Bridging Different Eras in Sports: Comment

Author(s): Jim Albert

Source: *Journal of the American Statistical Association* , Sep., 1999, Vol. 94, No. 447 (Sep., 1999), pp. 677-680

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: https://www.jstor.org/stable/2669974

Comment

Jim ALBERT

1. INTRODUCTION

This article addresses a fascinating question: How can one compare the accomplishments of players in a given sport from different eras? In the sport of baseball, is it possible to compare Babe Ruth's accomplishment of 60 home runs in the 1927 baseball season with Mark McGwire's historic 70–home run season in 1998? The authors propose a Bayesian hierarchical model for combining players' sports performances over a period of years. This model incorporates effects to account for the changing abilities of players, season effects, and the effect of aging on performance. The authors claim that they are able to compare players from different eras due to the fact that the careers of the players within each sport overlap (the so-called bridging effect).

This article is notable for several aspects. First, the authors are analyzing rich datasets. They are analyzing baseball hitting data from the turn of the century, all of the individual round scores for all players in the major golf championships for the past 60 years, and 50 years of hockey scoring data. Second, the hierarchical models proposed seem to be appropriate for simultaneously estimating the season, aging, and ability distributions of interest. Finally, the results described in Sections 6–8 from the fitted model are very interesting. The authors are able to contrast the effects of aging on performance in the three sports, and interesting comments are made about the changing ability distributions in each sport.

I focus my comments on the home run hitting analysis because I am most familiar with these data. First, I relate the results of the article with research by Bill James, who has written quite a bit on the comparison of baseball players. I next make some comments on the authors' particular choice of aging function and propose a simpler functional form that may be useful in constructing a hierarchical model to compare a set of players. Finally, I focus on the comparison of the 52 greatest home run hitters and question whether the authors really have succeeded in bridging the home run accomplishments of Babe Ruth and Mark McGwire.

2. BILL JAMES'S EVALUATION OF PLAYERS AND THEORY OF AGING

There is a lively research among baseball statisticians on the comparison of players. This research is probably best represented by the writings of Bill James. Although James is not a professional statistician, he has devoted his life to exploring baseball data, and he may be the most respected expert among baseball fans on the proper analysis of baseball records. Much of his research was described in his yearly *Baseball Abstract* books that were published in the 1980s. (James 1988 contains a summary of his basic methods and a listing of the best players in baseball history.)

James has a standard method of evaluating the performance of a baseball hitter during a particular season. One starts with the basic hitting data for a player, which includes the number of at bats, hits, doubles, triples, home runs, and walks. From these data, James computes a statistic, called *runs created*, which measures how many runs the player creates by getting on base and by advancing runners to home who are already on base. One of James' basic findings is the so-called *Pythagorean formula*, which states that the ratio of wins to losses for a team is approximately equal to the square of the ratio of the runs scored to the runs allowed. Using this formula, he finds a player's *win-loss ratio* by comparing his runs created (per game) to the average runs scored (per game) for his league in his ballpark in the particular season:

$$\frac{\text{wins}}{\text{losses}} = \frac{\text{runs created per game}}{\text{league runs per game}}.$$

This win–loss ratio is used to measure the number of games that a player is responsible for winning and losing for a particular season.

To compare the performance of different players in baseball history, James notes that a player's performance plotted against age typically increases until a peak value and then decreases. Thus to compare different players, one must decide whether one is interested in a player's peak performance or a player's total career performance. When ranking hitters, James has different lists based on "peak" and "career" performance. A simple measure of a player's career offensive performance is found by summing the number of player wins and losses over all the seasons of his career.

James (1982) has done an extensive analysis of the effect of aging on batting and pitching performance. In one analysis, he looked at the total contribution of nonpitchers and pitchers of different ages. He found that both pitchers and nonpitchers attain their greatest aggregate value at ages 26 and 27. James also tracked the career performances of groups of nonpitchers of superstar and lower status. He concluded that the higher-quality batters generally reached their peak value at age 27, and the lower-quality batters peaked at age 26.

James does not specifically analyze batting average and home run average data, because he is interested in a player's offensive performance, which is better measured by his runs-created formula. However, his writings empha-

Jim Albert is Professor, Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403.

Journal of the American Statistical Association
September 1999, Vol. 94, No. 447, Applications and Case Studies

size the importance of looking at hitting data in the proper context—James' win–loss ratio statistic adjusts a player's hitting data by the particular season and the ballpark where he plays. Also, James stresses the importance of looking at career performance when comparing players. His conclusion that a batter's performance generally peaks at age 27 mirrors the result in the article that a player's batting average reaches a peak at age 27.

3. USING A QUADRATIC AGING FUNCTION TO COMPARE THE BEST HOME RUN HITTERS

I am not convinced that the nonparametric aging function proposed in Section 3.4 is the best way of modeling the aging process. Although the function g is unspecified, the model makes the implicit assumption that each player within a given sport peaks at the same age. Also, the maturing period and declining period (choice of a_{M} and a_{D}) are assumed to be the same across all players. It seems that these restrictions are made primarily for computational reasons. Also the estimated aging functions appear to be too rough for large ages (see Figs. 5 and 6).

A convenient way of modeling the aging effect is through a quadratic function. (In Albert 1992 I used a Poisson log-linear model with a quadratic age component to model home run rates.) Suppose that the probability of a home run of the ith player during the jth season, π_{ij}, is modeled by means of the logistic model

$$\operatorname{logit}(\pi_{ij}) = \beta_{0i} + \beta_{1i}a_{ij} + \beta_{2i}a_{ij}^2.$$

From a modeling perspective, it is useful to reparameterize $(\beta_{0i}, \beta_{1i}, \beta_{2i})$ to the parameters

$$\gamma_{0i} = \beta_{0i} - \frac{\beta_{1i}^2}{4\beta_{2i}}, \qquad \gamma_{1i} = -\frac{\beta_{1i}}{2\beta_{2i}}, \qquad \gamma_{2i} = \beta_{2i}.$$

For an aging function for the ith player, the parameter γ_{0i} represents the player's peak value, γ_{1i} represents the age at which the peak is achieved, and γ_{2i} measures the curvature of the function or the rate of growth and decline. Players can be compared by their peak values $\{\gamma_{0i}\}$, and the parameters $\{\gamma_{1i}\}$ and $\{\gamma_{2i}\}$ are informative about the players' aging processes.

If one is simultaneously estimating the aging curves for a number of ballplayers, then an exchangeable prior distribution is useful for combining the data from the individual careers. Specifically, suppose that one is interested in estimating the aging function of the home run rates for the 52 best home run hitters in baseball. Here "best" means the 52 ballplayers with the largest number of career home runs. The number of home runs for this group ranges from 350 (Ken Griffey, Jr.) to 755 (Hank Aaron). Because the players in this group are all "home run hitters," it is reasonable to expect them to have similar aging functions. To model the belief that the players have similar peak values, we let $\gamma_{0,1}, \ldots, \gamma_{0,52}$ be a random sample from a $\mathrm{N}(\mu_0, v_0^2)$ distribution. We also believe that the players will have similar peak ages, which is represented by letting $\gamma_{1,1}, \ldots, \gamma_{1,52}$ be a random sample from $\mathrm{N}(\mu_1, v_1^2)$. Also, $\gamma_{2,1}, \ldots, \gamma_{2,52}$

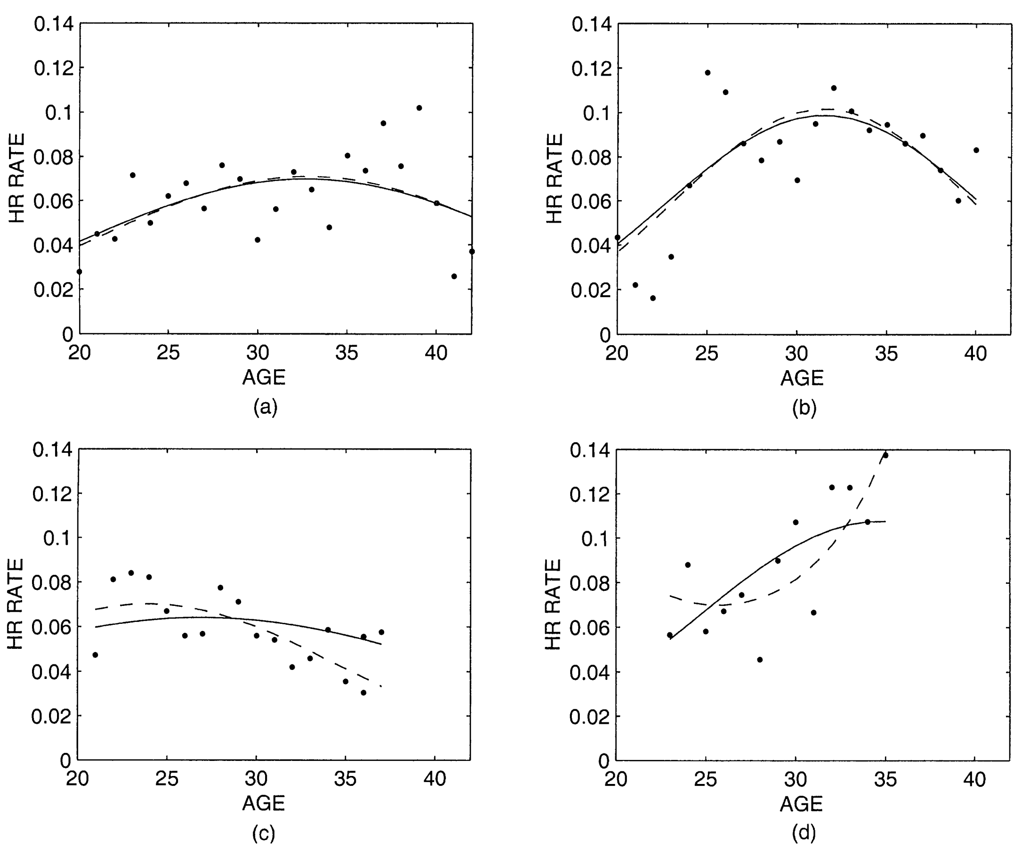

Figure 1. Observed Home Run Rates (Dots), Individual Logistic Estimates (Dashes), and Exchangeable Estimates (Solid Line) for Hank Aaron (a), Babe Ruth (b), Ernie Banks (c), and Mark McGwire (d).

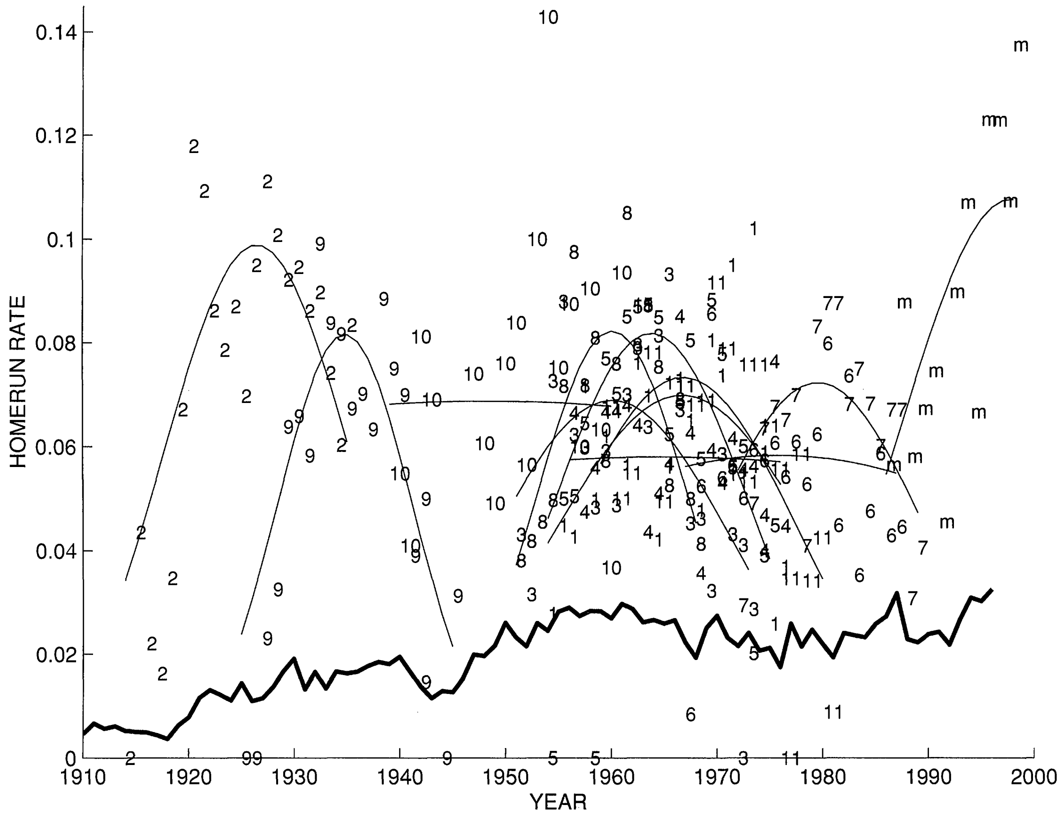

Figure 2. Observed Home Run Rates and Exchangeable Estimates (Solid Lines) for 12 Great Home Run Hitters. A plotted plot of 1 corresponds to Aaron, 2 to Ruth, 3 to Mays, 4 to Robinson, 5 to Killebrew, 6 to Jackson, 7 to Schmidt, 8 to Mantle, 9 to Foxx, 10 to Williams, 11 to McCovey and "m" to McGwire. The dark solid line corresponds to the season home run rate for all players.

are taken to be a random sample from $N(\mu_2, v_2^2)$, reflecting a belief that the 52 players have similar growth and decline rates. Vague priors are assigned to the hyperparameters $\{\mu_j, v_j^2\}$, reflecting little knowledge about an "average" aging function.

The posterior mean aging curves for four of the ballplayers are shown in Figure 1. The dashed line in each graph represents a logistic fit using only the data for the individual player, and the solid line in each graph represents the posterior mean aging curve using the exchangeable prior. For Hank Aaron and Babe Ruth, the observed patterns of home run rates are similar to a "typical" aging function, and the individual and exchangeable estimates are almost identical. The individual estimated aging curves are relatively poor for the other two players—the individual curve estimate predicts that Banks peaks at age 24 and that McGwire's home run rate is still increasing at age 35. For these players, the posterior mean aging curve seems to be a more reasonable fit that resembles the aging curves for the other players.

The quadratic aging curve used here may not be as flexible as the nonparametric form chosen in the article for modeling the hitting behavior for young or old players. However, this quadratic aging curve has a simpler interpretation than the aging function used in the article and can accommodate different peak ages for the ballplayers.

4. ADJUSTING FOR SEASON EFFECTS

The exchangeable model just described is not satisfactory, because it ignores the basic fact that these home run hitters played during different eras and the difficulty of hitting a home run has likely changed over the years. Figure 2 illustrates the basic problem of comparing 12 great home run hitters. This group includes the 11 hitters who have hit the most career home runs, and Mark McGwire who is the most famous hitter not included in this list. The graph plots the observed home run rates for these 12 players against the season in which the home runs were fit. The smooth curves represent the posterior aging curve estimates using the exchangeable model described earlier. The dark solid line at the bottom of the figure shows the actual rate of home run hitting for each season from 1910 to 1996.

We see from this figure that Ruth played in the 1920s, Foxx in the 1930s, seven of the hitters played during the 1960s, and McGwire is playing in the 1990s. Ruth and McGwire appear to stand out as the greatest home run hitters, because they both peak at a level between 10 and 12%. But home runs were much rarer during Ruth's day—the league home run rate in the mid-1920s was 1%, compared to 3% in 1998.

The article offers two possible explanations for the different home run rates between the 1920s and the 1990s. One explanation is that conditions were such that it was just harder to hit a home run in the 1920s. These conditions include the liveliness of the baseball, the quality of the pitchers, and the rules of the game. These conditions are modeled by the year effects $\{\delta_y\}$ in the article. A second explanation for the different rates is that the talent pool is substantially different in the two eras—there were few powerful hitters (besides Ruth) in the 1920s. These different talent pools are modeled using the decade specific parameters $\{\mu_\theta(d_i), \sigma_\theta(d_i)\}$.

The authors conclude from their analysis that the difficulty of hitting a home run has not changed significantly from the 1920s to the current day. But they conclude that the talent pool has dramatically changed, and there are more powerful hitters playing today than in the 1920s. Looking at the list of top 25 peak players (Table 11), we see that McGwire is rated significantly higher than Ruth. This is an interesting conclusion. Babe Ruth hit home runs at a much greater rate than his contemporaries, but he apparently was a shining light in a sea of mediocrity.

What seems to be missing from this article is a demonstration how the model is indeed bridging the careers of these players. From Figure 2, we see that there is a weak bridge connecting these great home run hitters who lived in an 80-year period. So it is difficult at best to compare Ruth and McGwire's accomplishments on this basis of this small dataset. Of course, the authors use the careers of all players to perform a bridge from the 1920s to the present day. However, it seems hard in my mind to separate the year effects from the changing talent pool of ballplayers—I suspect that they are highly correlated in the posterior analysis. Therefore, I am not convinced from this study that the dramatic increase in home run hitting from 1920 to the present day is primarily due to a change in the talent pool.

Although the results of this article are fascinating, it is important that one understand the limitations in the comparison of players. James (1988), in his chapter on the 100 greatest players in the last century, concludes that he has no idea whether baseball was tougher in the past or the present, and gives each period of baseball equal respect by choosing equal numbers of great players from each era for his greatest player list. Babe Ruth clearly was the dominant player during his era as McGwire appears to be in the modern era, but I am doubtful that any statistical model can really compare the two slugger's accomplishments.

ADDITIONAL REFERENCES

Albert, J. (1992), "A Bayesian Analysis of a Poisson Random Effects Model for Homerun Hitters," *The American Statistician*, 46, 246–253.

James, B. (1982), "The Bill James Baseball Abstract," New York: Ballantine Books.

James, B. (1988), *The Bill James Historical Baseball Abstract*, New York: Villard Books.

Comment

Joseph B. KADANE

This is a landmark article in applied statistics. It takes a subject of interest to many, who is the best—in hockey, golf and baseball—and applies a serious model to the data to get answers. The modeling technique addresses each of the important phenomena: age of player, year of play, comparative rapidity of increase of ability to optimal age or decrease after optimal age, as well as ability of the player. The results are very interesting to everyone at least mildly interested in sports.

Why should one care? The only connection that many people have with "statisticians" comes from sports, although the extent to which those who do it are statisticians in the ASA sense is open to question. More important, sports may be a route by which many of us come to understand that the averages really do apply to individual cases. Thinking back to my own youthful enthusiasm for the (Brooklyn) Dodgers, I wonder how much of my interest in statistics was begun long before I knew what a standard deviation was.

I do have a technical qualm with one aspect of the article. It seems to me that fixing parameters on the basis of "previous runs of the model," as was done with $a_{\mathrm{M}}, a_{\mathrm{D}}$, and $\breve{a}$, means that the results being reported are conditional on the chosen values for these parameters and hence do not incorporate uncertainty about them. Perhaps a reasonable joint prior on these parameters that reflects their necessary order ($a_{\mathrm{M}} < \breve{a} < a_{\mathrm{D}}$) would be more in the spirit of the rest of the article.

Also, I have some discomfort with grouping player prior abilities by decade. I suppose it is meant to incorporate the idea that players close to each other in birthyear might reasonably be supposed to have somewhat related abilities. I would have preferred an autoregressive model for this smoothing.

Why are Roger Maris and Sammy Sosa not in the list of top 25 home run hitters?

The next big step for this line of inquiry, I think, is team play. Some of the complexity of this issue is hinted at in the discussion of the fit of the model for hockey in Section 5.1. There are many kinds of contributions a player makes to a team, some of which are measured in the usual data collection and some of which are not. How can the combination of excellence of play and inspiration to others be measured? Perhaps if this next step is taken, we can see just how wonderful Jackie Robinson and Roy Campanella were in the Dodgers of the early 1950s.

Joseph B. Kadane is Leonard J. Savage Professor of Statistics and Social Sciences, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213.