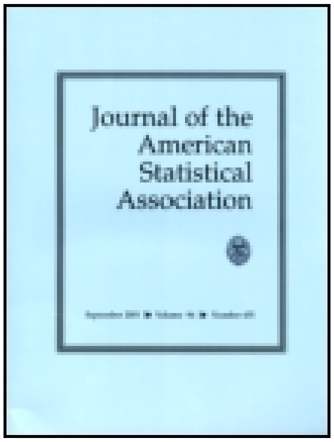

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Exploring Baseball Hitting Data: What about those Breakdown Statistics?

Jim Albert [a]

[a] Department of Mathematics and Statistics , Bowling Green State University , Bowling Green , OH , 43403

Published online: 27 Feb 2012.

Exploring Baseball Hitting Data: What About Those Breakdown Statistics?

Jim ALBERT*

During a broadcast of a baseball game, a fan hears how baseball hitters perform in various situations, such as at home and on the road, on grass and on turf, in clutch situations, and ahead and behind in the count. From this discussion by the media, fans get the misleading impression that much of the variability in players' hitting performance can be explained by one or more of these situational variables. For example, an announcer may state that a particular player struck out because he was behind in the count and was facing a left-handed pitcher. In baseball one can now investigate the effect of various situations, as hitting data is recorded in very fine detail. This article looks at the hitting performance of major league regulars during the 1992 baseball season to see which situational variables are "real" in the sense that they explain a significant amount of the variation in hitting of the group of players. Bayesian hierarchical models are used in measuring the size of a particular situational effect and in identifying players whose hitting performance is very different in a particular situation. Important situational variables are identified together with outstanding players who make the most of a given situation.

KEY WORDS: Hierarchical modeling; Outliers; Situational variables.

1. INTRODUCTION

After the end of every baseball season, books are published that give detailed statistical summaries of the batting and pitching performances of all major league players. In this article we analyze baseball hitting data that was recently published in Cramer and Dewan (1992). This book claims to be the "most detailed statistical account of every major league player ever published," which enables a fan to "determine the strengths and weaknesses of every player."

For hitters, this book breaks down the usual set of batting statistics (e.g., hits, runs, home runs, doubles) by numerous different situations. Here we restrict discussion to the fundamental hitting statistics—hits, official at-bats, and batting average (hits divided by at-bats)—and look at the variation of this data across situations. To understand the data that will be analyzed, consider the breakdowns for the 1992 season of Wade Boggs presented in Table 1. This table shows how Boggs performed against left- and right-handed pitchers and pitchers that induce mainly groundballs and flyballs. In addition, the table gives hitting statistics for day and night games, games played at and away from the batter's home ballpark, and games played on grass and artificial turf. The table also breaks down hits and at-bats by the pitch count ("ahead on count" includes 1-0, 2-0, 3-0, 2-1, and 3-1) and the game situation ("scoring position" is having at least one runner at either second or third, and "none on / out" is when there are no outs and the bases are empty). Finally, the table gives statistics for the batting position of the hitter and different time periods of the season.

What does a fan see from this particular set of statistical breakdowns? First, several situational variables do not seem very important. For example, Boggs appears to hit the same for day and night games and before and after the All-Star game. But other situations do appear to matter. For example, Boggs hit .243 in home games and .274 in away games—a 31-point difference. He appears to be more effective against flyball pitchers compared to groundball pitchers, as the difference in batting averages is 57 points. The most dramatic situation appears to be pitch count. He hit .379 on the first pitch, .290 when he was ahead in the count, but only .197 when he had two strikes on him.

What does a fan conclude from this glance at Boggs's batting statistics? First, it is difficult to gauge the significance of these observed situational differences. It seems that Boggs bats equally well during day or night games. It also appears that there are differences in Boggs's "true" batting behavior during different pitch counts (the 93-point difference between the "ahead in count" and "two strikes" averages described earlier). But consider the situation "home versus away." Because Boggs bats 31 points higher in away games than in home games, does this mean that he is a better hitter away from Fenway Park? Many baseball fans would answer "yes." Generally, people overstate the significance of seasonal breakdown differences. Observed differences in batting averages such as these are often mistakenly interpreted as real differences in true batting behavior. Why do people make these errors? Simply, they do not understand the general variation inherent in coin tossing experiments. There is much more variation in binomial outcomes than many people realize, and so it is easy to confuse this common form of random variation with the variation due to real situational differences in batting behavior.

How can fans gauge the significance of differences of situational batting averages? A simple way is to look at the 5-year hitting performance of a player for the same situations. If a particular observed seasonal situational effect is real for Boggs, then one might expect him to display a similar situational effect during recent years. Cramer and Dewan (1992) also gave the last 5 years' (including 1992) hitting performance of each major league player for all of the same situations of Table 1. Using these data, Table 2 gives situational differences in batting averages for Boggs for 1992 and the previous 4-year period (1988–1991).

Table 2 illustrates the volatility of the situational differences observed in the 1992 data. For example, in 1992 Boggs

* Jim Albert is Professor, Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403. The author is grateful to Bill James, the associate editor, and two referees for helpful comments and suggestions.

Table 1. Situational 1992 Batting Record of Wade Boggs

	AVG	AB	H
1992 season	.259	514	133
versus left	.272	158	43
versus right	.253	356	90
groundball	.235	136	32
flyball	.292	144	42
home	.243	251	61
away	.274	263	72
day	.259	193	50
night	.259	321	83
grass	.254	437	111
turf	.286	77	22
1st pitch	.379	29	11
ahead in count	.290	169	49
behind in count	.242	157	38
two strikes	.197	213	42
scoring position	.311	106	33
close and late	.322	90	29
none on/out	.254	142	36
batting #1	.222	221	49
batting #3	.287	289	83
other	.250	4	1
April	.253	75	19
May	.291	86	25
June	.242	95	23
July	.304	79	24
Aug	.198	96	19
Sept/Oct	.277	83	23
Pre-All-Star	.263	278	73
Post-All-Star	.254	236	60

NOTE: Data from Cramer and Dewan (1992).

hit flyball pitchers 57 points better than groundball pitchers. But in the 4-year period preceding 1992, he hit groundball pitchers 29 points higher than flyball pitchers. The 31-point 1992 home/away effect (favoring away) appears spurious, because Boggs was much better at Fenway Park for the previous 4-year period. Looking at the eight situations, only the night/day and pre- post-All-Star situational effects appear to be constant over time.

From this simple analysis, one concludes that it is difficult to interpret the significance of seasonal batting averages for a single player. In particular, it appears to be difficult to conclude that a player is a clutch or "two-strike" hitter based solely on batting statistics for a single season. But a large number of major league players bat in a particular season, and it may be easier to detect any significant situational patterns by pooling the hitting data from all of these players. The intent of this article is too look at situational effects in batting averages over the entire group of hitters for the 1992 season.

Here we look at the group of 154 regular major league players during the 1992 season. We define "regular" as a player who had at least 390 official at-bats; the number 400 was initially chosen as a cutoff, but it was lowered to 390 to accommodate Rob Deer, a hitter with unusual talents (power with a lot of strikeouts).

Using data from Cramer and Dewan (1992) for the 154 regulars, we investigate the effects of the following eight situations (with the abbreviation for the situation that we use given in parentheses):

- opposite side versus same side (OPP–SAME)
- groundball pitcher versus flyball pitcher (GBALL–FBALL)
- home versus away
- day versus night
- grass versus turf
- ahead in count versus two strikes in count (AHEAD–2 STRIKE)
- scoring position versus none on/out (SCORING–NONE ON/OUT)
- pre-All-Star game versus post-All-Star game (PRE/AS–POST/AS).

A few remarks should be made about this choice of situations. First, it is well known that batters hit better against pitchers who throw from the opposite side from the batter's hitting side. For this reason, I look at the situation "opposite side versus same side"; batters who switch-hit will be excluded from this comparison. Next, for ease of comparison of different situations, it seemed helpful to create two nonoverlapping categories for each situation. All of the situations of Table 1 are of this type except for pitch count, clutch situations, and time. Note that the pitch categories are overlapping (one can be behind in the count and have two strikes) and so it seemed simpler to just consider the nonoverlapping cases "ahead in count" and "two strikes." Likewise, one can simultaneously have runners in scoring position and the game be close and late, so I considered only the "scoring position" and "none on/out" categories. The batting data across months is interesting; however, because the primary interest is in comparing the time effect to other situational variables, I used only the pre- and post-All-Star game data.

When we look at these data across the group of 1992 regulars, there are two basic questions that we will try to answer.

Table 2. Situational Differences in Batting Averages (One Unit = .001) for Wade Boggs for 1992 and 1988–1991

Year	Right–left	Flyball–groundball	Home–away	Night–day
1992	−19	57	−31	0
1988–1991	61	−29	85	6
Year	**Turf–grass**	**Ahead–2 strikes**	**None out–scoring position**	**Pre-All Star–Post-All Star**
1992	32	93	−57	9
1988–1991	−30	137	1	−8

First, for a particular situation it is of interest to look for a general pattern across all players. Baseball people believe that most hitters perform better in various situations. In particular, managers often make decisions under the following assumptions:

- Batters hit better against pitchers throwing from the opposite side.
- Batters hit better at home.
- Batters hit better during day games (because it is harder to see the ball at night).
- Batters hit better when they are ahead in the count (instead of being behind two strikes).
- Batters hit better on artificial turf than on grass (because groundballs hit on turf move faster and have a better chance of reaching the outfield for a hit).

The other three situations in the list of eight above are not believed to be generally significant. One objective here is to measure and compare the general sizes of these situational effects across all players.

Once we understand the general situational effects, we can then focus on individuals. Although most players may display, say, a positive home effect, it is of interest to detect players who perform especially well or poorly at home. It is easy to recognize great hitters such as Wade Boggs partly because his success is measured by a well-known statistic, batting average. The baseball world is less familiar with players who bat especially well or poorly in given situations and the statistics that can be used to measure this unusual performance. So a second objective in this article is to detect these unusual situational players. These outliers are often the most interesting aspect of baseball statistics. Cramer and Dewan (1992), like many others, list the leading hitters with respect to many statistical criteria in the back of their book.

This article is outlined as follows. Section 2 sets up the basic statistical model and defines parameters that correspond to players' situational effects. The estimates of these situational effects from a single season can be unsatisfactory and often can be improved by adjusting or shrinking them towards a common value. This observation motivates the consideration of a prior distribution that reflects a belief in similarity of the set of true situational effects. Section 3 summarizes the results of fitting this Bayesian model to the group of 1992 regulars for each one of the eight situational variables. The focus, as explained earlier, is on looking for general situational patterns and then finding players that deviate significantly from the general patterns. Section 4 summarizes the analysis and contrasts it with the material presented by Cramer and Dewan (1992).

2. THE MODEL

2.1 Basic Notation

Consider one of the eight situational variables, say the home/away breakdown. From Cramer and Dewan (1992), we obtain hitting data for $N = 154$ players; for each player, we record the number of hits and official at-bats during home and away games. For the ith player, this data can be represented by a 2×2 contingency table,

	HITS	OUTS	AT-BATS
home	h_{i1}	o_{i1}	ab_{i1}
away	h_{i2}	o_{i2}	ab_{i2} ,

where h_{i1} denotes the number of hits, o_{i1} the number of outs, and ab_{i1} the number of at-bats during home games. (The variables h_{i2}, o_{i2}, and ab_{i2} are defined similarly for away games.) Let p_{i1} and p_{i2} denote the true probabilities that the ith hitter gets a hit home and away. If we assume that the batting attempts are independent Bernoulli trials with the aforementioned probabilities of success, then the number of hits h_{i1} and h_{i2} are independently distributed according to binomial distributions with parameters (ab_{i1}, p_{i1}) and (ab_{i2}, p_{i2}).

For ease of modeling and exposition, it will be convenient to transform these data to approximate normality using the well-known logit transformation. Define the observed logits

$$y_{ij} = \log\left(\frac{h_{ij}}{o_{ij}}\right), \qquad j = 1, 2.$$

Then, approximately, y_{i1} and y_{i2} are independent normal, where y_{ij} has mean $\mu_{ij} = \log(p_{ij}/(1 - p_{ij}))$ and variance $\sigma^2_{ij} = (ab_{ij}p_{ij}(1 - p_{ij}))^{-1}$. Because the sample sizes are large, we can accurately approximate σ^2_{ij} by an estimate where p_{ij} is replaced by the observed batting average h_{ij}/ab_{ij}. With this substitution, $\sigma^2_{ij} \approx 1/h_{ij} + 1/o_{ij}$.

Using the foregoing logistic transformation, we represent the complete data set for a particular situation, say home/away, as a $2 \times N$ table:

	Player 1	2	...	N
home	y_{11}	y_{21}	...	y_{N1}
away	y_{12}	y_{22}	...	y_{N2} .

The observation in the (i, j) cell, y_{ij}, is the logit of the observed batting average of the ith player during the jth situation. We model μ_{ij}, the mean of y_{ij}, as

$$\mu_{ij} = E(y_{ij}) = \mu_i + \alpha_{ij},$$

where μ_i measures the hitting ability of player i and α_{ij} is a situational effect that measures the change in this player's hitting ability due to the jth situation. The model as stated is overparameterized, so we express the situational effects as $\alpha_{i1} = \alpha_i$ and $\alpha_{i2} = -\alpha_i$. With this change, the parameter α_i represents the change in the hitting ability of the ith player due to the first situational category.

2.2 Shrinking Toward the Mean

For a given situational variable, it is of interest to estimate the player situational effects $\alpha_1, \ldots, \alpha_N$. These parameters represent the "true" situational effects of the players if they were able to play an infinite number of games.

Is it possible to estimate accurately a particular player's situational effect based on his hitting data from one season?

To answer this question, suppose that a player has a true batting average of .300 at home and .200 away, a 100-point differential. If he has 500 at-bats during a season, half home and half away, then one can compute that his observed seasonal home batting average will be between .252 and .348 with probability .9 and his away batting average will be between .158 and .242 with the same probability. So although the player's true differential is 100 points, his seasonal batting average differential can be between 10 and 190 points. Because this is a relatively wide interval, one concludes that it is difficult to estimate a player's situational effect using only seasonal data.

How can we combine the data from all players to obtain better estimates? In the situation where one is simultaneously estimating numerous parameters of similar size, it is well known in the statistics literature that improved estimates can be obtained by shrinking the parameters toward a common value. Efron and Morris (1975) illustrated the benefit of one form of these shrinkage estimators in the prediction of final season batting averages in 1970 for 18 ballplayers based on the first 45 at-bats. (See also Steffey 1992, for a discussion of the use of shrinkage estimates in estimating a number of batting averages.) Morris (1983) used a similar model to estimate the true batting average of Ty Cobb from his batting statistics across seasons.

In his analysis of team breakdown statistics, James (1986) discussed the related problem of detecting significant effects. He observed that many team distinctions for a season (e.g., how a team plays against right- and left-handed pitching) will disappear when studied over a number of seasons. A similar pattern is likely to hold for players' breakdowns. Some players will display unusually high or low situational statistics for one season, suggesting extreme values of the parameters α_i. But if these situational data are studied over a number of seasons, the players will generally have less extreme estimates of α_i. This "regression to the mean" phenomena is displayed for the "ahead in count–two strikes" situation in Figure 1. This figure plots the 1992 difference in batting averages (batting average ahead in count–batting average 2 strikes) against the 1988–1991 batting average difference for all of the players. Note that there is an upward tilt in the graph, indicating that players who have larger batting average differences in 1992 generally had larger differences in 1989–1991. But also note that there is less variability in the 4-year numbers (standard deviation .044 versus .055 for the 1992 data). One way of understanding this difference is by the least squares line placed on top of the graph. The equation of this line is $y = .126 + (1 - .683)(x - .122)$, with the interpretation that the 4-year batting average difference generally adjusts the 1992 batting average difference 68% toward the average difference (.12) across all players.

Suppose that the underlying batting abilities of the players do not change significantly over the 5-year period. Then the 4-year batting average differences (based on a greater number of at-bats) are better estimates than the 1-year differences of the situational abilities of the players. In that case, it is clear from Figure 1 that the observed seasonal estimates of the α_i should be shrunk toward some overall value to obtain more accurate estimates. In the next section we describe a prior distribution on the situation parameters that reflects a belief in similarity of the effects and results in sample estimates that will shrink the season values toward a pooled value.

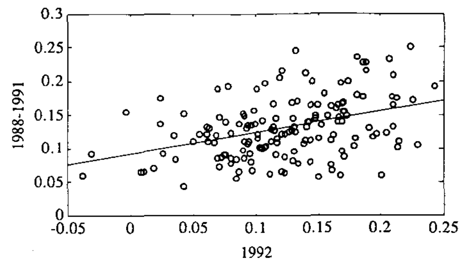

Figure 1. Scatterplot of 1992 Pitch Count Difference in Batting Averages (AHEAD-2 STRIKES) Against Previous 4-Year Difference for All 1992 Players.

2.3 The Prior Distribution

The model discussed in Section 2.1 contains $2N$ parameters, the hitting abilities $\mu_1, \ldots, \mu_N$, and the situational effects $\alpha_1, \ldots, \alpha_N$. Because the ability parameters in this setting are nuisance parameters, we assume for convenience that the μ_i are independently assigned flat noninformative priors.

Because the focus is on the estimation of the situational effects $\alpha_1, \ldots, \alpha_N$, we wish to assign a prior that reflects subjective beliefs about the locations of these parameters. Recall from our earlier discussion that it seems desirable for the parameter estimates to shrink the individual estimates toward some common value. This shrinkage can be accomplished by assuming a priori that the effects $\alpha_1, \ldots, \alpha_N$ are independently distributed from a common population $\pi(\alpha)$. This prior reflects the belief that the N effects are similar in size and come from one population of effects $\pi(\alpha)$.

Because the α_i are real-valued parameters, one reasonable form for this population model is a normal distribution with mean μ_α and variance σ_α^2. A slightly preferable form used in this article is a t distribution with mean μ_α, scale σ_α, and known degrees of freedom ν (here we use the relatively small value, $\nu = 4$). The parameters of this distribution are used to explain the general size of the situational effect and to identify particular players who have unusually high or low situational effects. The parameters μ_α and σ_α describe the location and spread of the distribution of effects across all players. To see how this model can identify outliers, note that a $t(\mu_\alpha, \sigma_\alpha, \nu)$ distribution can be represented as the mixture $\alpha_i \mid \lambda_i$ distributed $N(\mu_\alpha, \sigma_\alpha^2/\lambda_i)$, λ_i distributed gamma$(\nu/2, \nu/2)$. The new scale parameters $\lambda_1, \ldots, \lambda_N$ will be seen to be useful in picking out individuals who have situational effects set apart from the main group of players.

To complete our prior specification, we need to discuss what beliefs exist about the parameters μ_α and σ_α^2 that describe the t population of situational effects. First, we assume that we have little knowledge about the general size of the situational effect; to reflect this lack of knowledge, the mean μ_α is assigned a flat noninformative prior. We must be more careful about the assignment of the prior distribution on σ_α^2, because this parameter controls the size of the shrinkage

Table 3. Posterior Means of Parameters of Population Distribution of the Situation Effects and Summary Statistics of the Posterior Means of the Batting Average Differences $p_{i1} - p_{i2}$ Across All Players

Situation	$E(\mu_\alpha)$	$E(\sigma_\alpha)$	Summary statistics of $[E(p_{i1} - p_{i2})]$ (one unit = .001) Q_1	M	Q_3	Q_3-Q_1
GRASS–TURF	−.002	.107	−17	−3	15	32
SCORING–NONE ON/OUT	.000	.108	−13	0	17	30
DAY–NIGHT	.004	.105	−13	2	16	29
PRE/AS–POST/AS	.007	.101	−9	3	17	26
HOME–AWAY	.016	.103	−8	8	21	29
GBALL–FBALL	.023	.109	−7	11	24	31
OPP–SAME	.048	.106	5	20	32	27
AHEAD–2 STRIKES	.320	.110	104	123	142	38

of the individual player situational estimates toward the common value. In empirical work, it appears that the use of the standard noninformative prior for σ_α^2, $1/\sigma_\alpha^2$, can lead to too much shrinkage. So to construct an informative prior for σ_α^2, we take the home/away variable as one representative situational variable among the eight and base the prior of σ_α^2 on a posterior analysis of these parameters based on home/away data from an earlier season. So we first assume that μ_α, σ_α^2 are independent, with μ_α assigned a flat prior and σ_α^2 distributed according to an inverse gamma distribution with parameters $a = 1/2$ and $b = 1/2$ (vague prior specification), and then fit this model to 1991 home/away data for all of the major league regulars. From this preliminary analysis, we obtain posterior estimates for σ_α^2 that are matched to an inverse gamma distribution with parameters $a = 53.2$ and $b = .810$.

To summarize, the prior can be written as follows:

- $\mu_1, \ldots, \mu_N$ independent with $\pi(\mu_i) = 1$.
  $\alpha_1, \ldots, \alpha_N$ independent $t(\mu_\alpha, \sigma_\alpha, \nu)$, where $\nu = 4$.
- μ_α, σ_α independent with
  $\pi(\mu_\alpha) = 1$ and $\pi(\sigma_\alpha^2) = K(1/(\sigma_\alpha^2)^{a+1})\exp(-b/\sigma_\alpha^2)$, with $a = 53.2$ and $b = .810$.

This prior is used in the posterior analysis for each of the situational variables.

3. FITTING THE MODEL

3.1 General Behavior

The description of the joint posterior distribution and the use of the Gibbs sampler in summarizing the posterior are outlined in the Appendix. For each of the eight situational variables, a simulated sample of size 1,000 from the joint posterior distribution of $(\{\mu_i\}, \{\alpha_i\}, \mu_\alpha, \sigma_\alpha^2\}$ was obtained. These simulated values can be used to estimate any function of the parameters of interest. For example, if one is interested in the difference in breakdown probabilities for player i, $p_{i1} - p_{i2}$, then one can simulate values of this difference by noting that $p_{i1} - p_{i2} = \exp(\mu_i + \alpha_i)/(1 + \exp(\mu_i + \alpha_i)) - \exp(\mu_i - \alpha_i)/(1 + \exp(\mu_i - \alpha_i))$, and then performing this transformation on the simulated values of $\{\mu_i\}$ and $\{\alpha_i\}$.

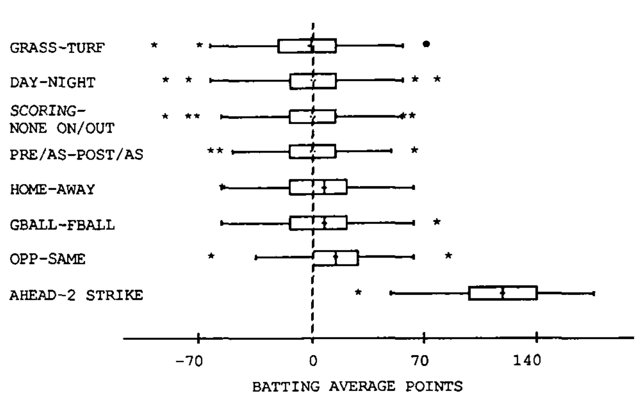

Figure 2. Boxplots of the Posterior Means of the Differences in Batting Averages for the Eight Situational Variables.

Our first question concerns the existence of a general situational effect. Do certain variables generally seem more important than others in explaining the variation in hitting performance? One can answer this question by inspection of the posterior distribution of the parameters μ_α and σ_α, which describe the population distribution of the situational effects $\{\alpha_i\}$. The posterior means of these hyperparameters for each of the eight situations are given in Table 3. Another way to look at the general patterns is to consider the posterior means of the true batting average differences $p_{i1} - p_{i2}$ across all players. Table 3 gives the median, quartiles, and interquartile range for this set of 154 posterior means; Figure 2 plots parallel boxplots of these differences for the eight situations.

Note from Table 3 and Figure 2 that there are significant differences between the average situational effects. The posterior mean of μ_α is a measure of the general size of the situational effect on a logit scale. Note from Table 3 that the "ahead in count–2 strikes" effect stands out. The posterior mean of μ_α is .32 on the logit scale; the corresponding median of the posterior means of the batting average differences is 123 points. Batters generally hit much better when ahead versus behind in the pitch count. Compared to this effect, the other seven effects are relatively insignificant. Closer examination reveals that "opposite–same arm" is the next most important variable, with a median batting average difference of 20 points. The "home–away" and "groundball–flyball" effects follow in importance, with median batting average differences of 8 and 11 batting average points. The remaining four situations appear generally insignificant, as the posterior mean of μ_α is close to 0 in each case.

The posterior means of σ_α give some indication of the relative spreads of the population of effects for the eight situations. The general impression from Table 3 and Figure 2 is that all of the situational populations have roughly the same spread, with the possible exception of the pitch count situation. So the differences between the sets of two different situational effects can be described by a simple shift. For example, the "groundball–flyball" effects are approximately 10 batting average points higher than the "day–night" effects.

3.2 Individual Effects—What is an Outlier?

In the preceding section we made some observations regarding the general shape of the population of situation effects $\{\alpha_i\}$. Here we want to focus on the situation effects for individual players. Figure 3 gives stem and leaf diagrams (corresponding to the Fig. 2 boxplots) for these differences in batting averages for each of the eight situations. These plots confirm the general comments made in Section 3.1 about the comparative locations and spreads of the situational effect distributions.

Recall from Section 2 that it is desirable to shrink the batting average differences that we observe for a single season toward a common value. Figure 4 plots the posterior means of the differences $p_{i1} - p_{i2}$ against the season batting average differences for the effect "ahead in count–2 strikes." The line $y = x$ is plotted on top of the graph for comparison purposes. This illustrates that these posterior estimates shrink the seasonal estimates approximately 50% toward the average effect size of 122 points. Note that some of the seasonal batting average differences are negative; these players actually hit better in 1992 with a pitch count of 2 strikes. But the posterior means shrink these values toward positive values. Thus we have little faith that these negative differences actually relate to true negative situational effects.

We next consider the issue of outliers. For instance, for the particular effect "home–away," are there players that bat particularly well or poor at home relative to away games? Returning back to the displays of the estimates of the batting average differences in Figure 4, are there any players whose estimates deviate significantly from the main group? We observe some particular estimates, say the −95 in the "grass–turf" variable, that are set apart from the main distribution of estimates. Are these values outliers? Are they particularly noteworthy?

We answer this question by first looking at some famous outliers in baseball history. With respect to batting average, a number of Hall of Fame or future Hall of Fame players have achieved a high batting average during a single season. In particular, we consider Ted Williams .406 batting average in 1941, Rod Carew's .388 average in 1977, George Brett's .390 average in 1980 and Wade Boggs's .361 average in 1983. Each of these batting averages can be considered an outlier, because these accomplishments received much media attention and these averages were much higher than the batting averages of other players during that particular season.

These unusually high batting averages were used to calibrate our Bayesian model. For each of the four data sets—Major League batting averages in 1941 and American League batting averages in 1977, 1980, and 1983—an exchangeable model was fit similar to that described in Section 2. In each model fit we computed the posterior mean of the scale parameter λ_i corresponding to the star's high batting average. A value $\lambda_i = 1$ corresponds to an observation consistent with the main body of data; an outlier corresponds to a small positive value of λ_i. The posterior mean of this scale parameter was for Williams .50, .59 for Carew, .63 for Brett, and .75 for Boggs. Thus Williams's accomplishment was the greatest outlier in the sense that it deviated the most from American League batting averages in 1941.

This brief analysis of famous outliers is helpful in understanding which individual situational effects deviate significantly from the main population of effects. For each of the eight situational analyses, the posterior means of the scale parameters λ_i were computed for all of the players. Table 4 lists players for all situations where the posterior mean of λ_i is smaller than .75. To better understand a player's unusual accomplishment, this table gives his 1992 batting average in each category of the situation and the batting average difference ("season difference"). Next, the table gives the posterior mean of the difference in true probabilities $p_{i1} - p_{i2}$. Finally, it gives the difference in batting averages for the previous 4 years (given in Cramer and Dewan 1992).

What do we learn from Table 4? First, relatively few players are outliers using our definition—approximately one per situational variable. Note that the posterior estimates shrink the observed season batting average differences approximately halfway toward the average situational effect. The amount of shrinkage is greatest for the situational variables (such as "grass–turf") where the number of at-bats for one of the categories is small. The last column addresses the question if these nine unusual players had exhibited similar situational effects the previous 4 years. The general answer to this question appears to be negative. Seven of the nine players had 1988–1991 effects that were opposite in sign from the 1992 effect. The only player who seems to display a constant situational effect over the last 5 years is Tony Gwynn; he hits for approximately the same batting average regardless of the pitch count.

4. SUMMARY AND DISCUSSION

What have we learned from the preceding analysis of this hitting data? First, if one looks at the entire group of 154 baseball regulars, some particular situational variables stand out. The variation in batting averages by the pitch count is dramatic—batters generally hit 123 points higher when ahead in the count than with 2 strikes. But three other variables appear important. Batters on average hit 20 points higher when facing a pitcher of the opposite arm, 11 points higher when facing a groundball pitcher (as opposed to a flyball pitcher), and 8 points higher when batting at home. Because these latter effects are rather subtle, one may ask if these patterns carry over to other seasons. Yes they do. The same model (with the same prior) was fit to data from the 1990 season and the median opposite arm, groundball pitcher, and home effects were 25, 9, and 5 points respectively, which are close to the 1992 effects discussed in Section 3.

Do players have different situational effects? Bill James (personal communication) views situational variables as either "biases" or "ability splits." A bias is a variable that has the same effect on all players, such as "grass–turf," "day–

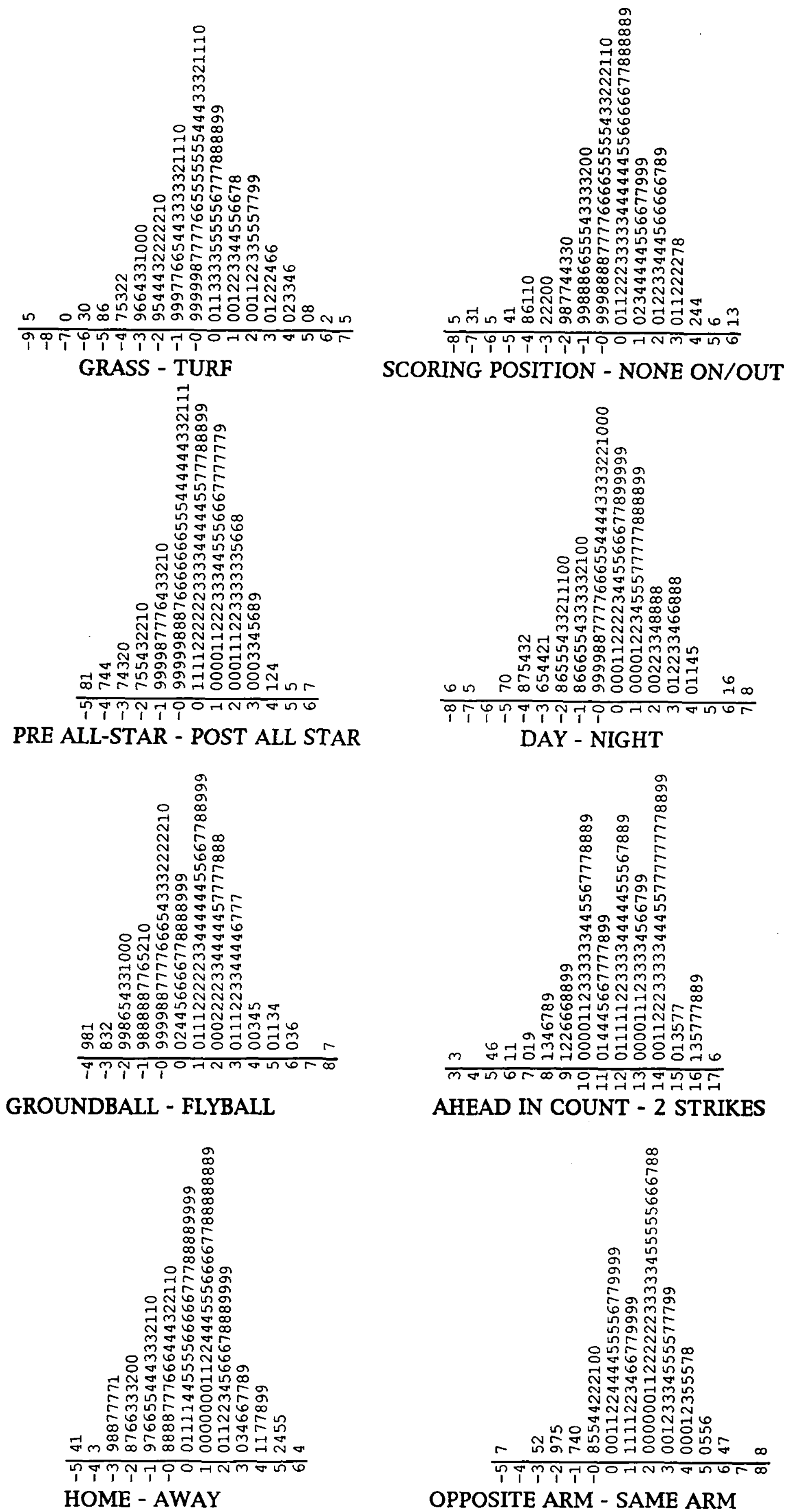

Figure 3. Stem-and-Leaf Diagrams of the Posterior Means of the Differences in Batting Averages for the Eight Situational Variables.

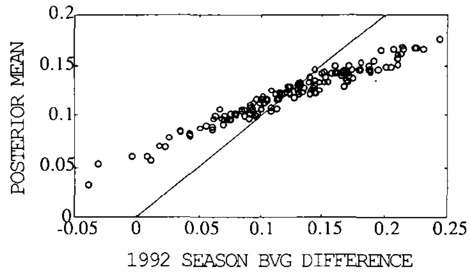

Figure 4. Posterior Means of the Differences in Batting Averages Plotted Against the Seasonal Difference in Batting Averages for the Situation "Ahead in Count/2 Strikes."

night", or "home–away." James argues that a player's ability to hit does not change just because he is playing on a different surface, or a different time of day, or at a particular ballpark, and so it is futile to look for individual differences in these situational variables. In contrast, the pitch count is an "ability split." There will exist individual differences in the "ahead in count–2 strikes" split, because one's batting average in the 2-strike setting is closely related to one's tendency to strike out. This statement is consistent with Figure 1, which indicates that players with high situational effects for pitch count during the previous 4-year period were likely to have high effects during the 1992 season. But there is much scatter in this graph, indicating that season performance is an imperfect measurement of this intrinsic ability.

Although there are clear situational patterns for the entire group of players, it is particularly difficult to see these patterns for individual players. We notice this in our brief study of nine unusual players in Section 3.2. These players had extreme estimated effects for the 1992 season, but many of them displayed effects of opposite sign for the previous 4-year period. The only player who seemed to have a clear outlying situational split was Tony Gwynn. But this does not mean that our search for players of high and low situational splits is futile. Rather, it means that we need more data to detect these patterns at an individual level.

Let us return to the book by Cramer and Dewan (1992), where the data were obtained. How does this book summarize the situational batting statistics that are listed? There is little discussion about the general size of the situational effects, so it is difficult to judge the significance of individual batting accomplishments. For example, suppose that a given hitter bats 100 points higher when ahead in the count compared with 2 strikes: Is that difference large or small? By our work, we would regard this as a relatively small difference, because it is smaller than the average of 123 batting average points for all players.

The book lists the 1992 batting leaders at the back. In the category of home games, we find that Gerry Sheffield had the highest batting average at home. But in this list, the players' hitting abilities and situational abilities are confounded; Sheffield's high average at home may reflect only the fact that Sheffield is a good hitter. In this article we have tried to isolate players' hitting abilities from their abilities to hit better or worse in different situations.

5. RELATED WORK

Because breakdown batting statistics are relatively new, there has been little statistical analysis of these data. There has been much discussion on one type of situational variable—hitting streaks or slumps during a season. Albright (1993) summarized recent literature on the detection of streaks and did his own analysis on season hitting data for 500 players. This data are notable, because a batter's hitting performance and the various situational categories are recorded for each plate appearance during this season. Albert (1993), in his discussion of Albright's paper, performed a number of stepwise regressions on this plate appearance data for 200 of the players. His results complement and extend the results described here. The "home–away" and "opposite arm–same arm" effects were found to be important for the aggregate of players. In addition, players generally hit for a higher batting average against weaker pitchers (deemed as such based on their high final season earned run averages). Other new variables that seemed to influence hitting were number of outs and runners on base, although the degree of these effects was much smaller than for the pitcher strength variable. The size of these latter effects appeared to be similar to the "home–away" effects. Players generally hit for a lower average with two outs in an inning and for a higher average with runners on base. This brief study suggests that there may be more to learn by looking at individual plate appearance data.

APPENDIX: DESCRIPTION OF THE POSTERIOR DISTRIBUTION OF THE SITUATIONAL EFFECTS AND SUMMARIZATION OF THE DISTRIBUTION USING THE GIBBS SAMPLER

The use of Bayesian hierarchical prior distributions to model structural beliefs about parameters has been described by Lindley and Smith (1972). The use of Gibbs sampling to simulate posterior distributions in hierarchical models was outlined by Gelfand, Hills, Racine-Poon, and Smith (1990). Albert (1992) used an outlier model, similar to that described in Section 2.3, to model homerun hitting data.

The complete model can be summarized as follows. For a given breakdown variable, we observe $\{(y_{i1}, y_{i2}), i = 1, \ldots, N\}$, where y_{ij} is the logit of the seasonal batting average of batter i in the jth category of the situation. We assume that the y_{ij} are independent, where y_{i1} is $N(\mu_i + \alpha_i, \sigma_{i1}^2)$ and y_{i2} is $N(\mu_i - \alpha_i, \sigma_{i2}^2)$, where the variances σ_{i1}^2 and σ_{i2}^2 are assumed known. The unknown parameters are $\alpha = (\alpha_1, \ldots, \alpha_N)$ and $\mu = (\mu_1, \ldots, \mu_N)$. Using the representation of a t density as a scale mixture of normals, the prior distribution for (α, μ) is written as the following two-stage distribution:

Stage 1. Conditional on the hyperparameters μ_α, σ_α, and $\lambda = (\lambda_1, \ldots, \lambda_N)$, α and μ are independent with μ distributed according to the vague prior $\pi(\mu) = c$ and the situational effect components $\alpha_1, \ldots, \alpha_N$ independent with α_i distributed $N(\mu_\alpha, \sigma_\alpha^2/\lambda_i)$.

Stage 2. The hyperparameters μ_α, σ_α, and λ are independent with μ_α distributed according to the vague prior $\pi(\mu_\alpha) = c$, σ_α^2 is distributed inverse gamma(a, b) density with kernel $(\sigma_\alpha^2)^{-(a+1)}\exp(-b/\sigma_\alpha^2)$, and $\lambda_1, \ldots, \lambda_N$ are iid from the gamma $(\nu/2, \nu/2)$ density with kernel $\lambda_i^{\nu/2-1} \exp(-\lambda_i\nu/2)$. The hyperparameters a, b, and ν are assumed known.

Table 4. Outlying Situational Players Where the Posterior Mean of the Scale Parameter $\lambda_i < .75$

Player	Situation	Batting average 1	Batting average 2	Season difference	Estimate of $p_1 - p_2$	Previous 4 years
Terry Steinbach	grass–turf	.251	.423	−.172	−.095	.033
Darrin Jackson	grass–turf	.278	.156	.122	.075	.020
Kevin Bass	scoring–none on/out	.205	.376	−.171	−.085	.063
Joe Oliver	scoring–none on/out	.172	.319	−.147	−.073	.079
Kent Hrbek	pre/AS–post/AS	.294	.184	.110	.074	−.007
Keith Miller	day–night	.167	.325	−.158	−.086	.022
Mike Devereaux	day–night	.193	.309	−.116	−.075	.035
Mickey Morandini	groundball–flyball	.324	.155	.169	.082	−.045
Tony Gwynn	ahead–2 strikes	.252	.291	−.039	.033	.061

Combining the likelihood and the prior, the joint posterior density of the parameters α, μ, μ_α, σ_α, and λ is given by

$$\exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left[\frac{(y_{i1}-\mu_i-\alpha_i)^2}{\sigma_{i1}^2}+\frac{(y_{i2}-\mu_i+\alpha_i)^2}{\sigma_{i2}^2}\right]\right\}$$
$$\times\left(\prod_{i=1}^{N}\lambda_i^{1/2}\right)\frac{1}{(\sigma_\alpha^2)^{N/2}}\exp\left\{-\frac{1}{2\sigma_\alpha^2}\sum_{i=1}^{N}\lambda_i(\alpha_i-\mu_\alpha)^2\right\}$$
$$\times\prod_{i=1}^{N}\left[\lambda_i^{\nu/2-1}\exp\left(-\frac{\lambda_i\nu}{2}\right)\right]\frac{1}{(\sigma_\alpha^2)^{a+1}}\exp\left(-\frac{b}{\sigma_\alpha^2}\right) \tag{A.1}$$

To implement the Gibbs sampler, we require the set of full conditional distributions; that is, the posterior distributions of each parameter conditional on all remaining parameters. From (A.1), these fully conditional distributions are given as follows:

a. $[\mu\,|\,\alpha, \mu_\alpha, \sigma_\alpha, \lambda]$. Define the variates $z_{i1} = y_{i1} - \alpha_i$ and $z_{i2} = y_{i2} + \alpha_i$. Then, conditional on all remaining parameters, the μ_i are independent normal with means $(z_{i1}/\sigma_{i1}^2 + z_{i2}/\sigma_{i2}^2)/(1/\sigma_{i1}^2 + 1/\sigma_{i2}^2)$ and variances $(1/\sigma_{i1}^2 + 1/\sigma_{i2}^2)^{-1}$.
b. $[\alpha\,|\,\mu, \mu_\alpha, \sigma_\alpha, \lambda]$. Define the variates $w_{i1} = y_{i1} - \mu_i$ and $w_{i2} = y_{i2} - \mu_i$. Then the α_i are independent normal with means $(w_{i1}/\sigma_{i1}^2 - w_{i2}/\sigma_{i2}^2 + \mu_\alpha\lambda_i/\sigma_\alpha^2)/(1/\sigma_{i1}^2 + 1/\sigma_{i2}^2)$ and variances $(1/\sigma_{i1}^2 + 1/\sigma_{i2}^2 + \lambda_i/\sigma_\alpha^2)^{-1}$.
c. $[\mu_\alpha\,|\,\mu, \alpha, \sigma_\alpha, \lambda]$ is $N(\Sigma\lambda_i\alpha_i/\Sigma\lambda_i, \sigma_\alpha^2/\Sigma\lambda_i)$.
d. $[\sigma_\alpha^2\,|\,\mu, \alpha, \mu_\alpha, \lambda]$ is inverse gamma with parameters $a_1 = N/2 + a$ and $b_1 = \Sigma_{i=1}^N \lambda_i(\alpha_i - \mu_\alpha)^2/2 + b$.
e. $[\lambda\,|\,\mu, \alpha, \mu_\alpha, \sigma_\alpha^2]$. The λ_i are independent from gamma distributions with parameters $a_i = (\nu + 1)/2$ and $b_i = \nu/2 + \lambda_i(\alpha_i - \mu_\alpha)^2/2\sigma_\alpha^2$.

To implement the Gibbs sampler, one starts with an initial guess at $(\mu, \alpha, \mu_\alpha, \sigma_\alpha, \lambda)$ and simulates in turn from the full conditional distributions a, b, c, d, and e, in that order. For a particular conditional simulation (say μ), one conditions on the most recent simulated values of the remaining parameters (α, μ_α, σ_α, and λ). One simulation of all of the parameters is referred to one cycle. One typically continues cycling until reaching a large number, say 1,000, of simulated values of all the parameters. Approximate convergence of the sampling to the joint posterior distribution is assessed by graphing the sequence of simulated values of each parameter and computing numerical standard errors for posterior means using the batch means method (Bratley, Fox, and Schrage 1987). For models such as these, the convergence (approximately) of this procedure to the joint posterior distribution takes only a small number of cycles, and the entire simulated sample generated can be regarded as an approximate sample from the distribution of interest.

[Received April 1993. Revised July 1993.]

REFERENCES

Albert, J. (1992), "A Bayesian analysis of a Poisson Random Effects Model for Homerun Hitters," *The American Statistician,* 46, 246–253.
—— (1993), Discussion of "A Statistical Analysis of Hitting Streaks in Baseball" by S. C. Albright, *Journal of the American Statistical Association,* 88, 1184–1188.
Albright, S. C. (1993), "A Statistical Analysis of Hitting Streaks in Baseball," *Journal of the American Statistical Association,* 88, 1175–1183.
Bratley, P., Fox, B., and Schrage, L. (1987), *A Guide to Simulation,* New York: Springer-Verlag.
Cramer, R., and Dewan, J. (1992), *STATS 1993 Player Profiles,* STATS, Inc.
Efron, B., and Morris, C. (1975), "Data Analysis Using Stein's Estimator and Its Generalizations," *Journal of the American Statistical Association,* 70, 311–319.
Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990), "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling," *Journal of the American Statistical Association,* 85, 972–985.
James, B. (1986), *The Bill James Baseball Abstract,* New York: Ballantine.
Lindley, D. V., and Smith, A. F. M. (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society,* Ser. B, 135, 370–384.
Morris, C. (1983), "Parametric Empirical Bayes Inference: Theory and Applications" (with discussion), *Journal of the American Statistical Association,* 78, 47–65.
Steffey, D. (1992), "Hierarchical Bayesian Modeling With Elicited Prior Information," *Communications in Statistics,* 21, 799–821.