Bayesian Analysis (2009) **4**, Number 4, pp. 669–674

Rejoinder

Shane T. Jensen*, Blakeley B. McShane† and Abraham J. Wyner‡

We thank each discussant for his insightful comments and suggestions for improvement. We are pleased by the positive reception of our current endeavor towards model-based prediction of hitting performance. It is our belief that academic statisticians can serve a leadership role in the transition of quantitative analysis of baseball from simple tabulations to sophisticated model-based approaches.

1 Alternative Models for Latent Variables

A clear theme of this discussion is the flexibility of the Bayesian hierarchical framework as a principled means for prediction in this application. Of course, the other side of that coin is that our model can always be improved by more sophisticated extensions. The discussants offer several great suggestions for improvements to our methodology. A first step in this effort is suggested by multiple discussants: extensions of the latent "elite" mixture model. These proposals are great directions for future research, and we briefly discuss the prospects of each below.

Albert & Birnbaum question our employment of a latent mixture model, citing the fact that these mixture components are not self-evident from the raw home-run rate distributions. However, they also note the presence of skewness and outliers. We argue that latent mixture models are a common strategy for addressing skewness and outliers. In fact, our original motivation for a latent mixture model was the observation that hitters with consistently high home run rates were over-shrunk in a model that did not allow for subpopulations of extreme home run performance.

Both Quintana & Müller and Glickman discuss the limitation of our mixture model to two latent states. In our original analysis, we experimented with the addition of a third latent state which was intended to capture players that showed inferior performance relative to their position. However, the estimated models that included this third state did not show any greater predictive power than the two-state model.

Quintana & Müller suggest a more comprehensive amelioration of our mixture model: allowing the number of latent states to be unknown and estimated. Certainly, this proposal is the most natural extension of the current approach and would help address the concerns raised by the discussants about the imposed "elite" vs. "non-elite" framework. The hurdle would be implementation of this more complicated model, as the reversible-jump approach proposed by Quintana & Muller could be complicated in practice.

*Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA, mailto:stjensen@wharton.upenn.edu

†Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA, mailto:mcshaneb@wharton.upenn.edu

‡Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA, mailto:ajw@wharton.upenn.edu

 DOI:10.1214/09-BA424REJ

Glickman proposes a model extension that is further afield. Instead of a discrete state space model, he proposes a latent state that evolves continuously in an autoregressive fashion. In our opinion, this continuous state space model would perform well for players with a long and consistent history of performance. However, we are skeptical there would be enough autocorrelated signal for younger players with very little personal history. For these cases with sparser information, we believe our simpler model is better able to pool information between players.

We have a similar concern about Albert & Birnbaum's proposal to fit random effects for each player. We concede that players (at the same position) can have very different trajectories, as illustrated by their comparison of Mickey Mantle and Hank Aaron. However, although there is enough information to model players with long careers in this way, we suspect that these random effects would be too variable for players who have only played a few seasons. For such players, the enforced shrinkage of our model is beneficial.

Furthermore, while the selection of Mantle and Aaron nicely illustrate the benefits of modeling trajectories individually, it also illustrates some of the pitfalls. Though Mantle and Aaron were both towering sluggers of their era, we contend that both players are unusually deviant from what is generally observed and their careers represent extreme points in the space of individual trajectories. Mantle suffered a precipitous decline due to debilitating injury while Aaron had an almost miraculously steady and lengthy career.

Thus, we are not sure it is a criticism to point out that we would have failed to predict Aaron's unusual performance into his forties or Mantle's steep early decline, unaided by health information. For the purposes of prediction, discounting unusual individual career trajectories and being guided mainly by position is a sound strategy, and we remind the reader that center fielders like Mantle are more likely to experience sharp declines in production than corner outfielders like Aaron. That said, the random effects framework is a great idea, and we are currently investigating extending our model to allow more flexible trajectories within each position.

There are of course many other generalizations and improvements not raised by discussants which we will consider in future work. Most promising is the extension of the usual first order Markov model to higher order or even variable order. This direction has the potential to more accurately model an individual player's trajectory.

2 Position and Other Potential Covariates

Beyond the latent mixture model, the discussants provide several suggestions for additional data and/or covariates that could further improve our predictions. Specifically, Albert & Birnbaum suggest the retrosheet database which provides more detailed within-season information for each player. We agree that the additional detail within the retrosheet database could improve our modeling efforts. One immediate advance, as proposed by Albert & Birnbaum, would be to divide each hitter's season into home

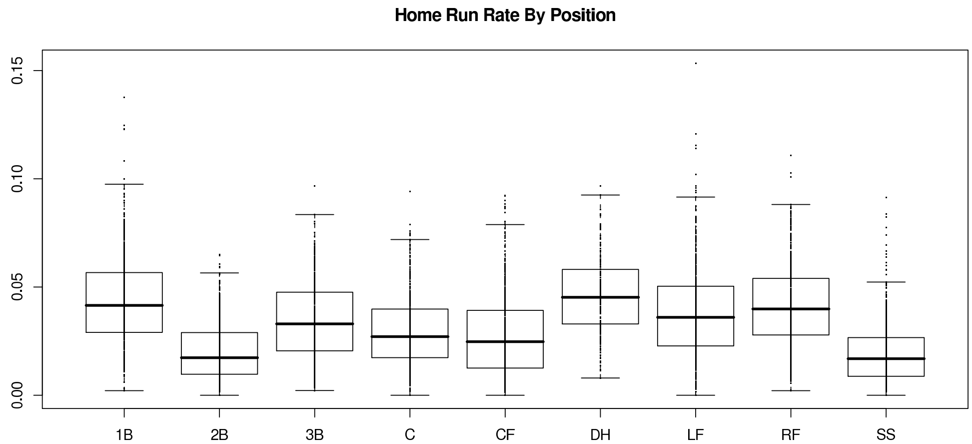

Figure 1: Boxplots of empirical home run rates by position. Each point gives HR/AB for a given player-season for all player-seasons with 300 or more at-bats from 1990-2006.

Table 1: Analysis of Variance Table

Source	DF	Sum Sq.	Mean Sq.	F Ratio	Prob > F
Position	8	0.31486	0.03936	120.6345	<2e-16
Year	16	0.05922	0.00370	11.3446	<2e-16
Age	25	0.00670	0.00027	0.8214	0.7178
Residuals	3801	1.24009	0.00033		

versus away games, thus enabling the estimation of true ballpark effects. We would favor estimation of ballpark effects in this way rather than the use of external park factors, which is also proposed by Albert & Birnbaum. In our experience, external park factors are highly inconsistent from year to year and do not seem to contain much signal except in some extreme cases (e.g., Coors Field or Citizens Bank Park).

Albert & Birnbaum question the use of position as a covariate in our model, claiming that it is not immediately evident what information is being added by position. They are correct to assert that there is heterogeneity of home run talent within each position, but there is large variation in home run rates across position as can be seen in Figure 1. In fact, we perform an analysis of the variance of home run rates by the nine positions, seventeen years, and twenty-six ages in our dataset in Table 1. Position accounts for 20% of the total variation in home run rates, far more than any other factor.

These results suggest that position is a very informative covariate for home run ability. In our view, position serves as a proxy variable for several player characteristics, such as body type and speed, that cannot be directly observed from the data. Scouts and managers incorporate many of these unobserved variables into their personnel decisions in terms of where to place players. By assigning a particular player to traditional power positions such as first base, managers are adding information about that player's propensity to hit home runs. We think this information is especially important for younger players who have less performance history upon which to base predictions.

Albert & Birnbaum also point out that our model does not address major shifts in hitting performance between different eras in baseball. We do not argue the point, as it was not the goal of our paper (though we note that Table 1 shows that the year factor accounts for a modest 3.6% of the variance in home run rates). Our priority is the prediction of future hitting performance, which motivated our focus on the current era. The comparison of hitting performance in different eras is also an interesting question, and has been addressed in the past with sophisticated Bayesian approaches (Berry et al. 1999).

We did investigate, somewhat indirectly, the possible effects of different eras on our predictions. We fit our full model on a larger dataset consisting of all seasons from 1970 to 2005, in addition to our presented analysis based on seasons from 1990 to 2005. We saw very little difference in the predictions between these two analyses, suggesting that any large-scale changes in hitting dynamics over the past forty years do not have a major impact on future hitting predictions.

Albert & Birnbaum also suggest using at-bats as a covariate for the modeling of home run rates. This is a good suggestion and we have investigated the modeling of at-bats as a means for improving the prediction of hitting totals. However, we need to correct one statement made by Albert & Birnbaum: we do not assume that each player's 2006 at-bats are the same as the at-bats in the previous season. Rather, we scale the predictions of hitting rates from our model (and the two external methods) by the actual 2006 at-bat totals in our comparisons.

3 Focus on Prediction

Glickman suggests that home run totals may not be the most interesting outcome to people in baseball. We certainly agree that home runs are not the best measure of overall hitting performance, and we emphasize that our methodology can be adapted to any other hitting event. Home runs were chosen for illustration since we believe that most readers have a good intuition about the scale and variation of home run totals. We also have experimented with a multinomial extension of our procedure that would model each hitting outcome (i.e., singles, doubles, etc.) simultaneously, and this remains an area of future research.

More generally, Albert & Birnbaum call for greater focus on model interpretation. Despite our emphasis on prediction, there are elements of our model that are interesting in their own right. The position-specific aging curves provide an interesting contrast in the aging process between players at these different positions. Our "elite" versus "non-elite" indicators also provide a means for separating out consistently over-performing players relative to their position.

Quintana & Müller also inquire about the predictive power of our model for *all* players, not just the subset of players examined in our analysis. Our primary motivation was to have a set of common players for comparison with the external methods. However, we concede that the players excluded from our analysis probably represent an even tougher challenge for prediction. Albert & Birnbaum also suggest that extra insight would be gained from a case-by-case exploration and comparison of our predictions. To this end, we have made available the entire set of our predictions for the 2006 season at the following website: `http://stat.wharton.upenn.edu/~stjensen/research/predictions.2006.xlsx`

References

Berry, S. M., Reese, S., and Larkey, P. D. (1999). "Bridging Different Eras in Sports." *Journal of the American Statistical Association*, 94: 661–686. 672