My goal for beating MARCEL was to attack it from essentially the same angle: keep it as simple as possible. However, in keeping with that spirit, I still wanted to use more advanced data and a slightly more advanced technique. The end result was building a simple multiple linear regression model using expected statistics pulled from Statcast.

Statcast has an expected Slugging (xSLG) and an expected On Base Percentage (xOBP) metric, so it follows that we can derive expected On Base Plus Slugging (xOPS). This xOPS became the basis of our model. By using expected statistics, which account for the quality of contact on batted balls, as well as plate discipline and swing decisions on a pitch by pitch basis, along with park factors, we remove a lot of the luck that is inherent in observed statistics, such as fielder positioning, the weather that day, or what the umpires strike zone is that game, giving us a more accurate representation of how we should expect the player to perform moving forward.

While we could have simply used the xOPS numbers in the existing MARCEL formula and achieved the desired outcome of beating the MARCEL projections (I did try this, it beat MARCEL), actually building out a linear regression model gives us two advantages over that process: it properly regresses the data back towards the mean, and it properly weights how much each previous year of data, as opposed to doing so arbitrarily like in the simple MARCEL projections. For example, while the first previous year remains about the same weight, the coefficient on the second past year (in our example for predicting 2021, this would be 2019) is weighted much more in my model as opposed to the MARCEL model, accounting for roughly 40% of the weight, as opposed to 33% in MARCEL. Correspondingly, this means that the third previous year is less weighted.

Also, the age factor works differently in my model versus MARCEL. In MARCEL, players improve until they turn 29 and then begin to fade after that. In my model, there is a continuous decline as players age, which actually is more in line with some of the newer aging curve studies, including ones from Neil Weinberg and Jeff Zimmerman at FanGraphs.com, which has players entering the league near peak wRC+ production, staying level until 27 or so, and then beginning their decline.

All in all, this is still a very simple projection model, which draws its inspiration from the simplicity of MARCEL and adds just a little extra bit of statistical theory to refine the results.