Josh Sellers

 MATH.421.01 - FYW: Mathematical Modeling

Dr. Cherry

 21 November, 2017

Most Valuable Player

INTRODUCTION

In the popular movie Money Ball, 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 perform 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; it only possible to make educated guesses as to whom 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 model would allow for some insights into what a baseball team will decide with regard to a given positional need without using a prohibitive amount of time and money. I made one possible model for this type of player analysis.

MODEL

The concept behind the model is to compute the impact of different players for a position over time. It provides data that will be used to compare the three options a team has: trade, free agency and bringing up a minor league player. This comparison will help fans answer the question of which player the team will probably chose. The impact of these three options to the team is based on Wins Above Replacement (WAR) per fifty thousand dollars of contract. WAR quantifies the number of wins a player adds to a team compared to an average player at that same position. It is an aggregate of offensive statistics from a given season. The WAR for a player will converge over time to their expected skill level, oscillating around that value to factor in better and worse seasons. As the player ages, there will be a decline in his WAR. Additionally, each year that a player is on the team, there is a probability that he 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 player’s 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 a pattern similar to a rostered player’s contracts and give the player a significant raise. If a player is a free agent, they will receive an 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.

ASSUMPTIONS

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 number 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. Scouting will be ascertained from previous performance, 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 mean rating for starters. Any other minor leaguer will be given the minimum rating for starters. The cutoff at starter is due to the assumption that a team will not promote a player for a crucial position unless they believe that player is capable of starting. 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 never miss more than a full season with an injury. The number by which their WAR is decreased 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-Reference.com.

Beyond the initial assumptions, there were several additional restrictions made in the creation of the equations and selection of parameters and constants for the model. Players are assumed to follow typical trends for their positions and not deviate from their contemporaries. This can be seen in the injury equations, where it is assumed that players will follow the average rates of injuries and missed time. For the decline equations, the same assumption of following normal player patterns is made (since most players start to decline at age 30). Additionally, the salary equations assume that players will not receive huge raises or pay cuts compared to the norm. Furthermore, players are not expected to have a sudden dramatic improvement in their performance. Essentially, the model works best if players have careers that are in line with their projections. Another assumption made was that players will play out their contract and not retire or demand a trade. This expected loyalty will allow anyone using the model to get a better idea of how different players will compare over time. One last assumption made was that teams will be willing to pay players the amount the model suggests. They will not be constrained by budgets or doubts about the players. The assumptions allow for a simple and easily utilized model.

EQUATIONS

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

SEE PDF

|  |  |  |  |
| --- | --- | --- | --- |
| Variables List | | | |
| Variable | Summary | Units | Range |
| Vn | Player value | WAR | WAR per Dollar |
| Wn | Player WAR | WAR | N/A |
| Sn | Player salary | Dollars | Sn ≥ League minimum |
| pn | Injury loss potential | WAR | N/A |
| Wexpected | Expected WAR | WAR | Value based on WAR table |
| An | Age during contract | Years | Must not be a student |
| S0 | Initial Salary | Dollars | S0 ≥ League minimum |
| W0 | Initial WAR | WAR | N/A |
| n | Number of years since first year of contract | Years | n≥0 |

5

|  |  |  |  |
| --- | --- | --- | --- |
| Parameters List | | | |
| Parameter | Summary | Units | Value |
| sp | Percentage of season missed for pitcher | N/A | 0.41 |
| sh | Percentage of season missed for hitter | N/A | 0.23 |
| dp | Rate of decline for pitchers | WAR | -0.187 |
| dh | Rate of decline for hitters | WAR | -0.576 |
| ip | Injury incidence rates for pitchers | N/A | 0.00416 |
| ih | Injury incidence rates for hitters | N/A | 0.00210 |
| rp | Rate of increase in WAR for pitchers | WAR | 0.695 |
| rh | Rate of increase in WAR for hitters | WAR | 0.651 |
| bf | Boost in salary for free agent | N/A | 1.81 |
| br | Boost in salary for a rookie | N/A | 1.33 |
| bc | Boost in salary for rostered player | N/A | 1.66 |
| b3 | Boost in salary for rookie arbitration | N/A | 4.5 |

For a player’s initial WAR, the estimation will be based on the explanation given by Baseball- Referance.com, which is summarized in the table below.

|  |  |
| --- | --- |
| 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, denoted in the model as Wexpected, will be based on if they had won a Cy Young Award, MVP award or had been voted onto the All Star team previously. The number of these accomplishments will dictate how high their estimated WAR will be within its range. For each Cy Young award and All Star appearance for non-MVP players, the WAR will be increased by 0.5, capping off at 8. For each MVP, the WAR will be increased by 0.2. If they have not received an All Star appearance or an MVP, then the WAR estimation will be based solely on the upper bound of whichever bracket their W0 belongs. For the case of rookie hitters, their expected WAR will be the mean amount for whatever category they are projected to attain and their initial WAR will be half of their expected WAR. For rookie pitchers, their expected WAR will be the lower bound of their projected category.

EXPLANATION OF EQUATIONS

The model’s three main equations are fairly simple and their construction was straightforward. The simplest are the Wn and pn equations. For the injury loss potential equation, it is simply the probability of getting injured and the average amount of time missed multiplied by the player’s WAR (Posner et al. 1678). That loss in WAR is split into two cases, pitchers and everyone else. For the WAR equation, there are four cases; before and after 26 years old for pitchers and before and after 28 years old for hitters (Fair 26). That is around the time that players start to decline. The rates of improvement and decline for pitchers and hitters were calculated by sampling 152 and 183 seasons, from a set of players, and finding the average change in WAR before and after the two set’s decline cutoff. The bounding function of Wexpected−Wn−1 uses the player’s expected WAR, |Wexpected −Wn−1 | whose calculation was discussed in the previous section, as an upper bound for the player’s WAR. Besides the WAR and injury equations, most complicated equation was the one for player salary (Sn ).

For the salary equation, there are six possible cases. The simplest is if a player is traded. In that case, his salary would remain the same as it was on his previous team (S0). If the player is a free agent, his new salary is calculated by multiplying his previous salary by the free agency boosting constant, bf . The constant was calculated by getting data from 50 players on the raises they received via free agency and averaged their increases in salary. The free agency boosting rate was found to be 0.807985381 ± 0.11 on a 90% Confidence Interval. Another 90% confidence interval was used to calculate the amount of salary increase for currently rostered players. For a sample of 62 players, the interval was found to be 0.661645563 ± 0.096. Finally for rookies, their S0 is expected to be the league minimum, since only a small number of rookies receive greater starting contracts. Based on a sample of 110 players, the increase in their salaries, br, was found to be 0.330120501 ± 0.13 on a 90% confidence interval. Rookies are eligible for salary arbitration after their third season (mlb.com). Based on a sample size of 23, the 90% confidence interval for this increase was found to be 3.500610892 0.97. After a rookie’s sixth season, he will follow a similar pattern to a rostered player, receiving a raise based on his sixth year and bc. The salary equation, the WAR equation and the injury equations were combined into the first equation, which calculates player value.

The player value equation, Vn, is a player’s WAR subtracted by the injury potential divided by the logarithm of his salary divided by $50,000. The league minimum salary is currently $507,500; the division by $50,000 means that the minimum logarithm for the denominator is going to be around one (the divisor was adjusted we testing against historical data with lesser minimums) (Associated Press). The logarithm was used because of the wide range of salaries for players; it corrects for that disparity in the values. This is output of the model and what will be used to compare different players. The main issue with the equation is in the constants, some of which have significant variability. This variability is most likely due to the wide range of abilities found in Major League Baseball. Since the constant represent the middle ground of the abilities in the league, that is what led to the assumption that players numbers will be in line with their projections. If they follow the average career growth and decline, the model will have a reasonable amount of accuracy. VALIDATION

In order to check the validity of the model, historical data for various players was compared to the model’s output. To guarantee a fair test, no players who experienced catastrophic injuries were selected. The first test was run on three players’ rookie years. Due to the known variability of rookie performance compared to expectations, only rookies that had a ranking from Baseball America were picked for the tests. The tree players chose were Craig Kimbrel, Jason Heyward and Madison Bumgarner. Their rookie rankings from Baseball America were 84, 1 and 14. The Table below shows their projected value compared to their actual value (a divisor of $40,000 was used against their league minimum).

|  |  |
| --- | --- |
| Craig Kimbrel Test | |
| Expected Vn | Actual Vn |
| 1.712501393 | 2.352585909 |
| 2.13358191 | 2.8234279 |
| 2.472395263 | 2.717880745 |
| 1.99287864 | 1.114559783 |
| 1.532892321 | 0.5526782 |
| 1.35927698 | 0.367483053 |
| 1.15588934 | 1.428483063 |
| Totals | |
| 12.35941585 | 11.35709865 |
| Error | |
| 0.088254688 | |
| Jason Heyward Test | |
| Expected Vn | Actual Vn |
| 3.24843025 | 6.4 |
| 3.469422186 | 2.285485974 |
| 3.646541169 | 5.043528887 |
| 2.735779756 | 1.887530839 |
| 2.889799792 | 3.022690869 |
| 3.026063648 | 2.838385048 |
| 2.470175799 | -0.109740088 |
| Totals | |
| 21.4862126 | 21.36788153 |
| Error | |
| 0.005537801 | |
| Madison Bumgarner Test | |
| Expected Vn | Actual Vn |
| 2.374285317 | 2.283208144 |
| 2.714501537 | 1.832256026 |
| 2.989835517 | 3.063626161 |
| 2.344784648 | 2.028427111 |
| 2.539112555 | 2.155130049 |
| 2.080752303 | 2.094728109 |
| 1.814345892 | 1.17951491 |
| Totals | |
| 16.85761777 | 14.63689051 |
| Error | |
| 0.151721246 | |

The model did a fairly good job of predicting the rookies’ value. The most divergent output was for Madison Bumgarner. He had an injury in the final year of time period tested and this caused the error.

Another three players were tested under the rostered scenario: Brandon Phillips, Joe Mauer and Felix Hernandez. The actual raises in their salaries were $12,000,000, $23,000,000 and $25,000,000 (spotrack.com). The predicted and actual outputs are shown below.

|  |  |
| --- | --- |
| Joe Mauer Test | |
| Expected Vn | Actual Vn |
| 2.411763129 | 0.56332573 |
| 2.317516243 | 1.614867093 |
| 2.129022469 | 1.99041758 |
| 1.846281809 | 0.788656022 |
| 1.469294262 | 0.56332573 |
| 0.998059829 | 0.86376612 |
| 0.432578508 | 1.276871655 |
| Totals | |
| 11.60451625 | 7.661229931 |
| Error | |
| 0.514706692 | |
| Brandon Phillips Test | |
| Expected Vn | Actual Vn |
| 1.611847605 | 1.680523111 |
| 1.372834085 | 0.7562354 |
| 1.038215156 | 0.714222322 |
| 0.607990819 | 1.470457722 |
| 0.082161073 | 0.336104622 |
| -0.53927408 | 0.336104622 |
| Totals | |
| 4.173774657 | 5.293647799 |
| Error | |
| 0.211550368 | |
| Felix Hernandez Test | |
| Expected Vn | Actual Vn |
| 1.83566295 | 1.926660908 |
| 1.791752497 | 2.519479649 |
| 1.660021138 | 1.630251538 |
| 1.440468873 | 0.592818741 |
| 1.133095702 | 0.296409371 |
| Totals | |
| 7.861001159 | 6.965620207 |
| Error | |
| 0.12854289 | |

There was more error for this test. This is due to the variability in when and how much a player declines. Since players don’t receive these types of contracts till the middle or the end of their careers, there is a high likelihood that they will begin to decline during their contracts. Thus, the variability speaks to the inadvisability of handing out these types of contracts.

The final test was on free agency. For this test, the three players were Robinson Cano, Albert Pujols and Darren O’Day. Their salaries were $24,000,000, $24,000,000 and $2,900,000 (spotrack.com). The predicted and actual outputs are shown below.

|  |  |
| --- | --- |
| Albert Pujols Test | |
| Expected Vn | Actual Vn |
| 1.698716092 | 1.790215641 |
| 1.426240358 | 0.559442388 |
| 1.062939379 | 1.454550208 |
| 0.608813156 | 1.156180935 |
| 0.063861688 | 0.522146229 |
| -0.571915024 | -0.671330865 |
| Totals | |
| 4.288655648 | 4.811204535 |
| Error | |
| 0.108610823 | |
| Robinson Cano Test | |
| Expected Vn | Actual Vn |
| 2.617647472 | 2.386954188 |
| 2.391744801 | 1.268069412 |
| 2.075481062 | 2.72261962 |
| 1.668856255 | 1.268069412 |
| Totals | |
| 8.75372959 | 7.645712633 |
| Error | |
| 0.144920037 | |
| Darren O’Day Test | |
| Expected Vn | Actual Vn |
| 1.244267985 | 1.134154617 |
| 0.930685622 | 1.304277809 |
| 0.512575805 | 1.587816463 |
| Totals | |
| 2.687529412 | 4.026248889 |
| Error | |
| 0.332497944 | |

For this test, the same decline-based error also occurred. Due to the shorter contracts, that error was mitigated. Hence, free agency will lead to some error, but not as much as giving a raise to a

rostered player.

APPLICATION

Overall, the model seems to do well with rookie contracts and free agency. It has more error when it comes to raises for rostered players. The model’s ability is due to the variability in performance as a player ages. Since rookies are no apt to decline, the model is most accurate for them. For free agency, the players are either younger or the contracts are not long enough to be affected by decline. Therefore, raises are the most sensitive to decline.

The best use for the model would be either for younger players or shorter contracts. Since it is usually unwise to give a long contract player anyway and pitchers tend not to recieve long contracts, due to the wear on the body that pitching brings, this restriction does not have a major impact on the model’s efficacy.

An example for using the model can be seen in the test data. Craig Kimbrel and Darren O’Day are both relief pitchers. Even though Kimbrel wound up earning more, via arbitration, at the end of his contract, he is still shown to be more valuable. That quantitative analysis can also be confirmed qualitatively by anyone who closely follows baseball; Kimbrel is a much better player. Beyond that obvious comparison, the model can also provide deeper insights. Felix Hernandez is a well-known pitcher with a track record of winning games. A team would not be faulted for picking him over most pitchers. However, the model shows that he is beginning to decline and that it would be better to pick a younger pitcher, like Madison Bumgarner.

CONCLUSION

Overall, this is an intriguing model. It will do a good job of describing player value for different positions. However, it is not perfect. Its limitation lies in the fact that it does not sufficiently factor in wild shifts in player skill-level, decline in skills and injury probability. As stated previously, 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 many situations. With the model, armchair general managers can understand some of the analysis done behind closed doors and attempt to predict their team’s roster moves.

Works Cited

BaseballAmerica.com, www.baseballamerica.com/.

Baseball-Reference.com, www.baseball-reference.com/.

Fair, Ray. ”Estimated Age Effects in Baseball.” Journal of Quantitative Analysis in Sports, vol. 4, no. 1, 2008, doi:10.2202/1559-0410.1074.

Posner, Matthew, et al. ”Epidemiology of Major League Baseball Injuries.” The American Journal of Sports Medicine, 27 June 2011, journals.sagepub.com/doi/abs/10.1177/0363546511411700.

Nowak, Joey. ”Everything You Need to Know about Service Time.” Major League Baseball, 1 Apr. 2015, m.mlb.com/news/article/115853014/everything-you-need-to-know-about-service- time/.

Press, Associated. ”MLB Minimum Salary Remains at $507,500 for 2016.” ESPN, ESPN Internet Ventures, 18 Nov. 2015, www.espn.com/mlb/story/ /id/14161690/mlb-minimum-salary-remains- 507500-2016.

Spotrac.com, www.spotrac.com/.