Michael Hymowitz

Prof. Fredrickson

Stats 406

27 April 2021

Exploring the Applicability of Price's Law in the Distribution of Wins Above Replacement Among Major League Baseball Clubs from 2000-2019

Introduction

Since the Moneyball-era 2002 Oakland Athletics sent shockwaves through Major League Baseball with their use of advanced baseball metrics to exploit inefficiencies in professional baseball, there has been an enormous growth in the use of metrics and techniques of this sort over the past 20 years (Grier and Cohen). One statistic that has dominated the baseball analytics community in this time period is Wins Above Replacement, or WAR for short, which attempts to measure the number of theoretical wins a player's performance adds to his team above what would be expected from a replacement level player ("Baseball-Reference.com…"). A replacement level player is theoretically defined to be "players easy to obtain when a starter goes down," with a team made up entirely of replacement level players expected to have a winning record of 0.294, which correlates to a 48-114 across a full, standard 162-game Major League season ("Baseball-Reference.com…"). In essence, WAR attempts to measure the value a player adds to their team. For context, according to the popular baseball website Baseball-Reference.com, whose version of WAR will be used in this analysis, on a single-season scale, a WAR of 8+ represents MVP quality, 5+ represents All-Star quality, 2+ represents starter quality, 0-2 represents reserve quality, and < 0 represents below replacement level ("Mike Trout Stats"). For example, Mike Trout accrued 8.2 WAR for the Los Angeles Angels in 2019, which should be interpreted as: "Mike Trout's performances earned the Angels an additional 8.2 more expected wins than they would have been predicted to accrue had a replacement level player played instead of Mike Trout for the entire 2019 season" ("Mike Trout Stats").

On the other hand, a common principle in the business world is Price's Law. Derek Price "was a British physicist, historian of science, and information scientist," whose law states that "50% of the work is done by the square root of the total number of people who participate in the work," meaning that contributions are not uniformly distributed among contributors (Foroux). Price's Law has been applied to many phenomena, such as the distribution of publications in a given field or contributions to a cause within a company (Foroux).

In this research project, I will be exploring the distribution of WARs among Major League clubs for each team from 2000-2019, determining if this distribution can be explained by a slight variation of Price's Law. To explain, Price's Law says that 50% of the work is done by (total number of people who participate in the work)$^{\theta}$ individuals, where $\theta = 0.5$, the research question that will be explored in this research is: what is the value of θ whereby 50% of the total WAR accrued by a Major League Baseball team is accounted for by just (total number of players on a given MLB team)$^{\theta}$ players on each team? The hypotheses for these tests, using a significance level of 5%, are as follows:

H_0: $\theta = 0.5$

H_A: $\theta \neq 0.5$, and if so, what is an interval for θ for which we would not reject the null hypothesis that $\theta = \theta^*$

This research paper will be structured as follows: the Data section provides insight onto the Baseball-Reference dataset used for this analysis and contains preliminary graphs, giving insight into the distribution of the sample data; the Method section will discuss the bootstrap and jackknife methods used to test the hypotheses stated, defining these methods, explaining their assumptions, and describing their applicability to this problem; the Simulations section will somewhat loosen the assumptions enforced by the bootstrap and jackknife methods to allow for some correlation between observations and prove that these methods are still appropriate to use; the Analysis section will describe and interpret the results from the research and formally reject or fail to reject the null hypothesis; and the Discussion section will provide a summary and context for the results as well as areas for future research.

Data

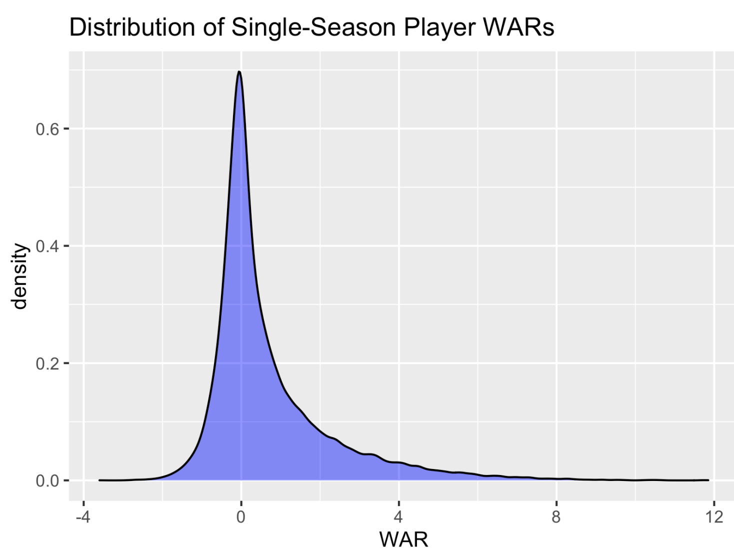
Data for this research paper came from Baseball-Reference.com's historical WAR archive, which contains single-season WAR data for every MLB player dating back to 1871 ("Baseball-Reference.com…"). Two datasets were downloaded from this webpage for this analysis: Baseball-Reference's "war_daily_bat" and "war_daily_pitch" datasets, which, as the names suggest, contain position players' and pitchers' WAR tallies, respectively. While these datasets contain 49 and 43 columns, respectively, the only variables that are needed for this analysis in both tables are the following:

- name_common – a player's name
- player_ID – a unique identifier for each individual player in the dataset
- year_ID – the year that the record corresponds to
- team_ID – a unique identifier for each MLB franchise in the dataset
- WAR – the amount of WAR accrued by a player for that team in that season
    - Note that while there are several existing versions of the WAR statistic, Baseball-Reference WAR, often referred to as bWAR or rWAR, is regarded as one of the

most popular and well-established versions of this metric (“WAR Comparison Chart”, Bingol).

These datasets contain all the information needed to conduct this analysis, and are nearly complete in terms of the cleaning of the data that is required. There are no missing entries for name_common, player_ID, year_ID, and team_ID, but for WAR, while there are no missing players, there are some records which have a WAR value of “NULL”. Fortunately, for the war_daily_bat table, all of these records correspond to entries which have a PA value of 0, meaning they did not have a single plate appearance the entire season, and for the war_daily_pitch table, all of these records correspond to entries which have an IPouts value of 0, meaning they did not record a single out while pitching the entire season. Thus, we can presume that those records which have a WAR value of “NULL” did not actually provide any significant contributions to their team, and thus we can safely assign these records a WAR value of 0.

Distribution of Single-Season Player WARs

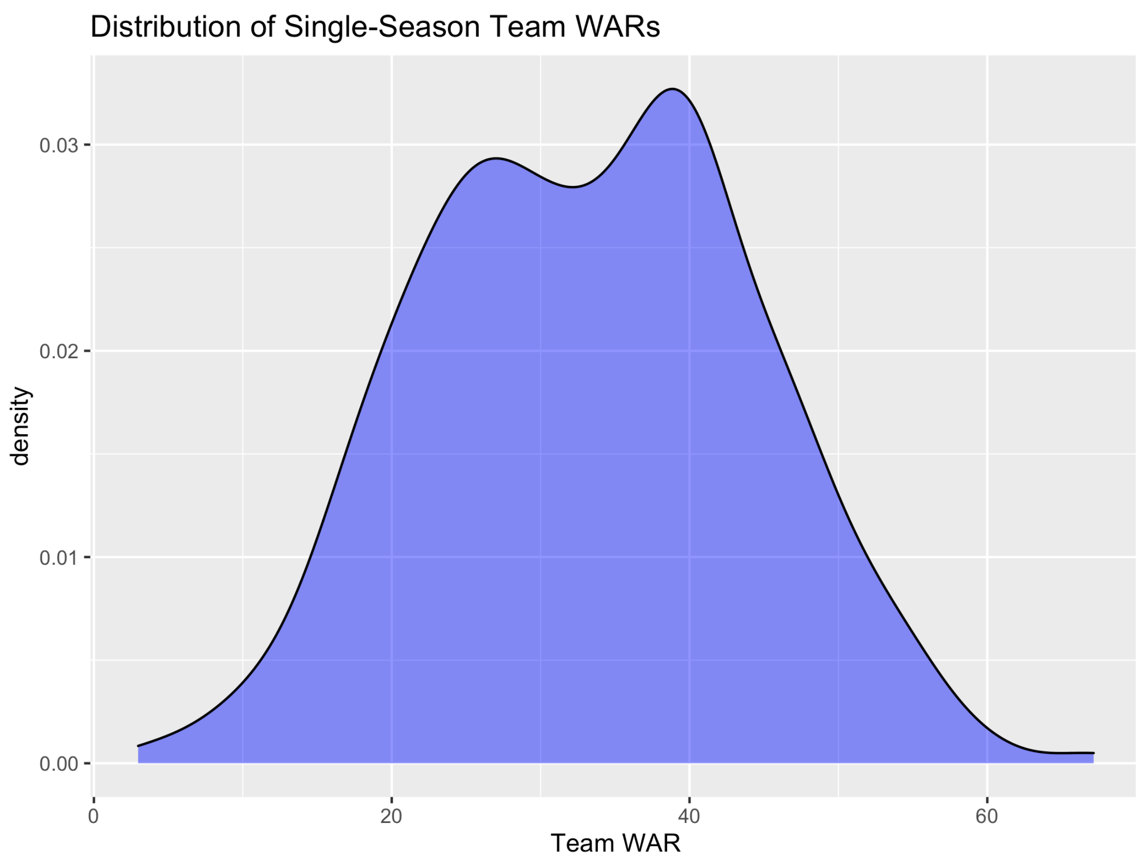


WAR	Team Size
< 0	10923
[0,2)	12908
[2,5)	3554
[5,8)	680
> 8	82

Figure 1; Source: Baseball-Reference; Distribution of single-season player WARs

These outputs show us that single-season player WARs have a right-skewed distribution, with the data spiking strongly at about 0.

Distribution of Single-Season Team WARs



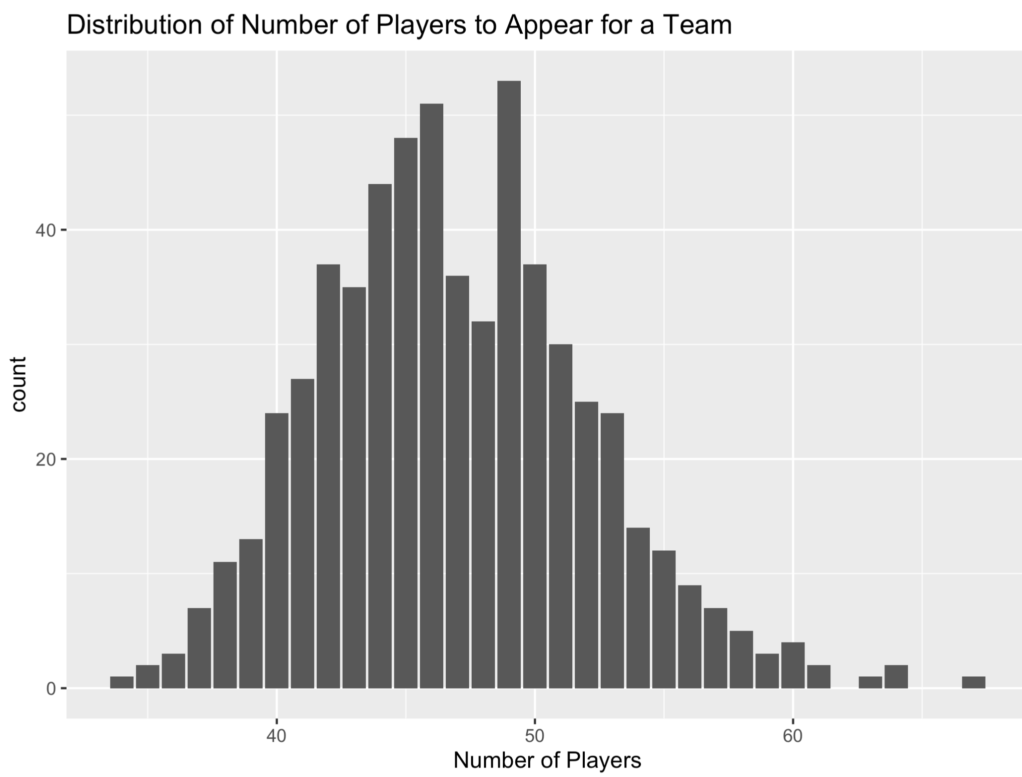
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.97	24.98	33.24	33.28	40.99	67.14

Figure 2; Source: Baseball-Reference; Distribution of team WARs

On the other hand, the distribution of team WAR, which is the total WAR accrued by a team's players in an entire season, is roughly symmetrically distributed, centered at about 33

WAR. This centering at nearly 33 makes sense, as the average record for a baseball team across a complete 162-game season is 81-81, as for each game there must be one winner and one loser. Therefore, if a team of replacement level players would accrue 48 wins, as discussed previously, then the addition of these 33 expected wins would take the average-team in a WAR-sense to the average record of 81 wins.

Distribution of Team Size



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
34	43	46	46.91	50	67

Figure 3; Source: Baseball-Reference; Distribution of number of players to appear for a team

From this distribution, we see that the distribution of the number of players to play for a Major League Baseball team from 2000-2019 is roughly normally distributed and centered at about 47.

Method

To estimate the parameter θ^* and its distribution, defined for which 50% of team$_i$'s total WAR is accumulated by just n_i^θ players, where n_i is the number of players to have played for the specified team in a season, the bootstrap and jackknife methods will be used. The bootstrap method with 1,000 replications will allow us to estimate the mean and median of the distribution of θ^*, while the standard sample variance formula will be used to estimate the population variance of θ^*, with the jackknife method giving us an estimation for the variance of this estimate. Through these bootstrap and jackknife procedures, 95% confidence intervals will be constructed for the mean, median, and variance of the distribution of θ^*.

Before proceeding, it is important to explain the bootstrap and jackknife methods and provide more insight into their procedures, assumptions, and the underlying statistical theory. The bootstrap method is a tool through which one can derive a reliable estimation for the sample distribution of a test statistic through the process of creating many new samples, with these "samples [being] generated by resampling from an observed sample," and thus it is "often used when the distribution of the target population is not specified; the sample is the only information available." (Rizzo 192, 213). The bootstrap proves itself to be a statistically reliable and powerful method as it "is based upon the notion that the empirical distribution of a sample $X_1, \ldots, X_n$ converges in n to the true distribution," and our dataset does provide a sufficiently large sample size of 600 teams (Robert and Casella 23). For example, to estimate the mean of the distribution of θ^* using the bootstrap method, the following procedure will be performed 1,000 times:

1. Select 600 teams from the sample with replacement.
2. Calculate θ from these 600 teams.
3. The average of these 600 θ values will constitute the bootstrap estimate for that given bootstrap sample.

   (Rizzo 215-216)

Finally, we can use these 1,000 bootstrap estimates to construct a 95% confidence interval for the true value of the mean of the distribution of θ^*.

The jackknife procedure is computationally more efficient than the bootstrap method, and is used to estimate the variance of an estimator for some test statistic through repeatedly leaving one observation out of the sample and recalculating the test statistic (Rizzo 220). The following is the procedure used to estimate the variance of the sample variance of the distribution of θ^*:

1. Calculate the sample variance of θ.
2. Repeat the following n = 600 times:
   a. Remove the i^{th} data point from the sample.
   b. Calculate the variance of θ_{-i} for the remaining 599 teams.
3. Calculate the sample variance of the 600 values calculated in step 2, and multiply this number by n = 600. This value is the jackknife estimator for the variance of the sample variance of the distribution of θ^*.

   (Rizzo 222)

With this estimate for the population variance from the standard sample variance formula and the jackknife estimator for the variance of the sample variance of the distribution for this estimate, we can proceed to calculate a 95% confidence interval for the population variance of θ^*.

To use these methods, we must first affirm that our assumptions have been sufficiently met. The first two assumptions for the bootstrap and jackknife methods are that our sample size is large and that the number of replications used is also large (Week 7: Bootstrap). Our sample size is 30 teams * 20 years = 600 and our number of replications is 1,000, both of which are large enough to make the bootstrap method applicable. The final assumption to use the bootstrap method is that our data is independent and identically distributed (Week 7: Bootstrap). Because the baseball teams in our sample compete against each other, and one player accruing positive WAR often, but not necessarily, requires that an opponent accrues some amount of negative WAR, it is unjust to conclude that this dataset is entirely independent. But, as a result of there being 30 teams in the sample for any given season, the scale of the impact one team and its players can surmount on any other team is limited, and thus it is fair to say that this dataset is roughly independent. Furthermore, in the Simulations section, it will be shown that even with some reasonable amount of negative correlation between the WAR totals of different teams, this method is still reliable.

Simulations

To go about demonstrating the applicability of the bootstrap and jackknife methods described in the Method section, we will run simulations of these methods on a simulated dataset both with the assumption of all teams having independent θ values, as well as simulations with this assumption relaxed, and demonstrate that this method is still reliable. In each of these different settings, we will create 1,000 samples of 30 θ values to mimic our dataset. To construct the θ values of any given sample, a multi-variate normal distribution random number generator will be used, with the correlation matrix coming from a first-order autoregressive model and an inputted non-positive correlation value ρ ("Generating Correlation…"). Note that it is not of the utmost importance how the θ values are correlated, just that there is some correlation being introduced between these values. These random numbers will then be squared to generate the simulated θ values, so that each θ generated follows a chi-squared distribution with 1 degree of freedom, which thus has an expected value of 1. Using the bootstrap method with 1,000 replications described previously, we will then calculate 95% confidence intervals for these simulated θ's to see if this method captures the true expected θ value of 1, and thus can deem this method to be reliable.

In the following table, we can note the results of running this simulation with varying values of ρ:

Rho	Lower Bound	Upper Bound	Interval Covers 1
0.0	0.9848	1.0170	TRUE
-0.1	0.9949	1.0290	TRUE
-0.2	0.9714	1.0020	TRUE
-0.3	0.9739	1.0050	TRUE
-0.4	0.9792	1.0110	TRUE
-0.5	0.9796	1.0120	TRUE
-0.6	0.9786	1.0110	TRUE
-0.7	0.9772	1.0080	TRUE
-0.8	0.9920	1.0230	TRUE
-0.9	0.9673	0.9996	FALSE

Figure 4; Source: random simulated data; Results of bootstrap simulations for estimating θ^* using different values of ρ

As we can see, the bootstrap confidence intervals constructed capture the expected value of the mean of the values of θ of 1 through a ρ value of -0.8, which is far more strongly correlated than any two team's θ values can practically be in reality by an enormous margin.

Therefore, despite weakening the assumptions somewhat to incorporate the reality that there does likely exist some negative correlation between the θ values for different teams in the same year, the procedures outlined in the Method section are still applicable and appropriate to use on this dataset.

Analysis

Through some preliminary analysis of the sample, it is very clear that the square root of the number of players on MLB teams generally contribute far more than half of the team's WAR, with this holding for 595 out of 600 teams in our sample, good for a 99% binomial test confidence interval of (0.977, 0 .998). It is therefore with an extreme degree of confidence that at our significance level of 5%, can reject the null hypothesis that $\theta = 0.5$, which in the context of this research topic means we reject the null hypothesis that the exponent θ whereby 50% of the total WAR accrued by a Major League Baseball team is accounted for by just (total number of players on a given MLB team)$^{\theta}$ players on each team is 0.5, and thus will explore the alternative hypothesis, which asks what is an interval for θ for which we would not reject a null hypothesis that $\theta = \theta^{*}$. To explore this question, we will utilize the bootstrap and jackknife methods discussed in the Simulations method in order to estimate the mean and median as well as the population variance of the distribution of θ^{*}.

Estimating the Mean of the Distribution of θ^*

Utilizing the boot package in the R programming language, we can undergo the bootstrap method to predict the mean value of the distribution of θ^*.

Using 1,000 replications, we attain a 95% bootstrap confidence interval using the Basic method for the mean value of θ^* of (0.3451, 0.3592). Our interpretation of this interval is thus that with 95% confidence, we can say that the true population mean of the distribution of θ^* is between 0.3451 and 0.3592. In the context of this problem, this means that with 95% confidence, we can say that the mean number of players per Major League Baseball team needed to accrue half of a team's total WAR is accrued by between (total number of players on a given MLB team)$^{0.3451}$ and (total number of players on a given MLB team)$^{0.3592}$ players. Plugging in the average number of players to make an appearance for our sample of teams for the total number of players on a given MLB team in the prior interval, which was demonstrated in the Data section to be 46.91, simplifies the latter boundaries to be 3.774 and 3.984 players, displaying how much of the WAR accrued by a team throughout a season is concentrated in the performance of just a few players.

Estimating the Median of the Distribution of θ^*

We will once again utilize the boot package in the R programming language in order to apply the bootstrap method to predict the median value of the distribution of θ^*, this time applying a nested bootstrap procedure in order to more accurately account for the variance of our estimates. Using 100 replications in the nested bootstrap and 1,000 replications in the outer bootstrap, we attain a 95% bootstrap confidence interval using the Basic method for the median value of θ^* of (0.3578, 0.3641). Our interpretation of this interval is thus that with 95%

confidence, we can say that the true population value for the median of the distribution of θ^* is between 0.3578 and 0.3641. In the context of this problem, this means that with 95% confidence, we can say that the median number of players per Major League Baseball team needed to accrue half of a team's total WAR is accrued by between (total number of players on a given MLB team)$^{0.3578}$ and (total number of players on a given MLB team)$^{0.3641}$. Plugging in the average number of players to make an appearance for our sample of teams for the total number of players on a given MLB team in the prior interval simplifies the latter boundaries to be 3.963 and 4.060 players, providing a similar conclusion to the one deduced when estimating the mean of the distribution of θ^*, which was that much of the WAR accrued by a team throughout a season is concentrated in the output of just a few players.

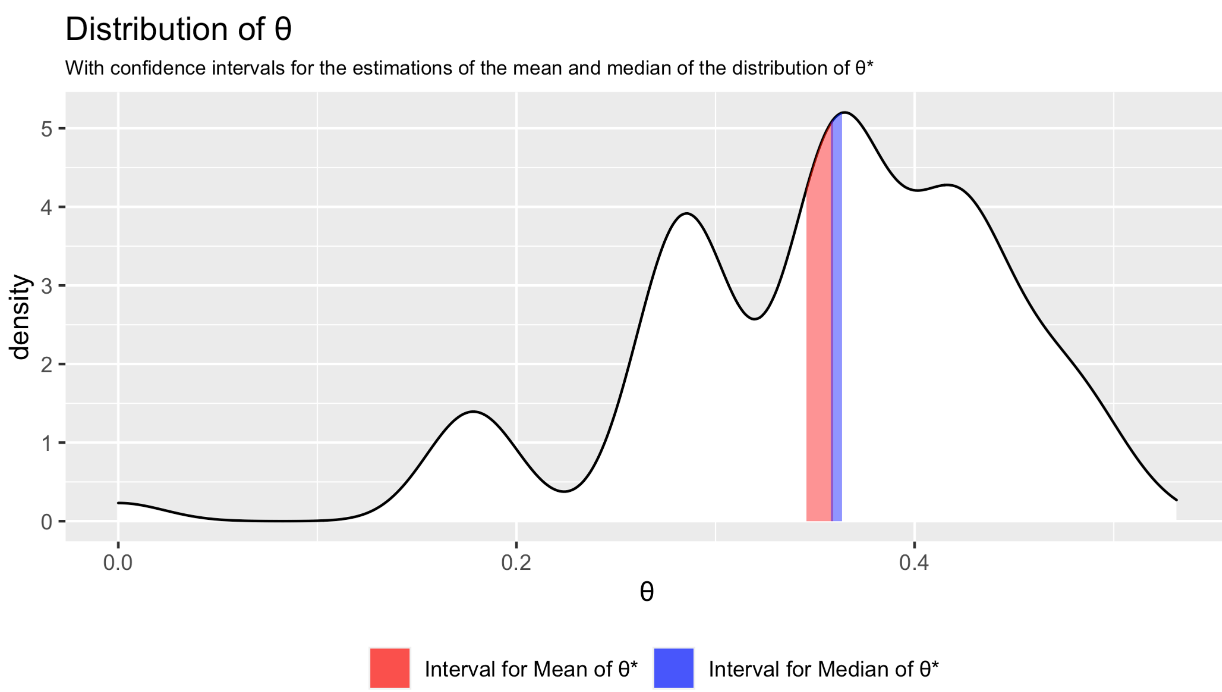


*Figure 5; Source: Baseball-Reference; Distribution of ϑ with confidence intervals for the estimations of the mean and median of ϑ**

It is important to note that the 95% confidence interval for the estimation of the true median of θ^* lies predominantly above (or graphically, to the right of) the 95% confidence interval for the estimation of the true mean of θ^*, indicating that this distribution is more than likely left-skewed.

Estimating the Variance of the Distribution of θ^*

Next, we will use the standard sample variance formula to produce an estimate for the population variance of the distribution of θ^*, and then use the jackknife method discussed previously to construct a 95% confidence interval for this estimate.

The standard sample variance formula of:

$$\hat{\sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

gives us an estimation for the population variance of θ^* to be 0.00840, indicating that there is relatively very little variance in the values of θ.

Next, we can apply the jackknife methods to calculate our estimate for the variance of this sample variance estimator to be $4.275 * 10^{-7}$. This means that a 95% confidence interval for the population variance of θ^* is

$$(0.00840 - 1.96 * \sqrt{4.275 * 10^{-7}}, 0.00840 + 1.96 * \sqrt{4.275 * 10^{-7}})$$

$$= (0.00712 \ , \ 0.00968)$$

We can interpret this confidence interval to mean that with 95% confidence, we can say that the true variance of the distribution of θ^* falls between 0.00712 and 0.00968, demonstrating the lack of variability and preciseness in these estimates regarding the distribution of θ^*.

Discussion

In conclusion, this paper questioned the application of Price's Law, which says that 50% of the work is done by (total number of people who participate in the work)$^{\theta}$ individuals, where $\theta = 0.5$, to the distribution of WAR on Major League Baseball teams from 2000-2019 (Foroux). The hypothesis tested was whether or not $\theta = 0.5$, defining θ to be the value whereby 50% of the total WAR accrued by a Major League Baseball team is accounted for by just (total number of players on a given MLB team)$^{\theta}$ players, and if not, what is the range of values for θ^* for which we would not reject the null hypothesis that $\theta = \theta^*$.

In the Data section, we explored the Baseball-Reference datasets we were utilizing and noted any shortcomings of the data, which were shown to be minimal and inconsequential. In the Methods section, we discussed the utilization of the bootstrap and jackknife methods to construct our estimates for the mean, median, and variance of the population distribution of θ^*, defining these methods and noting the assumptions they require, demonstrating that these assumptions were sufficiently met. In the Simulations section, we accounted for the reality that our data are inherently somewhat dependent, as frequently one player accruing more WAR means that an opposition player loses some WAR, and demonstrated that even an unreasonably high amount of correlation still renders our methods applicable and appropriate.

An interval for θ^* for which we would not reject a null hypothesis that $\theta = \theta^*$ at a significance level of 5% is (0.3451, 0.3592) using the bootstrap estimate for the mean of the distribution of the true θ^*, and is (0.3578, 0.3641) using the bootstrap estimate for the median of the distribution of the true θ^*. This means that assuming all teams had 46.91 players, the average number of players to appear on a team between 2000-2019, using the interval for the mean we would not reject a null hypothesis that half of the team's total WAR is concentrated in between the best 3.774 and 3.983 players, and using the interval for the median we would not reject a null hypothesis that half of the team's total WAR is concentrated in between the best 3.963 and 4.060 players, using WAR to define "best."

Practically, these results tell a very interesting story about where wins come from in Major League Baseball. Baseball is famously a sport where depth is a necessity, as the rules dictate teams have a lineup of 8 hitters (or 9, depending on the existence of the designated hitter), a starting rotation of about 5 pitchers, a bullpen, and bench depth, making all 26 roster spots on an MLB team incredibly valuable and leaving no room to waste (Sullivan, "26-man Roster"). Furthermore, as opposed to other sports, such as football and basketball, whereby one player can take part in every single play and nearly single-handedly dominate a game, hitters in baseball will only come up to bat about 4 or 5 times per game, and pitchers only start once every 5 games or so, limiting the total impact even the best players in the game can have. On top of this, over the stretch of a 162-game season, players are bound to face injuries both long- and short-term, as teams are forced to call up players from the roughly 250-person pool of minor league players teams employ in their respective minor league systems (Cooper). These facts may lead one to a seemingly logical conclusion that Major League Baseball clubs would likely construct their teams such that talent is roughly evenly distributed across the roster, given the obvious importance of depth. Through this research, we can categorically deny this hypothesis, as we see that the majority of the value produced by players for each team is generally concentrated in about the best four performers.

This raises the question of why teams elect to construct their rosters in this manner. I would hypothesize that this is likely due to the fact that the distribution of talent in Major League Baseball is heavily skewed, as while there is only a handful of superstars, there is an overwhelming abundance of mediocre players ("Baseball-Reference.com…"). This skewed distribution means that "average players are relatively rare and can be expensive to acquire," and thus assembling a team of entirely average players will likely be more expensive and a less efficient method to accumulate as much WAR, and thus win as many games, as possible, as opposed to using that money to employ just a couple superstars ("Baseball-Reference.com…"). Furthermore, due to the league's salary structure, which dictates how much money baseball players can make throughout their careers, young, talented players are paid the league minimum salary for the first few years of their career, regardless of how talented they actually are, and do not reach the free agent market and negotiate with all 30 teams until after playing in the league for six full seasons, with this being the first time players are truly paid what the market deems them to be valued (Rivera). This confluence of factors plays a critical role in dictating how Major League front offices choose to construct their teams, and likely is the reason MLB teams are constructed as they are with respect to WAR. Further research and access to the necessary salary data could expand upon the research question explored in this paper to better understand how the distribution of salaries on teams relates to the distribution of WAR accrued, noting if these distributions are similar and documenting how correlated salary and WAR are.

Bibliography

"26-Man Roster." MLB.com, www.mlb.com/glossary/transactions/26-man-roster.

"Baseball-Reference.com WAR Explained." Baseball-Reference.com, www.baseballreference.com/about/war_explained.shtml.

Bingol, Spencer. "The Strain of WAR You Use Can Dramatically Affect Your Perception of a Player." Beyond the Box Score, Beyond the Box Score, 15 Nov. 2015, www.beyondtheboxscore.com/2015/11/15/9733498/wins-above-replacement-2015-fangraphs-baseball-reference-zack-greinke-chris-sale.

Cooper, J.J. "How Many Players Are In Each MLB Organization? (Ask BA)." Baseball America, 21 Aug. 2018, www.baseballamerica.com/stories/ask-ba-how-many-players-are-in-each-organization/.

Fredrickson, Mark. "Week 7: Bootstrap." *University of Michigan STATS 306.*

Foroux, Darius. "Price's Law: What It Is and Why You Should Care." dariusforoux.com/prices-law/.

Grier, Kevin, and Tyler Cowen. "The Economics of Moneyball." Grantland, 9 Dec. 2011, grantland.com/features/the-economics-moneyball/.

"Generating Correlation Matrix for AR(1) Model: R-Bloggers." R-Bloggers, 7 Feb. 2020, www.r-bloggers.com/2020/02/generating-correlation-matrix-for-ar1-model/.

"Mike Trout Stats." Baseball-Reference.com, www.baseballreference.com/players/t/troutmi01.shtml.

Rivera, Joe. "What Is MLB Arbitration? Explaining the Rules, Eligibility, & How the

Process Works.” Sporting News, 13 Jan. 2020, www.sportingnews.com/us/mlb/news/what-is-arbitration-mlb-what-does-it-mean-baseball-eligibility-process/1atg6pycmf69o1x5yy2v31o03w.

Rizzo, Maria L. Statistical Computing with R. 2nd ed., Chapman Et Hall/CRC, 2019.

Robert, Christian, and George Casella. Introducing Monte Carlo Methods with R. Springer, 2010.

Sullivan, Jeff. “A Quick Attempted Measure of Team Depth, Again.” FanGraphs Baseball, 10 Mar. 2015, blogs.fangraphs.com/a-quick-attempted-measure-of-team-depth-again/.

“WAR Comparison Chart.” Baseball-Reference.com, www.baseballeference.com/about/war_explained_comparison.shtml.