

Beyond runs expectancy

Jim Albert*

Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH, USA

Received: 7 July 2014; revised 7 July 2014; accepted: 5 December 2014

Abstract. George Lindsey was one of the first to present run scoring distributions of teams of Major League Baseball. One drawback of Lindsey's approach is that his calculations represented a situation where the team is average. By use of a multinomial/multilevel modeling approach, we look more carefully how run-scoring distributions vary between teams and how run-scoring is affected by different covariates such as ballpark, pitcher quality, and clutch situations. By use of exchangeable models over ordinal regression coefficients, one gets a better understanding which covariates represent meaningful differences between run-scoring of teams.

Keywords: Multilevel modeling, clutch hitting, multinomial, ordinal regression modeling

1. Introduction

1.1. *George Lindsey and the runs expectancy matrix*

One of the pioneers in sabermetrics was George Lindsey, a defense consultant in Canada who had a great love for baseball. Lindsey wrote two remarkable papers in the 1960's that had a great influence on the quantitative analysis of baseball. In particular, Lindsey (1963) focused on several questions of baseball strategy such as stealing a base, sacrificing to sacrifice a runner to second base, and issuing an intentional walk. Lindsey observed that these questions could be answered by the collection of appropriate data.

“By collecting statistics from a large number of baseball games it should be possible to examine the probability distributions of the number of runs resulting from these various situations. Object of all choices is ... to maximize the probability ... of winning the game.”

Lindsay noted that the probability of winning at a point during the game depends on the current score and inning and presented tables of $W(I, H_i)$, the probability a team with a lead of I runs at the home half of the i th inning H_i will win the game.

In this paper, Lindsey also considered the runs scoring distribution of a team during an inning. He focused on the run potential, the number of runs a team will score in the remainder of the inning. He noted that the run potential depends on the current number of outs and the runners on base. Since there are three possible number of outs and eight possible runner configurations, there are $3 \times 8 = 24$ possible out/runner situations, and Lindsey focused on computing

$$Prob(R|T, B),$$

the probability of scoring exactly R runs in the remainder of the inning given T outs and runners on base B .

To learn about run potential, Lindsey's father collected run-scoring data for over 6000 half-innings in games during the 1959 and 1960 seasons, and Lindsey was able to find empirical run-scoring distributions for each of the 24 situations. By computing the mean runs for each situation, he produced the runs expectancy matrix as displayed in Table 1.

*Corresponding author: Jim Albert, Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH, USA. Tel.: +1 419 372 7456; Fax: +1 419 372 6092; E-mail: albert@bgsu.edu.

Table 1
Lindsey's runs expectancy matrix from Lindsey (1963)

Outs	Bases occupied							
	0	1	2	3	1,2	1,3	2,3	1,2,3
0 Outs	0.46	0.81	1.19	1.39	1.47	1.94	1.96	2.22
1 Out	0.24	0.50	0.67	0.98	0.94	1.12	1.56	1.64
2 Outs	0.10	0.22	0.30	0.36	0.40	0.53	0.69	0.82

At the beginning of an inning with no outs and bases empty, a team will score, on average, 0.46 runs in the remainder of the inning. In contrast, when the bases are loaded with one out, a team will score 1.64 runs in the rest of the inning.

The runs expectancy table has been illustrated in a number of sabermetrics books such as Thorn et al. (1984), Albert and Bennett (2003), Keri et al. (2007) and Tango et al. (2007) for deciding on proper baseball strategy, for learning about the value of plays, and for evaluating players. For example, Chapter 7 of Albert and Bennett (2003) shows the use of run expectancies in determining the values of singles, doubles, triples, and home runs that leads to use of linear weights to measure batting performances. Chapter 9 of Albert and Bennett (2003) uses the run expectancy table to assess the value of common baseball strategies such as a sacrifice bunt, intentional walk, and stealing a base.

Lindsay (1963) cautioned that the run expectancy table represented the situation where all players were "average".

"It must be reiterated that these calculations pertain to the mythical situation in which all players are 'average'. The allowance for the deviation from average performance of the batter at the plate, and those expected to follow him, or of runners on the bases, can be made by a shrewd manager who knows his players."

Also Lindsey (1963) mentioned the possibility of extending this analysis to run-scoring of teams with non-average players.

"If it were desired to provide a manager with a guide to the advisability of attempting various strategies in different situations, it would be possible to complete calculations of the type outlined here for all stages of the game and scores, pertaining to average players. It would also be possible, although onerous, to compute tables for nonaverage statistics, perhaps based on the past records of the individual players on the team."

Table 2

Basic probability model fits to runs scored in MLB. The Actual column gives the observed fractions of scoring 0, 1, 2, ... runs in the 2013 season and the Poisson and NB columns give the estimated fractions from fitting the Poisson and negative binomial distributions

Runs	Actual	Poisson	NB
		$\lambda = 0.462$	$n = 0.399, p = 0.464$
0	0.738	0.630	0.736
1	0.146	0.291	0.158
2	0.066	0.067	0.059
3	0.029	0.010	0.025
4	0.013	0.001	0.012
5	0.005	0.000	0.005
6	0.002	0.000	0.003
7	0.001	0.000	0.001

1.2. Modeling runs scored in an inning

A related line of research is the use of probability models to represent the number of runs scored in an inning or a game. Fits of basic discrete probability models do not accurately represent actual runs scored in an inning. To demonstrate this claim, the Poisson (λ) and negative binomial (n, p) models were separately fit to all inning runs scored in the 2013 season and the fitted probabilities for these two models and the actual run-scoring distribution are displayed in Table 2. Looking at the table, the Poisson fit underestimates the probability of a scoreless inning. Generally, the Poisson fit understates the actual variability in runs scored. The negative binomial fit is better than the Poisson, but this model overestimates the probability of scoring one run and underestimates the probability of scoring two runs.

Since basic probability models appear inaccurate, more sophisticated models for run scoring have been proposed in the literature. For example, Rosner et al. (1996) represents the probability of scoring x runs in an inning by the sum

$$h(x) = \sum_{N=x+3}^{x+6} f(N)g(x|N),$$

where N represents the number of batters in the inning. In this paper, the probability function $f(N)$ is represented by a modification of a negative binomial distribution, and the conditional distribution of the number of runs given the number of batters, $g(x|N)$, is modeled by a truncated binomial distribution. Another approach in Wollner (2000) provides an exponential formula for the probability a team averaging A runs per game scores R runs in a particular inning.

1.3. Multilevel modeling

Another research theme in modeling of sports data is the use of multilevel or hierarchical models to estimate parameters from several groups. Efron and Morris (1975) illustrate the use of an exchangeable model to estimate true batting averages of 18 players from the 1970 season. Albert (2004, 2006) provide further illustrations of the use of an exchangeable model to estimate a set of true batting rates or pitching rates. These models can be extended to the regression framework. Morris (1983) uses a multilevel regression model to assess if Ty Cobb was ever a true .400 hitter. Albert (2009) uses an exchangeable regression model to simultaneously estimate the true pitching trajectories for a number of pitchers.

1.4. Plan of the paper

This paper provides a multilevel modeling framework for understanding run scoring of teams in Major League Baseball, and understanding how team run scoring is affected by several factors, such as ballpark, pitcher quality, and runners in scoring position. The basic multilevel model, described in Section 2.1, simultaneously estimates means from several populations, say the average number of runs scored for 30 MLB teams, when the means are believed to follow a normal curve model with unknown mean and standard deviation.

As in Lindsay's work, we focus on the number of runs scored in a half-inning. A run scoring distribution for a particular team is represented by a multinomial distribution with underlying probabilities over the classes 0 runs, 1 run, 2 runs, 3 runs, and 4 or more runs. Section 2.2 describes a multilevel model for simultaneously estimating a set of multinomial probability vectors. By using this model to estimate the run scoring distributions for the 30 MLB teams, one obtains "improved" estimates at the teams' abilities to score runs.

After estimating teams' scoring distributions, we next explore team situational effects. An ordinal regression model, described in Section 2.3, is a useful method for describing how run-scoring probabilities change as one changes a predictor such as the quality of a pitcher. This model can be described in terms of the odds of scoring at least a particular number of runs. It is called a proportional odds model as a change in the predictor will result in the odds of scoring runs being

increased by a constant factor. Section 3.2 introduces the use of odds in comparing run scoring distributions of the National and American Leagues. Section 3.3 extends this approach to compare run scoring of the 30 teams and illustrates the value of multilevel modeling.

Section 4 focuses on the effect of the following covariates on team run scoring.

- **(Home Effects)** Teams generally score more runs at home compared to away games. What is the general size of the home/away effect for scoring runs and how does this effect vary among the teams?
- **(Pitcher Effects)** Clearly, the pitcher has a significant impact on run-scoring. A strong starting pitcher can neutralize the runs scored by even the best-scoring team in baseball. Generally, one expects that a team's run scoring is negatively associated with the quality of the pitcher. How can one quantify this pitcher effect, and how does this effect differ among teams?
- **(Runner Advancement)** Run scoring can be viewed as a two-step process – batters get on base and then other batters advance them to home. Do teams differ in their ability to get runners home?

We address each of these questions by a multilevel ordinal regression model where the effect of the covariate can vary across teams. Section 5 generalizes this approach to handle several covariates at once, Section 6 gives a brief discussion of the potential covariates relevant to run-scoring, and Section 7 summarizes the general findings.

2. Methods

2.1. Estimating a group of means

Consider the problem of estimating a collection of means from several populations. One observes sample means y_1, \dots, y_N , where the sample mean from the j th population, y_j is distributed normal with mean θ_j and known standard error σ_j . We describe this situation in a baseball setting, where $N = 30$, y_1, \dots, y_{30} correspond to the sample mean numbers of runs scored in a half-inning for the thirty teams for a season, and $\theta_1, \dots, \theta_N$ are the corresponding population means of runs scored.

One can obtain improved estimates at the population means by means of the following multilevel model. We assume that the means $\theta_1, \dots, \theta_{30}$ represent a

sample from a normal curve with mean μ and standard deviation τ . The locations of the normal curve parameters μ and τ are unknown, and so (under a Bayesian framework) vague or imprecise prior distributions are assigned to these parameters.

One can fit this multilevel model from the observed sample means of runs scored for the 30 teams. In particular, we estimate the population mean and standard deviation μ and τ by $\hat{\mu}$ and $\hat{\tau}$, respectively. The estimate $\hat{\mu}$ represents the overall or combined estimate of a population mean of the runs scored from the 30 teams and the estimate $\hat{\tau}$ represents the spread of these population means.

Three sets of estimates

There are three types of estimates of the population means. If we estimate each population mean only by using data from the corresponding sample, we obtain *individual* estimates

$$\hat{\theta}_j^I = y_j, \quad j = 1, \dots, N.$$

On the other extreme, if we assume that the population means are equal, that is, $\theta_1 = \dots = \theta_N$, then we'd estimate θ_j by the *combined* estimate

$$\hat{\theta}^C = \frac{\sum_{j=1}^N y_j / \sigma_j^2}{\sum_{j=1}^N 1 / \sigma_j^2},$$

where the sample mean y_j is weighted by the inverse of the sampling variance $1/\sigma_j^2$.

By use of the multilevel model, we obtain improved estimates of the population means that compromise between the individual and combined estimates. The multilevel (ML) estimate of the population mean for team j , $\hat{\theta}_j^{ML}$, is given by

$$\hat{\theta}_j^{ML} = \frac{1/\sigma_j^2}{1/\sigma_j^2 + 1/\hat{\tau}^2} y_j + \frac{1/\hat{\tau}^2}{1/\sigma_j^2 + 1/\hat{\tau}^2} \hat{\mu}.$$

The estimate of the mean runs scored for the j th team, $\hat{\theta}_j^{ML}$ shrinks the individual team estimate y_j towards the combined $\hat{\mu}$, where the size of the shrinkage depends on the ratio of the estimated population standard deviation $\hat{\tau}$ to the standard error σ_j . If $\hat{\tau}$ is small relative to the standard error σ_j , then the multilevel estimate $\hat{\theta}_j^{ML}$ will be close in value to the combined estimate. In contrast, if the estimate $\hat{\tau}$ is large (relative to the standard error), the multilevel estimate will be close to the individual estimate.

2.1.1. Example

This multilevel model was fit to the sample means of runs scored for the 30 teams in the 2013 season. We obtain the estimates $\hat{\mu} = 0.461$, $\hat{\tau} = 0.045$, so the population means of the runs scored are estimated by a normal curve with mean 0.461 and standard deviation 0.045. In the 2013 season, Anaheim averaged 0.500 runs scored per inning with a standard error of 0.027. The individual estimate of Anaheim's mean runs scored is $y_j = 0.500$ and the combined estimate is given by $\hat{\theta}^C = \hat{\mu} = 0.461$. The improved multilevel estimate of Anaheim's mean is given by

$$\begin{aligned} \hat{\theta}_{ANA}^{ML} &= \frac{1/0.027^2}{1/0.027^2 + 1/0.045^2} 0.500 \\ &+ \frac{1/0.045^2}{1/0.027^2 + 1/0.045^2} 0.461 = 0.490 \end{aligned}$$

Here the multilevel estimate of Anaheim mean shrinks the individual estimate 26% towards the combined estimate. Across all teams, the median shrinkage is 23%. Since this shrinkage percentage is a small value, this indicates that the observed differences in run scoring between the 30 teams correspond to meaningful differences in the teams' run-scoring abilities.

2.2. Estimating collections of multinomial data

A related problem is estimating a multinomial population. Suppose we classify data into k bins and observe the vector of frequencies $W = (w_1, \dots, w_k)$. The vector W is assumed to have a multinomial distribution with sample size $N = w_1 + \dots + w_k$ and probability vector $p = (p_1, \dots, p_k)$, where p_j represents the probability that an observation falls in the j th bin. In our setting, we classify the number of runs scored in an inning in the five bins 0, 1, 2, 3, and 4 or more runs, and the frequencies w_1, \dots, w_5 represent the number of innings where a particular team scores the different number of runs.

This type of multinomial data on runs scored is collected for each of the 30 baseball teams. Let W^1, \dots, W^{30} denote the vectors of frequencies of runs scored for the 30 teams. We assume the vector for the j th team W^j is multinomial with probability vector p^j . We are interested in estimating the probability vectors p^1, \dots, p^{30} . As in the population means case, we can consider individual, combined, and multilevel estimates for the probability vectors.

Individual estimates

Suppose we use the data from only the j th team to learn about its run scoring tendencies. Then the probabilities of falling in the different run groups are estimated by the individual estimates:

$$\hat{p}^{lj} = (\hat{p}_1^j, \dots, \hat{p}_k^j) = \left(\frac{w_1^j}{N^j}, \dots, \frac{w_k^j}{N^j} \right),$$

where w_1^j, \dots, w_k^j are the bin counts for the j th team and N^j is the number of innings for the j th team.

Combined estimates

Instead, suppose we assume that the probabilities of falling in the different runs group are the same for all teams; that is, $p^1 = \dots = p^{30}$. Then we can estimate each team's probability vector by the combined estimate

$$\hat{p}^C = (\hat{p}_1^C, \dots, \hat{p}_k^C) = \left(\frac{\sum w_1^j}{\sum N^j}, \dots, \frac{\sum w_k^j}{\sum N^j} \right).$$

Here we are pooling the bin counts for all teams to get the probability estimate for a particular bin.

Multilevel estimates

We wish to get improved estimates at the team probability vectors p^1, \dots, p^{30} that compromise between the individual estimates and the combined estimates. This is accomplished by means of the following multilevel model. We assume that the unknown probability vectors p^1, \dots, p^{30} are a random sample from a Dirichlet distribution with mean vector η and precision K with density proportional to

$$g(p) \propto \prod_{j=1}^5 p_j^{K\eta_j-1},$$

where $\eta = (\eta_1, \dots, \eta_5)$. The parameters η and K are assigned vague prior distributions.

One fits this multilevel model to the observed count data and one obtains estimates at K and η from the posterior distribution – call these estimates \hat{K} and $\hat{\eta}$. (Details of this computation are provided in the appendix.) The estimate $\hat{\eta}$ is approximately the combined estimate \hat{p}^C . Then the multilevel estimate at the probability vector of team j is given by

$$\hat{p}_{ML}^j = \frac{N_j}{N_j + \hat{K}} \hat{p}^j + \frac{\hat{K}}{N_j + \hat{K}} \hat{\eta}.$$

As anticipated, the multilevel model estimate for the probability vector of team j is approximately a weighted average of the individual estimate \hat{p}^j and the combined estimate \hat{p}^C . The weights depend on the ratio of the precision parameter estimate \hat{K} and the multinomial sample size N_j . We will illustrate the application of this model in simultaneously estimating the run-scoring distributions of the 30 teams in Section 3.3. (An early use of this Dirichlet distribution in modeling is given in Good (1967) and a good survey of smoothing estimates for multinomial data is provided in Simonoff (1995)).

2.3. *Groups of ordinal multinomial data with regression*

Ordinal logistic regression

Suppose one observes the vector of run frequencies $w = (w_1, \dots, w_k)$ which is multinomial with probability vector $p = (p_1, \dots, p_k)$. The categories 1, ..., k are ordinal – in our setting, these categories will be “0 runs”, “1 run”, “2 runs”, etc. Define the probability

$$\theta_c = p_c + \dots + p_k$$

which represents the probability of scoring at least c runs. The odds of scoring at least c runs is given by

$$odds = \frac{\theta_c}{1 - \theta_c}.$$

Suppose one observes a variable x that influences the run-scoring probabilities p . The ordinal logistic model (see McCullagh (1980) and Johnson and Albert (1999)) says that the log of the odds of scoring at least c runs is a linear function of x . This model is written as

$$\log \left(\frac{\theta_c}{1 - \theta_c} \right) = \gamma_c + x\beta,$$

where γ_c and β are unknown parameters. A unit increase in the covariate x results in the log odds of scoring c or more runs to be increased by β . By taking the exponential of both sides, this model can be written in the “proportional odds” form:

$$\frac{\theta_c}{1 - \theta_c} = \exp(\gamma_c + x\beta).$$

For each unit increase in the variable x , the odds of scoring at least c runs will increase by a factor of $\exp(\beta)$. This proportional odds model has the special property that the odds of scoring c or more runs will increase (with a unit increase of x) by a factor of $\exp(\beta)$ for all values of c . We illustrate this basic ordinal regression model in comparing the run-scoring distributions of the NL and AL in Section 3.2.

Ordinal logistic regression over groups

As a more general set-up, suppose we observe run-scoring frequencies for $N = 30$ teams, where the frequencies for the j th team w^j is multinomial with probability vector $p^j = (p_1^j, \dots, p_k^j)$. With the same covariate x , we fit the ordinal logistic model to w^j of the form

$$\log\left(\frac{\theta_c^j}{1 - \theta_c^j}\right) = \gamma_c^j + x\beta_j,$$

where $\theta_c^j = p_c^j + \dots, p_k^j$. In our setting, the regression coefficient β_j represents the additive increase in the logit of scoring at least c runs for the j team for each unit increase in x , and $\exp(\beta_j)$ represents the multiplicative increase in the odds of scoring c or more runs. After performing separate fits of the ordinal model to the run-scoring data for each of the N teams, we obtain the estimates $\hat{\beta}_1, \dots, \hat{\beta}_N$ with associated standard errors $\sigma_1, \dots, \sigma_N$.

One can now use the estimation of separate means methodology of Section 2.1 to get improved (multilevel) estimates at the regression effects. The true regression effects β_1, \dots, β_N are given a normal distribution with mean μ and standard deviation τ , we place vague priors on μ and τ . Improved estimates are provided by fitting this multilevel model – the estimates have the general form

$$\hat{\beta}_j^{ML} = \frac{1/\sigma_j^2}{1/\sigma_j^2 + 1/\hat{\tau}^2} \hat{\beta}_j + \frac{1/\hat{\tau}^2}{1/\sigma_j^2 + 1/\hat{\tau}^2} \hat{\mu}.$$

These multilevel estimates $\{\hat{\beta}_j^{ML}\}$ adjust the individual estimates $\{\hat{\beta}_j\}$ towards a combined estimate $\hat{\mu}$. As will be seen in a later section, the degree of adjustment depends on the sizes of the estimate of the standard deviation of the true effects $\hat{\tau}$ relative to the size of the standard error estimates $\{\sigma_j\}$.

3. League and team differences in scoring

3.1. Run scoring of all teams in 2013 season

We begin by collecting in Table 3 the runs scored for all complete innings (three outs) in the 2013 season. Since scoring five or more runs in an inning is relatively rare, Table 4 collapses the data from Table 3 by combining the counts of runs four or greater into the class “4+”.

3.2. Logits and comparing scoring of the NL and AL

To motivate our general approach, a useful re-expression of a proportion p is the logit $L = \log\left(\frac{p}{1-p}\right)$. For example, using data from Table 4, the logit of the proportion of innings where at least one run is scored is

$$L = \log\left(\frac{1 - 0.738}{0.738}\right) = -1.036.$$

This particular logit is computed by dividing the data by the breakpoint “0/1 runs” and comparing the proportions of the categories “1 or more runs” and “0 runs”. In a similar way, one can divide the data by each of the breakpoints “1/2 runs”, “2/3 runs”, and “3/4+ runs” and compute the logits of the resulting proportions. If we do this for all breakpoints, Table 5 is obtained.

Logits are useful in comparing groups of ordinal data such as runs scored. To illustrate, suppose we want to compare the runs scored per inning of National and

Table 3

Frequency table of runs scored for all complete innings in 2013 season

Runs	Count	Runs	Count
0	32315	6	81
1	6400	7	38
2	2899	8	14
3	1254	9	4
4	584	10	1
5	207	11	1

Table 4

Counts and percentages of inning runs scored with the new category “4+”

Runs	0	1	2	3	4+
Count	32315	6400	2899	1254	930
Percentage	73.80	14.60	6.60	2.90	2.10

Table 5
Logits of runs scored for all breakpoints

Breakpoint	0/1 runs	1/2 runs	2/3 runs	3/4+ runs
Proportion scoring “large” runs	0.262	0.116	0.050	0.021
Proportion scoring “small” runs	0.738	0.884	0.950	0.979
Logit = $\log \frac{\text{Proportion large}}{\text{Proportion small}}$	-1.036	-2.031	-2.944	-3.842

Table 6

Runs scored per inning by the American and National Leagues in the 2013 season

Runs	0	1	2	3	4+
American League	15963	3246	1497	654	500
National League	16352	3154	1402	600	430

Table 7

Logits of runs scored for all breakpoints for the two leagues

Breakpoint	0/1 runs	1/2 runs	2/3 runs	3/4+ runs	Median
American League	-0.996	-1.980	-2.887	-3.755	
National League	-1.074	-2.082	-3.011	-3.912	
Difference	0.078	0.102	0.124	0.157	0.113

American League teams in the 2013 season. First we obtain the counts of runs scored by the two leagues as displayed in Table 6. Each row of counts of the table is converted to proportions, and then logits are computed using each of the four breakpoints. The logits for each league are displayed in Table 7 and the “Difference” row gives the difference in the logits for the two leagues.

An ordinal regression model is a useful way to summarize the relationship seen in Table 7. If θ_c is the probability of scoring at least c runs, then the model is

$$\log \left(\frac{\theta_c}{1 - \theta_c} \right) = \gamma_c + x\beta,$$

where the $\{\gamma_c\}$ are parameters defining the cutpoints of the probabilities of scoring on the logit scale, x indicates the league ($x = 1$ if AL and $x = 0$ if NL), and β is the increase in the log odds of scoring c runs or more for the AL league. If we fit this model to the 2013 scoring data, we obtain the estimates

$$\hat{\gamma}_1 = -0.994, \hat{\gamma}_2 = -1.990, \hat{\gamma}_3 = -2.907, \\ \hat{\gamma}_4 = -3.790, \hat{\beta} = 0.0822.$$

Note that the estimates of γ_c are approximately equal to the logit values given in Table 7. Due to the designated

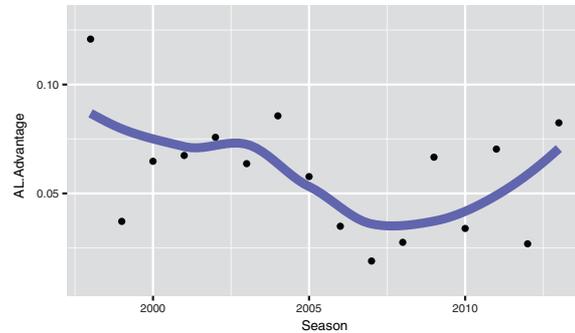


Fig. 1. American League run-scoring advantage over the National League for the seasons 1998 to 2013, the advantage is the estimate of the parameter β in the fit of the ordinal regression model. A smoothing curve indicates that the American League advantage in scoring runs has remained pretty constant over this time interval.

hitter, more runs are scored by American League teams and the size of this effect is measured by the coefficient estimate $\hat{\beta}$. On the logit scale, the the probability an American League team scores a large number of runs is 0.0822 greater than the probability a National League team scores a large number of runs. There is a nice approximation for differences in logits – for small values of x , an increase of x on the logit scale is approximately a $100 \times x$ percentage increase on the probability scale. Using this interpretation, we can say that the probability an American League team scores a large number of runs is $100 \times 0.0822 = 8.22\%$ greater than the probability a National League team scores a large number of runs.

To gain some historical perspective on the run-scoring advantage of the American League, this ordinal regression model was fit for each of the 16 seasons 1998 through 2013. Figure 1 displays a graph of the AL advantage (on the logit scale) against season. It is interesting to note that there is substantial variability in the AL advantage – for example the advantage was 0.121 in 1998 and decreased sharply to 0.019 in 2007. But there is no clear pattern in the changes in AL advantage over this time period. Generally, the probability an AL team scores a large number of runs is about 6% larger than the probability a NL team scores a large number of runs.

3.3. Comparing scoring of teams

How does run scoring vary across teams? If we break down this runs table by the batting team in Table 8, we see some interesting variation in runs scored.

Table 8
Percentages of different inning runs scored for all teams in the 2013 season

	BAT_TEAM	R0	R1	R2	R3	R4
1	ANA	72.4	15.5	6.5	2.8	2.8
2	ARI	74.6	13.1	7.1	3.3	1.9
3	ATL	74.2	13.5	6.9	2.9	2.5
4	BAL	71.1	16.0	7.2	3.4	2.3
5	BOS	68.9	15.9	7.9	4.1	3.2
6	CHA	75.7	14.2	6.4	2.0	1.6
7	CHN	76.3	12.9	6.3	2.7	1.7
8	CIN	72.7	16.2	6.2	2.7	2.2
9	CLE	71.4	15.4	7.4	3.5	2.4
10	COL	72.4	15.1	7.0	3.4	2.1
11	DET	72.0	13.6	7.4	3.4	3.5
12	HOU	76.7	13.0	5.3	3.2	1.8
13	KCA	74.4	14.0	5.6	3.4	1.9
14	LAN	73.2	15.3	7.5	2.3	1.7
15	MIA	79.1	12.5	4.9	2.2	1.4
16	MIL	74.6	14.4	6.6	2.5	1.9
17	MIN	74.6	15.2	6.0	2.7	1.5
18	NYA	74.6	14.2	6.1	2.9	2.2
s 19	NYN	75.8	13.6	6.3	2.5	1.7
20	OAK	72.0	14.4	7.9	2.6	3.1
21	PHI	75.6	13.8	6.5	2.1	1.9
22	PIT	74.4	15.4	5.8	2.6	1.8
23	SDN	74.8	15.1	5.7	2.5	1.8
24	SEA	74.6	14.7	6.9	2.2	1.6
25	SFN	74.6	14.9	5.6	3.4	1.6
26	SLN	71.6	15.5	6.8	3.0	3.2
27	TBA	71.9	16.1	7.2	2.9	1.8
28	TEX	72.5	14.8	7.6	2.6	2.5
29	TOR	72.4	14.8	7.4	3.2	2.2
30	WAS	74.1	14.4	6.7	3.0	1.9

If one looks at the percentage of big innings (four runs or more), Boston's percentage of big innings is 3.2, contrasted with Miami's percentage of 1.4. Boston's percentage of scoreless innings is 68.9, while the New York Mets were scoreless in 75.8 percent of their innings.

The logit approach described in Section 3 is used to facilitate comparisons in team scoring. Logits were computed for each team using the four breakpoints. For example, for Philadelphia, we computed the four logits

$$\log\left(\frac{P(\text{runs} \geq 1)}{P(0 \text{ runs})}\right), \log\left(\frac{P(\text{runs} \geq 2)}{P(0, 1 \text{ runs})}\right), \log\left(\frac{P(\text{runs} \geq 3)}{P(0, 1, 2 \text{ runs})}\right), \log\left(\frac{P(\text{runs} \geq 4)}{P(0, 1, 2, 3 \text{ runs})}\right)$$

To help in interpretation, the collection of logits for each team was converted to a residual by subtracting the MLB logits found in Table 5, and the team logit residuals are displayed in Fig. 2. Teams with lines above the horizontal line at zero represent above-average scoring teams. The run scoring for most of the

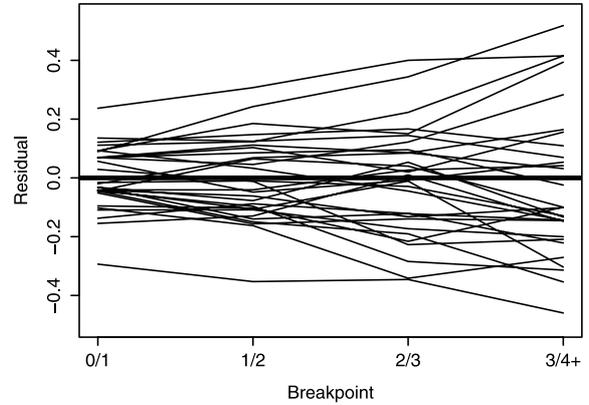


Fig. 2. Logit run scoring estimates for all teams. A team's residual logits were obtained by subtracting the overall logits.

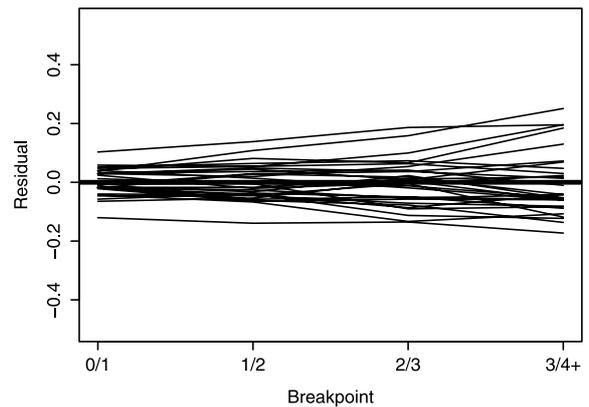


Fig. 3. Logit run scoring estimates for all teams using a multilevel model. The residual logits were obtained by subtracting the overall logits.

teams fall within 0.1 (on the logit scale) of the average. Note that some of the team lines are not monotone which indicate the presence of some chance variability in these run scoring estimates. Also, note there is more variability in the logit of 3 or more runs scored – that is a reflection of the small number of occurrences of big innings for the teams.

3.4. Multilevel modeling of team run distributions

Since there appears to be chance variability in the estimates of run scoring for the individual teams, there

may be some advantage to estimating the run scoring distributions simultaneously using a multilevel model. The exchangeable model of Section 2.2 is fit to the run distributions for the 30 teams. The estimate of the smoothing parameter K is 1973. Most teams play close to the median number of innings 1457. So the size of the shrinkage of the individual estimates towards the combined estimate is

$$\frac{1973}{1973 + 1457} = 0.57.$$

The multilevel estimates shrink the observed percentages approximately 57% towards the combined estimate. Figure 3 displays the logit reexpression of these multilevel estimates. We see these multilevel estimates adjust the individual estimates towards the combined estimate represented by the horizontal line at zero. One attractive feature of these estimates is that they remove some of the nonmonotone behavior of the individual estimates and reduces the variability in the logits of runs scored that we saw in Fig. 2. From similar experience with the use of multilevel models in estimating a collection of batting averages, we would expect these multilevel model estimates to provide better predictions than individual estimates of team run-scoring for future games.

4. Covariate effects

4.1. General method

Section 3 introduced the use of logits in the comparison of run-scoring distributions. To explore the effect of specific covariates like home/away, pitcher quality, and clutch hitting on scoring, we fit the general ordinal logistic model of the form

$$\log\left(\frac{\theta_c}{1 - \theta_c}\right) = \gamma_c + x\beta,$$

where θ_c is the probability of scoring at least c runs, and x is the covariate of interest (either of the categorical or measurement type). To understand how the covariate influences run scoring for all MLB teams, we use the following two-step modeling approach.

1. The ordinal logistic model is first fit separately to run-scoring data for each of the 30 teams. One obtains the regression effects $\hat{\beta}_1, \dots, \hat{\beta}_{30}$

with associated estimated standard errors $\hat{\sigma}_1, \dots, \hat{\sigma}_{30}$.

2. The multilevel exchangeable model in Section 2.1 is used to simultaneously estimate the regression effects. We assume the estimated covariate effect $\hat{\beta}_j$ is normal with mean β_j and standard error $\hat{\sigma}_j$. Once the estimates $\hat{\beta}_1, \dots, \hat{\beta}_{30}$ are computed, we are interested in simultaneously estimating the set of true effects $\beta_1, \dots, \beta_{30}$. The improved estimate of the covariate effect for the j th team

$$\hat{\beta}_j^{ML} = \frac{1/\sigma_j^2}{1/\sigma_j^2 + 1/\hat{\tau}^2} \hat{\beta}_j + \frac{1/\hat{\tau}^2}{1/\sigma_j^2 + 1/\hat{\tau}^2} \hat{\beta}^C$$

is a compromise between the individual estimate $\hat{\beta}_j$ and a combined estimate $\hat{\beta}^C$. The size of the shrinkage of the individual estimates towards the combined estimate depends on the size of $\hat{\tau}$, the spread of the true effects $\{\beta_j\}$.

4.2. Home versus away effects

To consider a team's scoring advantage of playing at home, write the logit of the probability of scoring at least c runs by

$$\log\left(\frac{\theta_c}{1 - \theta_c}\right) = \gamma_c + HOME \times \beta,$$

where the variable $HOME$ is equal to 1 if the team is playing at home, and $HOME = 0$ otherwise. If one computes the difference in the logits of the probability of scoring at least c runs at home θ_c^H , and away θ_c^A , one obtains

$$\log\left(\frac{\theta_c^H}{1 - \theta_c^H}\right) - \log\left(\frac{\theta_c^A}{1 - \theta_c^A}\right) = \beta,$$

or

$$\left(\frac{\theta_c^H}{1 - \theta_c^H}\right) / \left(\frac{\theta_c^A}{1 - \theta_c^A}\right) = \exp(\beta).$$

This ordinal regression model is initially fit to run-scoring data for all teams in the 2013 season and one obtains the covariate estimate $\hat{\beta} = 0.032$. The interpretation is that the ratio of the odds of scoring at least c runs at home and away is equal to $\exp(0.032) = 1.032$. Specifically, the probability that a team scores at least one run in an inning is increased by 3% at home, the probability a team scores at least two runs in an inning is increased by 3%, and so on.

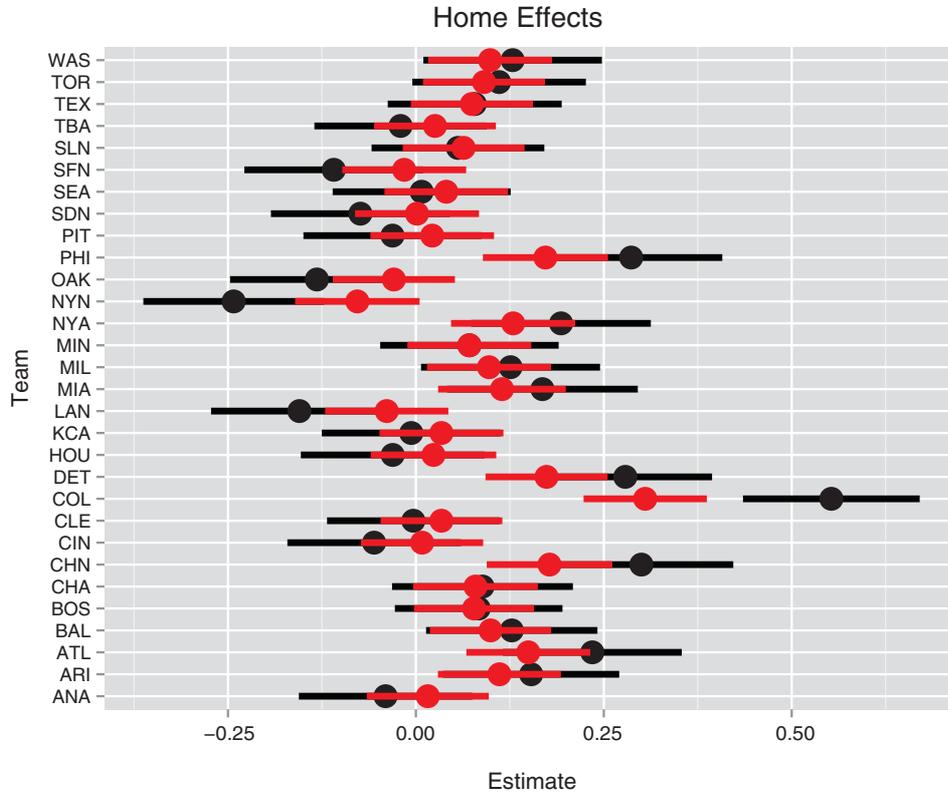


Fig. 4. Individual and multilevel home field run effects for all teams. The black line represents the individual estimate plus and minus the standard error and the red line represents the multilevel estimate plus and minus the standard error.

Since ballparks are known to have a significant impact on scoring, it is natural to fit this home/away model for each of the 30 teams. Figure 4 displays the individual estimates of the parameters $\beta_1, \dots, \beta_{30}$ as black dots where the endpoints of the bars correspond to the estimates plus and minus the standard errors. As expected, the individual home/away estimates show substantial variability from the New York Mets ($\hat{\beta}_j = -0.31$) to the Colorado Rockies ($\hat{\beta}_j = 0.52$).

Next we apply the multilevel model of Section 2.1 to simultaneously estimate the 30 home effects $\beta_1, \dots, \beta_{30}$. One obtains the multilevel estimates $\hat{\mu} = 0.071$ and $\hat{\tau} = 0.114$. The estimate $\hat{\mu}$ represents an average ballpark effect and $\hat{\tau}$ is an estimate at the spread of the true ballpark effects. Here the multilevel model estimates of $\{\beta_j\}$ shrink the individual estimates 52% of the way towards the combined estimate $\hat{\mu}$. The red bars in Fig. 4 show the posterior means plus and minus the posterior standard deviations. Since the size of the shrinkage is relatively small, this indicates that there is sizeable variation between the true home park effects.

4.3. Pitcher effects

The same ordinal regression approach can be used to learn about the pitcher effect on team run-scoring. The logit of the probability of scoring at least c runs in a half-inning is given by

$$\log\left(\frac{\theta_c}{1-\theta_c}\right) = \gamma_c + PITCHER \times \beta,$$

where the variable *PITCHER* is a measure of the quality of the pitcher who starts to pitch the half-inning.

The challenge in this approach is to find a good measure of pitcher quality. Many pitchers are of the relief type with a small number of plate appearances, and the measurements of pitcher quality typically show high variability. We use a multilevel model approach to obtain improved estimates of pitcher quality and these improved estimates are used as covariates in the ordinal model for run scoring.

To begin, we use a run value approach illustrated in Albert and Bennett (2003), Keri et al. (2007) and Tango

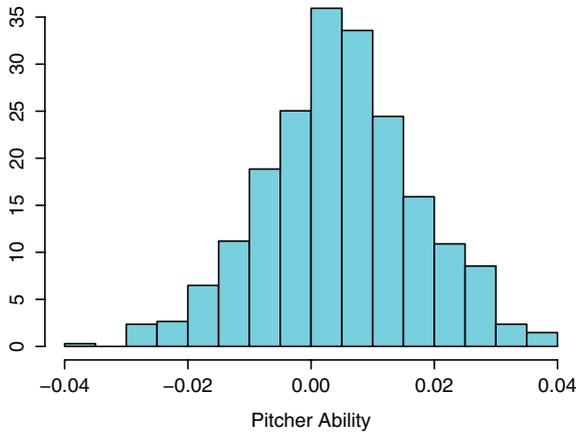


Fig. 5. Histogram of abilities of all pitchers in 2013 season.

et al. (2007) to measure the quality of all pitchers who played in the 2013 season. For each plate appearance (PA), we measure the run value as

$$\text{RUNS} = \text{Run Potential after PA} - \text{Run Potential before PA} + \text{Runs Scored},$$

where the run potential values come from the runs expectancy matrix (see Table 1) using 2013 season data. For the j th pitcher in the 2013 season, we compute the mean run value \bar{y}_j and the number of PA's n_j . We assume that \bar{y}_j is approximately normally distributed with mean μ_j and standard error $\sigma/\sqrt{n_j}$, where σ is estimated using all of the run values for the season. If we have N pitchers, then we estimate the population means μ_1, \dots, μ_N by use of an exchangeable multilevel model (see Section 2) and the corresponding multilevel estimates $\hat{\mu}_1, \dots, \hat{\mu}_N$ are used as surrogates for the qualities of the N pitchers in the ordinal regression model.

Figure 5 displays a histogram of the run value measures of the 679 pitchers in the 2013 season. These multilevel estimates shrink the individual mean run value estimates towards the overall value, the degree of shrinkage depending on the number of batters faced. For pitchers who faced fewer than 100 batters, the shrinkage exceeds 80%; for a starter facing 900 batters, the shrinkage was only 35%. In this particular season, Clayton Kershaw stands out with a run value measure of -0.036 – since he faced 934 batters, he saved $934 \times -0.036 = -33.6$ runs during the 2013 season.

If the ordinal regression model with pitcher quality covariate is fit to data for all teams, one obtains

the estimate $\hat{\beta} = 19.51$. The standard deviation of all pitcher abilities is 0.012. So if the pitcher's ability measure is increased by one standard deviation, one predicts a team's log odds of scoring runs to increase by $19.51 \times 0.012 = 0.23$.

By fitting the multilevel ordinal regression model, we learn about the pitcher effects for all 30 teams. Figure 6 displays individual (team) effects by black bars and the multilevel model estimates are displayed by red bars. The individual estimates show substantial variation. For example, Seattle's regression coefficient is $\hat{\beta}_j = 9.84$ and Philadelphia's coefficient is $\hat{\beta}_j = 28.24$, indicating that the Phillies were much more able to take advantage of poor pitching than Seattle in the 2013 season. The multilevel estimates in this case shrink the individual estimates about 76% towards the average value. Seattle and Philadelphia's pitcher effects, under the multilevel model, are corrected to 16.85 and 21.05, respectively. These multilevel estimates are better reflections of the true team pitcher effects than the individual team estimates, and would lead to better predictions of run-scoring against pitchers of different abilities.

4.4. Advancing baserunners home (clutch effects)

Run scoring generally is viewed as a two-step process. A team places runners on bases and then these runners are advanced to home. There is a special focus in the media on advancing runners who are in scoring position (on second and third base). One commonly hears about the number of runners left on base and the proportion of runners in scoring position who score. Do teams really differ in their abilities to advance runners from scoring position?

To address this clutch hitting question, an ordinal regression model is constructed. For each half-inning, one records the total number of (unique) runners SP who are in scoring position (either on second or third base) and the number of these runners R who eventually score. (Note that R can be smaller than the number of runs that score in the half-inning since runners who score don't need to be in scoring position.) As before, we classify R into the categories 0, 1, 2, 3, and "4 or more", and consider the ordinal regression model

$$\log \left(\frac{\theta_c}{1 - \theta_c} \right) = \gamma_c + SP \times \beta,$$

where θ_c is the probability that R is at least c . This model is first fit to data for all teams. One obtains

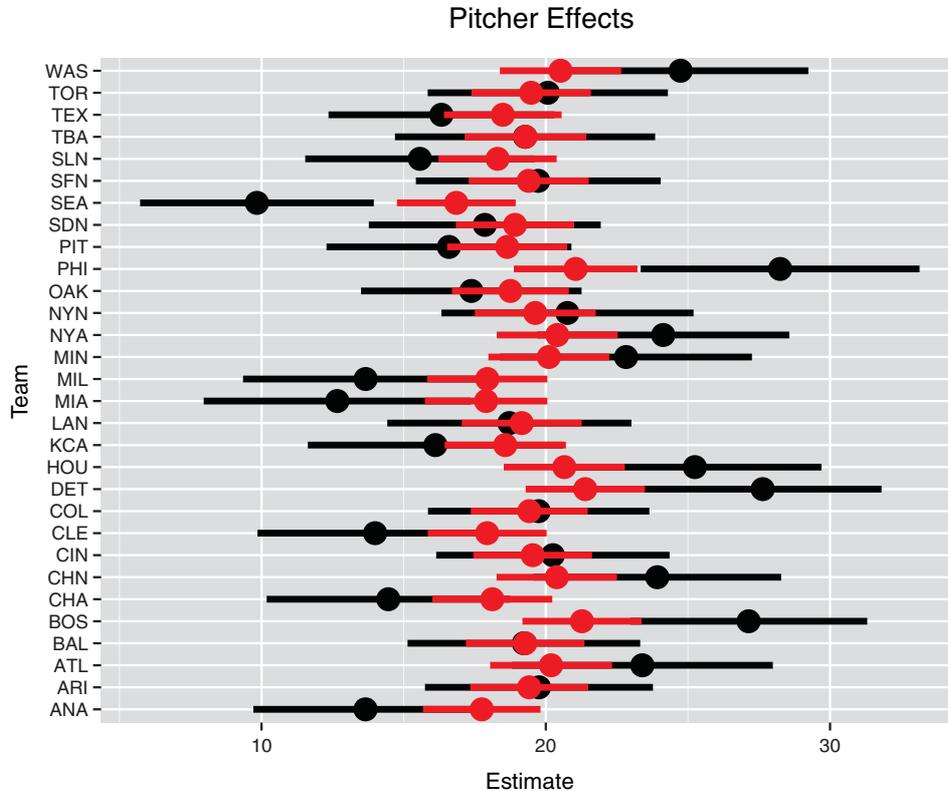


Fig. 6. Individual and multilevel pitcher effects for all teams. The black line represents the individual estimate plus and minus the standard error and the red line represents the multilevel estimate plus and minus the standard error.

the estimate $\hat{\beta} = 2.395$ which indicates that for each additional runner in scoring position, the log odds of scoring runs is increased by 2.395.

To see how this clutch hitting statistic varies between teams, this ordinal regression model was fit separately for all teams. Anaheim and the New York Mets had the smallest and largest regression estimates of $\hat{\beta}_j = 2.19$ and $\hat{\beta}_j = 2.60$, respectively. This indicates that the Mets were the best team in baseball in 2013 from a clutch-hitting perspective and Anaheim was the worst. However, when we use the multilevel model to simultaneously estimate the true clutch regression coefficients for all teams, we learn that this observed variability in clutch measures is primarily due to chance. Figure 7 displays the individual and multilevel clutch estimates for all teams. Here the shrinkage is about 93% and we see that the individual clutch estimates are shrunk almost entirely towards the common value in the multilevel model fit. The interpretation is that although teams obviously have different abilities to score runs,

teams appear to have similar abilities to advance runners in scoring position.

5. Multiple regression approach

5.1. Introduction

In our exploration of the effects of covariates on run-scoring, we focused on the application of an ordinal regression model with a single covariate. This was done since it is simpler to describe the covariate effects and the shrinkage behavior of the multilevel model. A natural extension of this approach is to model run-scoring of all teams by a large ordinal regression model where one includes all covariates that one believes have an impact on run scoring. Using our general notation, one can write the ordinal logistic model as

$$\log \left(\frac{\theta_c^j}{1 - \theta_c^j} \right) = \gamma_c^j + x_1 \beta_{j1} + \dots + x_k \beta_{jk},$$

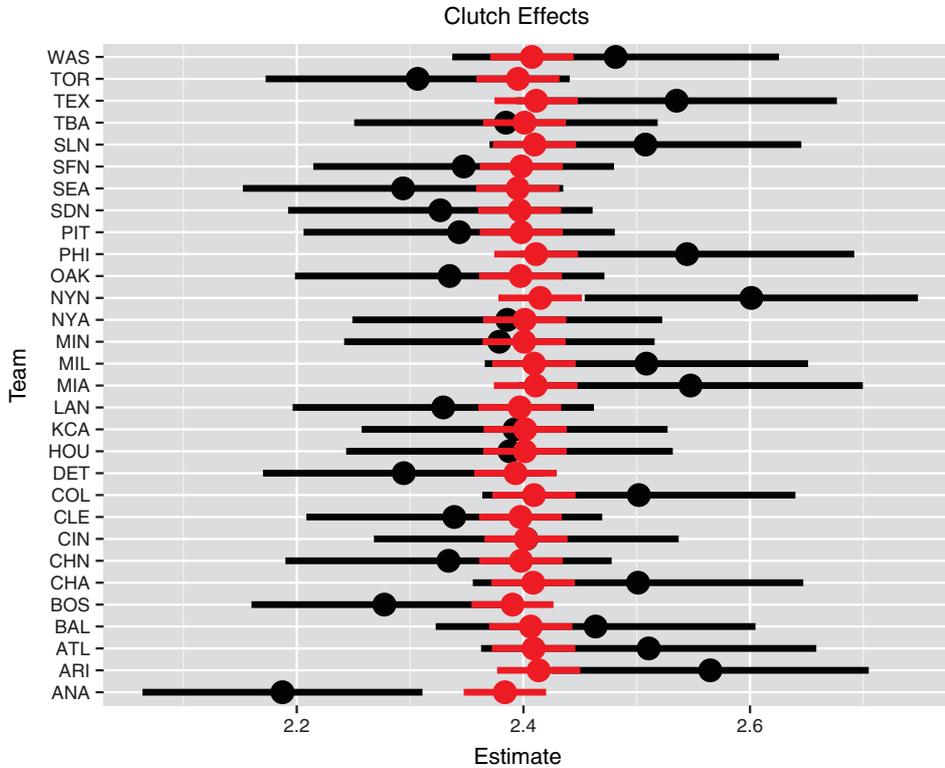


Fig. 7. Individual and multilevel clutch effects for all teams. The black line represents the individual estimate plus and minus the standard error and the red line represents the multilevel estimate plus and minus the standard error.

where θ_c^j represents the probability of scoring at least c runs for the j th team, x_1, \dots, x_k represent k possible covariates (such as league, ballpark, and pitcher quality), and $\beta_{j1}, \dots, \beta_{jk}$ represent the regression effects for the j th team. In the multilevel modeling framework, one could assign $\gamma_c^1, \dots, \gamma_c^N$ a common multivariate normal distribution, and likewise assume each of the sets of team covariate effects $\{\beta_{jk}, j = 1, \dots, N\}$ come from a common normal distribution. This model is more complicated to fit due to the large number of unknown parameters, but it would accomplish the same smoothing effect as demonstrated in this paper. Team scoring distributions would be shrunk towards an overall scoring distribution – this is accomplished by the common multivariate normal distribution played on the bin cutpoint parameters γ_c^j . In a similar fashion, team effects for a particular covariate such as pitcher quality would be shrunk or moved towards a common covariate effect. This particular approach is promising and can help us better understand relevant covariates that affect run scoring.

5.2. Example

To illustrate this multiple regression approach, suppose we are interested in learning about the home and inning effects on team run scoring. For the j th team, we write the ordinal regression model as

$$\log \left(\frac{\theta_c^j}{1 - \theta_c^j} \right) = \gamma_c^j + HOME \beta_{j1} + INNING \beta_{j2},$$

where $HOME = 1$ if the batting team is playing at home ($HOME = 0$ otherwise), and $INNING = 1$ if the team is batting in the first through fifth innings ($INNING = 0$ for later innings). We assume that the home and inning effects for the 30 teams, $(\beta_{11}, \beta_{12}), \dots, (\beta_{30,1}, \beta_{30,2})$ come from a common bivariate normal distribution.

This model was fit on the 2013 season data. One can represent an estimate of the pair (β_{j1}, β_{j2}) by an ellipse that contains the unknown pair of parameters with 95% probability. Figure 8 displays the estimates of the home and inning effects where one uses individual

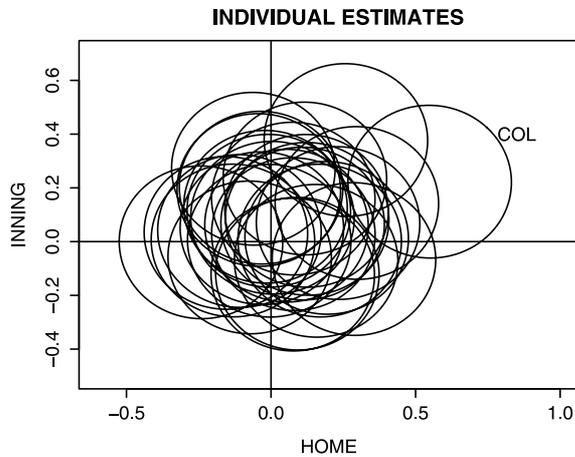


Fig. 8. Individual team estimates of home/away and early/late inning effects from ordinal multivariate regression model for 2013 season. Each ellipse represents the location of a 95% confidence region and the region corresponding to Colorado is labelled.

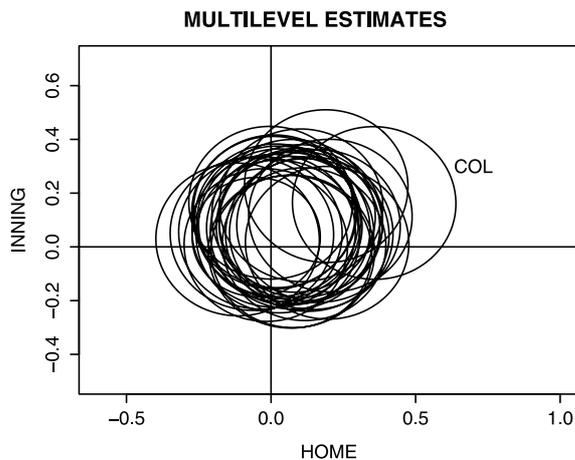


Fig. 9. Multilevel ordinal regression model estimates of home/away and early/late inning effects for 2013 season. Each ellipse represents the location of a 95% confidence region and the region corresponding to Colorado is labelled.

regression models for the 30 teams, and Fig. 9 displays estimates of the same effects using the multilevel regression model. Note that there are general tendencies for teams to score more at home and in early innings (the ellipses tend to be located in the upper-right section of the graph), but the locations of the ellipses for the multilevel model estimates tend to be shrunk towards a common estimate.

6. What covariates are relevant to run-scoring?

The ordinal regression approach seems advantageous to the expected runs approach for understanding team run-scoring, and we have illustrated the use of several covariates such as home/away, pitcher quality, and ability to advance runners in scoring position. But there needs to be a more careful look at all of the team attributes that affect run-scoring. One believes that this modeling approach can be extended to handle many other covariates.

Home/away affect. The home-away run-scoring effects documented actually have several dimensions. A team scores more in their ball park partly due to physical and weather attributes of their ballpark, but there is evidence to suggest that umpires may favor the home team in the calling of balls and strikes. How much of the home/away effect is due to each factor?

Individual batter effects. This paper has focused on run-scoring of teams. But a team's batting lineup consists of players who aid run-scoring by their ability to get on-base and other players contribute by advancing runners home by extra-base hits. One typically evaluates the contribution of individual hitters by run value, but these run value calculations are based on league averages. One would speculate that one could get better evaluations of hitter runs contributed by using run distributions individualized for the player's team.

Base-running effects. As documented in the recent World Series between the Royals and the Giants, runs are often scored by good runner advancement. Several of the runs were set up by a runner on second base who advanced to third on a sacrifice fly. Teams clearly differ with respect to their ability to steal and take the extra base, and one would be interested in measuring the effect of base-running on run scoring for all teams.

Situational effects. Runs are scored within a context of the opposing pitcher and defense, and the inning. The measurement of pitcher quality used here is actually a measurement of pitcher quality and team defense. How can one isolate the effect of these two effects in run scoring? There is a recent movement to shifting the infielders for specific hitters – what is the value of this shift from the viewpoint of runs? During different innings, teams may have different goals towards run-scoring. In close games, the objective might be to score a single run. In other situations, the goal might be to

score multiple runs. Do teams differ with their ability to score a single run?

7. Summary and concluding comments

The primary objective of this work is to provide a better understanding of the run-scoring patterns in Major League Baseball. Instead of focusing on the mean number of runs scored in a half-inning, we focus on the probability distribution of runs scored. We are interested in how this run-scoring distribution varies among teams and the effect of various covariates such as the ballpark, pitcher/defense, and “runners in scoring position” situations.

The primary method is ordinal regression of the multinomial run-scoring outcome and the use of multilevel modeling to estimate run-scoring over different groups. One finding is that there are some covariates such as ballpark where there are clear team-specific run-scoring advantages. This is not surprising since it is well-known that Coors Field (home of Colorado Rockies) is very advantageous for run-scoring and other parks such as Citi Field (home of the New York Mets) are more restrictive for scoring runs. But there are other covariates such as advancing runners in scoring position where there are little differences between teams in scoring runs. It should be clarified that we do observe differences in teams’ advancing runners in scoring position, but there is little evidence to suggest that teams have different clutch-hitting abilities. If the media understands this conclusion, then there would be less discussion about teams’ batting performances when runners are in scoring position. The goal of future work to find suitable covariates such as the ones suggested in Section 5.2 that have a significant impact on run-scoring, and the knowledge of these covariates would be helpful for teams building an offense.

References

Albert, J., 2004. A batting average: Does it represent ability or luck? Technical Report.

Albert, J., 2006. Pitching statistics, talent and luck, and the best strikeout seasons of all-time. *Journal of Quantitative Analysis in Sports*. 2(1), 2.

Albert, J., 2009. Is Roger Clemens’ WHIP trajectory unusual? *Chance*. 22(2), 8–20.

Albert, J., Bennett, J., 2003. *Curve Ball*, Springer.

Efron, B., Morris, C., 1975. Data analysis using Stein’s estimator and its generalizations. *Journal of the American Statistical Association*. 70(350), 311–319.

Good, I.J., 1967. A Bayesian significance test for multinomial distributions. *Journal of the Royal Statistical Society. Series B*, pp. 399–431.

Johnson, V., Albert, J., 1999. *Ordinal Data Modeling*, Springer.

Keri, J., *Baseball Prospectus*, 2007. *Baseball Between the Numbers: Why Everything You Know About the Game is Wrong*, Basic Books.

Lindsey, G., 1963. An investigation of strategies in baseball. *Operations Research*. 11(4), 477–501.

McCullagh, P., 1980. Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B*. 42(2), 109–142.

Morris, C., 1983. Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*. 78(381), 47–55.

Rosner, B., Mosteller, F., Youtz, C., 1996. Modeling pitcher performance and the distribution of runs per inning in major league baseball. *The American Statistician*. 50(4), 352–360.

Simonoff, J., 1995. Smoothing categorical data. *Journal of Statistical Planning and Inference*. 47(1), 41–69.

Tango, T., Lichtman, M., Dolphin, A., 2007. *The Book: Playing the Percent-ages in Baseball*, Potomac Books, Inc.

Thorn, J., Palmer, P., Reuther, D., 1984. *The Hidden Game of Baseball: A Revolutionary Approach to Baseball and Its Statistics*, Doubleday Garden City, New York.

Wollner, K., 2000. An analytic model for per-inning scoring distributions, *Baseball Prospectus*.

Appendix: Estimating collections of multinomial data using an exchangeable model

In Section 2.2, one observes W^1, \dots, W^{30} , vectors of runs scored for the 30 teams and one assumes that $W^j = (W_1^j, \dots, W_5^j)$, the number of runs scored in the bins 0, 1, 2, 3, and 4 or more is multinomial with probability vector $p^j = (p_1^j, \dots, p_5^j)$. Under the multilevel exchangeable model, we assume that p^1, \dots, p^{30} represent a random sample from a Dirichlet distribution of the form

$$g(p) = \frac{1}{D(K\eta)} \prod_{i=1}^5 p_i^{K\eta_i - 1},$$

where

$$D(K\eta) = \frac{\prod_{i=1}^5 \Gamma(K\eta_i)}{\Gamma(K)},$$

is the Dirichlet function and $\Gamma()$ is the gamma function. The model is completed by assigning the mean vector

$\eta = (\eta_1, \dots, \eta_5)$ and the “precision” parameter K priors – we assume that one has little information about these parameters and that leads to the prior density of the form

$$g(\eta, K) = \frac{1}{(1 + K)^2}, K > 0.$$

One estimates the precision parameter K from its marginal posterior distribution. One can show that the posterior density of (K, η) has the form

$$g(\eta, K|\text{data}) \propto \prod_{j=1}^{30} \left(\frac{D(K\eta + W^j)}{D(K\eta)} \right) \frac{1}{(1 + K)^2}.$$

One estimates (K, η) by the posterior mode, that is, the value $(\hat{K}, \hat{\eta})$ that maximizes the marginal posterior density $g(K, \eta|\text{data})$. These estimates are used in computing the multilevel estimate at the probability vector of team j given in Section 2.2 as

$$\hat{p}_{ML}^j = \frac{N_j}{N_j + \hat{K}} \hat{p}^j + \frac{\hat{K}}{N_j + \hat{K}} \hat{\eta}.$$

The multilevel estimate \hat{p}_{ML}^j is a compromise between the individual team estimate $\hat{p}^j = W^j/N^j$ and the combined estimate $\hat{\eta} \approx \sum W^j / \sum N^j$.