	1.068	0.735
2009	AZ	22	1.042	0.774
2010	AZ	23	1.063	0.815
2011	AZ	24	1.095	0.857
2012	AZ	25	1.192	0.903
2013	ATL	26	0.925	0.950
2014	ATL	27	1.122	1.000
2015	SD	28	1.085	0.950

Notes:
(1)-(2) Source: baseball-reference.com
(3) Source: espn.com/mlb
(4) Based on selected trend of 5.0%;
Assumes peak performance at age-27 year

Player Projection Model – Projected HR/AB

Exhibit 4

Justin Upton
2016 Home Run Projection Model

Projected 2016 Home Runs Per At-Bat

Year	Team	Age	Home Runs (HR)	At-Bats (AB)	Home Runs Per At-Bat (HR/AB)	Ballpark Factor	Age Factor	HR/AB Trend	Age-27 Ballpark-Adj Trended HR/AB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2007	AZ	20	2	140	0.0143	1.112	0.698	0.914	0.0168
2008	AZ	21	15	356	0.0421	1.068	0.735	0.923	0.0495
2009	AZ	22	26	526	0.0494	1.042	0.774	0.932	0.0571
2010	AZ	23	17	495	0.0343	1.063	0.815	0.941	0.0373
2011	AZ	24	31	592	0.0524	1.095	0.857	0.951	0.0530
2012	AZ	25	17	554	0.0307	1.192	0.903	0.961	0.0274
2013	ATL	26	27	558	0.0484	0.925	0.950	0.970	0.0534
2014	ATL	27	29	566	0.0512	1.122	1.000	0.980	0.0448
2015	SD	28	26	542	0.0480	1.085	0.950	0.990	0.0461
Total			190	4,329	0.0439	1.076	0.875	0.956	0.0451

	7-Yr Avg	0.0451
	7-Yr x-H/L	0.0468
	5-Yr Avg	0.0444
	4-Yr Avg	0.0422
	3-Yr Avg	0.0478
(10)	Selected	0.0460
(11)	Credibility	93.0%
(12)	Complement of Credibility	0.0273
(13)	Credibility Weighted	0.0447

Notes:

(1)-(4)	Source: baseball-reference.com
(5)	= (3) / (4)
(6)	Exhibit 3, Col (3)
(7)	Exhibit 3, Col (4)
(8)	Exhibit 2, Col (6)
(9)	= (5) / (6) / (7) x (8)
(11)	= [(4) Total / 5,000] ^ 0.5; Assumes full credibility standard of 5,000 at-bats
(12)	Exhibit 2, Col (7) Selected
(13)	= (10) x (11) + {(12) x [1.0 - (11)]}

Player Projection Model – Projected 2016 HR

Exhibit 5

Justin Upton
2016 Home Run Projection Model

Projected 2016 Home Runs

Year	Team (1)	Age (2)	Projected Home Runs Per At-Bat (HR/AB) (3)	2016 Ballpark Factor (4)	2016 Age Factor (5)	HR/AB Trend (6)	Projected 2016 HR/AB (7)
2016	DET	29	0.0447	0.791	0.903	1.000	0.0319

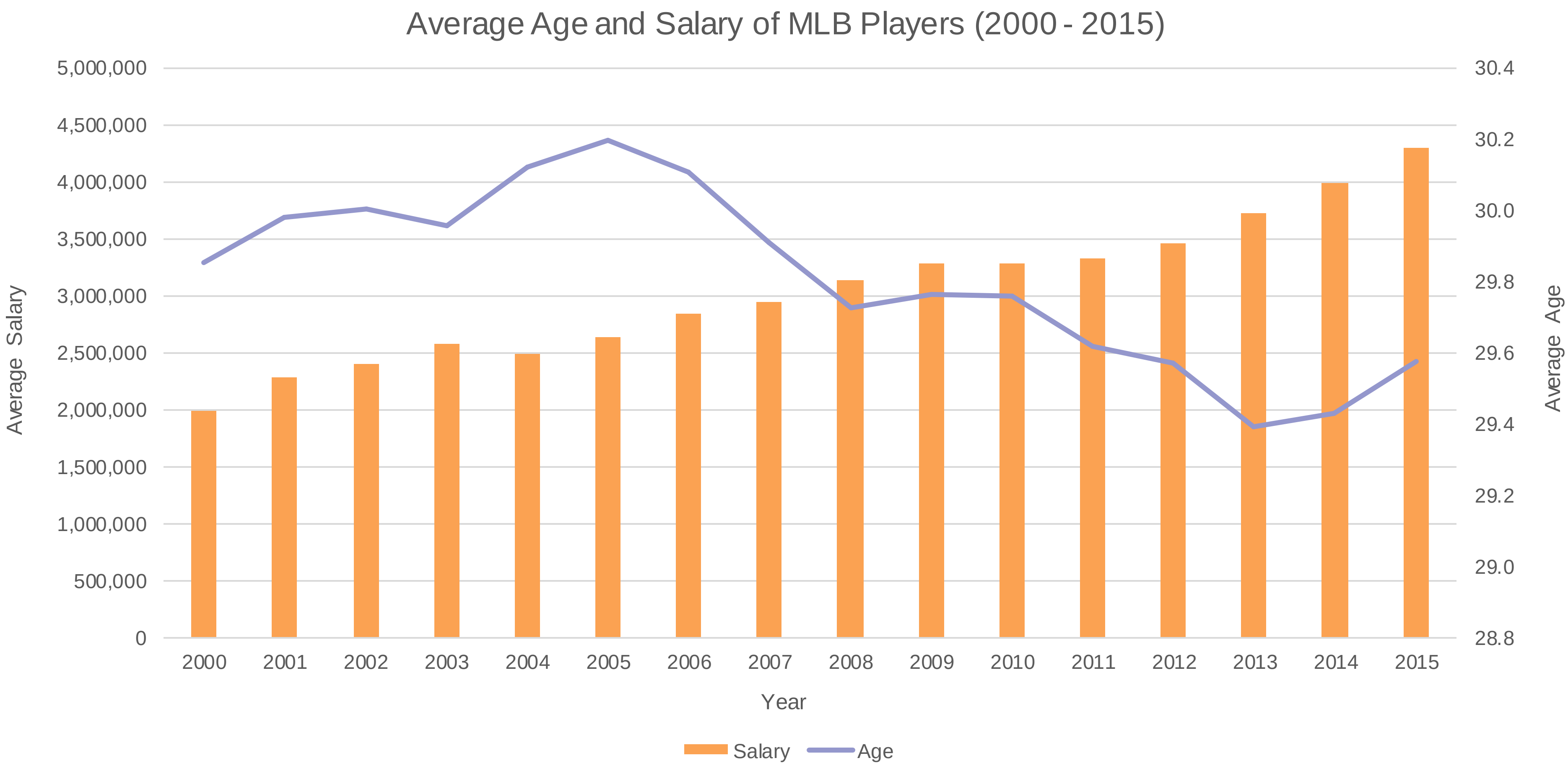
Notes:
(1)-(2) Source: baseball-reference.com
(3) Exhibit 4, Col (13)
(4) Source: espn.com/mlb
(5) Based on Exhibit 3
(6) Based on Exhibit 2
(7) = (3) x (4) x (5) x (6)

At-Bats	Projected Home Runs
450	14
500	16
550	18
600	19

Conclusion

- History of baseball statistical analysis (sabermetrics)
- Baseball and actuarial science
- Projection models using actuarial techniques
- Baseball insurance
- Recognizing trends and instances where projection model is a good/bad indicator of wins
  - Adjustments based on parameters that can identify this
    - Base Runs
    - Scoring rate

Baseball Insurance – Why is it needed?



Baseball Insurance

- Insurance policies purchased by teams of players with long-term contracts
  - Protection against value lost due to injury
  - Written for three-year intervals
  - Cover between 50-80% of contract value, premiums up to 10% of annual contract value
  - May include exclusions for pre-existing injuries
  - Typically include 60-90 day deductible periods to eliminate short-term injuries
- Explosion of player contract values in last 10-15 years
  - Baseball contracts are fully guaranteed, no salary cap
  - Insurance influences playing time, contract negotiations, etc.

Thank you for your attention

Jordan Paszek

309-807-2340

jpaszek@pinnacleactuaries.com

Chris Audino

caudino@ilstu.edu

PINNACLE ACTUARIAL RESOURCES, INC.