Predicting 2022 MLB Free Agents Market Price

Nicholas Kondo

Introduction

The purpose of this paper is to establish a market price for each 2022 free agent starting pitcher by using statistics in the years prior to a player signing a contract. The second objective is to introduce park factors to the model and adjust players accordingly.

To understand how much players should be valued at, we must first understand what inclines a front office to sign a player. There are many factors and variables that go into making the decision to sign a Major League Baseball player, and the decisions have become more and more critical over the years. The average baseball salary is up over 20,000% since the first Collective Bargaining Agreement in 1968. (1) As players are paid more and more each year, these decisions have turned from ten-thousand-dollar decisions to ten-million-dollar decisions. Baseball has turned into a business.

To find a market price on each player, we will conduct linear regression analysis by taking the 3 years of statistics prior to signing and testing which stats are most likely correlated with annual value. In other words, we want to see what statistics General Managers are paying the most for. In reality, there is no market price in the MLB. I talk about this in detail in my writing, "The Economic Inequality Problem in Major League Baseball," but there is no market price in baseball there is a huge wealth gap among teams. The Yankees will throw $30 million a year at a player that no one else in the league would ever spend. Therefore, there is no true market price in baseball because a wealthy team like the Yankees are willing to spend more than a team like the Rays, who operate on a tight budget.

Procedure

All performance data was collected from Baseball-Reference.com, all salary data was obtained from Spotrac.com, and park factors were taken from baseballsavant.com. First, we took every active contract in Major League Baseball and populated it with 3 years of statistics. Starting pitchers were filtered, and players still on rookie contracts were removed. All 'total stats' such as strikeouts, innings pitched, opponent hits, were all adjusted in 2020 to a 162-game season. A 3-year weighted average was calculated where the season before signing carried the most weight, while the season 3 years before signing carried less weight. If a pitcher didn't throw at least 50 innings (162-game season), then that season was thrown out for them and wasn't calculated into the 3-year average. If a pitcher didn't throw more than 50 innings in more than 2 out of 3 of their prior years to signings, they were removed.

A correlation analysis was ran on every basic, advanced, and value statistic that baseball-reference lists. We found that the 3-year weighted average of Strikeouts, WAR, opponent OBP, and the previous year's salary were the most highly correlated stats with average annual value (AAV). To accommodate with park factor adjustments, strikeouts, opponent OBP, and salary were used in a regression to formulate an algorithm for market price.

AAV = [Strikeouts] * $97,522 – [OBP] * $83,487,339 + [Salary] * $0.43 + $21,255,354

Reminder: These variables are 3-year weighted averages

This model generated an adjusted R Square of .76, which means that 76% of the differences in AAV can be explained by the differences in strikeouts, OBP, and salary. If we wanted to isolate these variables, for every strikeout a pitcher records, on average, their AAV increases by $97,522, holding everything else constant. On average, for every point a pitcher's opponents OBP decreases, their AAV increases by $83,487.

Park Factors

Coors Field has long been known as a park for offensive slugfests and pitcher's nightmare. To fairly value a player, park factors are needed, because all 30 ballparks have a different impact on the game. MLB is the only sport where none of the playing fields are the same. In the NHL, NBA, and NFL, there are things that might make the stadiums feel different than one another, but the measurements are the same. Yes, the bases are all 90 feet apart and the pitcher's mound is still 60 feet, 6 inches, but the fences all have different dimensions and heights. This brings up the problem that we now need to bring in an additional variable when evaluating the game. (https://library.fangraphs.com/principles/park-factors/)

It's not only the fences that make a difference in the game. A short porch doesn't automatically make a stadium a hitter's park, and deep fences don't always favor pitchers. The weather, air density, and topology of the surrounding area matters. The ball tends travel better in warm and thin air, and surrounding building structures can influence how well the ball carries. For example, Petco Park has a marine layer that doesn't let the ball carry. On the other hand, Coors Field is way above sea level, where the air is thin, allowing the ball to fly.

The chart below shows park factors for each ballpark on a 3-year rolling average from baseballsavant.mlb.com. Each number is set so that 100 is the average for that metric, and the park specific number is generated by comparing the frequency of that metric in the selected park compared to the performance of those players in other parks.

Team	Venue	Team	Year	Park Factor	wOBACon	BACON	R	OBP	H	1B	2B	3B	HR	BB	SO	PA
Rockies	Coors Field	COL	2019-2021	115	113	112	132	112	119	115	120	233	121	94	84	14573
Red Sox	Fenway Park	BOS	2019-2021	107	107	108	114	107	107	102	136	111	92	107	97	16232
Reds	Great American Ball Park	CIN	2019-2021	106	109	106	112	103	104	101	100	49	129	103	105	15762
Nationals	Nationals Park	WAS	2019-2021	104	104	102	108	103	104	102	106	93	114	101	95	15897
Orioles	Oriole Park at Camden Yards	BAL	2019-2021	104	104	101	108	102	103	100	98	98	124	96	95	14797
Phillies	Citizens Bank Park	PHI	2019-2021	103	104	103	106	101	102	103	90	133	114	99	100	15518
Braves	Truist Park	ATL	2019-2021	103	104	104	106	103	104	104	107	84	102	101	100	15829
Angels	Angel Stadium	LAA	2019-2021	103	104	103	106	102	102	103	96	91	108	105	103	15556
D-backs	Chase Field	ARI	2019-2021	101	102	103	102	102	103	103	111	139	87	99	100	15946
Indians	Progressive Field	CLE	2019-2021	101	103	103	102	101	101	98	113	74	101	101	103	15811
Royals	Kauffman Stadium	KC	2019-2021	100	98	102	100	102	108	112	118	129	76	85	90	14849
White Sox	Guaranteed Rate Field	CHW	2019-2021	100	99	97	100	99	95	94	83	47	125	105	101	14905
Astros	Minute Maid Park	HOU	2019-2021	99	100	99	98	99	98	99	89	103	105	104	103	15345
Brewers	American Family Field	MIL	2019-2021	99	100	99	98	99	96	95	89	102	106	104	106	16730
Pirates	PNC Park	PIT	2019-2021	99	97	98	98	100	99	94	125	111	84	103	95	15453
Dodgers	Dodger Stadium	LAD	2019-2021	99	101	99	98	96	100	97	95	50	124	87	104	15466
Yankees	Yankee Stadium	NYY	2019-2021	98	98	97	96	98	96	97	85	63	105	104	103	14613
Twins	Target Field	MIN	2019-2021	98	96	97	96	98	98	99	105	96	90	101	99	16278
Tigers	Comerica Park	DET	2019-2021	98	96	97	96	98	99	100	95	239	84	98	96	14225
Cubs	Wrigley Field	CHC	2019-2021	98	99	99	96	99	98	99	92	154	97	105	102	16179
Blue Jays	Rogers Centre	TOR	2019-2021	98	99	96	96	97	97	94	94	96	114	95	100	10133
Mets	Citi Field	NYM	2019-2021	97	97	96	94	97	94	94	89	58	104	104	106	15724
Marlins	loanDepot park	MIA	2019-2021	97	97	99	94	99	98	101	97	130	84	100	102	15364
Giants	Oracle Park	SF	2019-2021	95	95	98	90	97	96	100	95	146	75	99	105	16103
Cardinals	Busch Stadium	STL	2019-2021	95	91	94	90	97	97	103	91	101	81	98	93	14242
Mariners	T-Mobile Park	SEA	2019-2021	94	95	96	88	95	94	99	85	43	97	98	107	14407
Padres	Petco Park	SD	2019-2021	94	94	94	88	96	92	93	94	55	91	107	104	16834
Rays	Tropicana Field	TB	2019-2021	94	96	97	88	95	95	95	100	118	84	92	108	16486
Athletics	Oakland Coliseum	OAK	2019-2021	94	93	95	88	96	96	100	101	73	82	94	98	14937

For example, the Rockies have an '84' strikeout metric, which means that for batters and pitchers who both played at Coors Field and elsewhere, 16% less strikeouts were recorded at Coors Field. The OBP metric at Coors Field is 112— which means for batters and pitchers who both played at Coors Field and elsewhere, OBP was 12% higher at Coors. This means a pitcher like Jon Gray, who spends half the year pitching in Coors Field, needs to have his opponent OBP adjusted down as a result. Because a pitcher spends half his games at home, we will apply half these percentages, so Jon Gray will have his opponent OBP adjusted down by 6%, and his strikeouts will be adjusted up by 8%.

Results

After filtering out the free agents who did not play or did not play enough, Max Scherzer was the top valued 2022 free agent, valued at $36.6 million. Carlos Rodon and Robbie Ray are currently the most undervalued starting pitchers. According to the model, both of their AAV's should be around 19 million. In 2021, Carlos Rodon's AAV was 3 million while Robbie Ray's was 8 million. The highest premium in the 2022 free agent class is Jon Gray which is no surprise since he's in Colorado. After adjusting his OBP and SO's, the added premium that should be added on Jon Gray's salary is $2.56 million.

Limitations

There are endless factors that go into negotiating a players salary, and each player should be treated on a case-by-case basis. This model does not take age into account. The reason age wasn't used in the model is because it wasn't a significant variable to AAV and didn't have a strong correlation and doesn't predict AAV well. Of course, teams need to take a player's age into account, and a guy like Adam Wainwright, who's 40 years old and having a great year, may land a great market price in this model.

Another limitation of this model is the park factors are calculated for the player's ballpark that he is currently playing in. For example, if a player plays in Fenway for 2 years, and then gets traded to the Dodgers, the player is assigned Dodger Stadium park factors. This should be considered when evaluating a player who's been traded or switched teams within 3 years of signing. In the future, I hope to add this into the model.

One last thing to be considered when evaluating players is the adjustment for 2020. Many people argue that pro-rating the player's 60-game season statistics into a 162-game season is unrealistic. There are countless examples of players who have had amazing first few months of a season, only to cool off and return back to life. A pitcher can also have a really bad start to the season but get hot in the second half. It's important to look at the 2020 season stats and see if the 162-game season stat is an accurate measure of the player.

All in all, this model should not be used to make decisions on valuating a player, but rather serve as a piece of information and a benchmark for monetizing players based on performance. This model will best serve pitchers who have played 3 full seasons prior to signing, and pitchers who have stayed with the same team.

Excel Table of Contents:

2021 Stats
Data from the 2021 season, updated last Monday, 9/14. Only the statistics that were selected to be in the regression were populated.

Stats 2011-2020
Basic, value, and advanced pitching season statistics of all players from 2011 to 2020

Contracts
Every active contract in Major League Baseball taken from Spotrac. Pitchers contract information including the year they signed, AAV, and signing bonus. Stats from 1,2, and 3 years before signing are populated in the rows.

Correlation
Each statistic was calculated into a 3-year weighted average and a correlation was ran on all the variables and AAV. WAR, SO, and OBP were the most highly correlated variables with AAV.

Park Factors
Park factors from baseballsavant.com. Each number is set so that 100 is the average for that metric.

Regression
First, a regression analysis is conducted without salary data. This model can be used if a player is on a rookie contract or there's no sufficient salary data from the year prior. Then below, a regression analysis is conducted with salary data, and there is less standard error as a result.

SP Free Agents
2022 Free Agents filtered into starting pitchers and pitchers' who are eligible for analysis. The 3 year weighted average is shown with park factors and a market price is calculated.

Final
Values are copied and pasted and serves as a good place to sort and filter data.

Name	SO+	OBP+	Market Price+	2021 Salary
Max Scherzer	242	0.276	$ 36,600,123	$34,603,480
Trevor Bauer	205	0.260	$ 34,728,918	$35,333,333
Clayton Kershaw	156	0.253	$ 28,627,272	$31,000,000
Zack Greinke	151	0.283	$ 27,344,816	$35,000,000
Kevin Gausman	198	0.278	$ 25,429,609	$18,900,000
Charlie Morton	186	0.296	$ 21,067,396	$15,000,000
Robbie Ray	224	0.318	$ 19,975,937	$8,000,000
Carlos Rodon	196	0.261	$ 19,830,092	$3,000,000
Marcus Stroman	156	0.302	$ 19,344,000	$18,900,000
Adam Wainwright	172	0.288	$ 17,367,368	$8,000,000
Eduardo Rodriguez	190	0.315	$ 17,039,802	$8,300,000
Johnny Cueto	121	0.336	$ 14,487,287	$22,000,000
Yusei Kikuchi	145	0.326	$ 14,356,707	$14,500,000
Andrew Heaney	161	0.313	$ 13,745,395	$6,750,000
Dylan Bundy	137	0.304	$ 12,766,234	$8,325,000
Drew Smyly	120	0.304	$ 12,273,217	$11,000,000
Danny Duffy	105	0.314	$ 11,944,350	$15,500,000
Merrill Kelly	119	0.286	$ 11,192,477	$5,250,000
Alex Cobb	96	0.310	$ 11,155,836	$15,000,000
Zach Davies	134	0.321	$ 11,155,090	$8,630,000
Jon Gray	129	0.306	$ 10,894,460	$6,000,000
Garrett Richards	124	0.326	$ 10,429,398	$10,000,000
J.A. Happ	126	0.322	$ 10,012,078	$8,000,000
Michael Pineda	91	0.299	$ 9,431,042	$10,000,000
Anthony DeSclafani	124	0.319	$ 9,347,135	$6,000,000
Tyler Anderson	121	0.315	$ 7,828,738	$2,500,000
Steven Matz	130	0.342	$ 7,595,799	$5,200,000
Rich Hill	108	0.308	$ 7,112,677	$2,500,000
Martin Perez	117	0.336	$ 6,541,983	$4,500,000
Jose Quintana	106	0.352	$ 5,638,599	$8,000,000
Jon Lester	110	0.348	$ 3,726,157	$2,000,000
Jhoulys Chacin	64	0.292	$ 3,550,404	$1,000,000
Jake Arrieta	91	0.370	$ 3,533,259	$10,000,000
Michael Wacha	105	0.355	$ 3,150,068	$3,000,000
Carlos Martinez	59	0.375	$ 3,012,070	$17,000,000
Brett Anderson	73	0.322	$ 2,600,855	$2,500,000
Michael Lorenzen	53	0.311	$ 2,397,069	$4,437,500
Matt Harvey	92	0.343	$ 2,011,429	$1,000,000