A DRAFT PICK TRADE VALUE MODEL FOR ASSESSING OPPORTUNITY COST

Hank Cole and Jake Larson

Introduction: Every June, the thirty MLB teams gather for the Rule 4 draft to select the best young players to enter professional baseball. The point of the draft is to provide a regulatory structure for the fair distribution of talent. It maintains competitive balance by preventing large-payroll teams from repeatedly acquiring the best prospects, assigning higher pick preference to recently unsuccessful teams, and providing extra picks to teams that have lost key veteran players. The actual selection of players is guided by an established scouting system that provides recommendations to general managers. However, with the uncertainty of a prospect's adaptability to the professional environment, his possibility to sustain injury, and the differing positional needs between teams, drafting is a complex and strategic process.

Under current rules, teams draft until they spend their “bonus pool” which is a soft limit on how much teams are allowed to spend on signing bonuses to leverage player contracts without incurring league taxes or forfeiting future draft picks. Bonus pools are the sum of a team’s “slot values” which are assigned by the MLB and are designed to decay exponentially with pick so that teams with better picks get more purchasing power. MLB uses a fixed draft order from the worst to best team of the previous season record percentage. This order prohibits trading of picks. However, teams can effectively trade their first round picks in exchange for signing free agents and are allowed to trade compensatory picks awarded from losing free agents.

In order to make informed decisions regarding benefits from the forfeiture or trade of a draft pick it is important to understand the pick’s value. In this work we describe a draft pick trade value model which estimates the contribution of a pick to the revenue of a franchise.

Model: To determine draft pick trade value we develop a simple model that produces the estimated dollar value of each pick. Our model considers three factors: likelihood to reach the majors, cumulative WAR up to seven years, and dollar value per WAR of each position. We combined two datasets taken from baseball-reference.com as input to our model, both providing information about every player selected in the Rule 4 draft since 1990. One provides hitting and fielding performance metrics for each player to reach the majors and the other contains pitching performance for relevant players.

Quality control and data reduction steps:

1. Remove players drafted after 2010
2. Sum first seven WAR entries for players
3. Remove pitchers from batting data
4. Remove batters from pitching data
5. Remove players with recurring names

We do not use players drafted after 2010 because it would introduce a significant number of players that have not yet had a chance to reach the majors. Using an earlier year would ensure that each player contributes to our model equally, but we think using 2010 balances quantity and quality of the data. Since the average time a player spends in the minors is three years, we capture enough information about the likelihood of a player to reach the majors. The negative effect on our model quality is a slight understatement of WAR contribution per slot. The first seven entries for each player were summed to determine their total WAR during their first years in the majors. Since a traded player has two entries for one year, we add the seventh entry. This has the added effect of improving the WAR contribution of a player who will continue in the MLB as a free agent after their six year contract expires. We view this as a fair overstatement because those players are likely to hold more trade value anyway.

Our reduced dataset consists of ~13,000 draftees. To produce the input data to our model, we remove players that never reached the majors and use the difference to calculate the likelihood of players from each draft rank to make it to the majors. We then take the cumulative WAR of all remaining players and multiply it by the value corresponding to their drafted position shown in Table 1 to produce a dollar value of each player. We then take the dollar values of each player, multiply by likelihood to reach the majors, and calculate the mean and standard deviation for each pick, which is the input of our model.

Position	$M / WAR	LF	10.6
C	5.0	CF	6.3
1B	7.1	RF	6.3
2B	3.7	DH	8.1
3B	4.8	SP	7.0
SS	4.1	RP	7.9

Table 1. Estimated value of player position per WAR based on cost of a win. [HardballTimes]

We fit a scaled exponential decay model to the input data using this equation,

$$f(x) = a * e^{-b*x} + c$$

where a, b, and c are the fitting parameters and x is the slot number. A least squares approach was used to determine the optimal parameters that minimize residuals between the model and mean value of slots. The optimal fit parameters and their standard deviations are given in Table 2.

Parameter	a	b	c
Value	45.735	0.044	12.219
σ	6.546	0.011	1.960

Table 2. The optimal fitting parameters and their standard deviation (σ) for our input data applied to a scaled exponential decay model.

Our best fit model agrees well with a moving average of trade value, giving us confidence in its ability to estimate slot pick value. After pick ~150, the model levels out while the moving average of trade value appears to fluctuate. This is partially due to a smaller number of players making the major leagues after this point in the draft.

Results: The following plots show data from the intermediate steps in our data reduction process. Both show a best fit function based on the same equation as our model.

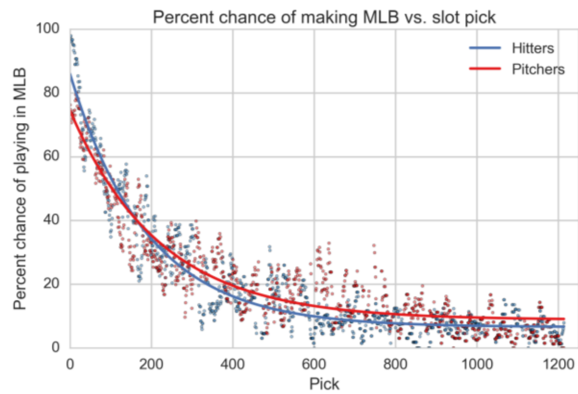

Figure 1. Plot of the percent chance of a player selected at a specific pick will reach MLB split by position players (blue) and pitchers (red).

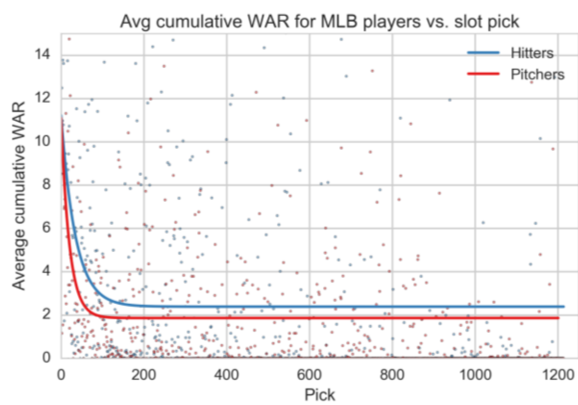

Figure 2. Plot of seven-entry cumulative WAR average for each slot split by position players (blue) and pitchers (red).

These show that position players are more likely to reach the majors prior to pick ~150 and pitchers seem to have substantial advantage between ~300-800. The average cumulative WAR shows a large degree of scatter with a small advantage for position players and shows superior performance for early selections up to pick ~150.

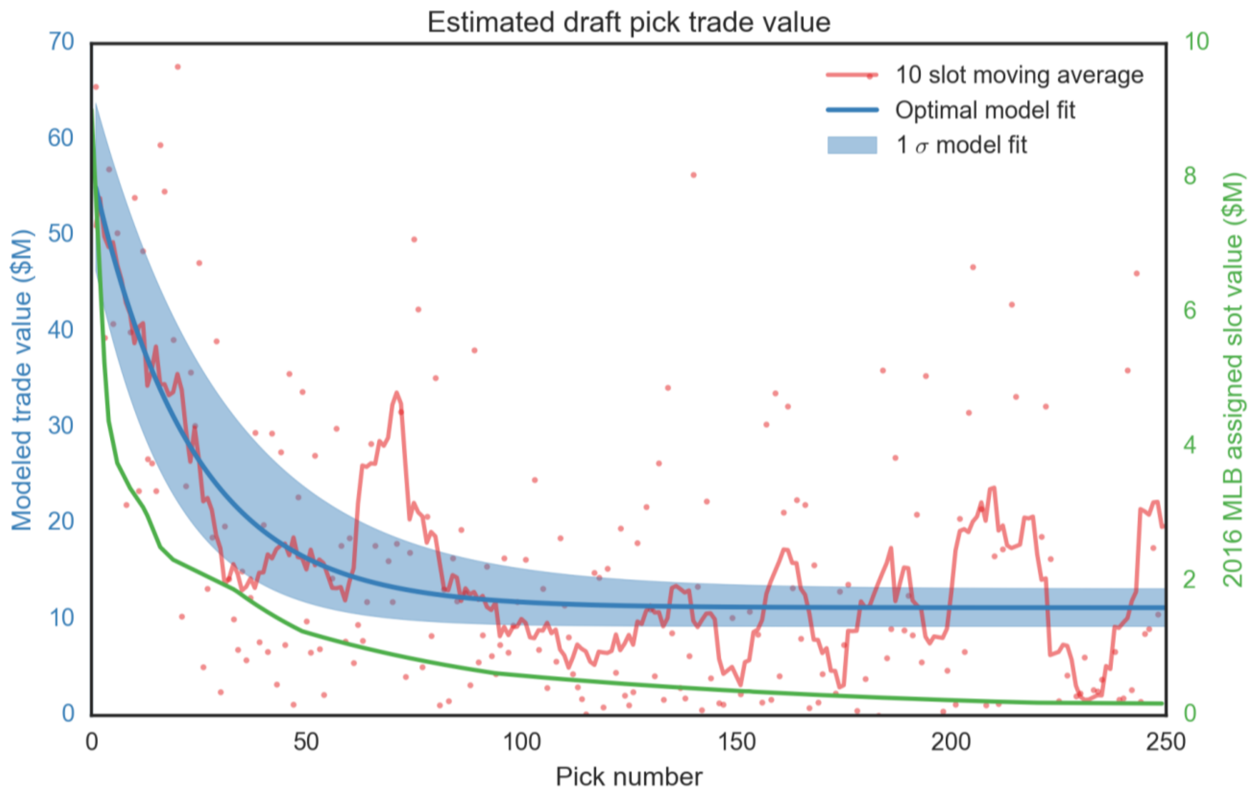

Figure 3. Plot of input data (scattered red) with model results (blue) and 2016 slot value (green).

Pick	Trade value ($M)	Pick	Trade value ($M)	Pick	Trade value ($M)	Pick	Trade value ($M)
1	55.00	21	29.52	41	18.87	61	14.28
2	53.13	22	28.74	42	18.54	62	14.15
3	51.35	23	27.99	43	18.23	63	14.03
4	49.63	24	27.28	44	17.93	64	13.91
5	47.99	25	26.59	45	17.65	65	13.79
6	46.42	26	25.94	46	17.37	66	13.68
7	44.92	27	25.31	47	17.11	67	13.58
8	43.48	28	24.71	48	16.86	68	13.48
9	42.11	29	24.13	49	16.62	69	13.38
10	40.79	30	23.58	50	16.39	70	13.29
11	39.53	31	23.05	51	16.17	71	13.20
12	38.32	32	22.55	52	15.95	72	13.11
13	37.16	33	22.06	53	15.75	73	13.03
14	36.06	34	21.60	54	15.56	74	12.96
15	35.00	35	21.16	55	15.20	75	12.88
16	33.98	36	20.73	56	15.03	76	12.81
17	33.01	37	20.33	57	14.86	77	12.74
18	32.08	38	19.94	58	14.71	78	12.68
19	31.19	39	19.57	59	14.56	79	12.62
20	30.34	40	19.21	60	14.42	80	12.56

Table 3. Trade value of optimal model fit for the first 80 picks.

Discussion: As expected, our results show that the highest ranked draft picks have greatest trade value. This trend continues until the ~150th pick where the average trade value of a pick levels out. We note that the early picks have a trade value up to $55M, but that a real value of a player can be much more than that. The discrepancy accounts for the probability that a player from that slot might have negligible (or even negative) real value to the franchise. In this way, our trade value accounts for risk.

We accept that there are a number of limitations to our model. Most importantly, it depends heavily of the WAR calculated for each player. Since WAR is an estimate of a player's impact on a team's winningness, the uncertainty is carried over into our model. However, WAR overestimates and underestimates players so we assume this uncertainty is acceptable because it is likely balanced out. Our model also uses an estimate of relative $M/WAR that was published in 2014. The difference in value likely changes substantially over time, but we accept this simplification because the methods of making this estimate vary with the individual publications. This likely has an impact on our model of overestimating the trade value of pitchers, since they are more highly valued in the modern game.

Admittedly, our model is quite simple and could be improved in various ways. First, we could recover the ~4200 players that were removed from our input because they shared a first and last name. The MLB does assign an ID to each draft pick, so matching a draftee to their major league performance would be trivial with complete data. We could also improve our model by replacing our chance to reach the majors component with a non-binary calculation. Ideally this would incorporate the number of years that a draftee plays in the majors, up to the contracted six years, placing greater weight on players who spent more time in the majors. This would likely skew the model for early picks to have even greater relative trade value since we should expect that late round picks are less likely to play six years in the majors. Considering minor league performance would also make our model more sensitive to the real trade value of a pick. While it would change the core interpretation of our model, we could also consider the bonus and salary paid to each player throughout their career. This would require that we change our definition of trade value to imply profitability. It is entirely possible that early-round picks are less profitable than mid-round picks because they are likely to draw a greater salary through their entire career. Picking a player of comparable productivity but with smaller bonus and salary leaves the team with a greater portion of their payroll to acquire other productive players.

Conclusions: Our trade value model is only useful if we can use it to make informed decisions to develop a drafting strategy. An important consideration is whether or not it is worth giving up a first round pick to another team to sign one of their free agents. Obviously this depends on the free agent, but we can quantify the minimum level free agent that justifies this trade-off. For example, the Rockies are slotted to have the 11th pick in the first round of the 2017 draft. Our model states that this draft slot has estimated trade value of $39.53M. Based on Table 1 we can determine the minimum WAR that a player of a certain position would need to be projected to produce to be worth losing the 11th pick. If the prospective free agent plays 1B, then he will need to produce at least 5.57 WAR over the length of his contract to be worth losing the 11th pick. Our trade value model also suggests that it is better to prefer position players until pick ~150 and prefer pitchers from ~300-800. After ~800, any bonus pool spent would likely have been better applied leveraging an earlier selection that is more likely to reach the majors.

The results from our model can likely be applied in many other ways too. It provides a simple static trade value that is independent of the assigned slot value, prospective player selection, his bonus, and his salary. It provides an estimated value of the pick’s probabilistic opportunity cost.