The Rise in Infield Hits

Parker Phillips
Harry Simon

December 10, 2014

Abstract

For the project, we looked at infield hits in major league baseball. Our first question was whether or not infield hits have been on the rise since 2005. We found through analyzation of the percentage of infield hits, number of infield hits, and percentage of ground balls, we can definitively say that infield hits have become more common in major league baseball. From there, we looked at why this has been occurring. Our findings suggest that one important aspect of why this has happened is due to a decrease in the overall ERA of the league. From there, we looked at if IFH% is at all an important statistic in evaluating the offensive worth of a player. This result was inconclusive, but suggest that IFH% is more important that it has been accredited for.

1 Introduction

First, why is a player's Infield Hit% (IFH%) important? IFH% is calculated by taking the total number of hits that do not leave the infield divided by the total number of ground balls. A player with a high IFH% tells us that the player is turning would-be-groundball-outs into singles. A player with a low IFH% tells us that out of all their ground balls, the player is generating outs (potentially double plays) instead of singles.

The first question we had to address is the following: Has the number of infield hits been increasing since 2005? To look at this we needed to look at both the infield hits statistic, and any statistic related to infield hits to isolate whether or not it has in fact been increasing on its own, and not as a result of something else.

All of the data we worked with was obtained from fan graphs.com We grabbed the information that pertained to the entire MLB in relation to IFH%.

After graphing IFH% versus the year (see Figure 1), it became quite clear that the percentage of infield hits has been on a definite upward trend since 2005. From there, to make sure that this trend was not a result of there being less ground balls, we analyzed the number of ground balls in that time as well. Figure 2 also has an upward trend. These two figures together proved to us that the number of infield hits in the major leagues has definitely been increasing since 2005.

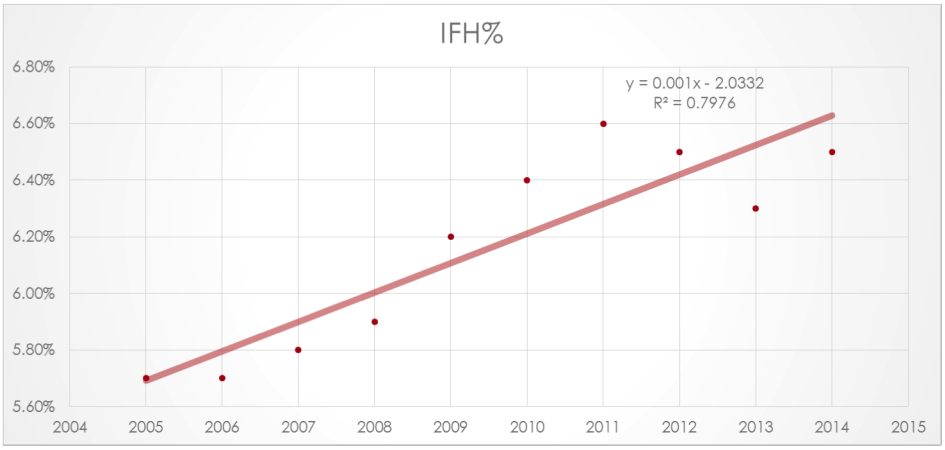

Figure 1: IFH% over time

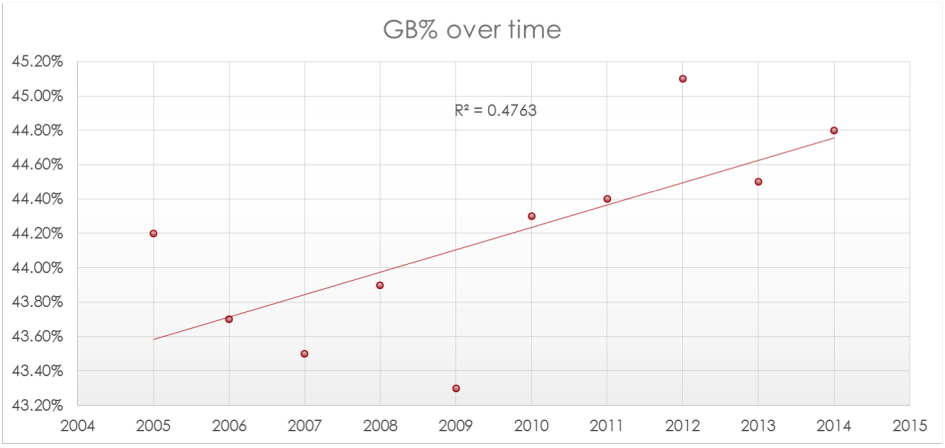

Figure 2: GB% over time

The data holds that IFH% has been on a rise in the past years. In the following sections, we discuss different methods of evaluating players by their IFH%.

2 Over Time

Once we found that IFH% has definitely been increasing, we asked if IFH% is an important tool in evaluating the worth of a player. To do this, we wanted to see any correlation between IFH% and WAR. We chose WAR because it is a subjective, widely accepted tool used to evaluate the worth of a player. If IFH% has a strong correlation with WAR, we would be able to say that it is an important statistic. However, since it is not a strong correlation, we created a new offensive statistic to show that although IFH% by itself is not strong in predicting WAR, adding it to other stats increases its value.

We look at data for ERA over time. Because ERA is a statistic where the higher the number is, the worse the pitcher is, we changed it to negative ERA in order to clearly visualize the relationship it has with other statistics. First, we graphed it against time. If pitching has been either improving or declining over time, then it would be possible for it to be affecting IFH%. (See Figure 3

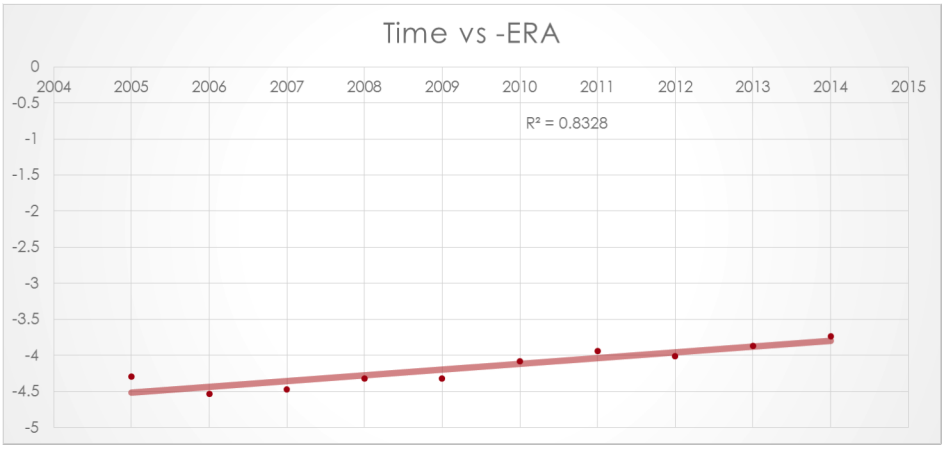

Figure 3: Negative ERA over time

In Figure eraOvertime, it is easy to see that pitching has definitely been improving over time. The R^2 value between season and ERA is the IFH% statistic is 0.83. With this information, we could then directly compare IFH% and ERA and know that any correlation would be matched over time as well which would support our question (see Figure 4). After graphing the two against each other, it became clear of a definite correlation. The R^2 between the two is 0.7365, which is a strong correlation.

To see if this correlation is unique, ERA was compared to other hitting relating statistics, including Average, R^2 = .87, with a negative slope. Intuitively, this statistic makes sense. As pitching

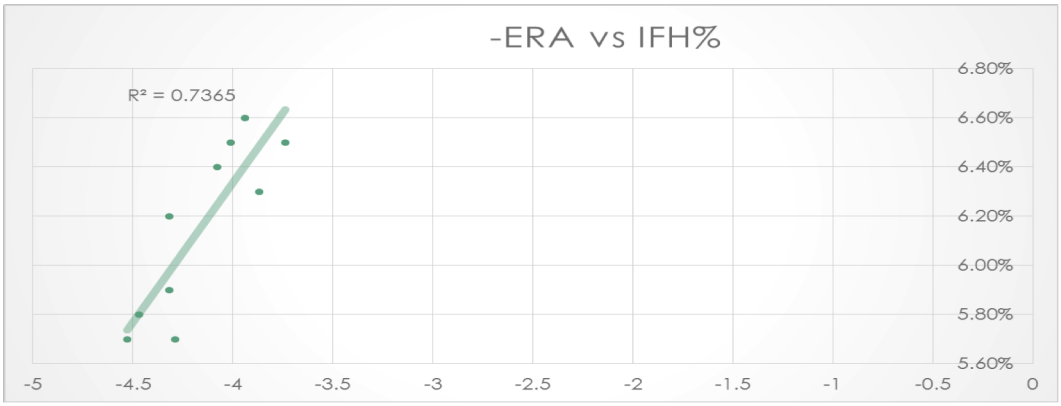

Figure 4: IFH% versus Negative ERA

increases, the batting average of players should be decreasing. When compared to slugging average, the R^2 value equals .83 with a negative slope, and for number of RBIs, it is also .83 with a negative slope.

What these comparisons showed us is that hitting statistics normally correlated with stronger hitters have been getting worse as ERA gets better. Although this makes sense, it made us question why then, would Infield Hit Percentage increase with better pitching.

3 Hitting versus Pitching

Hypothesis: As pitching gets better, players are hitting solid hits less and less, including top tier players. These top tier players however, have speed, and are still able to beat out the throw to first and get the infield hit. Infield hits are far less effective tools that a normal base hit, which explains why this is happening as ERA decreases.

To see if this was true, we created a statistic based on IFH% and SLG% to see if IFH does play any role in predicting WAR (see Figure 5). SLG% was chosen because, when compared with WAR, it had the highest R^2 value of .4. When added together in a 5 to 4 ratio (5*SLG% + 4*IFH%) the combination was better at predicting WAR than either by itself, with an R^2 value of .44 (see Figure 6).

By adding even more unlikely statistics that relate to IFH%, such as GB% and LD%, we were even able to push the R^2 value up to .477 (see Figure 7). This shows that players who have both a high SLG% and a high IFH% tend to have higher WARs than players who simply have a high SLG%. A simple example would be to compare Andrew McCutchen to Chris Davis. McCutchen has an 8.1 IFH% and a .508 SLG%. Davis has a 3.1 IFH% and a whopping .634 SLG%. However, of the two, McCutchen has a better WAR value by 1.4 (8.2 to 6.8).

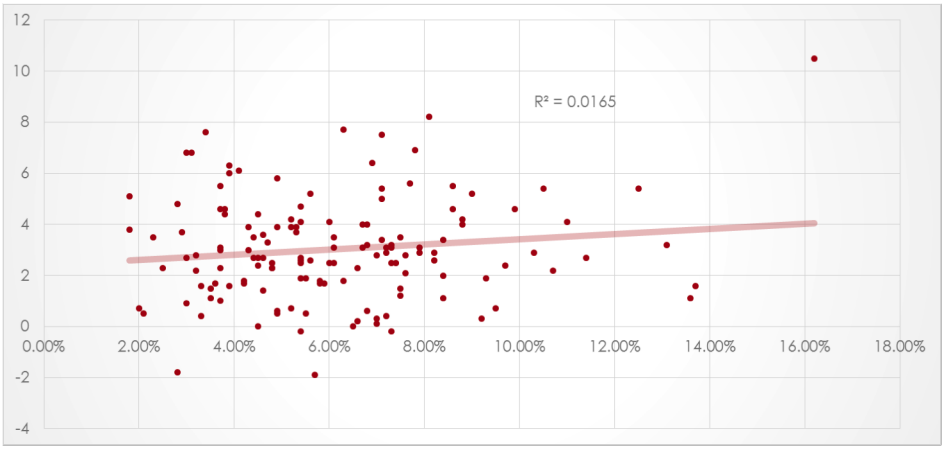

Figure 5: IFH% as a predictor of WAR

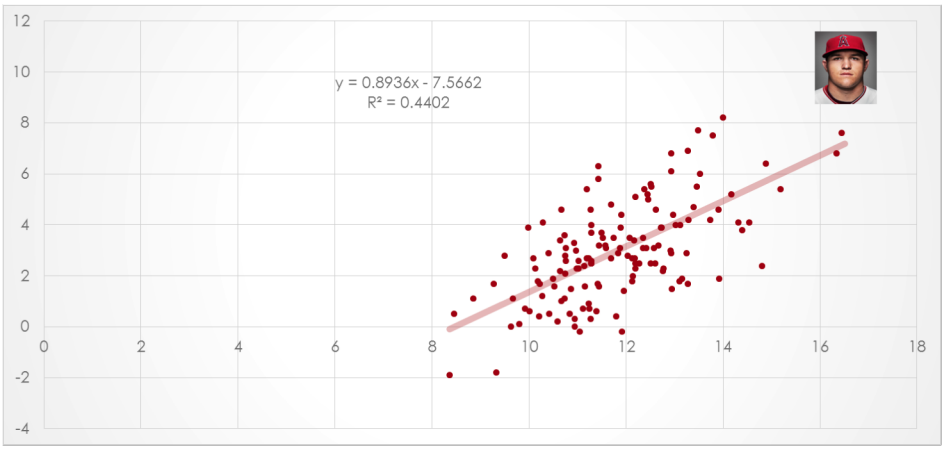

Figure 6: IFH% and SLG% as predictors of WAR

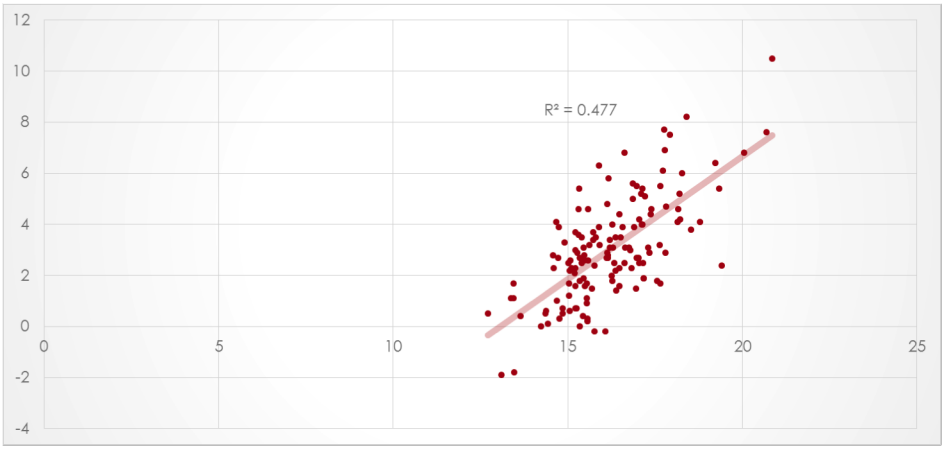

Figure 7: IFH% + SLG% + GB% + LD% as predictors of WAR

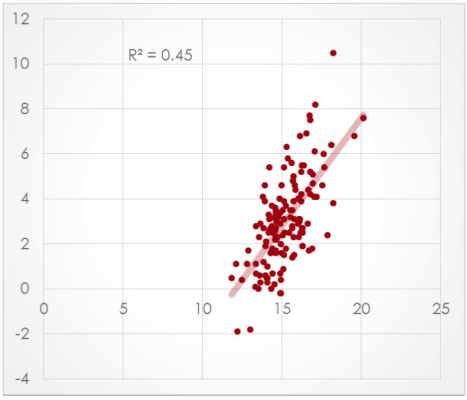

Figure 8: SLG% + GB% + LD% as predictors of WAR

4 Speedy Players

Currently there are not many statistics that predict how valuable a player's speed is to the game of baseball. In the 1970's, Bill James formulated a statistic called Speed Score or Spd. On Fangraphs, Spd is calculated using Stolen Base Percentage (SB%), Frequency of Stolen Base Attempts, Percentage of Triples, and Runs Scored Percentage. Moreover, Fangraphs has a Base-Running Value statistic (BsR), which they consider to be an "all encompassing base running statistic that turns stolen bases, caught stealings, and other base running plays into runs above and below average".

However, neither statistic incorporates IFH% into their calculations. Stolen bases and triples are rather clear indicators of speed. From a stolen base, a speedy player turns a single into a double based off of speed alone. A speedy player can turn a double into a triple perchance (and even sometimes a single into a double). However, a player who can turn a would-be-groundball-out into a single should also be deemed as a speedy player. From this, it is more than reasonable to believe that IFH% should be incorporated in determining the speed of a player.

The new statistic we formulated takes into account IFH%, the player's Base Running value, Stolen Base %, and Triples. We call this statistic NewSpeed. It is defined as:

$$\text{NewSpeed} = c_1 \times \frac{IFH\%}{LeagueAverageIFH\%} + c_2 \times \frac{BsR}{LeagueAverageBsR} + c_3 \times \frac{SB\%}{LeagueAverageSB\%} + c_4 \times \frac{3B}{LeagueAverage3B}$$

Consider Table 1 comparing outfielders Starling Marte and Carlos Gomez. Both players have nearly identical stats, except Starling Marte has a 2% lead in IFH% and nearly twice the value in Base Running. The average for IFH% was 7.9% among outfielders. Spd rates Marte as the speedier player, which NewSpeed accounts for as well. Yet, Marte's and Gomez's Spd's are nearly identical, despite the gaps in Base Running value and IFH%. Our stat NewSpeed suggests Marte is a much speedier player, accounting for Marte's abilities on the base paths.

Player	Offense	Base Running	IFH	IFH%	SB	SB%	3B	Spd	NewSpeed
Starling Marte	25.3	5.8	24	14.5 %	30	73.2%	6	7.1	6.09
Carlos Gomez	26.8	3.3	20	12.7%	34	73.9%	4	6.5	4.58
Who wins?	C.G.	S.M.	S.M.	S.M	C.G	C.G	S.M.	S.M.	S.M.

Table 1: Comparing Players

4.1 Analysis

In order to assess our new statistic NewSpeed, we assessed how it measured up to the other speed statistics in predicting WAR.

Figure 9 shows Spd as a predictor of WAR. However, it possesses the lowest R^2 value. Spd is not a good predictor of WAR.

Figure 10 shows Base Running Value as a predictor of WAR and possesses a R^2 value of 0.1309, double that of Spd.

Figure 11 shows an unweighted NewSpeed as a predictor of WAR and possesses the greatest R^2 value of 0.17. This suggests of the three statistics, it is the best predictor of WAR. When NewSpeed is weighted with the following weights:

$$c_1 = 5,\ c_2 = 2,\ c_3 = 5,\ c_4 = 0$$

The weighted NewSpeed has an R^2 value of .28.

Figure 13 factors out IFH% as part of the NewSpeed statistic. The R^2 value decreases here significantly, almost identical with that of Base Running value. Figure 13 shows that IFH% does better help predict WAR values.

Our new speed statistic shows that players who possess good IFH%, SB%, general Base Running Value, and are able to generate Triples will be the more valuable *speedier* players.

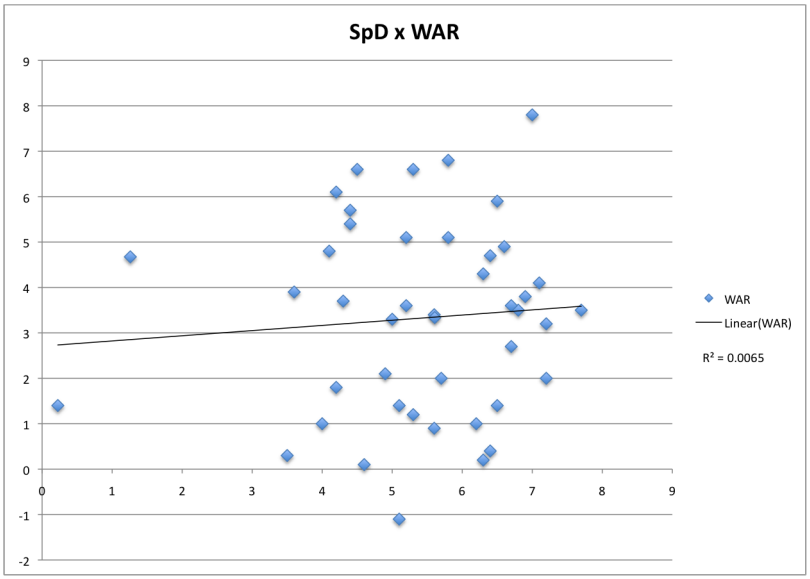

Figure 9: SpD Rating x WAR

5 Conclusion

Although our findings show that IFH% has potential to be a useful statistic for predicting value, the best statistic we could create with IFH% was still fairly poor at predicting the WAR of a player with an R^2 value of only 0.47. What we did find to be useful is that IFH% is a good statistic for predicting ERA of a pitcher. If a team keeps track of the number of infield hits given up by a pitcher, they can have a pretty good idea of what the ERA of the pitcher might be.

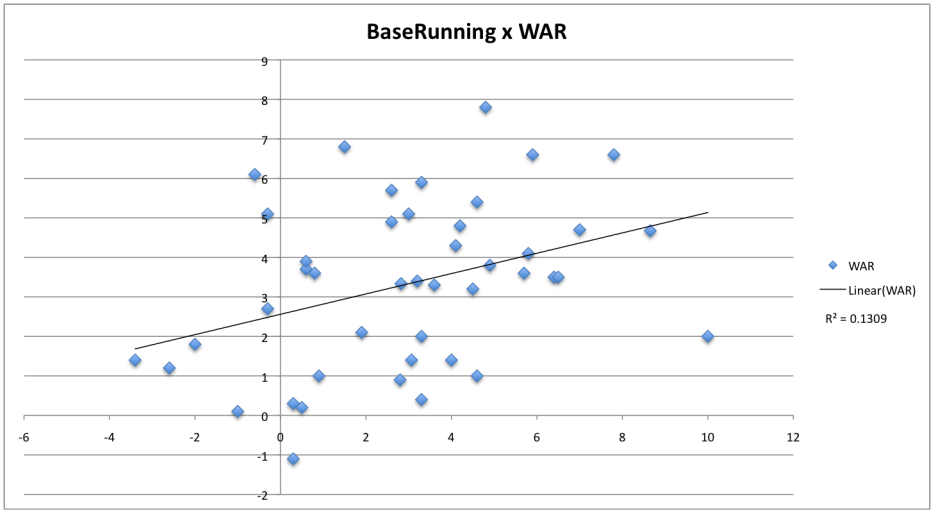

Figure 10: BaseRunning x WAR

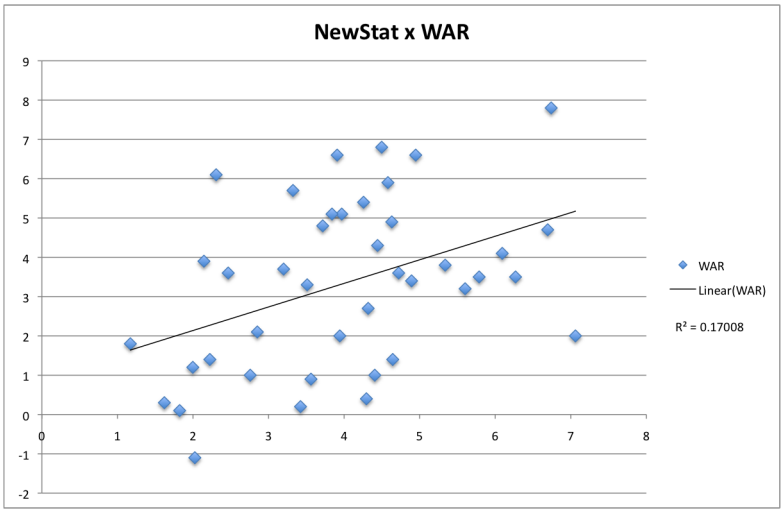

Figure 11: NewSpeed x WAR

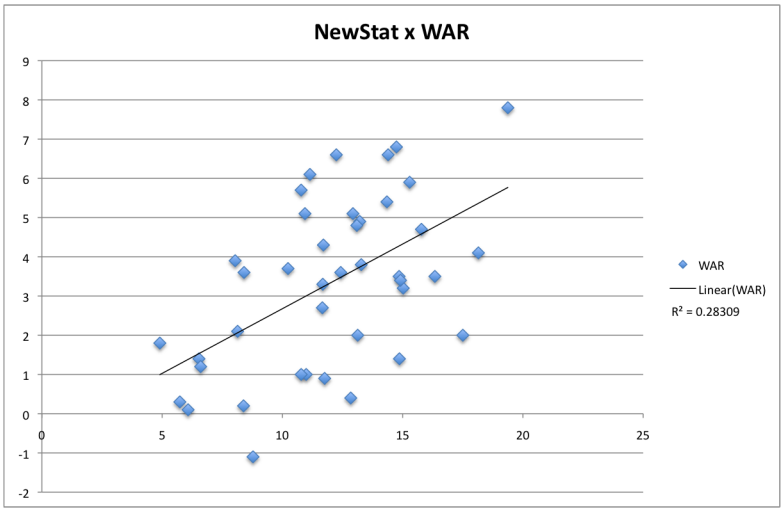

Figure 12: weightedNewSpeed x WAR

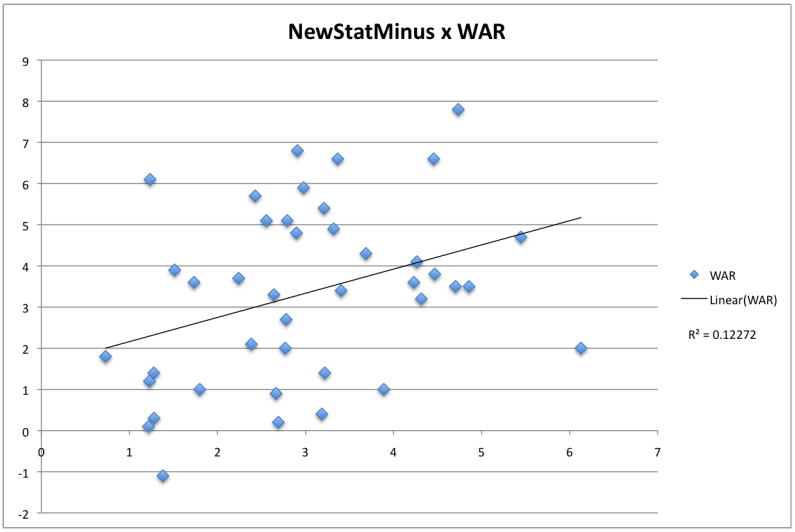

Figure 13: NewSpeedMinus x WAR

Overall, given more time and resources, we believe that it is definitely possible to create a statistic utilizing IFH% that accurately reflect either the value of a position player, or the value of a pitcher, and that IFH% can be used as an important statistic in the future.