Josh Sellers

MATH.421.01 - FYW: Mathematical Modeling

Dr. Cherry

2 November, 2017

<div align="center">Most Valuable Player</div>

A popular movie and book that came out recently was *Money Ball*. In it, the general manager for the Oakland Athletics, a baseball team, used mathematics and statistics to find underrated players that had been overlooked by other baseball organizations. That film is emblematic of the current trends in baseball; to find the best possible player for a given team's budget. This can be seen in the construction of current successful teams like the Cleveland Indians and the Chicago Cubs. Through some combination of good drafting and player development, smart trades and free agent pick-ups these teams have been able to create rosters that consistently win. These choices are informed by an analytical strategy.

Each team has its own unique and protected strategy for choosing players. They have dedicated staff that preform the necessary mathematical analysis on potential and current players to see who is the best option. For the average fan the mystery behind their favorite team's choices are exciting; you can only make educated guesses as to who the team will choose. However, baseball diehards, who want to know as much as their team's general manager, do not desire this mystery. Regular fans do not have same access to the huge sets of data, the numerous analysts and the computational resources as their favorite teams. While it is not possible to completely recreate the research capabilities that baseball teams have, it is possible to make a simple model for choosing players that emulates what teams have. This would allow for some insights into what a baseball team will decide with regard to a given positional need. In fact, this model has been made and can now be used for real-life analysis of players.

The concept behind the model is to compute the impact of different players for a position over time. It will compare the three options a team has: trade, free agency and bringing up a minor league player. The impact of these three options to the team is based on Wins Above Replacement (WAR) per five hundred thousand dollars of contract. WAR quantifies the amount of wins a player adds to a team compared to an average player at that same position. The WAR for a player will converge over time on what their expected skill level should play at, oscillating around that value to factor in better and worse seasons. As the player ages there will be a decline in their WAR. The amount of this decrease is based on the findings of Ray C. Fair of Yale. Additionally, each year that a player is on the team, there is a probability that they will be injured. If they are injured, they will produce less than their expected WAR for the team that year. Hence, the probability for

injury, and the amount of WAR lost due to an injury, is incorporated by subtracting an amount based on the potential for the player to get injured combined with the percentage of the season an injured player misses, on average, due to injury.

The players' salary will follow three scenarios over time. If the player is a minor leaguer, their salary will first follow the rookie salary standards for 6 years, gradually rising from its initial value to account for increases in the league minimum salary and rookie salary arbitration. Once the rookie contract is complete, the contract will follow the pattern of free agent's contracts and match what players of that caliber tend to receive. If a player is a free agent, they will receive a slight increase in their salary compared to what they earned on their previous team. In the case of a trade, the player's salary will be whatever it was on their previous team.

In order to use this model, there are certain restrictions that need to be incorporated. First, since anything that happens after a player's current contract runs out is impossible to predict, all players compared are assumed to be given the same length of years on their potential contracts. This mitigates any uncertainty between a player, with a contract that ends sooner, who get more or less money in a subsequent contract and a player who has a longer contract. Additionally, with the exception of the linearly increasing rookie contracts, contracts are assumed to be constant over time (ignoring incentives and bonuses). Second, scouting is used to inform how well a player will do during his time on the team. This will be ascertained from previous performance and scouting reports, if available, for a starter and the ranking that they are given by Baseball America, generally considered to be a good benchmark for predicting professional success, for minor league players. Since Baseball America only lists the top 100 prospects, the top 50 of those will be given the rating of All Star and the bottom 50 will be given the rating of starter. Any other minor leaguer will be given the rating of reserve. This is due to the variability in predicting how minor leaguers will do in the transition to the major leagues. Once in the model, players will oscillate around their expected value and not diverge (this assumes that players do not dramatically improve or regress). Third, the only epidemiology on player injuries found gave information for only pitcher and fielders, so all fielders will be assumed to have the same risk of injury. Since catastrophic injuries are rare, the players are assumed to not miss more that a full season with an injury. The number that their WAR is decreased by is an estimate based on the incidence rates of injuries for pitchers and for fielders per 1000 players combined with the mean time that injured pitchers and fielders miss due to injuries (Posner et al. 1678). Finally, with the exception of minor leaguers' rankings, all players initial parameter values are assumed to come from Baseball-Referance.com.

The model will be comprised of the following equations and parameters.

Model for player value

$$V_n = \left(\frac{W_n - p_n - D_n}{(S_n/500,000)} \right)$$

$$W_n = \left(\frac{W_{expected} - W_{n-1}}{|W_{expected} - W_{n-1}|} \right) \left(\frac{r_w}{A+n} \right) + W_{n-1}$$

$$S_n = \begin{cases} S_0 + (\frac{1}{2})(S_0) & \text{if free agent} \\ \frac{1}{3}S_{n-1} + S_{n-1} & \text{if } n \leq 6 \text{ and rookie} \\ S_6 + \frac{1}{2}S_6 & \text{if } n > 6 \text{ and rookie} \\ S_0 & \text{if trade} \end{cases} \tag{1}$$

$$D_n = \begin{cases} 0 & \text{if } A+n < 30 \\ -(0.01)(A+n) & \text{if } A+n \geq 30 \end{cases}$$

$$p_n = \begin{cases} (0.41)(0.00416)(W_n) & \text{if pitcher} \\ (0.23)(0.00210)(W_n) & \text{if fielder} \end{cases}$$

Parameters List		
Parameter	Summary	Units
V_n	Player value	WAR per $500,000
W_n	Player WAR	WAR
S_n	Player salary	Dollars
p_n	Injury probability	Percentage
$W_{expected}$	Expected WAR	WAR
r_w	Rate of increase in WAR	Percentage
A	Age	years
S_n	Salary	Dollars
S_0	Salary before current model	Dollars
W_0	WAR before current model	WAR

WAR Explanation	
WAR Range	Equivalent
8+	MVP
[5,8)	All Star
[2,5)	Starter
[0,2)	Reserve
0-	Replacement

The estimation for WAR for major leaguers will be based on if they had won a MVP award or had been voted on to the All Star team previously. The number of these will dictate how high their

estimated WAR will be within its range. For each All Star appearance for non-MVP players, the WAR will be increase by 0.25, capping off at 8. For each MVP, the WAR will be increased by 0.25. If they have not received an All Star appearance or an MVP, then the WAR estimation will be based on their W_0.

The model has not gone under validation testing and analysis. As such, the constants for incrementing MVPs' and All Stars' values may change with testing. Additionally, injury risk and aging decline may be increased or decreased after further review. Finally, salaries need to be researched as well to determine if the salary equation matches the majority of players. While these equation should work, many of their constants and coefficients are based on research papers and estimations. Hence, practical testing may reveal different values.

Overall, this is an intriguing model. It will do a good job of describing player value for different positions. Its limitation lies in the fact that it does not factor in wild shifts in player skill-level. The best scenario for this model to be used is for players that follow expected trends and do not exceed or fall below expectations. Since most players fall into this category, it should be useful for most situations. As previously stated, the model still needs to be validated on historical data and this validation may result in some coefficients being modified. However, it does not seem likely like that the actual model itself will need to be changed. If there is extra time, the model could be improved by adding in a comparative equation to the teams overall salary, to see how much of a restriction a player would be on the team's finances. With that being said, the model is in excellent shape and should progress well in the future.

References

Baseball America

This site provides the rankings for the top 100 minor league players. It will be used in the salary equation.

http://www.baseballamerica.com/

Baseball-Referance.com

This is a non-scholarly reference that is used as a first stop for getting player data.

https://www.baseball-reference.com/

Epidemiology of Major League Baseball Injuries

This scholarly article from *The American Journal of Sports Medicine* by Posner et al. provides insight into injuries in pitchers and fielders.

http://journals.sagepub.com/doi/abs/10.1177/0363546511411700

Estimated Age Affects in Baseball

This scholarly article by Ray C. Fair of Yale describes how players decline as they age.

https://fairmodel.econ.yale.edu/rayfair/pdf/2005D.PDF

Everything you need to know about service time

This article from Major League Baseball's website, mlb.com, explains rookie contracts.

http://m.mlb.com/news/article/115853014/everything-you-need-to-know-about-service-time/