



University of Pennsylvania  
ScholarlyCommons

---

Statistics Papers

Wharton Faculty Research

---


7-2007

# Evaluating Throwing Ability in Baseball

Matthew Carruth  
*University of Pennsylvania*

Shane T. Jensen  
*University of Pennsylvania*

Follow this and additional works at: [http://repository.upenn.edu/statistics\\_papers](http://repository.upenn.edu/statistics_papers)

 Part of the [Engineering Commons](#), and the [Statistics and Probability Commons](#)

---

## Recommended Citation

Carruth, M., & Jensen, S. T. (2007). Evaluating Throwing Ability in Baseball. *Journal of Quantitative Analysis in Sports*, 3 (3), <http://dx.doi.org/10.2202/1559-0410.1079>

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/statistics\\_papers/445](http://repository.upenn.edu/statistics_papers/445)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# Evaluating Throwing Ability in Baseball

## **Abstract**

We present a quantitative analysis of throwing ability for major league outfielders and catchers. We use detailed game event data to tabulate success and failure events in outfielder and catcher throwing opportunities. We attribute a run contribution to each success or failure which is tabulated for each player in each season. We use four seasons of data to estimate the overall throwing ability of each player using a Bayesian hierarchical model. This model allows us to shrink individual player estimates towards an overall population mean depending on the number of opportunities for each player. We use the posterior distribution of player abilities from this model to identify players with significant positive and negative throwing contributions. Our results for both outfielders and catchers are publicly available.

## **Keywords**

baseball, Bayesian hierarchical model, shrinkage

## **Disciplines**

Engineering | Statistics and Probability

# *Journal of Quantitative Analysis in Sports*

---

*Volume 3, Issue 3*

2007

*Article 2*

---

## Evaluating Throwing Ability in Baseball

**Matthew Carruth**, *School of Engineering and Applied Science, University of Pennsylvania*

**Shane Jensen**, *Department of Statistics, The Wharton School, University of Pennsylvania*

**Recommended Citation:**

Carruth, Matthew and Jensen, Shane (2007) "Evaluating Throwing Ability in Baseball," *Journal of Quantitative Analysis in Sports*: Vol. 3: Iss. 3, Article 2.

**DOI:** 10.2202/1559-0410.1079

©2007 American Statistical Association. All rights reserved.

# Evaluating Throwing Ability in Baseball

Matthew Carruth and Shane Jensen

## Abstract

We present a quantitative analysis of throwing ability for major league outfielders and catchers. We use detailed game event data to tabulate success and failure events in outfielder and catcher throwing opportunities. We attribute a run contribution to each success or failure which is tabulated for each player in each season. We use four seasons of data to estimate the overall throwing ability of each player using a Bayesian hierarchical model. This model allows us to shrink individual player estimates towards an overall population mean depending on the number of opportunities for each player. We use the posterior distribution of player abilities from this model to identify players with significant positive and negative throwing contributions. Our results for both outfielders and catchers are publicly available.

**KEYWORDS:** baseball, Bayesian hierarchical model, shrinkage

**Author Notes:** The authors thank Abraham Wyner and Dylan Small for helpful discussion and suggestions.

# 1 Introduction

The impact of a fielders arm strength on their respective defensive rating has long been neglected and unmeasured. Research into outfielder ability has tended to focus on estimation of differences in fielding range, such as the Ultimate Zone Rating (Lichtman, 2003) and the recent work by David Pinto (Pinto, 2006). Despite this recent advancement in methods for quantifying the range of outfielders, there has been less development of sophisticated methods of quantifying an outfielders throwing ability as a defensive tool. The *put-outs* and *assists* statistics are a common but unacceptable summary of outfielder throwing ability since it only quantifies successful events, and outfielders are rarely given errors for unsuccessful ability to throw out players as a balancing measure of unsuccessful events. In addition, there is a more subtle effect of throwing ability that is not captured by current measures: an outfielder with a reputation of a strong arm will be tested far less often and as a result, will save runs over the course of the season by reducing baserunner attempts to take extra bases. More recent work has begun to address this need by quantifying both *hold* and *kill* events (Walsh, 2007) but does not consider the influence of outfield ball-in-play location on these events.

Research into catcher fielding also shows a lack of sophisticated analyses of throwing ability, which is even more necessary since there is limited information available for other aspects of catcher fielding. Fielding range on pop-ups, bunts and short groundballs has been examined (Pinto, 2006), but these are relatively rare events, and in the case of pop-ups, the vast majority have large enough hang times that every catcher makes a successful play. More success has been achieved in the study of passed balls and wild pitches, due to the work by David Gassko (Gassko, 2005) and others. Studies by Keith Woolner (Woolner, 1999) have attempted to quantify differences between catchers in terms of pitch calling, but these effects have not been shown to be statistically significant. Previous studies into throwing ability for catchers have not been satisfactory. Most studies of throwing ability (Tippett, 1997) have used broad categorizations such as caught stealing percentage, which does capture both successful and unsuccessful events to throw out baserunners. However, the more subtle issue of attempt prevention is not captured by this statistic: a catcher with a reputation for high throwing ability will have less attempts to throw baserunners out, since baserunners will be less likely to attempt a stolen base. The prevention of baserunning also has positive value to a team (though not as much as throwing out a baserunner), but this effect is not captured by statistics based

only on baserunning attempts.

In this paper, we quantify the relative throwing ability of each individual catchers and outfielders by tracking all throwing opportunities in the 2002-2005 seasons and evaluating the relative success of each fielder relative to the league average on the scale of runs saved/cost. We capture both individual ability to throw out baserunners as well as individual tendencies to prevent baserunning. We describe our overall methodology for catchers and outfielders in Section 2. We focus on catchers and outfielders since the length of throws for infielders are so short that differences in throwing ability are impossible to separate from fielding ability. In Section 3, we average individual player throwing ability across multiple seasons with a simple hierarchical model (and Gibbs sampling implementation) that allows for different variances and sample sizes between different players. We examine several interesting results from this approach for catchers in Section 4 and outfielders in Section 5. We conclude with a brief discussion of our analysis.

## **2 Evaluation of Throwing Ability**

To track throwing events, we used the Baseball Info Solutions database (BIS, 2007) which tabulates detailed play-by-play data for all games within the 2002-2005 MLB seasons. The BIS data contains all hitting and base-running events in each game, and any changes in scoring and baserunner configurations as a consequence of each event. In addition, additional fielding information is provided for balls put into play: fielders involved in the play as well as the location where the ball was fielded. Our evaluation is based on an initial categorization of each event as a baserunning opportunity or not depending on the game situation. In order to provide an interpretable and comparable measure for comparing individual players, we convert the successes or failures in these opportunities into a runs saved/cost measure. The scale of runs saved/cost is a natural one and has been used previously by Gassko (Gassko, 2005) in his work on the effects of passed balls and wild pitches. Our overall strategy will be to apply the Expected Runs Matrix (REM, 2007) to calculate the run contribution of a successful vs. unsuccessful plays, and reward/punish individual players accordingly. These run rewards for individual players will be calculated while taking into account the averages across all players in order to come up with a runs saved/cost for each player.

## 2.1 Evaluation for Catchers

Consider all events in the 2002-2005 seasons that involved a base-stealing opportunity. For catchers, a successful play could either be catching a base runner during a stolen base attempt, or preventing a baserunner from attempting a stolen base. Steal opportunities consisted of five categories: runner on first base, runner on second base, runners on first and third base, runners on second base with a runner on first, and runner on first base with a runner on second (a distinction is made between the last two categories in order to properly track double steals). These five categories can be further divided into fifteen subcategories ( $C = 1, \dots, 15$ ) based on how many outs there were prior to the play in question. Steals of home plate are not considered by our analyses since they are usually the consequence of the pitcher, not of the catcher. For each catcher  $P$ , we tabulated all base-stealing opportunities  $N(P, C)$  within each subcategory  $C$  as well as the number of opportunities  $A(P, C)$  where the baserunner did attempt a stolen base. If a stolen base was attempted, we tabulated the number of attempts that resulted in a successful steal  $S(P, C)$  and the number of attempts where the baserunner was thrown out  $F(P, C)$ . We also totaled these counts across all catchers in order to establish total values for the entire set of catchers. An example of our tabulation is given in Table 1 below. For each game situation

Table 1: Tabulation of Base-Stealing Opportunities for 2002

Catcher	Situation $C$	Opportunities $O_C$	Attempts $A_C$	Stolen Bases $S_C$	Caught Stealings $F_C$
J. Lopez	man on 1st, 0 outs	241	25	14	11
All	man on 1st, 0 outs	12361	831	519	312

$C$ , we want to focus on two important concepts: the situation's propensity for steal attempts and the success rate of those steal attempts. We use the totals over all catchers together with the opportunities for catcher  $P$  to cal-

culate expected counts of stolen bases and caught stealings for catcher  $P$ :

$$E[S(P, C)] = N(P, C) \times \frac{\sum_P S(P, C)}{\sum_P N(P, C)}$$

$$E[F(P, C)] = N(P, C) \times \frac{\sum_P F(P, C)}{\sum_P N(P, C)}$$

Returning to the situation in Table 1, our calculated expectations for J. Lopez were  $E[S(\text{J.Lopez}, C = 1)] = 10.08$  stolen bases and  $E[F(\text{J.Lopez}, C = 1)] = 6.07$  caught stealings. Comparing to his observed totals, we see that J. Lopez had 3.92 more stolen bases and 4.93 more caught stealings than expected in 2002. How should J. Lopez be compensated for these individual differences in expectation for both stolen bases and caught stealings?

The Expected Runs Matrix (REM, 2007) provides us with expected runs for each game situation (baserunner configuration  $\times$  number of outs), which we can use to calculate a runs saved/cost value for a particular successful or unsuccessful play. As an example, consider again the game situation from Table 1. From the Expected Runs Matrix, the expected run value  $R$  for a (1st base alone, 0 outs) situation is 0.90. If the baserunner attempts to steal and is thrown out by the catcher, then the game situation changes to (no baserunners, 1 outs) with a corresponding expected run value  $R$  of 0.28, which means that the catcher has, in expectation, saved his team 0.62 runs. However, if the baserunner successfully steals the base, then the game situation changes to (2nd base alone, 0 outs) with a corresponding expected run value  $R$  of 1.14, meaning that the catcher has saved his team -0.24 expected runs (ie. cost his team 0.24 runs).

More generally, for any game situation  $C$  involving a baserunner, we have the run value  $R(C)$  for that situation. If a stolen base is attempted and a runner is caught stealing, the game situation changes from  $C \rightarrow C'$ , and the positive run value for the catcher is  $R(C') - R(C)$ . If a stolen base is attempted and the runner gets a stolen base, the game situation changes from  $C \rightarrow C''$ , and the negative run value for the catcher is  $R(C'') - R(C)$ . Thus, the total number of runs saved by a catcher  $P$  for a particular game situation  $C$  is:

$$CV(P, C) = \{F(P, C) - E[F(P, C)]\} \times \{R(C') - R(C)\} + \{S(P, C) - E[S(P, C)]\} \times \{R(C'') - R(C)\} \quad (1)$$



where  $C'$  is the change to situation  $C$  from a caught stealing event, and  $C''$  is the change to situation  $C$  from a stolen base event. Revisiting the example in Table 1, we see that in 2002,  $\text{CV}(\text{J.Lopez}, C = 1) = 4.93 \times 0.62 + 3.92 \times -0.24 = 2.12$  total runs saved in that game situation. Equation (1) must be evaluated for all fifteen game situations  $C$ , giving us a total runs saved/cost of

$$\begin{aligned} \text{CV}(P) = & \sum_C \{F(P, C) - E[F(P, C)]\} \times \{R(C') - R(C)\} \\ & + \{S(P, C) - E[S(P, C)]\} \times \{R(C'') - R(C)\} \end{aligned} \quad (2)$$

which is evaluated for each player in each year.

## 2.2 Evaluation for Outfielders

The tabulation of outfielder throwing opportunities is somewhat more complicated than catchers. We want to examine all ball-in-play (BIP) events to an outfielder that also had potential baserunning consequences. For example, if a BIP event is a hit into the outfield, then it is a throwing opportunity only if there were baserunners on first and/or second base. Hits with baserunners only on third base were not included since it is assumed that any baserunner can score from third base on a hit. However, if the BIP event was an out (but not the third out), then the event can still be a throwing opportunity if there were baserunners on second and/or third base that could attempt to advance on the play. We assume that baserunners will not advance from first base on an out unless there is another throwing event involved on the same play. We categorize each outfielder throwing opportunity into a set of categories  $C$  depending on the configuration of baserunners and whether the BIP was a hit or an out ( $H = 1$  for hit,  $H = 0$  for out). In order to account for distance of the outfielder throw, the outfield surface was divided into a grid of 12 feet (X) by 10 feet (Y) zones  $Z$ , and each outfielder throwing opportunity was also categorized into a particular zone.

Within each zone  $Z$ , and for every combination of baserunner configuration  $C$  and hit vs. out  $H$ , we can tabulate the number of throwing opportunities  $N(P, Z, C, H)$  for each player  $P$ . We break these opportunities down into the number that resulted in thrown out baserunners  $S(P, Z, C, H)$  and the number that resulted in runner advancements  $F(P, Z, C, H)$ . Similar to our procedure for catchers, we can compare the actual counts for outfielder  $P$  to their expected counts based on their number of opportunities and the

league totals:

$$E[S(P, Z, C, H)] = N(P, Z, C, H) \cdot \frac{\sum_P S(P, Z, C, H)}{\sum_P N(P, Z, C, H)}$$

$$E[F(P, Z, C, H)] = N(P, Z, C, H) \cdot \frac{\sum_P F(P, Z, C, H)}{\sum_P N(P, Z, C, H)}$$

Similar to the catcher tabulation, we want to assign run values to the actions of an outfielder in each throwing opportunity, which again involves the use of the Expected Runs Matrix (REM, 2007). For each throwing opportunity, we have the starting configuration of baserunners  $C$ , which has a certain run value  $R(C)$ . If the throwing opportunity results in a thrown out baserunner, then our configuration changes to  $C'$  with run value  $R(C')$ , and the outfielder has contributed a positive run value of  $R(C) - R(C')$ . However, if the throwing opportunity results in a runner advancement, then our configuration changes to  $C''$  with run value  $R(C'')$ , then the outfielder has contributed a negative run value of  $R(C) - R(C'')$ . We can thus come up with the following total run contribution of an outfielder  $P$  relative to the average:

$$\begin{aligned} \text{OV}(P) = & \sum_{Z,C,H} \{S(P, Z, C, H) - E[S(P, Z, C, H)]\} \times \{R(C) - R(C')\} \\ & + \sum_{Z,C,H} \{F(P, Z, C, H) - E[F(P, Z, C, H)]\} \times \{R(C) - R(C'')\} \quad (3) \end{aligned}$$

This same evaluation was repeated for each player in each year.

### 3 Hierarchical Model for Multiple Seasons

Our tabulation procedure in section 2 gives us yearly throwing tabulations for 133 catchers and 500 outfielders. In addition to evaluating catchers and outfielders on a seasonal basis, we also use these yearly totals to estimate the overall throwing ability of each catcher and outfielder. We use a simple hierarchical normal model designed to shares information between players while allowing for differences in variances between players. We provide a short introduction to this model and also refer the reader to more detailed discussions in Gelman *et al.* (2003). Consider the general situation

of grouped data ie.  $Y_{ij}$  where  $j = 1, \dots, m_i$  indexes observations within group  $i$  and  $i = 1, \dots, N$  indexes the groups. We model our data as noisy observations centered at a group-specific mean  $\mu_i$  and with a group-specific variance  $\sigma_i^2$ ,

$$Y_{ij} \sim \text{Normal}(\mu_i, \sigma_i^2)$$

These group-specific means  $\mu_i$  are also modeled as coming from a normal distribution,

$$\mu_i \sim \text{Normal}(\mu_0, \tau^2)$$

The main parameters of interest are the unobserved means  $\mu_i$  for each group  $i$ , which we will refer to collectively with the vector  $\boldsymbol{\mu}$ . Inference for each  $\mu_i$  is based on a balance between the observed mean  $\bar{Y}_i = \sum_j Y_{ij}/m_i$  for that group and the population mean  $\mu_0$  across all groups. The details of that balance are determined by the within-player variance and between-player variance parameters,  $\sigma_i^2$  and  $\tau^2$ . Assuming that the parameters  $\sigma_i^2$ ,  $\tau^2$  and  $\mu_0$  are known, then the best estimate of  $\mu_i$  is:

$$\hat{\mu}_i = \frac{\frac{m_i}{\sigma_i^2} \bar{Y}_i + \frac{1}{\tau^2} \mu_0}{\frac{m_i}{\sigma_i^2} + \frac{1}{\tau^2}} \quad (4)$$

A key consequence of this model is that the estimate  $\hat{\mu}_i$  for a particular group is a compromise between the shared mean  $\mu_0$  across groups and the group-specific mean of observed data  $\bar{Y}_i$ . The number of observations  $m_i$  and amount of variance within the group  $\sigma_i^2$  controls how much the resulting estimate  $\hat{\mu}_i$  is *shrunk* towards the population mean  $\mu_0$ .

In reality, the parameters  $\sigma_i^2$ ,  $\tau^2$  and  $\mu_0$  are not known themselves. The Bayesian approach to this problem assumes prior distributions for these additional parameters. Instead of focussing on a single point estimate of these parameters (such as the maximum likelihood estimate), we want to calculate the full posterior distribution of all unknown parameters  $\Theta = (\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \tau^2, \mu_0)$  given all observed data  $\mathbf{Y}$ . Bayes rule is used to calculate this full posterior distribution:

$$p(\Theta|\mathbf{Y}) = \frac{p(\mathbf{Y}|\Theta) \cdot p(\Theta)}{p(\mathbf{Y})} \quad (5)$$

The posterior distribution provides the entire range of reasonable values for our unknown parameters, but we need to summarize this distribution in a principled way. We will use two summaries of each parameter in this paper:

- a. the posterior mean:  $\hat{\mu}_i = E(\mu_i | \mathbf{Y}_i)$  and
- b. the 95% posterior interval:  $(A, B)$  such that  $P(A \leq \mu_i \leq B) = 0.95$ .

Posterior intervals are a similar concept to confidence intervals except that under the Bayesian approach, the parameter  $\mu_i$  is considered to be a random variable taking the range of values given by the posterior interval. In contrast, under the classical approach  $\mu_i$  would be considered a fixed (but unknown) constant, and the range of values in a confidence interval refers to the coverage of this fixed constant across repeated samples. Unfortunately, the posterior distribution  $p(\Theta | \mathbf{Y})$  for this model is too complicated for posterior means and posterior intervals to be calculated analytically, and so we instead use a simulation-based approach called the Gibbs sampler to approximate the full posterior distribution  $p(\Theta | \mathbf{Y})$ . The Gibbs sampler (Geman and Geman, 1984) samples values from the full posterior distribution  $p(\Theta | \mathbf{Y})$  by iteratively sampling one parameter at a time from the conditional distribution of that parameter given the current values of all other parameters. Specific details about the Gibbs sampler for our model are given in Appendix A, and we again refer to Gelman *et al.* (2003) for a more involved discussion of Gibbs sampling and other simulation-based techniques.

We now present the actual hierarchical model used for our analysis of catcher throwing ability, and the same model is also used for outfielder throwing ability. In this application, the “groups” described above represent individual players, and the observations within each group are the seasonal arm values for a particular player. Our implemented model has the additional complication that we must also account for differences in the number of opportunities between player seasons. Let  $X_{ij}$  be the catcher run value CV for catcher  $i$  in season  $j$ , and let  $n_{ij}$  be the number of opportunities for catcher  $i$  in season  $j$ . In order to compare different catchers on the same scale of opportunities, we calculate the average number of opportunities  $\bar{n}$  in a season, and scale each run value by a factor of  $n_{ij}^* = n_{ij} / \bar{n}$ :

$$Y_{ij} = \frac{X_{ij}}{n_{ij}^*} = X_{ij} \cdot \frac{\bar{n}}{n_{ij}}$$

We model these re-scaled season run values as noisy observations from a underlying catcher-specific throwing talent  $\mu_i$ :

$$Y_{ij} \sim \text{Normal}(\mu_i, \sigma_i^2 / n_{ij}^*) \tag{6}$$

In addition to allowing player-specific variances  $\sigma_i^2$  in model (6), we are also using  $n_{ij}^*$  to account for the fact that our observations  $Y_{ij}$  should be

more precise in seasons  $j$  with greater numbers of opportunities  $n_{ij}$ . The second level of our model allows information to be shared between catchers by assuming common distributions for the catcher-specific means  $\mu$  and variances  $\sigma^2$ :

$$\mu_i \sim \text{Normal}(\mu_0, \tau^2) \quad \sigma_i^2 \sim \text{Inv-}\chi_\nu^2 \quad (7)$$

The parameters  $\mu_0$  and  $\tau^2$  capture the center and spread of the latent catcher-specific throwing abilities  $\mu_i$ . As discussed before, we will use a Bayesian approach that treats our catcher-specific latent throwing abilities  $\mu_i$  as random variables, and our inferential goal is the posterior distribution of each  $\mu_i$ . As mentioned in our introduction to the normal hierarchical model, an important consequence of our model is that our resulting estimates of  $\mu_i$  will be a compromise between the shared mean  $\mu_0$  and the catcher-specific mean  $\bar{Y}_i$  of scaled run values. The amount of *shrinkage* towards the shared mean for a particular  $\mu_i$  will be a function of the catcher-specific variance  $\sigma_i^2$  and the number of opportunities  $n_{ij}$  for that player.

To complete our model, we posit prior distributions  $\mu_0 \sim N(0, \beta)$  and  $\tau^2 \sim \text{Inv-}\chi_\gamma^2$ . Hyper-parameter  $(\beta, \nu, \gamma)$  values are used that make these prior assumptions *non-influential* on our inference, as discussed in Appendix A. In Section 4 below, we examine the results from our model implementation for catchers. We also implemented this same model for our outfielder run values, and the results are given in Section 5.

## 4 Results for Catchers

We focus our inference on the marginal posterior distribution of each  $\mu_i$ , the latent throwing ability for each catcher  $i$ . We calculated the posterior mean and 95% posterior interval of  $\mu_i$  for all 133 catchers in our dataset. However, these  $\mu_i$ 's were estimated from our scaled run values  $Y_{ij}$  that assumed the same number of opportunities for each catcher. In order to take into account differences in playing time, we also converted back to the scale of the original catcher-specific totals  $X_{ij}$  by multiplying each posterior mean and posterior interval by the average number of opportunities for that catcher. We use  $\mu_i^*$  to denote our re-scaled throwing contributions for each catcher  $i$ , which we call the player's *individual run contribution*. For comparison, we call the scaled posterior mean  $\mu_i$  the *scaled run contribution* for each player  $i$ . Our posterior means and posterior intervals for the individual run contributions of all 133 catchers are publicly available<sup>1</sup>. In Table 2, we give the

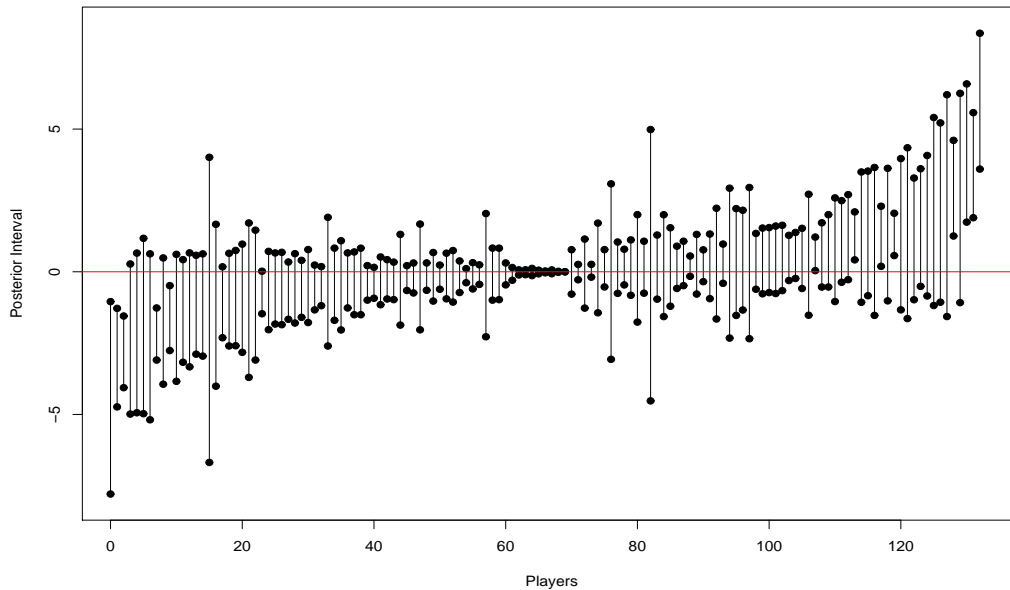
<sup>1</sup><http://whartonball.blogspot.com/2007/04/evaluating-catcher-throwing-ability.html>

Table 2: Catchers in 2002-05 with best and worst individual run contributions  $\mu_i^*$

Five Best Catchers			Five Worst Catchers		
Name	Mean	Interval	Name	Mean	Interval
Schneider, B.	6.55	(3.60, 8.36)	Piazza, M.	-4.98	(-7.79, -1.05)
Molina, Y.	4.41	(1.89, 5.58)	Varitek, J.	-3.20	(-4.74, -1.28)
Hall, T.	4.23	(1.74, 6.58)	Martinez, V.	-3.00	(-4.06, -1.55)
Ardoin, D.	3.67	(-1.08, 6.25)	Zaun, G.	-2.78	(-4.99, 0.27)
Miller, D.	3.00	(1.25, 4.61)	Fordyce, B.	-2.68	(-4.94, 0.66)

five best and five worst catchers in terms of the posterior mean of their individual run contributions  $\mu_i^*$ . We also provide the 95% posterior intervals for each of these catchers, and we observe that there is a large amount of variance. This is not unexpected, considering that there are at most four seasons of observations for each catcher. Even some of the best and worst catchers have posterior intervals that overlap with zero. In Figure 1, we plot the 95% posterior interval for all 109 catchers as a function of the posterior mean. Only 13 of 133 catchers have 95% posterior intervals that do not contain zero (indicated by the red line). We also see that players with larger magnitudes of their run contributions also tend to have wider intervals for their individual run contributions.

Figure 1: 95% Posterior intervals for each catcher ordered by the posterior mean



## 5 Results for Outfielders

Similar to our catcher analyses, we focus our outfielder inference on the *scaled run contribution*  $\mu_i$  and *individual run contribution*  $\mu_i^*$  for each outfielder, which were calculated using the same methodology presented in Section 3. Just as before, the *scaled run contribution*  $\mu_i$  are scaled to the same number of opportunities for each outfielder, whereas the *individual run contribution*  $\mu_i^*$  is re-scaled by the average number of opportunities faced by that particular outfielder  $i$ . Our posterior means and posterior intervals for the individual run contributions of all outfielders are publicly available<sup>2</sup>.

In Table 3, we give the ten best and ten worst outfielders in terms of the posterior mean of their individual run contributions  $\mu_i^*$ , along with 95% posterior intervals. We again observe a large amount of variance in the posterior intervals, and the magnitude of the run contribution for the best/worst outfielders is substantially greater than the magnitude of the

<sup>2</sup><http://whartonball.blogspot.com/2007/03/evaluating-fielder-throwing-ability.html>

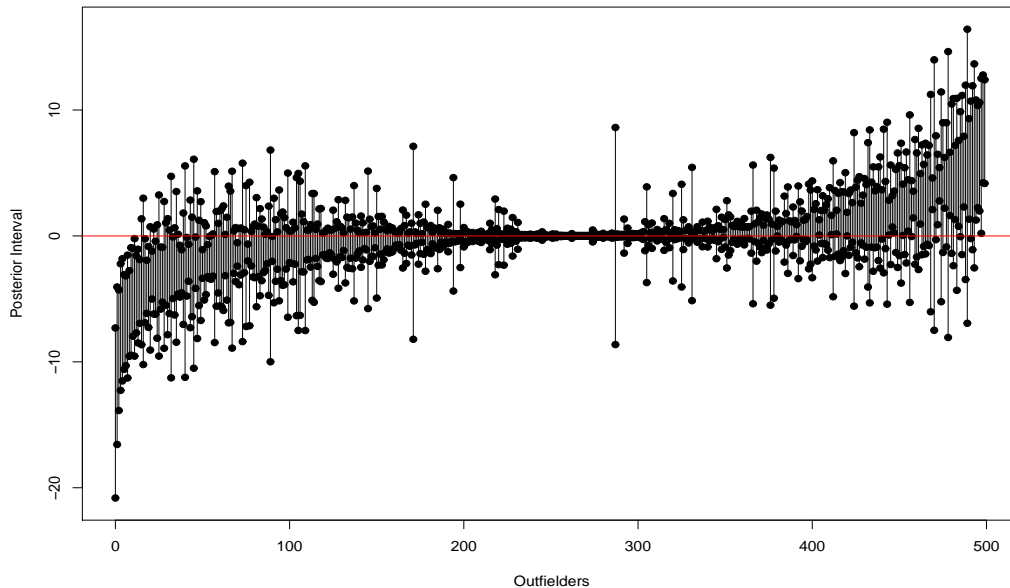
Table 3: Outfielders in 2002-05 with best and worst individual run contributions  $\mu_i^*$

Ten Best Outfielders			Ten Worst Outfielders		
Name	Mean	Interval	Name	Mean	Interval
Edmonds, J.	8.72	(4.2, 12.4)	Brown, E.	-16.78	(-20.8, -7.3)
Jones, J.	8.66	(4.2, 12.8)	Pierre, J.	-10.77	(-16.6, -4.1)
Taveras, W.	8.07	(0.2, 12.5)	Lawton, M.	-9.43	(-13.9, -4.3)
Johnson, K.	7.77	(2.00, 10.6)	Sanchez, A.	-7.79	(-12.3, -2.2)
Sullivan, C.	7.37	(2.2, 10.4)	Holliday, M.	-7.34	(-11.5, -1.8)
Chavez, E.	6.52	(1.3, 10.8)	Crawford, C.	-7.11	(-10.6, -3.1)
Guerrero, V.	6.20	(-2.5, 13.7)	DeJesus, D.	-7.00	(-10.3, -3.2)
Hidalgo, R.	6.18	(-1.1, 11.9)	Williams, B.	-6.79	(-11.3, -1.5)
Hunter, T.	5.97	(-0.2, 10.7)	White, R.	-6.56	(-9.6, -2.8)
Walker, L.	5.85	(1.3, 9.3)	Magee, W.	-5.97	(-9.5, -0.9)

run contribution for the best/worst catchers. In Figure 2, we plot the 95% posterior interval for all 500 outfielders as a function of the posterior mean. Only 60 of 500 catchers have 95% posterior intervals that do not contain zero (indicated by the red line). We again see that outfielders with larger magnitudes (highly positive or negative) of their run contributions also tend to have wider intervals for their individual run contributions.



Figure 2: 95% Posterior intervals for each outfielder ordered by the posterior mean



## 6 Discussion

In this paper, we have presented an evaluation of throwing ability for both major league catchers and outfielders. For outfielders, we focus on hits or outs on balls in play to the outfield when there are runners on base, whereas we focus our evaluation on base-stealing situations for catchers. For each player, our methodology tabulates the outcomes of their throwing successes and failures while taking into account the game situation for each opportunity. As described in Section 2, we convert the performance of each player into runs saved/cost by tabulating the change in expected runs as a consequence of each of their throwing actions. The run contribution for each player is calculated relative the average player, so a perfectly average defender would have a run contribution of zero. The magnitude of run contributions is substantially higher for outfielders compared to catchers, which is a consequence of the greater number of throwing opportunities for outfielders as well as the generally greater run consequence of those opportunities.

We use a hierarchical Bayesian model to estimate each player's innate throwing ability while accounting for differences between players in terms of variability and number of opportunities. This model acts to shrink the run contribution of players with high variability or low numbers of opportunities towards the common mean of all players. Based on this model, very few outfielders or catchers show significantly superior or inferior throwing performance, as defined by their 95% posterior interval excluding zero. The small number of statistically significant players is partly explained by the limitation of only having a maximum of four years of detailed game play data for each player. Additional seasons of data for these players would reduce their posterior variance, which would likely lead to additional players with statistically significant abilities (i.e. 95% posterior intervals excluding zero). This additional data would also permit the extension of our hierarchical model to allow the throwing ability of individual players to change over time. It would be difficult to model any time trends with our current four seasons of data, but this is a promising area of future research.

The use of the expected run matrix to evaluate the run consequence of stolen bases leads to an interesting consequence for our catcher evaluations. In terms of change in expected runs, it is much more valuable for a catcher to throw out a baserunner that is attempting to steal than it is to prevent a baserunner from attempting a stolen base. Some catchers that have a reputation for throwing out baserunners will not have as great of a run contribution because baserunners will attempt to steal less often, which is not rewarded as highly as throwing out a baserunner who does attempt a steal. An example is Ivan Rodriguez who is considered to be the best catcher in the game, as evidenced by his 12 gold gloves, the most awarded to a individual catcher in the history of the award. However, Rodriguez is ranked as only the eighth best catcher by our analyses, in part because he has one of the lowest proportion of steal attempts against him (3.47% of baserunning opportunities) among regular catchers. In Table 4, we compare Ivan Rodriguez to Brian Schneider, the top MLB catcher by our analysis. We see that Brian Schneider's overall run contribution is aided by the fact that he has a higher attempt percentage (4.54% of baserunning opportunities) relative to Rodriguez. Clearly, the optimal situation for a catcher is to have a high success rate on throwing out baserunners but without the reputation for doing so, so that baserunners still attempt to steal at a substantial rate. An extreme (and not recommended) implementation of this strategy would suggest that catchers could deliberately fail on a throwing attempt in an relatively unimportant game situation in the hopes that baserunners would then be more likely to attempt (and be thrown out) in a more important

situation.

Table 4: Comparison of Brian Schneider and Ivan Rodriguez

Name	Attempt %	Mean	Interval
Schneider, Brian	4.54 %	6.55	(3.60, 8.36)
Rodriguez, Ivan	3.47%	2.37	(-1.18, 5.41)

Considering further the issue of situational importance, a potential extension of our model would be the incorporation of some measure of leverage into the tabulation of throwing events. One could argue that catchers or outfielders should be rewarded more for making a successful throw or penalized more for failure in a key situation. It is not clear, however, whether the innate throwing ability that we are attempting to capture with our method should be dependent on the importance of the situation. The question of whether high leverage situations lead to a measurable difference in the performance of individual baseball players is a subject of ongoing speculation (eg. Tango (2004)).

## A Hierarchical Model Implementation

As outlined in Section 3, we have an observed number of opportunities  $n_{ij}$  and run value  $X_{ij}$  for catcher  $i$  in season  $j$ . Although the following model implementation is presented for our catcher evaluations, we use the same methodology for our outfielder analyses. We scale our run values  $X_{ij}$  to be on the same scale of opportunities,

$$Y_{ij} = \frac{X_{ij}}{n_{ij}^*} = X_{ij} \cdot \frac{\bar{n}}{n_{ij}}$$

and then model these scaled run values as

$$Y_{ij} \sim \text{Normal}(\mu_i, \sigma_i^2/n_{ij}^*) \quad (8)$$

where the parameters of interest are the underlying catcher-specific throwing talent  $\mu_i$ . We share information between catchers by assuming common distributions for the catcher-specific means  $\mu$  and variances  $\sigma^2$ :

$$\mu_i \sim \text{Normal}(\mu_0, \tau^2) \quad \sigma_i^2 \sim \text{Inv-}\chi_\nu^2 \quad (9)$$

Finally, we have the following prior distributions for our common mean  $\mu_0$  and variance  $\tau^2$ :

$$\mu_0 \sim \text{Normal}(0, \beta) \quad \tau^2 \sim \text{Inverse-}\chi_{\gamma}^2 \quad (10)$$

The hyper-parameters  $\nu$ ,  $\beta$ , and  $\gamma$  are assumed to be fixed and known, which means that we need to calculate the posterior distribution of our remaining unknown parameters  $\Theta = (\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \mu_0, \tau^2)$ ,

$$\begin{aligned} p(\Theta|\mathbf{Y}) &\propto \prod_{i=1}^N \prod_{j=1}^{m_i} p(Y_{ij}|\mu_i, \sigma_i^2, n_{ij}) \cdot p(\mu_i|\mu_0, \tau^2) \cdot p(\sigma_i^2|\nu) \cdot p(\mu_0|\beta) \cdot p(\tau^2|\gamma) \\ &\propto \prod_{i=1}^N \left[ (\sigma_i^2)^{-\left(\frac{m_i+\nu}{2}+1\right)} \exp \left\{ -\sum_{i=1}^N \frac{1}{2\sigma_i^2} \sum_{j=1}^{m_i} (n_{ij}^* (Y_{ij} - \mu_i)^2 + 1) \right\} \right] \\ &\times (\tau^2)^{-\left(\frac{m_i}{2}+\frac{\gamma}{2}+1\right)} \exp \left\{ \frac{-1}{2\tau^2} \left( \sum_{i=1}^N (\mu_i - \mu_0)^2 + 1 \right) + \frac{1}{2\beta} \mu_0^2 \right\} \quad (11) \end{aligned}$$

where  $N$  is the number of catchers and  $m_i$  is the number of seasons with a non-zero number of opportunities for catcher  $i$ . We will estimate this posterior distribution with a Gibbs sampling strategy (Geman and Geman, 1984) which consists of iteratively sampling from the following conditional distributions:

1.  $p(\boldsymbol{\mu}|\boldsymbol{\sigma}^2, \mu_0, \tau^2, \mathbf{Y})$
2.  $p(\boldsymbol{\sigma}^2|\boldsymbol{\mu}, \mu_0, \tau^2, \mathbf{Y})$
3.  $p(\mu_0|\boldsymbol{\sigma}^2, \boldsymbol{\mu}, \tau^2, \mathbf{Y})$
4.  $p(\tau^2|\mu_0, \boldsymbol{\sigma}^2, \boldsymbol{\mu}, \mathbf{Y})$

Step 1 can be done individually for each  $\mu_i$ . The conditional distribution of each  $\mu_i$  given the other parameters is

$$\mu_i \sim \text{Normal} \left( \frac{\frac{1}{\sigma_i^2} \sum_j^{m_i} n_{ij}^* Y_{ij} + \frac{1}{\tau^2} \mu_0}{\frac{1}{\sigma_i^2} \sum_j^{m_i} n_{ij}^* + \frac{1}{\tau^2}}, \frac{1}{\frac{1}{\sigma_i^2} \sum_j^{m_i} n_{ij}^* + \frac{1}{\tau^2}} \right) \quad (12)$$

We see that each catcher-specific throwing talent  $\mu_i$  is a weighted compromise between the observed data  $\sum_j^{m_i} n_{ij}^* Y_{ij}$  and the common mean  $\mu_0$ . The

amount of *shrinkage* towards this common mean  $\mu_0$  for a particular catcher is based on their number of opportunities and variance  $\sigma_i^2$ . Catchers with low variance  $\sigma_i^2$  will not be pulled as much towards the common mean  $\mu_0$ .

Step 2 can be done individually for each  $\sigma_i^2$ . The conditional distribution of each  $\sigma_i^2$  given the other parameters is

$$\sigma_i^2 \sim \text{InverseGamma} \left( \frac{m_i + \nu}{2}, \frac{\sum_{j=1}^{m_i} (n_{ij}^* (Y_{ij} - \mu_i)^2 + 1)}{2} \right) \quad (13)$$

From equation (13), we see that our prior distribution on  $\sigma_i^2$  are made non-influential relative to the data by letting  $\nu \rightarrow 0$ .

For step 3, we use the following conditional distribution of  $\mu_0$  given the other parameters,

$$\mu_0 \sim \text{Normal} \left( \frac{\frac{N}{\tau^2} \bar{\mu}}{\frac{N}{\tau^2} + \frac{1}{\beta}}, \frac{1}{\frac{N}{\tau^2} + \frac{1}{\beta}} \right) \quad (14)$$

From equation (14), we see that our prior distribution on  $\mu_0$  are made non-influential relative to the data by letting  $\beta \rightarrow \infty$ .

For step 4, we use the following conditional distribution of  $\tau^2$  given the other parameters,

$$\tau^2 \sim \text{InverseGamma} \left( \frac{N + \gamma}{2}, \frac{\sum_{i=1}^N (\mu_i - \mu_0)^2 + 1}{2} \right) \quad (15)$$

From equation (15), we see that our prior distribution on  $\tau^2$  are made non-influential relative to the data by letting  $\gamma \rightarrow 0$ .

## References

- BIS (2007). Baseball info solutions. *www.baseballinfosolution.com* .
- Gassko, D. (2005). Quantifying catcher defense, and other stuff like that. *The Hardball Times* **November 17, 2005**.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2003). *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL, 2nd edn.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **6**, 721–741.
- Lichtman, M. (2003). Ultimate zone rating. *The Baseball Think Factory* **March 14, 2003**.
- Pinto, D. (2006). Probabilistic models of range. *Baseball Musings* **December 11, 2006**.
- REM (2007). Run expectancy matrix. *http://www.tangotiger.net/RE9902.html* .
- Tango, T. (2004). Does clutch hitting exist? *Tangotiger.net* **February, 2004**.
- Tippett, T. (1997). Catcher throwing and pitcher hold ratings. *Diamond Mind Baseball* **November, 1997**.
- Walsh, J. (2007). Best outfield arms of 2006. *The Hardball Times* **February, 2007**.
- Woolner, K. (1999). Field general or backstop? *Baseball Prospectus* **January 10, 2000**.