Estimating Pitcher Win Loss Record Using Multinomial Logistic Regression and Quantifying Pitcher Luck in Major League Baseball

Aidan Andersen

September 2025

1 Introduction

The pitcher win-loss record is now nearly unanimously acknowledged in the baseball community to be a flawed statistic. Its flaw, of course, is that a starting pitcher's run support and bullpen affects their decision yet is out of their control, and although this obvious flaw has been recognized for much longer [1], only recently have seasons such as Felix Hernandez's 13-12 season in 2010 [2] and Jacob DeGrom's 10-9 season in 2018 exaggerated the role of "luck" in starting pitcher's decisions [3][4] to the point where sportswriters no longer seriously consider win-loss record when evaluating starting pitchers.

However, despite its glaring flaws, the concept of a pitcher winning-loss makes some sense. Good starts put teams in positions to win games, and if team offenses and bullpens *were* identical, then if starter A pitched better than starter B, then almost always team A would win, and we could more confidently base the starter's performance based on the outcome of games they started in.

Comparing this hypothetical universe to the our own (one which is unfair to starting pitchers!) allows us to consider some interesting questions. Given starter A allows R runs over IP innings against an average starter B, what is the probability that team A wins? We can use this hypothetical universe to give us an expectation on how many games starter A *should* have won in a season based on their RA/GS and IP/GS alone. We then can compare this expectation to starter A's *actual* record to give us a measure of how lucky or unlucky they were. This is what

we'll attempt to do in this paper.
One of the advantages of a pitcher win-loss record is that it offers a (as mentioned above, very flawed) way of quantifying how many wins and losses their performance was responsible for, which is an easy-to-understand way of encapsulating the entirety of their performance into one number (as opposed to, say, strikeouts or ERA, which must be looked at with other stats to give context). Indeed, encapsulating the entirety of a player's performance into one number based on wins and losses has been arguably the main objective of sabermetrics with the invention of metrics like Win Shares and WAR. So, if we were to be able to accurately say how many wins and losses we *expect* a pitcher to have, we answer a question that is answered by WAR with the added advantage of not having to compare to the vaguely defined "replacement level" player. We'll consider using our estimate in this way in this paper, too.
Finally, we'll note that, although attempts to correct pitcher win-loss record have been considered in the past (see, for example, Finlan (1989)), we believe our approach here is novel in its axiomatic approach and being based on well studied statistical principles, which we'll explain in the next section.

2 Our Model

We want to answer the following question. Given pitcher A starts a game against an average opposing starter, what is the probability that he receives a win, loss, or no decision? We need quantifiable information about pitcher A to get such an estimate, but luckily for us, deciding exactly *which* statistics to use here to get this estimate is obvious. Since the pitcher's decision is dependent on, by definition, only the amount of runs they allow and the amount of innings pitched, we'll attempt to estimate the relationship between $(RA/GS, IP/GS)$ and the starter's decision.
We'll use regression analysis to answer our question. Specifically, we'll use *multinomial logistic regression* [5] to give us the probabilities of a win, loss, or no decision as a function of $(RA/GS, 3 \times IP/GS)$; That is, we'll assume, if our independent variables $(RA/GS, 3 \times IP/GS)$ are given by the vector X, that we have a probability distribution of the form

$$P(Y = L) = \frac{e^{\beta_0 \cdot X}}{1 + \sum_{k=0}^{2} e^{\beta_k \cdot X}}$$

$$P(Y = W) = \frac{e^{\beta_1 \cdot X}}{1 + \sum_{k=0}^{2} e^{\beta_k \cdot X}}$$

$$P(Y = ND) = \frac{e^{\beta_2 \cdot X}}{1 + \sum_{k=0}^{2} e^{\beta_k \cdot X}}$$

, where $\beta_0, \beta_1, \beta_2$ are unknown vectors of coefficients that we aim to choose in a way such that our above function best fits our data set. The above model is what's known as the (as previously noted) multinomial logistic regression model, and we justify its use here by noting that it is perhaps the most well studied model in cases like ours where we have a *categorical* dependent variable (that is, a dependent variable that takes on finitely many values, which is contrasted with a *continuous* dependent variable, if, for example, we were modeling heart rate as a function of age). We further justify its use by noting that it's specifically seen applications in the realm of sports analytics. See, for example, Benter (2008), who uses the model to predict the outcome of horse races [6].

Now, as with any statistical regression model, there are several ways in which we may estimate our coefficients $\beta_0, \beta_1, \beta_2$. To decide the appropriate method of estimation of coefficients here, we'll note that *in a given start, runs allowed and innings pitched are highly correlated*! A pitcher who lets up a lot of runs in likely to be pulled from the game early, while a pitcher who lets up few runs is likely to be allowed to pitch for longer. As such, we'll use *ridge regression*, which was invented for problems like ours with highly correlated covariates.

We'll use pitching data from Retrosheet [7] and use the GLMNET package in R for our regression, a package which provides support for fitting a dataset to the multinomial logistic model using ridge regression to estimate the coefficients like we want. [8]

For each season, starting in 1901, we'll do the above outlined regression using all the pitching starts that season as our sample data. This will give us coefficients (for season S) β_0^S, β_1^S, β_2^S, which we'll use to estimate a starting pitcher's win loss record *in that season* in a way which we'll illustrate by example.

Say our regression shows us that our coefficients for the 1901 season are β_0^{1901}, β_1^{1901}, β_2^{1901}. If we know that Cy Young in 1901 started 41 games, and in those 41 games started allowed 111 runs and pitched 364 $\frac{2}{3}$ innings, we can compute

$$RA/GS = 26.6829...$$

$$3 \times IP/GS = 2.7073...$$

So we'll take vector $X = (2.7073..., 26.6829...)$ and compute the probabilities

$$P_{1901}(Y = L) = \frac{e^{\beta_0^{1901} \cdot X}}{1 + \sum_{k=0}^{2} e^{\beta_k^{1901} \cdot X}} = 0.2557655650249262$$

$$P_{1901}(Y = W) = \frac{e^{\beta_1^{1901} \cdot X}}{1 + \sum_{k=0}^{2} e^{\beta_k^{1901} \cdot X}} = 0.7162514710357424$$

$$P_{1901}(Y = ND) = \frac{e^{\beta_2^{1901} \cdot X}}{1 + \sum_{k=0}^{2} e^{\beta_k^{1901} \cdot X}} = 0.02798296393933147$$

. Now we multiply by our 41 games started to get

$$est\_wins_{1901} = 41 \times 0.716... = 29.366310312465437$$

$$est\_losss_{1901} = 41 \times 0.255... = 10.486388166021975$$

. We'll use this estimated record to determine what we'll define as *raw luck* in attempt to measure how "lucky" a pitcher was that season.

$$raw\_luck_{season} = (wins_{season} - est\_wins_{season}) + (est\_loss_{season} - loss_{season})$$

. (We justify this is a measure of luck using the following simple reasoning. If a pitcher gets more wins than we'd expect them to based on their other metrics, we might chalk those extra wins up to "luck", and the opposite for losses.) We then can sum over all seasons to give *career* estimates

$$est\_win_{career} = \sum_{seasons} est\_win_{season}$$

$$est\_loss_{career} = \sum_{seasons} est\_loss_{season}$$

$$raw\_luck_{career} = \sum_{seasons} raw\_luck_{season}$$

This now gives us a way to determine the luckiest starting pitching seasons and careers, but we'll note that our estimated wins and losses don't yet allow us to compare pitchers across eras like WAR does, because we don't adjust for the changes over time in the frequency of no decisions. To be able to use estimated wins and losses as an era independent metric of ability, we'll need to consider the estimated win loss percentage.

$$est\_WL\% = \frac{est\_wins}{est\_wins + est\_loss}$$

We're now in a position to answer all the questions outlined in the introduction.

3 Results and Analysis

3.1 Results for Pitcher Luck

Below are tables showing the top ten and bottom careers by raw luck in games started. Note that this is *only* for games started, and the years displayed on the career table are given accordingly. For example, Jim Kaat pitched from 1959 to 1983, but since he was only used as a reliever in the 1983 season, the table lists his terminal season as the 1983 season.

Table 1: Top 10 Careers by Raw Luck (1901-2024)

Pitcher	Wins	Losses	Estimated Wins	Estimated Losses	Raw Luck
1. Earl Whitehill (1923-1939)	208	181	142.22	249.25	134.03
2. Jim Kaat (1959-1982)	261	220	185.94	252.55	107.61
3. Herb Pennock (1912-1934)	222	145	166.18	192.28	103.12
4. Freddie Fitzsimmons (1925-1943)	200	140	150.67	192.60	101.93
5. Whitey Ford (1950-1967)	237	108	178.75	151.53	101.78
6. Burleigh Grimes (1916-1934)	252	196	195.07	237.45	98.38
7. Early Wynn (1939-1963)	288	234	216.86	259.36	96.51
8. Sad Sam Jones (1915-1935)	213	202	159.76	243.87	95.10
9. Mike Torrez (1967-1984)	213	202	159.76	246.87	95.10
10. Jesse Haines (1920-1937)	187	146	137.61	186.99	90.37

Table 2: Bottom 10 Careers by Raw Luck (1901-2024)

Pitcher	Wins	Losses	Estimated Wins	Estimated Losses	Raw Luck
1. Matt Cain (2005-2017)	109	120	127.24	93.60	-44.65
2. Sandy Alcantara (2017-2023)	42	56	58.71	34.48	-38.23
3. Jose DeLeon (1983-1993)	73	110	85.80	86.98	-35.82
4. Jacob deGrom (2014-2024)	88	58	103.86	39.76	-34.09
5. Steve Rogers (1973-1985)	162	152	165.24	125.84	-29.40
6. Curt Schilling (1988-2007)	217	137	214.62	110.82	-23.80
7. Madison Bumgarner (2009-2023)	142	127	136.85	101.22	-20.62
8. Mike Morgan (1978-2001)	127	178	116.69	147.33	-20.35
9. Jeff Samardzija (2009-2020)	70	101	73.78	84.78	-19.99
10. Luis Castillo (2017-2024)	74	78	70.10	55.21	-18.89

The next two tables display the top ten and bottom ten starting pitching seasons by raw luck. Again, relief appearances are not considered.

Table 3: Top 10 Seasons by Raw Luck (1901-2024)

Pitcher	Wins	Losses	Estimated Wins	Estimated Losses	Raw Luck
1. Jack Coombs (1911, Philadelphia Athletics)	26	11	14.33	22.06	22.73
2. Russ Meyer (1953, Brooklyn Dodgers)	15	4	5.59	16.53	21.94
3. Ray Caldwell (1920, Cleveland Indians)	20	11	8.47	20.26	20.78
4. Hooks Dauss (1919, Detroit Tigers)	20	9	9.74	18.87	20.14
5. Bert Husting (1902, Boston Americans/Philadelphia Athletics)	13	6	6.29	18.68	19.39
6. Monte Weaver (1932, Washington Nationals)	20	8	8.27	15.13	18.86
7. Nick Maddox (1908, Pittsburgh Pirates)	22	8	12.12	16.88	18.75
8. Jack Morris (1992, Toronto Blue Jays)	21	9	9.98	16.72	18.73
9. Lee Meadows (1926, Pittsburgh Pirates)	17	9	6.71	17.36	18.65
10. Whitey Ford (1961, New York Yankees)	27	4	17.41	12.92	18.51

Table 4: Bottom 10 Seasons by Raw Luck (1901-2024)

Pitcher	Wins	Losses	Estimated Wins	Estimated Losses	Raw Luck
1. Ed Walsh (1910, Chicago White Sox)	16	18	25.04	8.41	-18.64
2. Fred Glade (1905, St Louis Browns)	6	25	13.73	16.08	-16.64
3. Nolan Ryan (1987, Houston Astros)	8	16	14.14	7.26	-14.88
4. Jacob deGrom (2018, New York Mets)	10	9	19.25	3.52	-14.72
5. Paul Derringer (1933, St Louis Cardinals/Cincinnati Reds)	7	26	11.18	15.61	-14.57
6. Matt Cain (2007, San Francisco Giants)	7	16	13.00	7.58	-14.42
7. Shelby Miller (2015, Atlanta Braves)	6	17	12.01	8.99	-14.01
8. Ben Cantwell (1935, Boston Braves)	3	20	7.81	11.04	-13.77
9. Bob Gibson (1969, St Louis Cardinals)	20	13	25.50	5.10	-13.40
10. Jim Abbott (1992, California Angels)	7	15	12.81	7.60	-13.20

3.2 Results for Overall Estimated Wins and Estimated Winning Percentage

We might consider using our estimated wins and estimated winning percentage stats as metrics for player value, as mentioned in the introduction. Below are tables displaying the career and single season leaders in estimated wins and estimated winning percentage, which might be thought of as representing the most ”valuable” pitching careers and seasons. We note again that considering estimated winning percentage gives us a more context independent way of comparing pitchers across eras.

Table 5: Top 10 Careers in Total Estimated Wins (1901-2024)

Pitcher	Estimated Wins	Actual Wins
1. Walter Johnson	368.80	378
2. Greg Maddux	334.60	366
3. Roger Clemens	332.95	366
4. Grover Alexander	324.78	353
5. Warren Spahn	318.78	361
6. Tom Seaver	304.01	313
7. Christy Mathewson	299.53	348
8. Nolan Ryan	295.36	319
9. Steve Carlton	290.06	334
10. Gaylord Perry	289.85	306

Table 6: Top 10 Seasons in Total Estimated Wins (1901-2024)

Pitcher	Estimated Record	Actual Record
1. Ed Walsh (1908, Chicago White Sox)	31.2-14.7	35-14
2. Jack Chesbro (1904, New York Yankees)	31.1-16.9	38-12
3. Grover Alexander (1915, Philadelphia Phillies)	30.3-10.4	31-10
4. Cy Young (1901, Boston Americans)	29.4-10.5	31-10
5. Joe McGinnity (1901, Baltimore Orioles)	28.8-12.8	31-8
6. Bob Gibson (1968, St Louis Cardinals)	28.7-4.2	24-10
7. Joe McGinnity (1904, New York Giants)	28.6-8.9	29-18
8. Bob Feller (1946, Cleveland Indians)	28.3-18.4	31-16
9. Vic Willis (1902, Pittsburgh Pirates)	28.1-15.5	26-19
10. Jack Coombs (1910, Philadelphia Athletics)	27.6-12.3	30-8

Table 7: Top 10 Careers in Estimated Win-Loss % (1901-2024) (Minimum 100 Decisions)

Pitcher	Estimated W-L%
1. Jacob deGrom (2014-2024)	0.723
2. Clayton Kershaw (2008-2024)	0.710
3. Pedro Martinez (1992-2009)	0.696
4. Johan Santana (2000-2012)	0.676
5. Curt Schilling (1988-2007)	0.659
6. Roger Clemens (1984-2007)	0.645
7. Roy Halladay (1988-2013)	0.642
8. Roy Oswalt (2001-2013)	0.640
9. Gerrit Cole (2013-2024)	0.634
10. Chris Sale (2012-2024)	0.634

Table 8: Top 10 Seasons in Estimated Win-Loss % (1901-2024) (Minimum 15 Decisions)

Pitcher	Estimated W-L%
1. Dwight Gooden (1985, New York Mets)	0.878
2. Greg Maddux (1994, Atlanta Braves)	0.874
3. Bob Gibson (1968, St Louis Cardinals)	0.871
4. Greg Maddux (1995, Atlanta Braves)	0.867
5. Jacob deGrom (2018, New York Mets)	0.845
6. Roger Clemens (1997, Toronto Blue Jays)	0.840
7. Pedro Matrinez (1999, Boston Red Sox)	0.834
8. Bob Gibson (1969, St Louis Cardinals)	0.833
9. Jason Schmidt (2003, San Francisco Giants)	0.826
10. Zack Greinke (2015, Los Angeles Dodgers)	0.825

3.3 Open Questions

We might have hoped that our statistical analysis would also allow us to determine how much of a pitcher's decision is under their control. That is, how much of wins and losses are attributable to "luck" (factors outside the starter's control) vs the performance of the starting pitcher. To do this, we'd need to do what's called testing the *significance* of the relationship between our covariates RA/GS, $3 \times IP/GS$ and our independent variable. Unfortunately,testing for significance of coefficients in a multinomial logistic regression is difficult, and there is no easy extension of significance tests for other models which is widely accepted here. Additionally, we might want to consider questions of goodness of fit. That is, how well does the logistic function fit our data set? Again, unlike for other models,there is no *universally* agreed upon way of testing for goodness of fit here, although unlike for significance testing, there are goodness of fit measures here that can and have been used. The GLMNET package, which we used here, uses *deviance* to measure goodness of fit. We'll note its omission from the analysis because we feel that, without a way of testing for significance of coefficients, that goodness of fit alone tells us little about the relationship between pitcher wins and luck.

As far as our measure of luck is consider the fact that, because our raw luck stat is cumulative, pitchers with more starts will have more opportunities to accrue luck. Might we consider scaling raw luck based on number of decisions to give us a new measure of luck?

Finally, we'll consider again the question of potentially scaling our estimated win statistic so that it may better be a measure of pitcher value. What might be a proper scaling based on era, as to be able to use a "scaled estimated wins" statistic as a context independent measure of player value? Additionally, we may consider scaling estimated winning percentage by games started.

4 Data Availibility

The complete data set will be available here (personal website – omitted for review purposes), as will the coefficients obtained from our regression.

5 Notes

1 Finlan, Steve, "Evaluating Pitchers' Won-Lost Records," *The Baseball Research Journal* (1989).

2 All player statistics and career information (teams played for and years played) throughout this paper is obtained from Baseball Reference, www.baseball-reference.com

3 Neil Greenberg, "Jacob deGrom is one of the unluckiest aces in baseball history." *Washington Post*, April 13, 2021. Accessed September 10, 2025:

https://link.gale.com/apps/doc/A658322886/AONE?u=nysl_oweb&sid=sitemap&xid=884dd6d

4 David Schoenfield, “Numbers Don’t Lie: Felix Hernandez Is Unluckiest Pitcher of All Time.” ESPN, April 11, 2016. Accessed September 10, 2025: https://www.espn.com/blog/sweetspot/post/_/id/69650/numbers-dont-lie-felix-hernandez-is-unluckiest-pitcher-of-all-time.

5 David Hosmer Jr, Stanley Lemeshow, and Rodney X. Sturdivant. *Applied logistic regression* (John Wiley & Sons, 2013).

6 William Benter, ”Computer based horse race handicapping and wagering systems: a report.” In *Efficiency of racetrack betting markets*, pp. 183-198. 2008.

7 Data is available at www.retrosheet.org

8 Trevor Hastie, Junyang Qian, and Kenneth Tay. “An Introduction to ‘glmnet‘.” An Introduction to ‘glmnet‘ • glmnet, May 5, 2025. Accessed September 10, 2025 https://glmnet.stanford.edu/articles/glmnet.html.