Data 8 Connector: Sports Analytics

...

WAR and Win Shares

Course Stuff

- ➔ This lecture: A sketch of what WAR and Win Shares do
- ➔ No Demo: not sure what to show. Mostly mechanical and very complex Methodologies/ideas more important, in my opinion
- ➔ Will also talk about how it applies in basketball

Quantifying Wins

Pythagorean Expectation already gave us the notion of Runs per Win

At times I've related run production to wins

- Bonds' ~120 runs above average is about 12 wins above average

So we already know how to convert runs to wins, what else is there?

WAR and Win Shares

WAR: Wins Above Replacement

Win Shares (from Bill James): credited wins to a player

Two competing statistics/philosophies for summarizing contributions to winning

WAR and Win Shares

We already know Runs per Win, so what do we need to do to get WAR/WS?

The real challenges:

- ➔ Divying up credit
  - With wRAA, we already have run credit. WS uses RunsCreated
  - WS is different. Actually divy up wins (hence the name...)
- ➔ Accounting for Parks/Positions/League
  - Parks: dimensions, weather, etc
  - Positions: need to adjust for difficulty
- ➔ For WAR specifically: what is “replacement” and why?
  - Concept is easy, but the implementation?

WAR: Overview

4 different (but very similar) approaches to computing WAR

- ➔ Baseball Reference
- ➔ FanGraphs
- ➔ Baseball Prospectus
- ➔ openWAR (open source version)
  We likely will explore oWAR due to its extensive/creative/beautiful use of regression modeling

Going to look at FanGraphs approach but BR is very close

WAR: Overview

Fundamental Formula (for position players)

$$\text{WAR} = (\text{Batting Runs} + \text{Base Running Runs} + \text{Fielding Runs} + \text{Positional Adjustment} + \text{League Adjustment} + \text{Replacement Runs}) \ / \ \text{Runs Per Win}$$

- ➔ We probably will not cover fielding/defense or pitching
  - Tools/analyses are the same, it's just quite a bit more difficult, especially fielding
  - Also compressed due to time/desire to cover other topics
- ➔ So we know a hefty chunk of this formula, but there's still some meat on the bone

WAR: Overview

Still to address:

- ➔ Park effects on run scoring
- ➔ Positional adjustments
- ➔ League adjustments
- ➔ Replacement level

Win Shares: Overview

This is a *really* loose overview: see Bill James' book on WS for the details

I just want to hit on some salient points

The overall goal: divy up wins between offense/defense and then among players

To start: compute the *marginal runs scored* and *marginal runs saved* for a team

$$\text{xWin\%} = \frac{MRSc + MRSa}{2 \cdot lgR} = \frac{(R - .5 \cdot lgR) + (1.5 \cdot lgR - RA)}{2 \cdot lgR} = \frac{1}{2}\left(1 + \frac{R - RA}{lgR}\right)$$

Win Shares: Overview

Total offensive and defensive win shares for a team:

$$OWS = \frac{MR}{MR + MRA} \cdot W \cdot 3, \quad DWS = \frac{MRA}{MR + MRA} \cdot W \cdot 3$$

- ➔ Divy up the win shares to offense and defense
- ➔ Why 3? A win share is now indicative of a third of a win.
- ➔ And it makes the presentation look good

Win Shares: Overview

How to divy up OWS?

1. Compute the Runs Created for all players on a team
2. Compute the outs made by each batter.
3. Compute the player's marginal runs (no less than zero)

$$playerMR = RC - lgRperOuts \cdot Outs \ / \ 2$$

4. Player's share of OWS is playerMR / sum of playerMR

Left it out but we need to adjust runs for ballparks

Win Shares: Overview

For DWS, do similar with computations for fielding/pitching

So what does WS try to tell you?

- ➔ Divy up the *actual* wins according to run contributions
- ➔ Run contributions yield win contributions (or shares)
- ➔ Unlike a context neutral approach (ie. everything we've done so far), WS accounts for Pythagorean Luck (lower/higher Win% than runs suggest)

We'll return to WS as a philosophy in a bit

Park Factors

So far, we've just handwaved away the notion of adjusting for ballparks

Let's remedy that by getting to the bottom of what a park factor is

Park Factors: The Obvious Way

The easiest/simplest park factor is a simple ratio:

$$PF = \frac{\text{Total Runs per Game Home}}{\text{Total Runs per Game Away}}$$

From my understanding, this is the park factor ESPN uses

- ➔ You can improve on this by averaging over a multi-year period
- ➔ You can also improve it by “regressing to the mean”

$$PF^* = \alpha PF + (1 - \alpha), \quad 0 < \alpha < 1$$

Park Factors: Other Approaches

The simple factor has some issues and baseball reference has written up their approach

FanGraphs has their own method too (involving multiple years and regressing to the mean)

Generally, approaches are based on that initial ratio

Park Factors: Why?

First, why park factors at all?

We want to account for variability in the parks

This is known as *heterogeneity*.

- ➔ From heterogeneous: hetero = different, gene = kind
- ➔ It refers to observed data coming from different sources with different characteristics

Park Factors: Why?

Second, why is the ratio of run scoring the park factor?

In an “ideal” world, we’d know how every single ball hit is affected by the ballpark and “correct” the ball to account for the park

Realistically, that is just not possible

Instead, we opt for a “statistical correction” to account for overall increase/decrease in run scoring

Park Factors: Why?

As an aside, sometimes folks build park factors for events/components

E.g. a park factor to measure increase/decrease in HR

There are some issues/controversies with this

- ➔ If you get too granular (break the data down into lots of categories), you will shrink your sample sizes in any particular category and hurt your ability to estimate anything
- ➔ Modeling-wise, this isn't really what you want anyhow
  If a player with a certain style is affected by a ballpark differently than the generic park factor indicates (hitter friendly overall but not for this player), should we try to correct for that?
  Maybe, but not likely. This would be like correcting for handedness or height.

Park Factors: How to use them

Okay, computing the park factor is easy. How do you use them?

The most illustrative is by introducing a stat called *Weighted Runs Created Plus*

wRC+ measures how a player's weighted runs created (using our old friend wRAA) compares with league average after controlling for park effects

wRC+ is great because we can compare across era and the numbers are very intuitive

wRC+: Park-Adjusted Offensive Performance

The formula for *wRC+*

$$wRC/PA = (wRAA/PA + lgR/lgPA),$$
$$wRC+ = \frac{wRC/PA + (1 - PF) \cdot lgR/lgPA}{\text{League average } wRC/PA} \cdot 100$$

- ➔ Start with runs created: add the average to runs above average
- ➔ Add in the park factor correction
- ➔ Divide by league average and scale by 100
- ➔ *wRC+*=110 => 10% above average. *wRC+*=90 => 10% below average

wRC+: Park-Adjusted Offensive Performance

$$wRC/PA = (wRAA/PA + lgR/lgPA),$$
$$wRC+ = \frac{wRC/PA + (1 - PF) \cdot lgR/lgPA}{\text{League average } wRC/PA} \cdot 100$$

The park factor modifies the average in “runs above average”

- ➔ If PF = 1, no modification
- ➔ If PF = .9 (pitcher friendly), the ballpark suppressed run scoring, the average in that park is low, so adjusted RAA needs to be higher
- ➔ If PF = 1.1 (hitter friendly), the ballpark increased run scoring, the average in the park is high, so adjusted RAA needs to be lower

Park Factors

And that's kind of it for park factors

It's a bit opaque when you hear about adjusting for park

But the actual mechanism is simple and intuitive

Positional Adjustments

WAR uses positional adjustments to better rate players

The positional adjustment accounts for defensive difficulty

➔ Fielding values will rate against average at the position

➔ So how do you compare average *between* positions

The values were computed years ago and still are used (it's really debatable whether these values are any good)

Positional Adjustments

Adjustments are all per 162 defensive games

- ➔ Catcher: +12.5 runs
- ➔ First Base: -12.5 runs
- ➔ Second Base: +2.5 runs
- ➔ Third Base: +2.5 runs
- ➔ Shortstop: +7.5 runs
- ➔ Left Field: -7.5 runs
- ➔ Center Field: +2.5 runs
- ➔ Right Field: -7.5 runs
- ➔ Designated Hitter: -17.5 runs

League Adjustment

The league adjustment is just a simple adjustment for the average level of the league

It has a small impact but is meant to correct for the overall run scoring of the league

Subtract this from total runs for a player

$$\text{League Adjustment} = \frac{lg\text{BattingRuns} + lg\text{BaserunningRuns} + lg\text{FieldingRuns} + lg\text{PositionAdj}}{lgPA} \cdot PA$$

Replacement Level

Okay, now we get to the big idea behind WAR

First, what is replacement level and why should we care about it?

From FanGraphs

> Replacement level is simply the level of production you could get from a player that would cost you nothing but the league minimum salary to acquire. Minor league free agents, quad-A players, you get the idea. The concept is pretty tidy. These are the players that are freely available and if five of your MLB level players came down with the flu, you could go out and acquire replacement level players without really giving up anything you value other than their union mandated payday.

Replacement Level

Easy enough, so why should we care? Isn't Above Average good enough?

There are two issues with Above Average, one presentational, one practical

- ➔ Why should average be 0? Average players are helpful and important in baseball
- ➔ If player A produces 0 runs above average in 1 PA and player B produces 0 runs above average in 600 PA, who's done better? How do you tell the difference? They both netted 0 runs above average overall and per PA
- ➔ We want to reward playing average over long stretches
- ➔ If we peg players to a different "0 level", then average players will accumulate positive results and player B will look better than player A

Replacement Level

Okay, so how do we compute it?

Well, it's going to be a Runs Above Average value but it will be negative

Also, it's going to be on a per PA basis so we can pro-rate players to build WAR

So here's how we start (this is the FanGraphs approach, identical to B-R)

1. Assume that a team full of replacement players will win 29.7% of it's games in MLB (about 47-48 wins)

This is a modeling choice based on reasonable belief about ability of replacement-type players

Replacement Level

1. Assume that a team full of replacement players will win 29.7% of it's games in MLB
2. 30 replacement teams would yield about 1430 wins total
3. There are still 1000 wins left (2430 total wins in MLB)
4. Those 1000 wins are the "wins above replacement" to divy out to the players
5. For offensive players, they get 570 wins. Defense gets 430. (Modeling choice)
6. For offense, divy up the wins by PA

$$\text{Replacement Level Runs} = 570 \cdot \text{RunsPerWin} \cdot \frac{PA}{lgPA}$$

Replacement Level

1. Assume that a team full of replacement players will win 29.7% of it's games in MLB
2. 30 replacement teams would yield about 1430 wins total
3. There are still 1000 wins left (2430 total wins in MLB)
4. Those 1000 wins are the "wins above replacement" to divy out to the players
5. For offensive players, they get 570 wins. Defense gets 430. (Modeling choice)
6. For offense, divy up the wins by PA

$$\text{Replacement Level Runs} = 570 \cdot \text{RunsPerWin} \cdot \frac{PA}{lgPA}$$

Return to WAR

Let's look at the WAR formula again

$$\text{WAR} = (\text{Batting Runs} + \text{Base Running Runs} + \text{Fielding Runs} + \text{Positional Adjustment} + \text{League Adjustment} + \text{Replacement Runs}) \ / \ \text{Runs Per Win}$$

Aside from glossing over fielding runs, now we know what everything is!

WAR vs Win Shares: Philosophical Differences

Nov 17, 2017: Bill James wrote a wonderful screed (rant?) on WAR, Win Shares, and Aaron Judge vs Jose Altuve for MVP

https://www.billjamesonline.com/judge_and_altuve/

I will try to summarize the issue to illustrate the differences

WAR vs Win Shares: Philosophical Differences

Baseball-Reference had Altuve at 8.3 WAR and Judge at 8.1 WAR

That's basically identical

According to Bill James:

> It's nonsense. Aaron Judge was nowhere near as valuable as Jose Altuve. Why? Because he didn't do nearly as much to win games for his team as Altuve did. It is NOT close.

WAR vs Win Shares: Philosophical Differences

The issue comes down to context and neutrality

> ...there is a ***general*** relationship between runs and wins, a normal relationship, and there is a specific relationship, based on this specific player and this specific team. If you evaluate Altuve and Judge by the general and normal relationship of runs to wins, then it appears that Judge is almost even with Altuve. But if you evaluate them by the specific relationship of Altuve's runs to the Astros wins and Judge's runs to the Yankees wins, then Altuve moves up and Judge moves down, and a significant gap opens up between

WAR vs Win Shares: Philosophical Differences

The Yankees scored in such a way that they should have won 102 games. They won 91.

The Yankees played poorly in one-run games (18-26) and other close games.

Allow Bill James to explain why this might be a problem

> ***...it is not right to give the Yankee players credit for winning 102 games when in fact they won only 91 games.*** To give the Yankee players credit for winning 102 games when in fact they won only 91 games is what we would call an "error". It is not a "choice"; it is not an "option". It is an error.

Basically:

- ➔ Context neutrality looks at components to build runs.
- ➔ BJ argues if your components don't win games, you shouldn't be credited.

WAR vs Win Shares: Philosophical Differences

The gap between context neutrality and dependence widens because Judge performed poorly in important situations

- ➔ Runners in scoring position: Judge hit .262, 22 points less than his overall avg
- ➔ Judge's OPS was 90 points higher with the bases empty
- ➔ Altuve's OPS was 1 point higher with men on base.
- ➔ In late innings of close games (100 plate appearances), Judge hit .216 with a .780 OPS.
- ➔ With the Yankees 4 or more runs ahead or 4 or more runs behind (112 plate appearances), he hit .382 with an OPS of 1.500.
- ➔ In the late innings of close games, Jose Altuve hit .441 with a 1.190 OPS
- ➔ When the Astros were 4 or more runs ahead or 4 or more runs behind, Altuve hit .313 with a .942 OPS.

And so on...

WAR vs Win Shares: Philosophical Differences

Bill James' big question:

> Is it appropriate, in assigning the individual player credit for wins, to do so based on the usual and normal relationship of runs to wins, or based on the actual and specific relationship for this player and this team?

In summary, BJ has a very strong take: WAR is wrong and bad

As a comment on this, it's pretty rich that an analyst who is supposed to be practical and nuanced could get so rigid and dogmatic about something

WAR vs Win Shares: Philosophical Differences

Dave Cameron from FanGraphs responded

https://www.fangraphs.com/blogs/putting-war-in-context-a-response-to-bill-james/

> My primary guiding principle on the usefulness of a metric is twofold:
>
> 1. What question is it answering, and is the answer to that question interesting? If yes, proceed. If no, ignore.
>
> 2. Does the metric answer that question accurately?
>
> ...
>
> WAR, on the other hand, attempts to address a question that a lot of people seem interested in answering. If the WAR leaderboards were posed as a question, they might be written as something like this:
>
> “What did each player do, as an individual, to help his team try to win games?”

WAR vs Win Shares: Philosophical Differences

Another nice quote from Cameron:

> ...the general consensus in the sabermetric community has been that we want to reward (or penalize) hitters for what they can control. And the context of the situations in which they play is just not something players can create.

WAR vs Win Shares: Philosophical Differences

On MVP voting:

> I think the answer is that it depends on how you're using WAR. In the case of MVP voting, I do think there is a case to be made for looking at the circumstances under which a player performed, and I did use context-dependent metrics when I was an MVP voter. WAR is an imperfect tool, and it's particularly imperfect for things like the MVP award, which is why even those of us who host sites that promote WAR fairly extensively suggest not relying solely on its results when filling out a ballot.

WAR vs Win Shares: Philosophical Differences

Dave Cameron is a really good writer and he outlines a lot of issues with context neutrality (and FanGraphs has some interesting context-dependent stats too!)

In short, this comes full circle to analytical thinking and knowing your questions

They're both right:

- ➔ BJ is right WAR is not great for MVP
- ➔ Dave Cameron is right WS actually doesn't do enough with context dependence and WAR is good for what it's meant for.

What about basketball?

WARP in Basketball

I'll briefly go over WARP (from Kevin Pelton)

The meat of the computations can be found on his site, the main point is, as before, the methodology

The goal with WARP is to first get team offense and defense ratings for a hypothetical team featuring a particular player and 4 league average players

WARP in Basketball

To do, this we emphasize the offensive computationn(concepts carry over from PER actually)

- ➔ First compute points created (FGs and assists and adjusting for assisted fgs)
  - ◆ This is very similar to how PER computes points while adjusting for assists
- ➔ Possessions used by the player
  - ◆ Again, very similar to the formula for possessions pioneered by Dean Oliver
- ➔ Now you have an offensive rating for the player on the possessions used
  - ◆ To build the team, we need to account for how many possessions were used

WARP in Basketball

- Usage rate
  - Player possessions over total team possessions during player's time on court
- Hypothetical team offense rating
  - Weighted combination (by usage rate) of player rating + 4 average players (league off. rating)
- Factor in offensive rebounding to adjust team off. Rating up
- Similar ideas for defense but not nearly as crisp/clear

End results: hypothetical team offense and defense ratings

- Use Pythag Exp to get xWin%
- RL is .415

WARP in Basketball

Converting to wins

- ➔ We could use the points to wins from Pythagorean expectation
- ➔ We can also compute the difference between the expected win % of the hypothetical team and a replacement team
  - ◆ Compute expected win % from Pythagorean Expectation
  - ◆ Replacement level is a team with a replacement player and 4 average players
    Set at 415 win %
- ➔ Finally we get WARP:

$$WARP = (\text{xWin\%} - RL) \cdot \frac{MP}{48}$$

Win Shares in Basketball

The steps for WS in basketball is also not that difficult and like with baseball

1. Compute points produced for each player (Dean Oliver's stat)
2. Compute offensive possessions for each player
3. Compute marginal offense for each player.

$$\text{Marg. Off.} = \text{Points Produced} - 0.92 \cdot \text{League Points per Poss} \cdot \text{Off. Poss}$$

4. Compute marginal points per win

$$\text{Marg. Points per Win} = 0.32 \cdot \text{League Points per G} \cdot \frac{\text{Team Pace}}{\text{Lg Pace}}$$

Win Shares in Basketball

Finally credit Offensive Win Shares to the players

$$OWS = \frac{\text{Marg. Off.}}{\text{Marg. Points per Win}}$$

Similar approach holds for defense using the concepts we saw with WARP

Summary

WAR(P) and WS are two (very different) approaches to quantifying wins

WAR is context neutral whereas WS is not: this is an important analytical difference

Both rely on Park Factors but in general, we looked at ways to account for *heterogeneity* when trying to build a comprehensive metric

- ➔ Park factors: adjustments for hitter/pitcher friendly parks
- ➔ League adjustment: split by league instead of overall
- ➔ Positional adjustment: account for difficulty of defensive positions

Summary

Finally, it's important to remember Replacement Level as a very useful concept

➔ We want to fairly grade average players, especially players who can play a lot at average level. 0 is not a fair value.

And of course all this can be carried over to basketball.