Wins Above Replacement Career Trajectory Curves

Drew Jordan and Ehsan Bokhari
Department of Statistics, University of Illinois at Urbana-Champaign

1. Introduction

Baseball is a game of statistics where every event is categorized, counted, and analyzed. The common baseball statistics such as Hits, Homeruns, Runs Batted In, and Batting Average are as old as the sport itself. Until more recently, these were the primary stats that analysts of the sport used to establish which players performed the best. However, sometimes these statistics fall short in establishing how good a player is and how much they contribute to a team's ability to win games. To combat this problem, baseball analysts have created an all-encompassing statistic called Wins Above Replacement (WAR) that attempts to capture how many wins a particular player adds to his team per season above the typical available AAA (minor league) level player. In this research, we have reconstructed the WAR statistic using unique methods and have created a career trajectory curve for Albert Pujols, widely regarded as the best baseball player of the current generation. We believe that the plot of Albert Pujols' career WAR trajectory is very indicative of the progression of his career and that WAR trajectory curves can be used to predict future performance of Major League Baseball players.

4. Conclusion

From the results of this research, we can see that it is indeed possible to calculate a player's WAR on a play-by-play basis and then use this information to produce a career WAR trajectory curve. We have also shown that, at the very least for Albert Pujols, it is possible to predict a player's future performance based on what he has done in the past within a reasonable level of error. This research could prove to be useful in determining how a player's career will pan out, and help draw a better picture of how a player has performed in the past. The next step in this research is producing WAR career trajectory curves for every player in baseball history and using these curves to find similar curves to young, up and coming players today with the intent of using these similar curves to predict the future performance of young players. This would be immensely valuable to Major League Baseball teams because it would help them assess the value of young, unproven players. Overall, we feel that our calculation of WAR and the construction of WAR career trajectory curves can be a useful tool to assess player performance.

2. Methods

To construct a fitted career WAR curve for Albert Pujols, we proceeded in the following manner:

1. We collected play-by-play Major League Baseball data from every game from 1934-2014 from Retrosheet.com.
2. For each season, we computed a Run Expectancy Matrix, which gives the expected runs scored in a half-inning for each different possible state of the game (e.g., base runners on first and second with two outs).
3. Based on how Albert Pujols performed, we calculated the number of runs he was expected to add to the team by subtracting the expected runs in each half-inning before his plate appearance from the expected runs to be scored in each half-inning after his plate appearance. Additionally, we performed the same process for his base-running. We then added all of his expected runs up for each season to find his total for each season.
4. For each season of Pujols' career, we calculated how many wins each of his runs were worth using Bill James' formula for calculating a team's number of wins

$$Wins = \frac{R^x}{(R-RA)^x}$$

where R is runs scored and RA is runs allowed, and finding x using cross-validation. We then found the difference between the number of expected wins with Pujols' runs included and without his runs included to get an estimate of Pujols' WAR. We divided Pujols' WAR by his runs created to get an estimate of how much a win was worth per run created.
5. We then divided his career into chunks of 100 Plate Appearances giving us more points on which to model his career WAR.
6. Finally, we fit a model to these points using smoothing splines.

All analyses were carried out using R.

3. Results

Here, we display three plots of our smoothing spline models. The first one to the right, is the model we fit for Albert Pujols' WAR per 100 Plate Appearances for the 2001-2011 seasons, corresponding to Albert Pujols' career with the St. Louis Cardinals before leaving via free agency. As we can see, Pujols appears to have gradually gotten better throughout his career, peaking somewhere between the 2009 and 2010 seasons. The fitted curve fits the data rather well, and gives evidence of a sharp decline in performance.

WAR Trajectory Curve for Albert Pujols

WAR Trajectory Curve for Albert Pujols
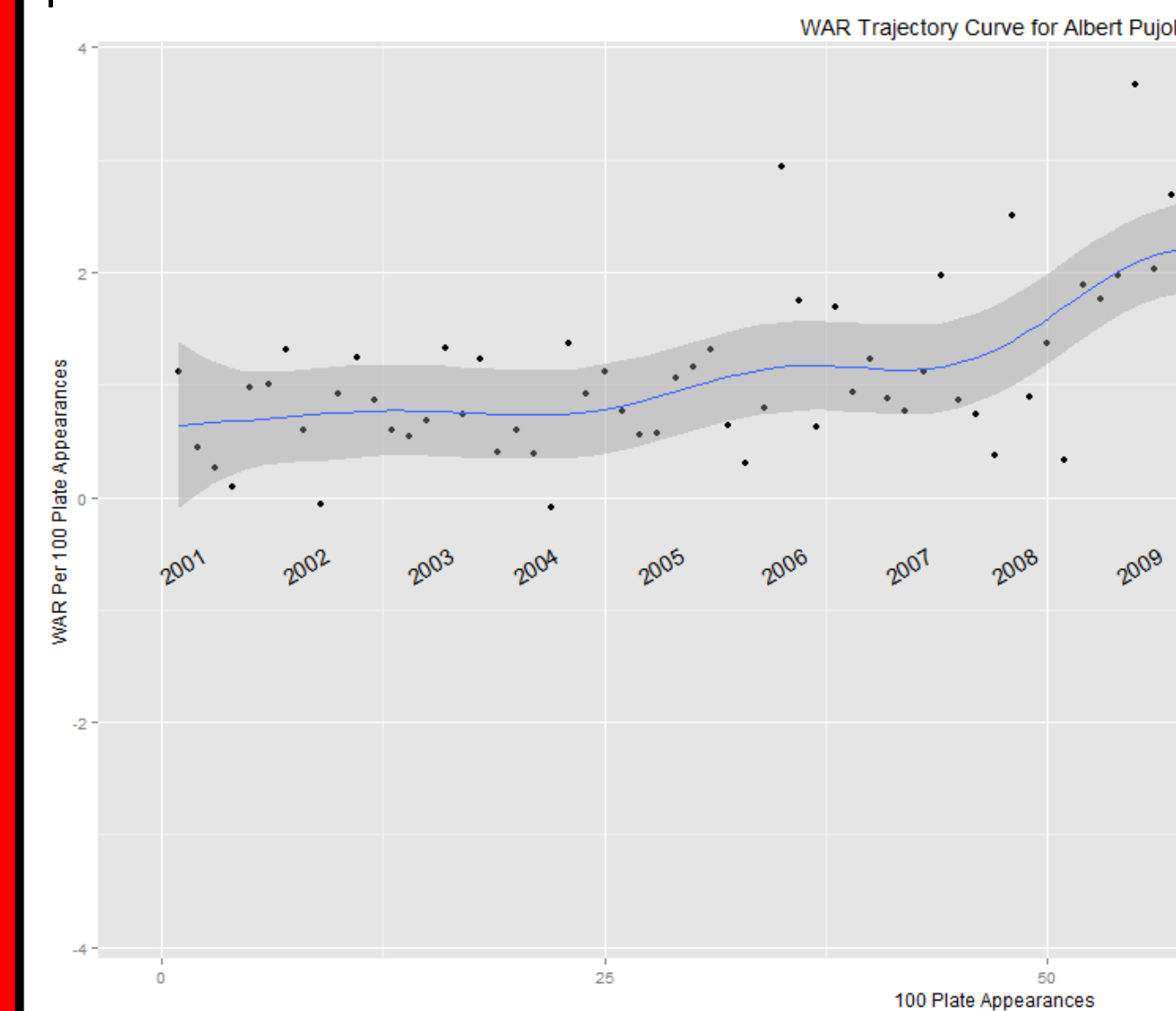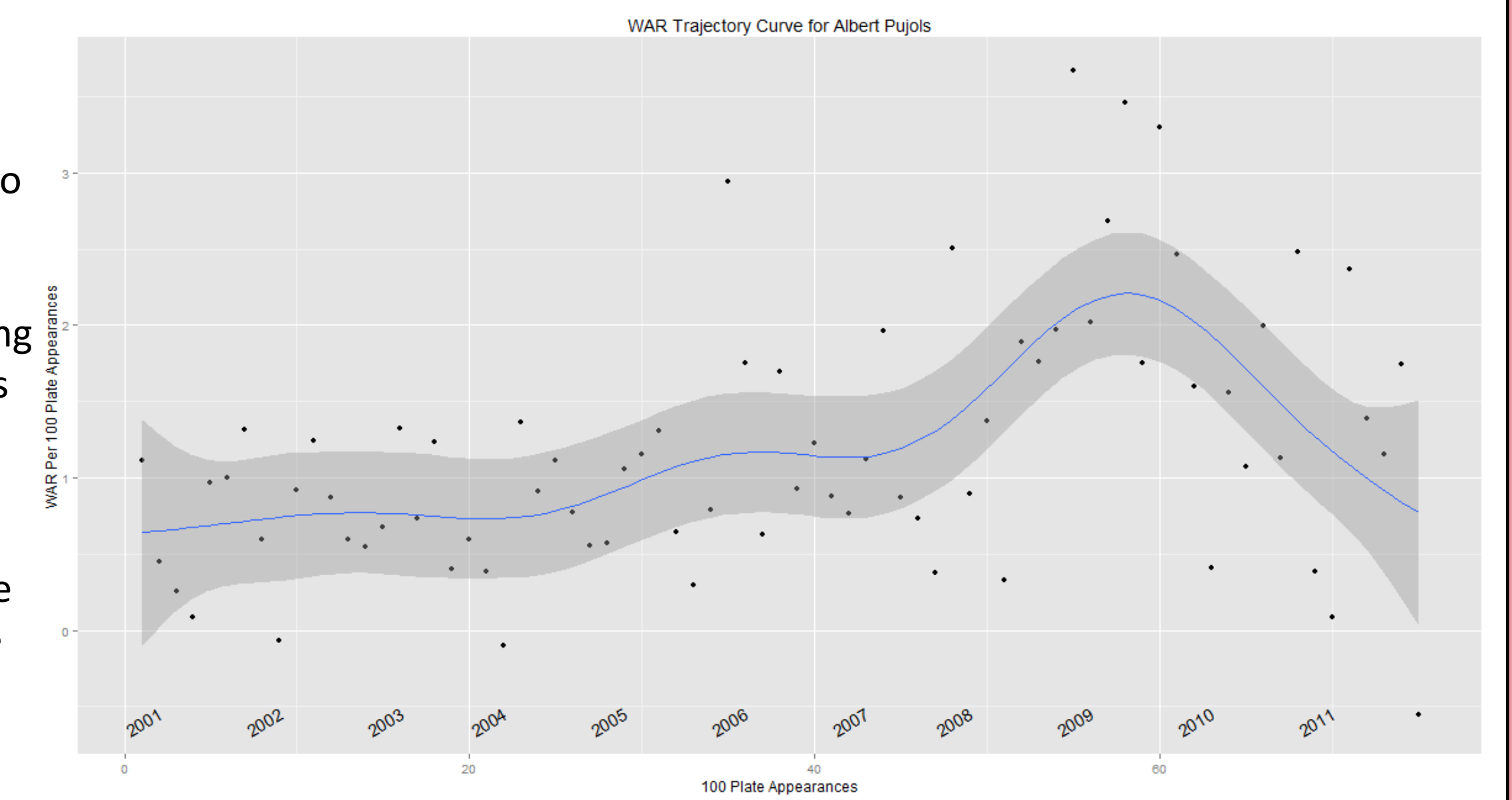
To the left, we have the same graph as above, but we have included a prediction for his performance for the next three seasons based on Pujols' performance from 2001-2011 (red line). Additionally, we have constructed a 95% prediction band around this prediction to account for random variation in our prediction. The model predicts that Pujols would continue his career decline based on WAR throughout the next three seasons. In 2011, Albert Pujols signed a ten year, $240M contract with the Los Angeles Angels of Anaheim.

Finally, to the right, we have all of the information contained on the previous two graphs, but this time we have included Albert Pujols' actual performance for seasons 2012-2014. We can see that the best fitting regression spline is within our 95% prediction band and is quite close to our prediction. From these three graphs, we can see that Pujols clearly peaked between 2009 and 2010 and that his performance is gradually declining. This follows the opinions of the baseball analyst community regarding his career performance.

WAR Trajectory Curve for Albert Pujols
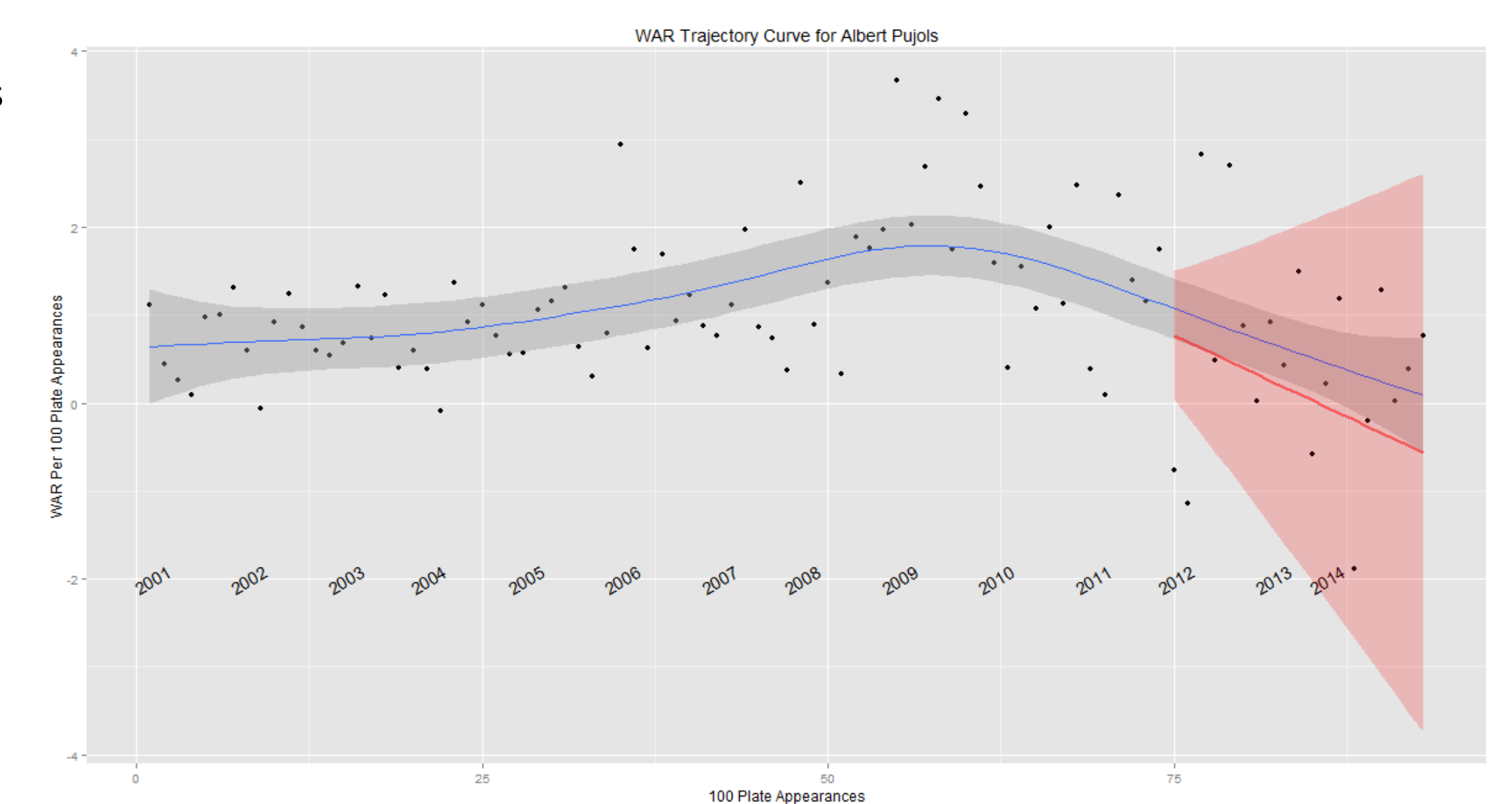
Works Cited

Friendly, M. (2014). Lahman: Sean Lahman's Baseball Database. R package version 3.0-1. http://CRAN.R-project.org/package=Lahman.

Lahman, S. (2015, April). The Lahman baseball database. Retrieved from http://www.seanlahman.com/baseball-archive/statistics/.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Retrosheet.org (2015, January). Retrosheet. Retrieved from http://www.retrosheet.org.

Acknowledgments:

Further Information:

If you would like further information you can contact me at my email: drjorda2@illinois.edu.