

Introducing *Grid WAR*: Rethinking WAR for Starting Pitchers

Ryan Brill, Justin Lipitz, Emma Segerman, Ezra Troy, Adi Wyner

March 1, 2022

Abstract

Traditional methods of computing WAR (wins above replacement) for pitchers are flawed. Specifically, Fangraphs and Baseball Reference compute a pitcher’s WAR as a function of his performance averaged over the entire season, which is problematic because they ignore a pitcher’s game-by-game variance. Hence we propose a new way to compute WAR for starting pitchers: *Grid WAR* (*gWAR*). The idea is to compute a starter’s *gWAR* for each of his individual games. A starter’s seasonal *gWAR* is thus the sum of the *gWAR* of each of his games. We find that *gWAR* as a function of the number of runs R a pitcher allows during a game is convex, which by Jensen’s inequality implies that current implementations of WAR undervalue high variance pitchers.

1 Introduction

WAR (wins above replacement) is a fundamental statistic for valuing baseball players, and has recently been proposed to determine arbitration salaries (Perry, 2021). So, it is of utmost importance to use a WAR statistic that accurately captures a player’s contribution to his team. However, current popular implementations of WAR for starting pitchers, implemented by Fangraphs (Slowinski, 2012) and Baseball Reference (2011), have flaws. Hence in this paper we propose a new way to compute WAR for starting pitchers, *Grid WAR*.

2 Problems with Current Implementations of WAR

2.1 The Problem: Ignoring Variance

The primary flaw of traditional methods for computing WAR for pitchers, as implemented by Baseball Reference and Fangraphs, is WAR is calculated as a function of a pitcher’s *average* performance. Baseball Reference averages a pitcher’s performance over the course of a season via *xRA*, or “expected runs allowed” (Reference, 2011). *xRA* is a function of a pitcher’s average number of runs allowed per out. Fangraphs averages a pitcher’s performance over the course of a season via *ifFIP*, or “fielding independent pitching (with infield flies)” (Slowinski, 2012). *ifFIP*

Table 1: Max Scherzer’s performance over six games prior to the 2014 all star break.

game	1	2	3	4	5	6	total
earned runs	0	10	1	2	1	1	15
innings pitched	9	4	6	7	8	7	41

is defined by

$$ifFIP := \frac{13 \cdot HR + 3 \cdot (BB + HBP) - 2 \cdot (K + IFFB)}{IP} + ifFIPconstant,$$

which involves averaging some of a pitcher’s statistics over his innings pitched.

To see why using a pitcher’s *average* performance to calculate his WAR is a bad way to measure his value on the mound, we consider Max Scherzer’s six game stretch from June 12, 2014 through the 2014 all star game, shown in table 1 (ESPN, 2014). In Scherzer’s six game stretch, he averages 15 runs over 41 innings, or 0.366 runs per inning. So, on average, Scherzer pitches 3.3 runs per complete game, which isn’t great. If we look at each of Scherzer’s individual games separately, however, we see that he has four dominant performances, one decent game, and one “blowup”. This should be worth something close to four wins, significantly more than we would attribute to Scherzer by looking only at his 3.3 average runs per complete game. In other words, averaging Scherzer’s performances significantly devalues his contributions during this six game stretch.

On this view, averaging over a pitcher’s games is problematic because it ignores his game-by-game variance. Because “*you can only lose a game once*”, it makes more sense to give Scherzer zero credit for his one bad game than to distribute his one poor performance over all his other games via averaging. Hence we should not compute WAR as a function of a pitcher’s average performance, and should compute season-long WAR as the summation of the WAR of his individual games.

2.2 Mathematical Intuition: The Convexity of WAR

We mathematically ground this intuition as follows: WAR should be a convex function in the number of runs allowed. To see why, let’s think of a starting pitcher’s WAR during the j^{th} game as a function of his runs allowed during that game, R_j . If a pitcher throws the same number of runs in each complete game ($R_j \equiv R$ for all j), then summing the WAR over his individual games is equivalent to computing the WAR of his average performance,

$$WAR\left(\frac{1}{n} \sum_{j=1}^n R\right) = \frac{1}{n} \sum_{j=1}^n WAR(R). \quad (1)$$

However, we’ve already conjectured that averaging over game-by-game performances undervalues pitchers such as Scherzer who exhibit some game-by-game variance in performance. Expressed mathematically, we expect

$$WAR\left(\frac{1}{n} \sum_{j=1}^n R_j\right) < \frac{1}{n} \sum_{j=1}^n WAR(R_j). \quad (2)$$

Mathematically, this means we expect WAR to be a convex function in the number of runs allowed. Traditional WAR metrics compute a WAR function reminiscent of the left side of equations (1) and (2), ignoring the right side of these equations. In this paper, by computing season-long WAR as the summation of the WAR of his individual games, as done on the right side of equation (2), we allow WAR to be a convex function.

3 Defining *Grid WAR* for Starting Pitchers

We wish to create a metric which computes a starting pitcher’s WAR for an individual game. The idea is to compute a context-neutral and offense-invariant version of win-probability-added that is derived only from a pitcher’s performance.

First, we define a starting pitcher’s *Grid WAR* ($gWAR$) for a game in which he exits at the end of an inning. We begin by creating the function $f = f(I, R)$ which, assuming both teams have league-average offenses and league-average defenses, computes the probability a team wins a game after giving up R runs through I innings. f is a context-neutral version of win probability, as it depends only on the starter’s performance.

To compute a wins *above replacement* metric, however, we need to compare this context-neutral win-contribution to that of a potential replacement-level pitcher. We use a constant w_{rep} which denotes the probability a team wins a game with a replacement-level starting pitcher, assuming both teams have a league-average offense and defense. We expect $w_{rep} < 0.5$ since replacement-level pitchers are worse than league-average pitchers.

Then, we define a starter’s *Grid WAR* during a game in which he gives up R runs through I complete innings as

$$f(I, R) - w_{rep}. \quad (3)$$

We call our metric *Grid WAR* because the function $f = f(I, R)$ is defined on the 2D grid $\{1, \dots, 9\} \times \{1, \dots, 25\}$.

Next, we define a starting pitcher’s *Grid WAR* for a game in which he exits midway through an inning. We begin by creating another function $g = g(R, S, O)$ which, assuming both teams have a league-average offense and defense, computes the probability that, starting midway through an inning with $O \in \{0, 1, 2\}$ outs and base-state

$$S \in \{000, 100, 010, 001, 110, 101, 011, 111\},$$

a team scores exactly R runs through the end of the inning. Then we define a starter’s *Grid WAR* during a game in which he gives up R runs and leaves midway through inning I with O outs and base-state S as the expected *Grid WAR* at the end of the inning,

$$\sum_{r \geq 0} g(r, S, O) f(I, r + R) - w_{rep}. \quad (4)$$

Finally, we define a starting pitcher’s *Grid WAR* for an entire season as the sum of the *Grid WAR* of his individual games. In order to compute *Grid WAR* for each starting pitcher, we need only estimate the grid functions f and g .

4 Estimating the Grid Functions f and g

4.1 Estimating f

First, we estimate the function $f = f(I, R)$ which, assuming both teams have league-average offenses, computes the probability a team wins a game after giving up R runs through I innings. For $I \in \{1, \dots, 8\}$, we define $f(I, R)$ as the empirical proportion of games in which a team wins after giving up R runs through I innings. We use all games from 2010 to 2019. For $I \in \{1, \dots, 8\}$ and $R \in \{1, \dots, 15\}$, each (I, R) has over [1000] occurrences, which is enough to use the empirical proportion alone.

In figure [PLOT], we plot the functions $R \mapsto f(I, R)$ for each inning I . We see that [].

[PLOT]

[DISCUSS WAR AS A CONVEX FUNCTION]

4.2 Estimating w_{rep}

To compute a wins *above replacement* metric, we need to compare the context-neutral win-contribution to that of a potential replacement-level pitcher. Thus we define a constant w_{rep} which denotes the probability a team wins a game with a replacement-level starting pitcher, assuming both teams have a league-average offense and defense. We expect $w_{rep} < 0.5$ since replacement-level pitchers are worse than league-average pitchers.

It is difficult to estimate w_{rep} because it is difficult to compile a list of replacement-level pitchers. According to Fangraphs (2010), *replacement-level* is the “level of production you could get from a player that would cost you nothing but the league minimum salary to acquire.” Since we are not members of an MLB front office, this level of production is difficult to estimate. Ultimately, the value of w_{rep} doesn’t matter too much because we rescale all pitcher’s *Grid WAR* to sum to a fixed amount, to compare our results to those of Fangraphs. So, we arbitrarily set $w_{rep} = 0.41$.

4.3 Estimating g

Second, we estimate the function $g = g(R, S, O)$ which, assuming both teams have league-average offenses, computes the probability that, starting midway through an inning with $O \in \{0, 1, 2\}$ outs and base-state

$$S \in \{000, 100, 010, 001, 110, 101, 011, 111\},$$

a team scores exactly R runs through the end of the inning. We define $g(R, S, O)$ by the empirical proportion, for $R \in \{1, \dots, X\}$. We use the first 8 innings of every game from 2010 to 2019. We average over the first 8 innings for simplicity, and because g isn’t significantly different across innings.

[PLOT CURVES of $g(R, S, O)$ as a function of R for some (S, O) combinations]

5 Comparing *Grid WAR* to Fangraphs WAR

[2D PLOT]

6 Example Seasons

7 Conclusion

8 Future Work

WAR for relievers...

References

ESPN. Max scherzer 2014 game log. https://www.espn.com/mlb/player/gamelog/_/id/28976/year/2014/category/pitching, 2014.

Fangraphs. Replacement level. <https://library.fangraphs.com/misc/war/replacement-level/>, 2010.

Dayn Perry. Mlb proposes determining arbitration salaries by using the war statistic, per report. <https://www.cbssports.com/mlb/news/mlb-proposes-determining-arbitration-salaries-by-using-the-war-statistic-per-report/>, 2021.

Baseball Reference. Pitcher war calculations and details. https://www.baseball-reference.com/about/war_explained_pitch.shtml, 2011.

Piper Slowinski. War for pitchers. <https://library.fangraphs.com/war/calculating-war-pitchers/>, 2012.