Journal of Quantitative Analysis in Sports

Manuscript 1463

Estimating Fielding Ability in Baseball Players Over Time

James Piette, *University of Pennsylvania*
Shane T. Jensen, *University of Pennsylvania*

DOI: 10.1515/1559-0410.1463

Estimating Fielding Ability in Baseball Players Over Time

James Piette and Shane T. Jensen

Abstract

Quantitative evaluation of fielding ability in baseball has been an ongoing challenge for statisticians. Detailed recording of ball-in-play data in recent years has spurred the development of sophisticated fielding models. Foremost among these approaches, Jensen et al. (2009) used a hierarchical Bayesian model to estimate spatial fielding curves for individual players. These previous efforts have not addressed evolution in a player's fielding ability over time. We expand the work of Jensen et al. (2009) to model the fielding ability of individual players over multiple seasons. Several different models are implemented and compared via posterior predictive validation on hold-out data. Among our choices, we find that a model which imposes shrinkage towards an age-specific average gives the best performance. Our temporal models allow us to delineate the performance of a fielder on a season-to-season basis versus their entire career.

KEYWORDS: fielding, hierarchical Bayesian models, time series

1 Introduction

Fielding ability is an aspect of player performance in baseball that is difficult to estimate numerically. For over a century, the official method of tracking fielding performance in major league baseball has been fielding percentage: the proportion of successfully fielded balls hit into play. A ball in play (BIP) that is "unsuccessfully fielded" is called an error. It is up to the scorekeeper, a person in charge of keeping an official tally of a game, to judge every failed attempt by the fielding team as either the negligence of a specific fielder or just circumstance. The error statistic is considered to be a highly debatable and subjective measure (Kalist and Spurr, 2006; Plaschke, 1993).

All balls in play can be categorized as one of three BIP types: flyball, liner or groundball. A flyball is any ball hit into the air that takes a high trajectory (greater than 45 degrees from the ground), whereas a liner is a ball hit into the air with a low trajectory (less than 45 degrees from the ground). Groundballs, or grounders, are balls that are batted into play which begin on the ground. It has been shown that a BIP hit on the ground is fielded by a team with a much higher success rate than a ball hit in the air (Dutton and Bendix, 2008).

Over the past decade, the availability of BIP location data has improved our ability to measure fielding. One improved approach is the Ultimate Zone Rating (Lichtman, 2003) where the playing surface is divided up into zones, and within each zone, we tabulate the difference in the out rate for an individual player and the league average out rate in that zone. These differences are weighted by run expectancies for that zone and then summed over all zones to give that player's UZR value[1].

A similar fielding measure is the Plus-Minus system (Dewan, 2006) where fielders are credited with a "plus" for each successful play and penalized with a "minus" for each unsuccessful play within a zone. A fielder's plus/minus is adjusted relative to the league average within each zone and then summed over all zones to get an overall plus/minus measure. Defensive runs saved (DRS) is an extension of the Plus-Minus system that weights each plus and minus by its run value. Another fielding measure is the Probabilistic Model of Range (Pinto, 2003), where the playing surface is divided into pie slices radiating out from home plate instead of zones.

All of these advanced fielding measures are a substantial improvement over the subjective error-based metrics (Schwarz, 2006). However, these measures collectively suffer from a similar drawback: they impose an ar-

[1]There are other minor adjustments in the UZR calculation for factors such as ball speed.

bitrary discretization of the playing surface into zones. Smooth parametric curves are an alternative that have been used to model other sports, such as shot chart data in basketball (Reich *et al.*, 2006).

Jensen *et al.* (2009) introduce a spatial probit model for individual fielding performance which avoids the need for discretization into zones. Their overall method, Spatial Aggregate Fielding Evaluation (SAFE), employs a Bayesian hierarchical model which also provides variance estimates for each individual fielders ability. We briefly review this approach in Section 2.

These fielding measures are typically fit to individual players separately for each season. Substantial variability has been observed in these measures for an individual player across multiple seasons. Zimmerman (2009) and Lichtman (2010) have noted that large sample sizes are critical to getting an accurate assessment.

In this paper, we explore several hierarchical models for fielding performance of individual players *across multiple seasons*. Sharing information across several seasons allows us to reduce variability in our estimates of fielding ability for individual players, which leads to improved predictive performance on held-out data.

Player trends over time have been well-studied by baseball researchers, but the work is exclusive to batting and pitching ability. Kaplan (2008) provides a thorough treatment of available time series models on batting performance data in a frequentist setting. The variety of models tested in Kaplan (2008) include both univariate (e.g. moving average) and multivariate (e.g. vector autoregressive) versions of common time series models. Null (2009) takes a Bayesian hierarchical approach for predicting offensive abilities of baseball players.

In our temporal modeling of fielding ability, we consider three approaches: 1. a model where player ability is constant over time, 2. a model where player ability is centered around an age-specific average, and 3. an autoregressive model where player ability evolves over time. These models are compared with an in-depth study of individual players and an overall posterior predictive validation. We find that our model where player ability is centered around an age-specific average gives the generally best performance on hold-out data.

2 Original Hierarchical Model for Fielding

Fielding estimation in this paper is based upon a high-resolution data source from Baseball Info Solutions (Dewan, 2009). The data consists of all balls-in-

play (BIP) hit into the field that occurred in every MLB game across seven seasons of play (2002 - 2008). Over the seven season span, nearly 930,000 such balls in play are observed of the three BIP types. For each BIP, we have three available covariates: the (x, y) landing location, velocity[2] and BIP type (groundball, flyball or liner).

For flyballs and liners, the (x, y) landing location represents the point at which the ball either (a) landed on the field or (b) was caught by the fielder. We calculate the distance traveled by the fielder to that landing location as well as whether the player was moving forward or backward towards that BIP.

For grounders, the (x, y) coordinates correspond to the location where a fielding attempt was made in the infield. We calculate the angle traveled by the fielder towards that BIP as well as whether the player was moving left or right. The differences in derived covariates between flyballs/liners and grounders are illustrated in Figure 1.

We do not have data on where each fielder was standing before the ball was hit. We estimate each fielder's starting location as the coordinates in the field where the highest proportion of outs are made for each position, which means that the distances/angle for each BIP is also an estimated quantity. The accuracy of those estimates will be affected by shifts in fielder position for specific hitters[3]. In Section 6, we discuss the next generation of video-based data that will have measured starting positions.

The approach of Jensen *et al.* (2009) is to model the outcome of every BIP as a binary variable: whether or not that BIP resulted in an out. Let T_i be the number of seasons played by fielder i. Let n_{it} be the number of BIPs hit while player i is fielding in season t. The outcome of the play on the j^{th} BIP is denoted by S_{itj}:

$$S_{itj} = \begin{cases} 1 & \text{if } j^{th} \text{ ball hit to } i^{th} \text{ player in season } t \text{ is fielded for an out,} \\ 0 & \text{if } j^{th} \text{ ball hit to } i^{th} \text{ player in season } t \text{ is not fielded for an out.} \end{cases}$$

These observed variables are modeled as Bernoulli realizations from an underlying, event-specific probability:

$$S_{itj} \sim \text{Bernoulli}(p_{itj}). \tag{1}$$

[2]In our data, this is coded as an integer from 1 to 3, where 1 corresponds to balls hit at the lowest velocity and 3 corresponds to balls hit at the highest velocity.

[3]We also investigated using different starting locations for fielders based on left- versus right-handed batters, but found that the starting locations did not differ substantially. We used the same fielder positions for both left- and right-handed batters in this paper.

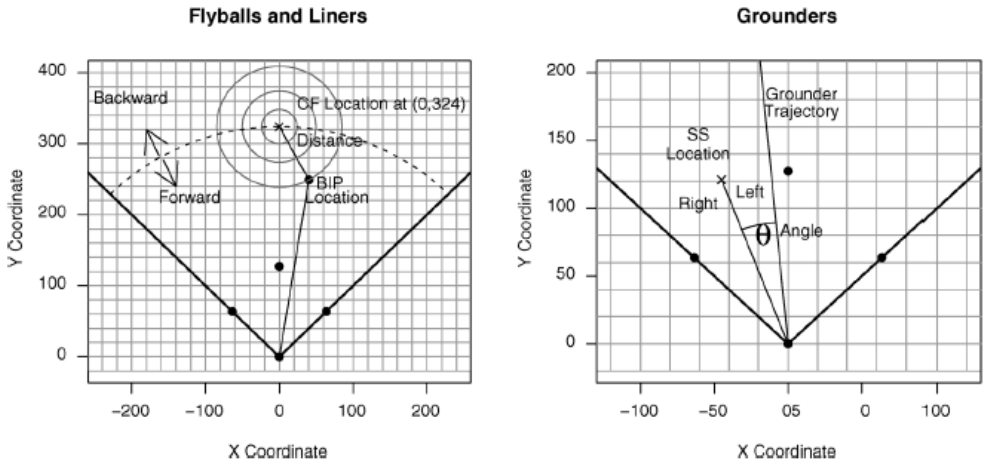

Figure 1: A visual representation behind the procedures used to calculate the distance to a flyball or liner (left) and the angle away to a grounder (right). Also shown is how moving forward on a flyball or liner is determined and how ranging to the right is found for grounders.

The BIP-specific probabilities, p_{itj}, are modeled as a probit function of covariates:

$$p_{itj} \;=\; \Phi(\boldsymbol{X}_{itj} \cdot \boldsymbol{\beta}_{it}), \tag{2}$$

where $\Phi(\cdot)$ is the cumulative distribution function for the Normal distribution and X_{itj} is the vector of covariates for BIP j. We have five different covariates related to each BIP.

For flyballs/liners, we have an intercept, the distance to the BIP from the fielder's starting position D_{itj}, the interaction between distance and velocity $D_{itj} \cdot V_{itj}$, the interaction between distance and a binary variable for whether the fielder is moving forward $D_{itj} \cdot F_{itj}$ and the interaction between all three covariates $D_{itj} \cdot V_{itj} \cdot F_{itj}$.

For grounders, we have an intercept, the angle to the BIP from the fielder's starting position A_{itj}, the interaction between angle and velocity $A_{itj} \cdot V_{itj}$, the interaction between angle and a binary variable for whether the fielder the fielder is moving to their right $A_{itj} \cdot R_{itj}$ and the interaction between all three covariates $A_{itj} \cdot V_{itj} \cdot R_{itj}$.

The coefficient vector $\boldsymbol{\beta}_{it}$ represents the fielding ability of player i in season t, and differences in $\boldsymbol{\beta}_{it}$ lead to different probabilities of making an out on a BIP between different players. These coefficients are allowed to vary between players i and across seasons t, but Jensen *et al.* (2009) posit a com-

Table 1: Valid position and BIP type combinations

	Flyballs	Grounders	Liners
1B		✓	✓
2B		✓	✓
3B		✓	✓
SS		✓	✓
LF	✓		✓
CF	✓		✓
RF	✓		✓

mon prior distribution,

$$\boldsymbol{\beta}_{it} \sim \text{Normal}(\boldsymbol{\mu}_t, \boldsymbol{\sigma_t^2}).$$

where $\boldsymbol{\sigma^2}_t$ is a 5×5 matrix with diagonal elements σ^2_{tk} and off-diagonal elements of zero. The population mean $\boldsymbol{\mu}_t$ and variance $\boldsymbol{\sigma_t^2}$ are indexed by season t since Jensen *et al.* (2009) implemented their model separately for each season.

For these population parameters, a non-informative prior distribution suggested by Gelman *et al.* (2003) is used,

$$p(\mu_{tk}, \sigma^2_{tk}) \propto \sigma_k^{-1}, \text{ for } k = 0, \ldots, 4.$$

In Appendix A, we outline the Markov Chain Monte Carlo implementation for obtaining samples of the posterior distribution of the unknown parameters $(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma^2})$ as given in Jensen *et al.* (2009).

This model is implemented separately on seven of the nine fielding positions[4]. However, some BIP types are not modeled at certain fielding positions. Table 1 provides a listing of all 14 eligible BIP and position combinations.

An important aspect of the approach of Jensen *et al.* (2009) is that the parameters of each player-season, $\boldsymbol{\beta}_{it}$, are treated separately despite the fact that we are actually observing the same players i over multiple seasons t. The fact that some of these $\boldsymbol{\beta}_{it}$ vectors represent the same player across multiple years is ignored. There is the potential to gain even more information about players i by sharing information across their observed seasons t and across players, as well as accounting for the age of players in each season. The temporal modeling of player ability is the goal of this paper, and we outline several different models in Section 3.

[4]Catchers and pitchers are not modeled since they have very few fielding chances.

3 Temporal Models for Fielding Ability

We propose three different models for sharing information within a player and across players over time. We first extend the approach of Jensen *et al.* (2009) in the simplest possible way by assuming fielding ability for each player is constant over time. We then consider a more sophisticated model where fielding ability is shrunk towards an age-specific moving average over all players. Finally, we examine an autoregressive approach where fielding ability for each player evolves over time via a state-space model.

3.1 Model 1: Constant Over Time Fielding Ability

To share information across multiple seasons by the same player, we propose an additional level to the Jensen *et al.* (2009) model in which a player's seasons are realizations from an underlying player ability that is constant over time. The observed BIP outcomes are still modeled as Bernoulli realizations from an underlying, event-specific probability:

$$S_{itj} \sim \text{Bernoulli}(p_{itj}).$$

The BIP-specific probabilities, p_{itj}, are still modeled as a probit function of covariates:

$$p_{itj} = \Phi(\boldsymbol{X}_{itj} \cdot \boldsymbol{\beta}_{it}),$$

where the $\boldsymbol{\beta}_{it}$ are season-specific coefficients for player i in season t. These season-specific parameters $\boldsymbol{\beta}_{it}$ are now drawn from a Normal distribution around underlying player abilities $\boldsymbol{\gamma}_i$,

$$\boldsymbol{\beta}_{it} \sim \text{Normal}(\boldsymbol{\gamma}_i, \boldsymbol{\tau^2}).$$

where $\boldsymbol{\tau^2}$ is the variance in season-to-season player abilities around the underlying player ability $\boldsymbol{\gamma}_i$ that does not vary over time. Note that if $\boldsymbol{\tau^2} \to \infty$, this model reduces to the original model of Jensen *et al.* (2009) where there is no sharing of information between different seasons t for each player i.

We do not consider the constant ability assumption to be particularly realistic, but it may be the case that with just seven seasons, we lack sufficient observed data to estimate a more complicated ability trajectory within each player (though we also consider an autoregressive approach in Section 3.3).

At the minimum, this constant ability model allows the sharing of information across seasons within a player. We also share information between players through a common prior distribution,

$$\boldsymbol{\gamma}_i \sim \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\sigma^2}),$$

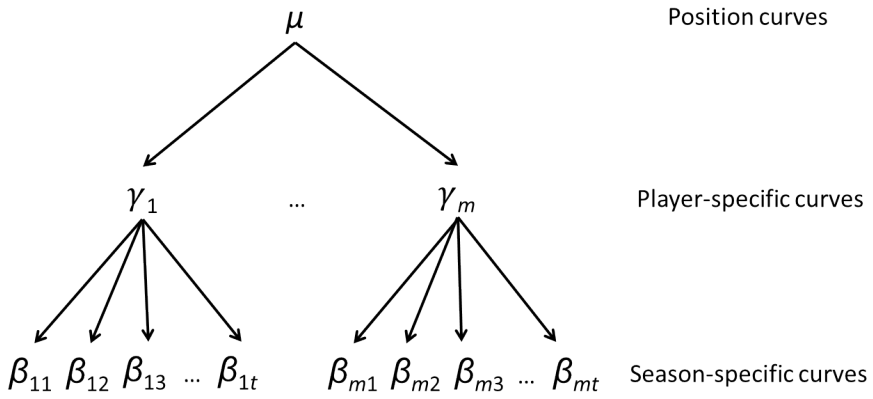

Figure 2: Hierarchy of the constant over time fielding ability model.

where we use a non-informative prior for our population parameters, $p(\mu_k, \sigma_k^2, \tau_k^2) \propto \sigma_k^{-1}\tau_k^{-1}$ for $k = 0, \ldots, 4$. In Sections 3.2 and 3.3, we consider more complicated models for sharing information across players.

We present a pictorial representation of this hierarchy in Figure 2. In Appendix B, we outline the Markov Chain Monte Carlo implementation for obtaining samples of the posterior distribution of the unknown parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{\tau^2})$.

3.2 Model 2: Age-Specific Average for Fielding Ability

One unrealistic aspect of Model 1 is that player ability remains constant over time which ignores the potential effects of aging. In our second model, player ability is allowed to change over time as a function of player age. Instead of promoting sharing between seasons within a player as in the constant over time model, this second model focuses on this time-dependent sharing across players.

In this new model, we track the age of each fielder in each of their observed seasons, replacing our season-specific parameters $\boldsymbol{\beta}_{it}$ with age-specific parameters $\boldsymbol{\beta}_{i,a_{it}}$ where a_{it} is the age of player i in season t. Otherwise, the first two levels of the model remain as before less indexing by ages a instead of seasons t.

$$\begin{aligned} S_{iaj} &\sim \text{Bernoulli}(p_{iaj}), \\ p_{iaj} &= \Phi(\boldsymbol{X}_{iaj} \cdot \boldsymbol{\beta}_{i,a_{it}}), \end{aligned}$$

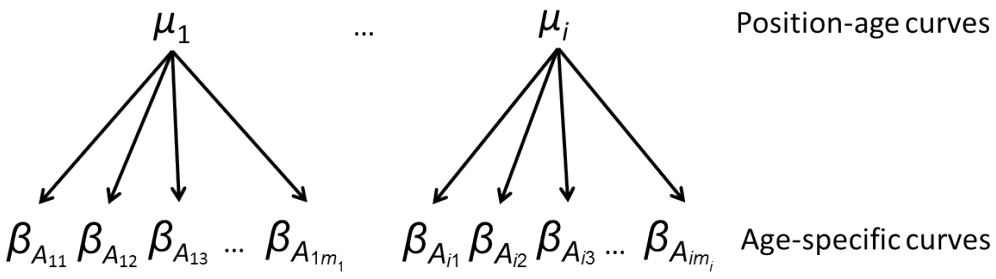

Figure 3: Hierarchy of the age-specific moving average fielding ability model

We place a Normal prior distribution on these age-specific coefficients $\boldsymbol{\beta}_{i,a_{it}}$,

$$\boldsymbol{\beta}_{i,a_{it}} \sim \text{Normal}(\boldsymbol{\mu}_a, \boldsymbol{\sigma^2})$$

with an age-specific, population average $\boldsymbol{\mu}_a$ which is shared by all fielders that have observed seasons at age a. This model extends the approach of Jensen *et al.* (2009) to factor player age a_{it} into the estimation of player ability: individual player parameters $\boldsymbol{\beta}_{i,a_{it}}$ will be shrunk towards an age-specific average $\boldsymbol{\mu}_a$ instead of an average $\boldsymbol{\mu}$ shared by players across all ages at that position, as in the original model (Section 2).

Finally, a non-informative prior is placed on the remaining unknown parameters,

$$p(\mu_{ak}, \sigma_k^2) \propto \sigma_k^{-1}, \text{ for all } k = 0, \ldots, 4.$$

We present a pictoral representation of this hierarchy in Figure 3. In Appendix C, we outline the Markov Chain Monte Carlo implementation for obtaining samples of the posterior distribution of the unknown parameters $(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma^2})$.

3.3 Model 3: Autoregressive Age Model for Fielding Ability

The average-age model in Section 3.2 has the advantage of incorporating knowledge of player age into the model. However, the average-age approach does not incorporate the previous fielding performance of each player. The constant ability model (Section 3.1) does use a player's previous performance, but does not allow their overall ability to change over time. Our final approach, an autoregressive age model, allows individual player ability to evolve over time.

We begin with our same model on our observed data,

$$\begin{aligned} S_{i,a_{it},j} &\sim \text{Bernoulli}(p_{i,a_{it},j}) \text{ and} \\ p_{i,a_{it}j} &= \Phi(\boldsymbol{X}_{i,a_{it},j} \cdot \boldsymbol{\beta}_{i,a_{it}}). \end{aligned}$$

where a_{it} is the age of player i in season t. We use the state-space approach of Carter and Kohn (1994) to model season-specific player parameters $\boldsymbol{\beta}_{i,a_{it}}$. We have an "emission" level of our model, where we again consider augmented variables $Z_{i,a_{it},j} \sim \text{Normal}(\boldsymbol{X}_{i,a_{it},j}\boldsymbol{\beta}_{i,a_{it}}\ ,\ 1)$ as the emitted variables. The player parameters $\boldsymbol{\beta}_{i,a_{it}}$ are modeled as underlying states that evolve linearly[5],

$$\begin{aligned} \text{Emission:} \qquad \boldsymbol{Z}_{i,a_{it}} &= \boldsymbol{X}_{i,a_{it}}\boldsymbol{\beta}_{i,a_{it}} + e_{i,a_{it}}, \\ \text{State Evolution:} \qquad \boldsymbol{\beta}_{i,a_{it}} &= \boldsymbol{\phi}_{a_{it}}\boldsymbol{\beta}_{i,a_{i(t-1)}} + u_{i,a_{it}}. \end{aligned}$$

The emission error variables $e_{ia_{it}}$ are standard Normals, while the evolution error variables are distributed as $\text{Normal}(0, \boldsymbol{\sigma^2})$. The $\boldsymbol{\phi}_{a_{it}}$ and $\boldsymbol{\sigma^2}$ are both diagonal matrices with diagonal terms $\boldsymbol{\phi}_a[k,k] = \phi_a$ and $\boldsymbol{\sigma^2}[k,k] = \sigma_k^2$ for $k = 1, \ldots, 5$. These $\boldsymbol{\phi}_a$ autoregressive parameters are shared across all players of the same age a, and can be interpreted as the cost/discount due to age on each player's previous season performance.

In this model, our estimates of the ability parameters $\boldsymbol{\beta}_{i,a}$ of player i at age a will be a function of their observed data at age a, as well as our estimates $\boldsymbol{\beta}_{i,a-1}$ of their ability at age $a-1$, with the discount $\boldsymbol{\phi}_a$ parameter on that previous age shared across all players at that position. Note that if $\boldsymbol{\phi}_a = 1$ for all ages a and $\boldsymbol{\sigma^2} = 0$ then this autoregressive model reduces to the constant ability model of Section 3.1.

We also define initial state distributions,

$$\text{Initial State:} \qquad \boldsymbol{\beta}_{i,a_{i1}} \sim \text{Normal}(\boldsymbol{\phi}_{a_{i1}}\boldsymbol{\alpha}_i, \boldsymbol{\tau^2}),$$

where $\boldsymbol{\alpha}_i$ and $\boldsymbol{\tau^2}$ are 5×1 vectors. Finally, we specify non-informative priors for the remaining parameters,

$$\begin{aligned} p(\phi_a, \sigma_k^2) &\propto \sigma_k^{-1} \qquad \text{for all } k = 0, \ldots, 4 \text{ and all } a \in Age, \\ p(\alpha_{ik}, \tau_k^2) &\propto \tau_k^{-1} \qquad \text{for all } k = 0, \ldots, 4, \end{aligned}$$

where Age is the set of all ages at which there is at least one fielder who played that position at that particular age.

A visualization of the autoregressive age model is provided in Figure 4. In Appendix D, we outline the Markov Chain Monte Carlo implementation

[5]Note that if a player misses an entire season a, their $a+1$ season parameters will be a linear function of their $a-1$ season parameters, but with a product of two coefficients $\boldsymbol{\phi}_a\boldsymbol{\phi}_{a+1}$.

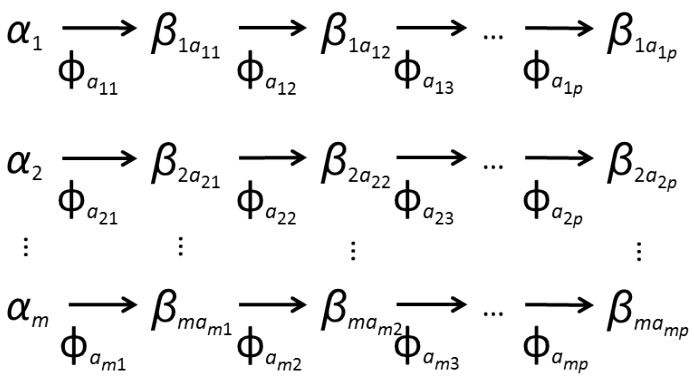

Figure 4: Evolution of parameters for m fielders as they age over time.

for obtaining samples of the posterior distribution of the unknown parameters $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\tau^2}, \boldsymbol{\sigma^2})$.

4 Posterior Predictive Internal Model Evaluation

The fit of any model to observed data can be judged by the ability of that fitted model to predict new data. For our particular data situation, we can judge how well the fielding curves produced by each model predict the individual ball-in-play outcomes on a season of data that is held out from the model fitting. In our context, each fielding curve is a posterior sample from a Bayesian model, so we refer to the predictive performance of each model as a *posterior predictive* validation of that model.

We employ a posterior predictive validation to compare the three models proposed in Section 3 to each other as well as to the original model of Jensen *et al.* (2009). We begin by splitting our ball-in-play data into a training set and a hold out testing set. Specifically, we use the 2002 - 2007 seasons for the fitting of each model, and then we test the predictive ability of these models on the 2008 season.

Our posterior predictive validation for each model consists of the following steps:

1. Sample player parameters $\boldsymbol{\beta}_{i,08}$ for the 2008 holdout season from their posterior distribution estimated using data from the 2002-2007 seasons.

2. For each BIP j to fielder i in 2008, estimate the probability of a success

p_{ij} on that BIP using the sampled $\boldsymbol{\beta}_{i,08}$,

$$\hat{p}_{ij} \;=\; \Phi(S_{ij}|X_{ij}, \boldsymbol{\beta}_{i,08})$$

3. Calculate the *predictive deviation* for player i as the average absolute differences between the predicted probability of success p_{ij} and the actual binary outcome S_{ij},

$$D_i \;=\; \frac{\sum_j |S_{ij} - \hat{p}_{ij}|}{n_i}$$

where n_i is the number of BIPs to player i in the holdout 2008 season[6].

4. Repeat steps 1-3 for $m = 1000$ different samples of $\boldsymbol{\beta}_{i,08}$, resulting in $m = 1000$ values of the predictive deviation D_i for each player i. The average of these values is the *average predictive deviation*, $\overline{D_i}$, for player i from that model.

Step 1 of this posterior predictive scheme differs depending on the particular model being considered. For the original model from Jensen *et al.* (2009), we sample $\boldsymbol{\beta}_{i,08}$ from a Normal distribution centered at the posterior values from the previous season,

$$\text{Original Model}: \qquad \boldsymbol{\beta}_{i,08} \;\sim\; \text{Normal}(\boldsymbol{\beta}_{i,07}\,,\, \boldsymbol{\sigma^2}_{07})$$

where $\boldsymbol{\beta}_{i,07}$ and $\boldsymbol{\sigma^2}_{07}$ are posterior samples generated from the Gibbs sampler outlined in Section 2, using the 2007 data.

We also consider a simpler version of the Jensen *et al.* (2009) model, where we set $\boldsymbol{\beta}_{i,08}$ equal to the maximum likelihood estimate $\text{MLE}(\boldsymbol{\beta})_{i,02:07}$ calculated by fitting the probit model in equations (1)-(2) to the 2002-2007 data.

$$\text{MLE}: \qquad \boldsymbol{\beta}_{i,08} = \text{MLE}(\boldsymbol{\beta})_{i,02:07}$$

For our first temporal model (Section 3.1), we sample $\boldsymbol{\beta}_{i,08}$ from a normal distribution centered at the underlying player ability $\boldsymbol{\gamma}_i$,

$$\text{Constant Over Time Model}: \qquad \boldsymbol{\beta}_{i,08} \;\sim\; \text{Normal}(\boldsymbol{\gamma}_i\,,\, \boldsymbol{\tau^2}).$$

where $\boldsymbol{\gamma}_i$ and $\boldsymbol{\tau^2}$ are posterior samples generated from the Gibbs sampler outlined in Section 3.1, using the 2002-2007 data.

[6]We only calculate the predicted deviation for players that had at least one BIP opportunity in the holdout 2008 season.

For our second temporal model (Section 3.2), we sample $\boldsymbol{\beta}_{i,08}$ from a normal distribution centered at the age-specific mean for the age $a_{i,08}$ for player i in 2008,

$$\text{Age Specific Average Model}: \qquad \boldsymbol{\beta}_{i,08} \sim \text{Normal}(\boldsymbol{\mu}_{a_{i,08}}, \boldsymbol{\sigma^2})$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$ are posterior samples generated from the Gibbs sampler outlined in Section 3.2, using the 2002-2007 data.

For our third temporal model (Section 3.3), we sample $\boldsymbol{\beta}_{i,08}$ from a Normal distribution centered at the previous years value $\boldsymbol{\beta}_{i,a_{i,07}}$, discounted by the age-specific autoregressive parameter $\boldsymbol{\phi}_{a_{i,08}}$

$$\text{Autoregressive Model}: \qquad \boldsymbol{\beta}_{i,08} \sim \text{Normal}(\boldsymbol{\phi}_{a_{i,08}} \cdot \boldsymbol{\beta}_{i,a_{i,07}}, \boldsymbol{\sigma^2})$$

The output of our posterior predictive procedure is a set of predicted deviations $\overline{D_i}$ for each player $i = 1, \ldots, m$ from each of the models outlined above. Clearly, we favor models that have lower values of $\overline{D_i}$ across most (if not all) players.

One intuitive way to compare predicted deviations between models is with a winning percentage $(\text{Win}\%)_L$ which we define as the weighted (by sample size) proportion of players i for which model L has the smallest $\overline{D_i}^L$ among the five models being examined, i.e.

$$(\text{Win}\%)_L = \frac{\sum_i n_{i,08} \cdot \text{I}\left[\overline{D_i}^L = \min_{l \in \mathcal{L}}(\overline{D_i}^l)\right]}{\sum_i n_{i,08}},$$

where $n_{i,08}$ is the number of BIPs for player i in the holdout 2008 season, and the five models in $\mathcal{L}$ are ("Original", "MLE", "Constant Over Time", "Age-Average" and "Autoregressive").

Figure 5 gives a mosiac plot of the $(\text{Win}\%)_L$ values for our five models across all position $\times$ BIP type combinations given in Table 1. The column widths are proportional to the number of BIPs for that position $\times$ BIP type, and within each column, a greater area for any particular color indicates better performance for the model corresponding to that color.

Red and orange represent the original Jensen *et al.* (2009) model and its simpler MLE version. The other colors represent our three temporal models from Section 3, with blue being the constant over time model, green being the age-average model (coded MAA in the figure) and purple being the autoregressive model.

The most obvious feature of Figure 5 is that there is substantial variation in performance depending on the position $\times$ BIP type combination. For fly

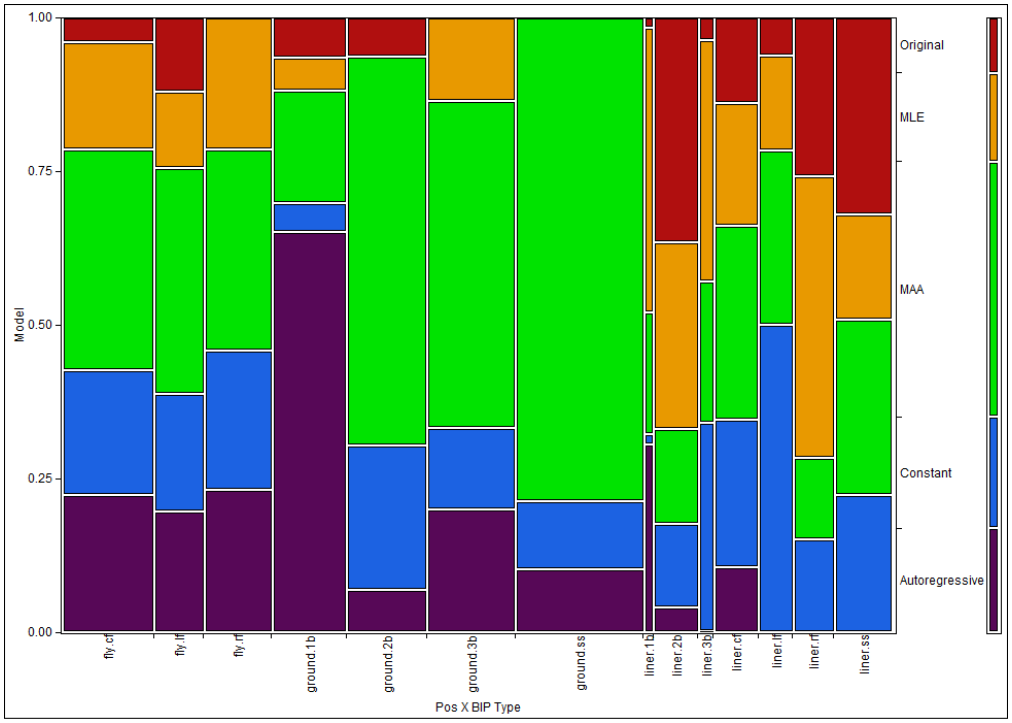

Figure 5: Mosaic Plot of $(\mathrm{Win}\%)_L$ for each BIP type $\times$ position combination. Colors correspond to different models: red = original model, orange = MLE, green = MAA (age-average) model, blue = constant over time model and purple = autoregressive model. Column width is proportional to the observed number of BIPs for that BIP type $\times$ position combination.

BIPs (columns 1-3) and grounder BIPs (columns 4-7), we see that the three temporal models introduced in Section 3 completely dominate the original Jensen *et al.* (2009) model (and its MLE version). Within those three models, the age-average model (green) from Section 3.2 seems to have the generally best performance with the exception of grounders to first basemen.

It is also interesting to note that our proposed temporal models do not dominate over the original Jensen *et al.* (2009) model for the liner BIP (at least compared to grounders and flys). Figure 5 also shows that there are far fewer observed liners for each position compared to flys and grounders. Perhaps the temporal models outlined in Section 3 are too complicated to be estimated accurately for the smaller number of liner BIPs, whereas grounders and flys have adequate sample sizes for these temporal models to dominate the original Jensen *et al.* (2009) formulation.

The temporal model that has the best overall posterior predictive performance is the age-average model from Section 3.2. We suspect that the constant-over-time (Section 3.1) and autoregressive (Section 3.3) models are hampered by the limited number of seasons of observed data for each player. In contrast, the age-average model shares information across players (rather than within a player) and so performs relatively better with a limited number of observed seasons.

5 Numerical Summaries of Individual Fielders

Each of our temporal fielding models produces player-specific parameter vectors that can be used to compare ability between players, but this comparison is difficult in a multi-dimensional parameter space. The spatial aggregate fielding evaluation (SAFE) measure of Jensen *et al.* (2009) provides a mechanism for reducing these player-specific parameters down to a single numerical estimate of fielding ability.

The first step in this process is calculating a player-specific fielding curve $p_{it}(x, y, v)$ that, for player i in year t, gives the probability that a BIP hit to (x, y) in the field with velocity v is successfully fielded. For our constant over time model (Section 3.1), we get seasonal probability curves p_{it} for each player from the probit function on seasonal parameters $\boldsymbol{\beta}_{it}$ as well as overall probability curves p_i for each player from the probit function on overall parameters $\boldsymbol{\gamma}_i$.

For our age-specific average model (Section 3.2), we get seasonal probability curves p_{it} for each player from the probit function on their age-specific parameters $\boldsymbol{\beta}_{i,a_{it}}$. For our autoregressive model (Section 3.3), we also get

seasonal probability curves p_{it} for each player from the probit function on their age-specific parameters $\boldsymbol{\beta}_{i,a_{it}}$.

For each of the temporal models in Section 3, we can also calculate an average probability curve $p_+(x, y, v)$ across all players at each position, which will be our baseline for each player's fielding performance. Following Jensen *et al.* (2009), we evaluate the curve differences $[p_{it}(x, y, v) - \hat{p_+}(x, y, v)]$ across all coordinates (x, y) and velocities v for each player i and season t under our three temporal models in Section 3.

These curve differences are incorporated into a weighted integration for each BIP type to produce an overall numerical estimate of fielding ability,

$$\begin{aligned} \text{SAFE}_{it} &= \int \hat{f}(x, y, v) \cdot \hat{r}(x, y, v) \cdot \hat{s}(x, y, v) \\ &\quad \cdot [p_{it}(x, y, v) - \hat{p_+}(x, y, v)] \, dx \, dy \, dv \end{aligned} \tag{3}$$

over all coordinates (x, y) and velocities v. This integration is weighted by the BIP frequency, $\hat{f}(x, y, v)$, estimated based on all BIPs to each coordinate (x, y) and velocity v. We also weight the integration by the run value, $\hat{r}(x, y, v)$, which we estimate by calculating the proportion of singles, doubles and triples for each coordinate (x, y) and velocity v, and then using the linear run formula of Tango *et al.* (2007).

Finally, we also weight the integration by the shared responsibility, $\hat{s}(x, y, v)$, which is estimated as the proportion of outs made by each position on BIPs to that coordinate (x, y) and velocity v. The purpose of this weighting is to avoid punishing a particular fielder completely for missing an out on a BIP that could have also been fielded by another position.

The overall SAFE_{it} can be interpreted as the runs saved (if positive) or runs cost (if negative) of player i in season t, relative to an average player at their position[7]. The SAFE_{it} value is aggregated over all valid BIP types (flys, liners, and grounders). For infielders, the SAFE value is aggregated over liners and grounders[8], whereas for outfielders the SAFE value is aggregated over liners and flys.

The SAFE_{it} for all players i across seven seasons t (2002-2008) are available at:

http://www-stat.wharton.upenn.edu/∼stjensen/research/safe.html

[7]To reflect an entire season of play, we scale each SAFE value by the average number of balls in play of that type seen in a given season.

[8]For grounders, the integration performed in (3) is over all angles θ instead of coordinates (x, y).

We provide posterior means and 95% posterior intervals for each SAFE_{it} from each of our three temporal models outlined in Section 3.

5.1 Distribution of SAFE Values By Position

We first examine the distribution of SAFE values over all players at each position in Figure 6. These SAFE values are the posterior means of SAFE_{it} from our age-specific average model (Section 3.2), which was the model with the best overall performance in our posterior predictive comparison (Section 4).

The differences in magnitude between the SAFE estimates for each position are stark. The position with highest variance appears to be shortstop, with some players saving up to 13 runs and other players costing up to 17 runs over an entire season. The other infielder position with large magnitudes is second base, whereas third and first basemen have much lower magnitudes. The outfielder positions have magnitudes that are less than SS and 2B but clearly larger than 1B.

If we consider greater magnitudes of SAFE values as evidence of a more difficult fielding position, then our results would rank the positions in approximately the same order as the Defensive Spectrum of James (2001), which ranks the positions as SS, 2B, CF, 3B, RF, LF, 1B in order of decreasing difficulty.

In terms of the overall value of fielding ability, it is apparent that batting skill is a much more crucial element to a player's repertoire than their fielding ability. Some elite batters are estimated to consistently deliver seasons of 60+ runs added above average Tango *et al.* (2007), while our results place the best fielders around 10 to 12 runs saved during their peak years. This is further substantiated by the left skew in nearly every distribution: very poor fielders can survive to play enough at each position because their batting skill greatly outweighs their shortcomings in fielding.

5.2 Comparison to External Fielding Measures

We also compared our model results to two popular external measures of fielding ability: UZR (Ultimate Zone Rating) and DRS (Defensive Runs Saved). UZR and DRS are briefly described in Section 1 and both measures are also on the scale of runs saved/cost over an entire season. The data on these two metrics is provided by the FanGraphs website (Appelman and Lichtman, 2011). For this comparison, we again use SAFE values calculated from our age-specific average model (Section 3.2).

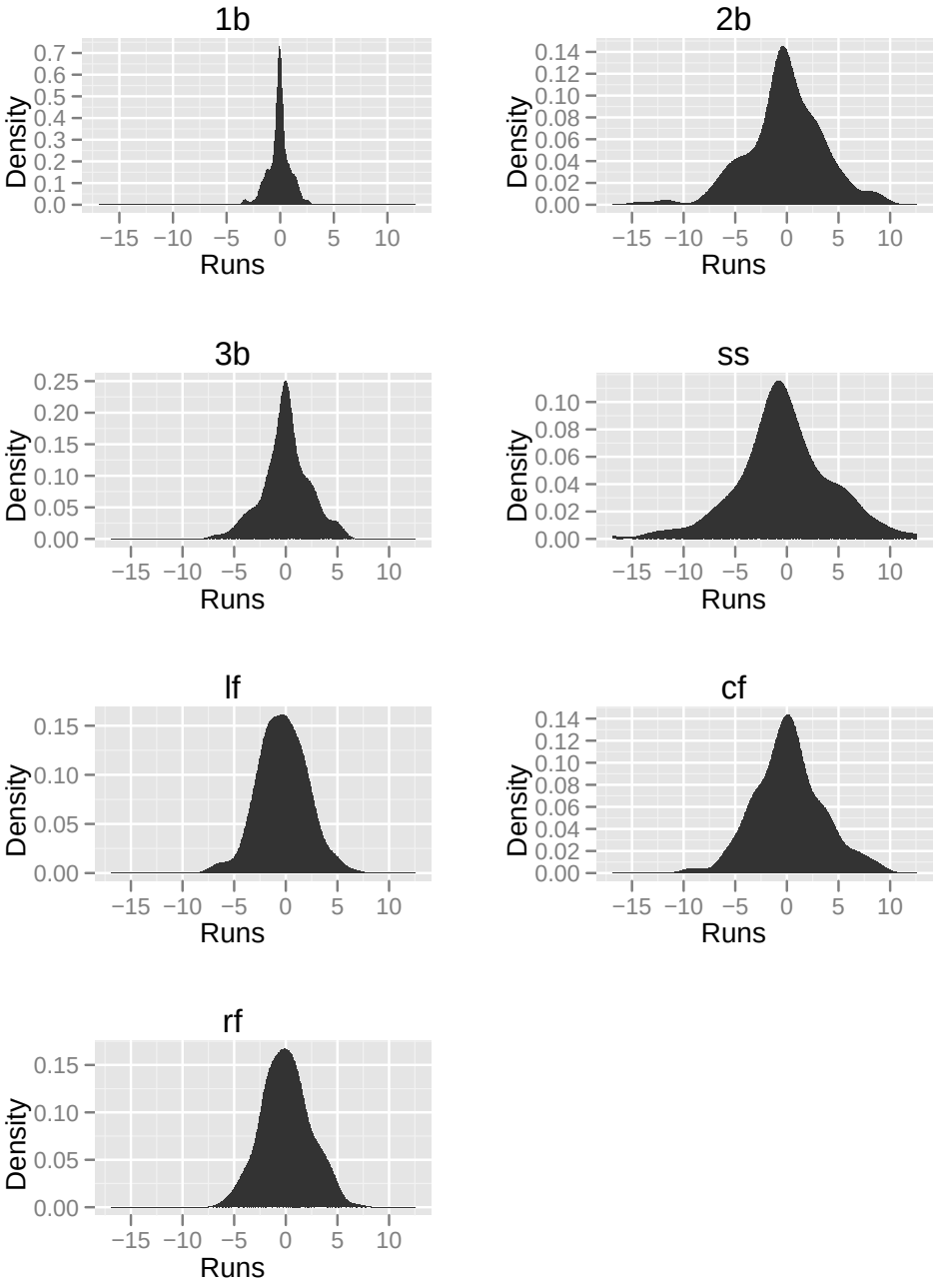

Figure 6: Distribution of SAFE posterior means for each position from our age-specific average model (Section 3.2).

Table 2: Correlations between each metric and standard deviations within each metric, all on the scale of runs saved/cost over average

Correlation	**SAFE**	**DRS**	**UZR**
SAFE	-	0.55	0.64
DRS	0.55	-	0.78
UZR	0.64	0.78	-
Standard Deviation	3.26	10.03	10.63

Correlations between each of the three metrics (SAFE, DRS and UZR) as well as standard deviations across all players are given in Table 2. Although SAFE is positively correlated with both of these external measures, it is less correlated with either UZR or DRS than these measures are correlated with each other. These correlations are also lower than the correlations between SAFE values calculated from each of our temporal models: $\hat{\rho}_{12} = 0.97, \hat{\rho}_{13} = 0.85$, and $\hat{\rho}_{23} = 0.84$ where $\hat{\rho}_{ij}$ is the correlation between model i and model j.

Another important observation is that our SAFE measure has dramatically lower standard deviations than either UZR or DRS. This result can partly be explained by the fact that our approach involves several levels of smoothing: we smooth over continuous coordinates (x, y) instead of using discrete zones, and we smooth performance over several years of individual performance with our temporal models.

Figure 7 further illustrates of the shrinkage imposed by our age-specific average model (Section 3.2). The SAFE values based on the MLE have a much larger range of values than the SAFE values based on the posterior distribution from our constant-over-time model. Although not shown, very similar shrinkage effects are observed in our constant-over-time model (Section 3.1) and autoregressive model (Section 3.3).

5.3 Case Study of Specific Players

In this subsection, we visualize the SAFE values from our three temporal models (Section 3) for one selected player at each of our seven fielding positions. Specifically, we will examine the infielders Albert Pujols (1B), Dan Uggla (2B), Derek Jeter (SS), and Adrian Beltre (3B), as well as the outfielders Adam Dunn (LF), Andruw Jones (CF) and Vladimir Guerrero (RF).

These particular players were chosen so that each position was represented by one well-known player who saw a large number of fielding op-

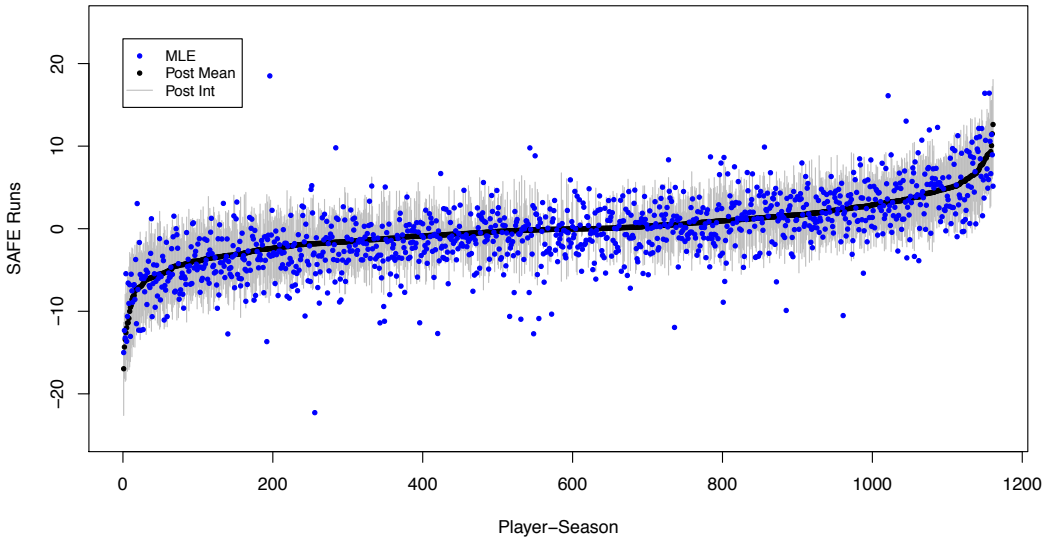

Figure 7: Comparing SAFE values based on MLE to posterior SAFE values based from age-specific average model. X-axis is all player-seasons in our observed data across all positions, ranked by the posterior SAFE value.

portunities in multiple years. Bernie Williams (CF) and Carlos Guillen (SS) were added to this case study as players who saw more limited fielding opportunities in our data.

In Figure 8, we examine the SAFE values from our constant over time model (Section 3.1). The solid black lines in figure 8 represent SAFE values based on player-specific parameters γ_i, while the gray lines correspond to season-specific parameters β_{it}. Blue lines represent SAFE values based on the MLE from the original Jensen *et al.* (2009) model.

We see that some players, such as Albert Pujols and Adrian Beltre, show very consistent season-to-season SAFE values (gray lines), and this leads to small 95% posterior intervals for their overall fielding ability (black lines). For these two players, we can conclude that they have significantly positive fielding ability (as determined by posterior intervals on overall ability that do not include zero), though the magnitudes of these values is not large.

We also see that for Albert Pujols and Adrian Beltre the model imposes shrinkage of their yearly SAFE values towards their overall career average, relative to the SAFE value based on the MLE from the original model. We did not see a systematic difference in the results for players with smaller amounts of data (Bernie Williams and Carlos Guillen) compared to players with larger amounts of data. The more important factor in the shrinkage

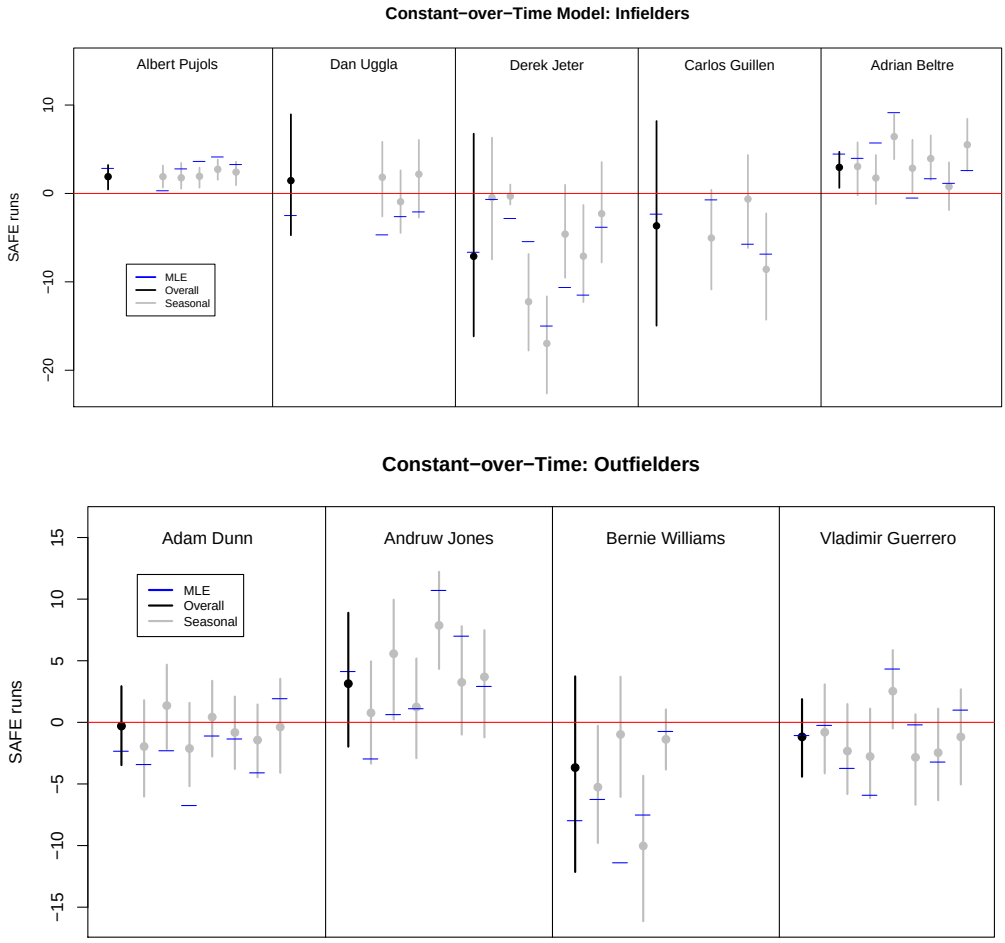

Figure 8: SAFE values for selected fielders under the constant over time model. Vertical lines represent 95% posterior intervals and dots represent posterior means. Solid black colors represent SAFE values based on player-specific parameters $\boldsymbol{\gamma}_i$. Gray color represent SAFE values based on season-specific parameters $\boldsymbol{\beta}_{it}$. Blue lines represent SAFE values based on the MLE from the original Jensen *et al.* (2009) model.

we observe in Figure 8 is the consistency in performance of specific players from year to year.

The other players in Figure 8 are much less consistent in their season-to-season SAFE values (gray lines), which leads to much more uncertainty in their overall fielding ability (black lines). For example, Derek Jeter shows generally poor performance (negative SAFE values) in most of his seasons, but the large variation in these seasonal values leads to a large posterior interval for his overall ability.

Even though the 95% posterior interval for Derek Jeter's overall ability contains zero, the large negative value for his posterior mean does still suggest that Derek Jeter was not a good fielding shortstop. A similar case on the positive side is Andruw Jones, whose overall posterior mean suggests he was a decent centerfielder, though he was not significantly above average in the sense that his 95% posterior interval does still contain zero.

In Figure 9, we examine the SAFE values from our age-specific average model (Section 3.2). We provide the SAFE values based on player-specific parameters $\boldsymbol{\beta}_{i,a_{it}}$ for each season of the seven players in our case study (red lines). For comparison, we also provide the SAFE values for the age-specific average parameters $\boldsymbol{\mu}_a$ of all players at that position (yellow lines).

We draw attention to a few specific cases where the individual player performance differed substantially from the age-specific average across all players at that position. Derek Jeter has a couple of seasons (ages 30 and 31) of poor performance relative to other shortstops at those ages, even accounting for the large variation among other players at his position. Andruw Jones has a couple of season (ages 26 and 28) of very good performance relative to other centerfielders at those ages. Although the magnitudes of their values are smaller, both Albert Pujols and Adrian Beltre have seasons where they substantially out-perform the other players at their position.

In Figure 10, we examine the SAFE values from our autoregressive model (Section 3.3). We provide the SAFE values based on player-specific parameters $\boldsymbol{\beta}_{i,a_{it}}$ for each season of the seven players in our case study (yellow lines).

For several of the players in our case study, the autoregressive model results in a smoothing of the season-to-season estimates of individual fielding ability. Specifically, for Adam Dunn and Andruw Jones, we see more autoregressive trend to their seasonal SAFE values in Figure 10 than we see in Figure 8. For other players, such as Derek Jeter, the autoregressive assumption does not seem to have much of an effect on their highly variable season-to-season values.

We also explore overall trends in fielding ability as a function of age

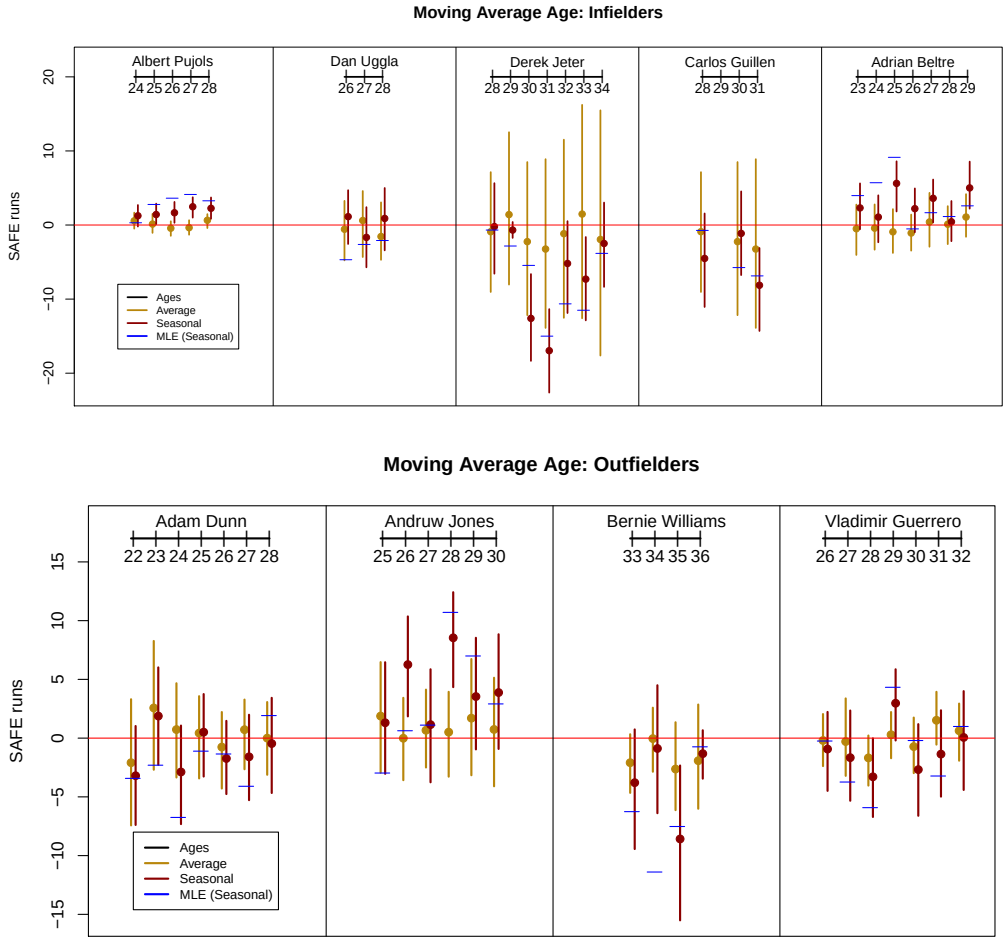

Figure 9: SAFE values for selected fielders under the age-specific average model. Vertical lines represent 95% posterior intervals and dots represent posterior means. Yellow lines correspond to the SAFE values based on age-specific average parameters $\boldsymbol{\mu}_a$, whereas red lines are the SAFE values for player-specific parameters $\boldsymbol{\beta}_{i,a_{it}}$. Blue lines represent SAFE values based on the MLE from the original Jensen *et al.* (2009) model. The numbered time line below each player's name refers to the age at which that player was during each of the seven observed seasons.

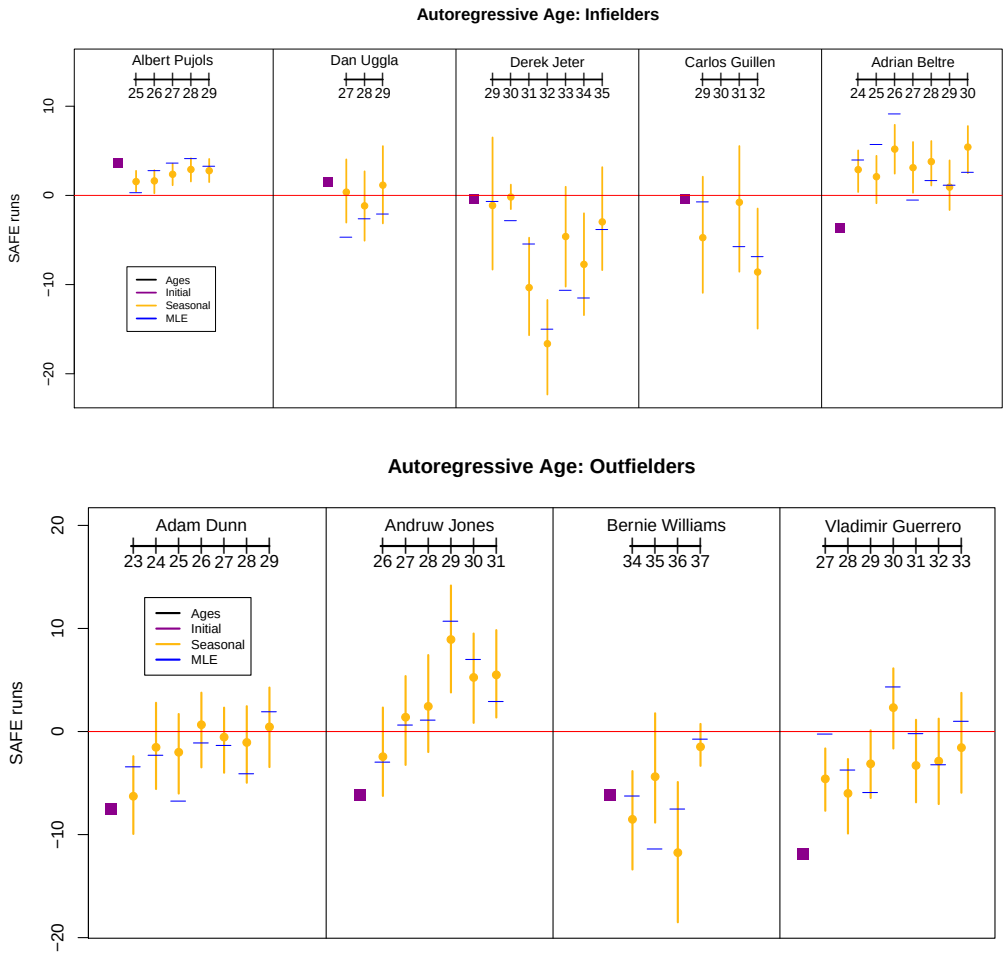

Figure 10: SAFE values for selected fielders under the autoregressive model. Vertical lines represent 95% posterior intervals and dots represent posterior means. Yellow lines correspond to the SAFE values for player-specific parameters $\boldsymbol{\beta}_{i,a_{it}}$. Purple squares represent the initial parameters $\boldsymbol{\alpha}_i$. Blue lines represent SAFE values based on the MLE from the original Jensen *et al.* (2009) model.The numbered time line below each player's name refers to the age at which that player was during each of the seven observed seasons.

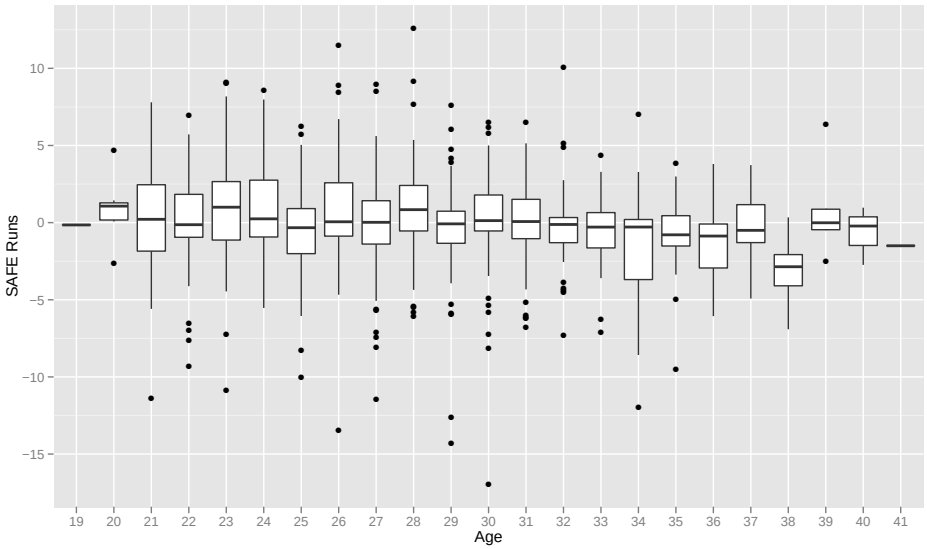

Figure 11: Distribution of SAFE posterior means from our age-specific average model for each player age

in Figure 11, where side-by-side box plots are given of the SAFE posterior means from our age-specific average model aggregating across all positions. We observe a minor decreasing trend with age in Figure 11 (with a peak fielding age of around 28) but the trend is not as dramatic as one might expect. One reason we do not see a stronger drop off in fielding performance with advanced age is that older players are usually shifted to easier fielding positions as they age.

6 Discussion

Previous statistical models for fielding ability in major league baseball have not taken advantage of the repeated observations of individual fielding performance over multiple seasons. Building off the sophisticated model-based approach of Jensen *et al.* (2009), we have presented three temporal models for measuring fielding ability over time.

The constant over time model (Section 3.1) assumes that seasonal fielding performance is an independent realization from an underlying constant ability for each player. The age-specific average model (Section 3.2) shrinks seasonal fielding performance for individual players to an age-specific average over all players at that position. The autoregressive model (Section 3.3)

uses a state-space structure to evolve fielding ability over time for individual players.

We compared these three models with an extensive posterior predictive validation on held-out data from the 2008 season. For the fly ball and grounder BIP types, our temporal models clearly dominate the previous approach of Jensen *et al.* (2009). Results are more mixed for the liner BIP type which occurs less frequently than fly balls and grounders. For liners, our temporal models are not clearly dominant over Jensen *et al.* (2009), which may suggest that larger sample sizes (such as we have for flys and grounders) are needed for our temporal models to be effective. Among our three temporal models, the age-specific average model (Section 3.2) gives the generally best predictive performance and thus would be the model we recommend to practitioners for evaluating the fielding ability of players over multiple seasons.

We used each of our temporal models to derive overall numerical estimates of ability, following the SAFE approach of Jensen *et al.* (2009). We compared these SAFE values to two external fielding metrics and also confirmed that the relative magnitudes of the different fielding positions match the defensive spectrum suggested by James (2001). Finally, we illustrated model differences by examining the results for an individual players at each position. In this case study, we see players with very consistent fielding performance over time (such as Adrian Beltre and Albert Pujols) contrasted with players that are highly variable in their performance (such as Derek Jeter).

Our models for the effects of age on fielding ability are limited by the fact that our estimates for each player is based on, at most, seven seasons of data. Future analyses will be improved by the availability of fielding data over the entire career of individual players. However, even with additional seasons of ball-in-play data, our estimation of the effects of age on fielding ability would be confounded by several factors. Survival bias is an issue, as team managers often force their players to move to other fielding positions when players begin to show signs of decline. In addition, some players who show poor fielding performance at any age might still be playing that position because of exceptional performance in other baseball skills (batting and/or base running).

Within each season, the estimation of fielding ability will be dramatically improved by the future availability of video data. The Field F/X system (Carey, 2010) will track the movement of every player on field, along with the trajectory, speed and direction of each BIP. This higher resolution data would address several issues with current fielding approaches, including

the lack of information about fielder positioning and hang time for BIPs. This higher resolution data will also aid the examination of effects for the different shapes and sizes of current ballparks.

A Jensen *et al.* (2009) Details and Implementation

In this section, we outline our Gibbs sampling (Geman and Geman, 1984) algorithm for the Jensen *et al.* (2009) model described in Section 2.

For probit models, Albert and Chib (1993) suggest augmenting our data with random variables $\boldsymbol{Z}$, where $Z_{ijt} \sim \text{Normal}(\boldsymbol{X}_{ijt} \cdot \boldsymbol{\beta}_{it}, 1)$ and observe that:

$$P(S_{ijt} = 1) \;=\; \Phi(\boldsymbol{X}_{ijt} \cdot \boldsymbol{\beta}_i) \;=\; P(Z_{ijt} \geq 0).$$

The posterior distribution for our model integrates over these augmented variables

$$p(\boldsymbol{\beta}, \boldsymbol{\mu}_t, \boldsymbol{\sigma^2} | \boldsymbol{S}, \boldsymbol{X}) \;\propto\; \int p(\boldsymbol{S}|\boldsymbol{Z}) \cdot p(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{X}) \cdot p(\boldsymbol{\beta}|\boldsymbol{\mu}_t, \boldsymbol{\sigma_t^2}) \cdot p(\boldsymbol{\mu}_t, \boldsymbol{\sigma_t^2})\, d\boldsymbol{Z},$$

where:

$$\begin{aligned}
p(\boldsymbol{S}|\boldsymbol{Z}) &= \prod_{i=1}^{m} \prod_{t=1}^{T_i} \prod_{j=1}^{n_{it}} \left(I(Y_{ijt} = 1, Z_{ijt} \geq 0) + I(Y_{ijt} = 0, Z_{ijt} \leq 0) \right), \\
p(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{X}) &\propto \prod_{i=1}^{m} \prod_{t=1}^{T_i} \prod_{j=1}^{n_{it}} \exp\left(-\frac{1}{2} (Z_{ijt} - \boldsymbol{X}_{ijt} \cdot \boldsymbol{\beta}_{it})^2 \right), \\
p(\boldsymbol{\beta}|\boldsymbol{\mu}_t, \boldsymbol{\sigma_t^2}) &\propto \prod_{k=0}^{4} \left[(\sigma_{tk}^2)^{-\frac{m}{2}} \cdot \prod_{i=1}^{m} \prod_{t=1}^{T} \exp\left(-\frac{1}{2\sigma_{tk}^2} (\beta_{itk} - \mu_{tk})^2 \right) \right], \\
p(\boldsymbol{\mu}, \boldsymbol{\sigma^2}) &\propto \prod_{k=0}^{4} (\sigma_{tk}^2)^{-\frac{1}{2}}.
\end{aligned}$$

A Gibbs sampler (Geman and Geman, 1984) is used to estimate the full posterior distribution of all unknown parameters by iteratively sampling from,

1. $p(Z_{ijt}|\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{X}_{ijt})$, for each $i = 1, \ldots, m$ and $t = 1, \ldots, T$ and $j = 1, \ldots, n_{it}$,
2. $p(\boldsymbol{\beta}_{it}|\boldsymbol{Z}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{X})$, for each $i = 1, \ldots, m$ and $t = 1, \ldots, T$,
3. $p(\boldsymbol{\mu}_t|\boldsymbol{\beta}, \boldsymbol{\sigma_t^2})$,
4. $p(\boldsymbol{\sigma_t^2}|\boldsymbol{\beta}, \boldsymbol{\mu})$.

For details on these conditional distributions, refer to Jensen *et al.* (2009). The Gibbs sampler was run for 30000 iterations from multiple starting points with convergence occurring within the first 5000 iterations. Those first 5000 iterations were discarded as burn-in and the remaining samples were thinned (taking only every 100th sample) to remove any autocorrelation between samples.

B Model 1 Details and MCMC Implementation

In this section, we outline our Gibbs sampling (Geman and Geman, 1984) algorithm for Model 1, Constant Over Time Fielding Ability, presented in Section 3.1. We implement this model by augmenting our data with random variables $\boldsymbol{Z}$ just as in Appendix A. The posterior distribution for our model integrates over these augmented variables

$$\begin{aligned} p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{\tau^2} | \boldsymbol{S}, \boldsymbol{X}) \quad \propto \quad & \int p(\boldsymbol{S}|\boldsymbol{Z}) \cdot p(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{X}) \cdot p(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\tau^2}) \cdot \\ & p(\boldsymbol{\gamma}|\boldsymbol{\mu}, \boldsymbol{\sigma^2}) \cdot p(\boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{\tau^2}) \; d\boldsymbol{Z} \end{aligned} \tag{4}$$

where:

$$\begin{aligned} p(\boldsymbol{S}|\boldsymbol{Z}) &= \prod_{i=1}^{m} \prod_{t=1}^{T_i} \prod_{j=1}^{n_{it}} \left(I(S_{itj} = 1, Z_{itj} \geq 0) + I(S_{itj} = 0, Z_{itj} \leq 0) \right), \\ p(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{X}) &\propto \prod_{i=1}^{m} \prod_{t=1}^{T_i} \prod_{j=1}^{n_{it}} \exp\left(-\frac{1}{2} (Z_{itj} - \boldsymbol{X}_{itj} \cdot \boldsymbol{\beta}_{it})^2 \right), \\ p(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\tau^2}) &\propto \prod_{k=0}^{4} (\tau_k^2)^{-\frac{T}{2}} \cdot \prod_{i=1}^{m} \prod_{t=1}^{T_i} \exp\left(-\frac{1}{2\tau_k^2} (\beta_{itk} - \gamma_{ik})^2 \right), \\ p(\boldsymbol{\gamma}|\boldsymbol{\mu}, \boldsymbol{\sigma^2}) &\propto \prod_{k=0}^{4} (\sigma_k^2)^{-\frac{m}{2}} \cdot \prod_{i=1}^{m} \exp\left(-\frac{1}{2\sigma_k^2} (\gamma_{ik} - \mu_k)^2 \right), \\ p(\boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{\tau^2}) &\propto \prod_{k=0}^{4} (\tau_k \cdot \sigma_k)^{-1}. \end{aligned}$$

We obtain samples $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}$ and $\boldsymbol{\tau^2}$ from the posterior distribution (4) by iteratively sampling from:

1. $p(Z_{itj}|\boldsymbol{\beta}, S_{itj}, \boldsymbol{X}_{itj})$ for each $i = 1, \ldots, m$ and $t = 1, \ldots, T_i$ and $j = 1, \ldots, n_{it}$

2. $p(\boldsymbol{\beta}_{it}|\boldsymbol{Z}_{it}, \boldsymbol{X}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{\tau^2})$ for each $i = 1, \ldots, m$ and $t = 1, \ldots, T_i$
3. $p(\boldsymbol{\gamma}_i|\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma^2})$ for each $i = 1, \ldots, m$
4. $p(\boldsymbol{\mu}|\boldsymbol{\gamma}, \boldsymbol{\tau^2}, \boldsymbol{\sigma^2})$
5. $p(\boldsymbol{\tau^2}|\boldsymbol{\beta}, \boldsymbol{\gamma})$
6. $p(\boldsymbol{\sigma^2}|\boldsymbol{\gamma}, \boldsymbol{\mu})$

Step 1 is a sample from a truncated normal distribution for each player i, season t and BIP j:

$$\begin{aligned} p(Z_{itj}|\boldsymbol{\beta}_{it}, \boldsymbol{S}, \boldsymbol{X}) \;\; \propto \;\; & \exp\left(-\frac{1}{2}(Z_{itj} - \boldsymbol{X}_{itj} \cdot \boldsymbol{\beta}_{it})^2\right) \\ & \cdot [\mathrm{I}(S_{itj} = 1, Z_{itj} \geq 0) + \mathrm{I}(S_{itj} = 0, Z_{itj} \leq 0)] \end{aligned}$$

Step 2 samples the season-specific parameters $\boldsymbol{\beta}_{it}$,

$$\boldsymbol{\beta}_{it} \;\sim\; \text{Normal}\left((\boldsymbol{T}^{-1} + \boldsymbol{X}_i'\boldsymbol{X}_i)^{-1} \cdot (\boldsymbol{T}^{-1}\boldsymbol{\mu} + \boldsymbol{X}_i'\boldsymbol{Z}_i), (\boldsymbol{T}^{-1} + \boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\right),$$

where $\boldsymbol{T}$ is a 5×5 matrix with diagonal elements τ_k^2 and zeroes in the off-diagonals. Step 3 samples the underlying player-specific parameters,

$$\gamma_{ik} \;\sim\; \text{Normal}\left(\frac{\overline{\beta_{ik}} \cdot \frac{T_i}{\tau_k^2} + \mu_k \cdot \frac{1}{\sigma_k^2}}{\frac{T_i}{\tau_k^2} + \frac{1}{\sigma_k^2}}, \frac{1}{\frac{T_i}{\tau_k^2} + \frac{1}{\sigma_k^2}}\right),$$

where $\overline{\beta_{ik}} = \sum_{t=1}^{T_i} \beta_{itk}/T_i$. Step 4 samples the population means μ_k as:

$$\mu_k \;\sim\; \text{Normal}\left(\overline{\gamma_k}, \frac{\sigma_k^2}{m}\right),$$

where $\overline{\gamma_k} = \sum_{i=1}^{m} \gamma_i/m$. We sample τ_k^2 as A_k^{-1} in step 5, where A_k is distributed as:

$$A_k \;\sim\; \text{Gamma}\left(\frac{T-1}{2}, \frac{\sum_{i=1}^{m}\sum_{t=1}^{T_i}(\beta_{itk} - \gamma_{ik})^2}{2}\right).$$

and in step 6, we sample σ_k^2 as B_k^{-1}, where B_k is distributed as:

$$B_k \;\sim\; \text{Gamma}\left(\frac{m-1}{2}, \frac{\sum_{i=1}^{m}(\gamma_{ik} - \mu_k)^2}{2}\right).$$

The Gibbs sampler for Model 1 was run for 60000 iterations from multiple starting points with convergence occurring within the first 10000 iterations. Those first 10000 iterations were discarded as burn-in and the remaining samples were thinned (taking only every 100th sample) to remove any autocorrelation between samples.

C Model 2 Details and MCMC Implementation

In this section, we outline our Gibbs sampling (Geman and Geman, 1984) algorithm for Model 2, Age-Specific Average for Fielding Ability, presented in Section 3.2.

We implement this model by again augmenting our data with random variables $\boldsymbol{Z}$ just as in Appendix A, except that again we index by age rather than season, so that $Z_{iaj} \sim \text{Normal}(\boldsymbol{X}_{iaj} \cdot \boldsymbol{\beta}_{i,a_{it}}, 1)$. The posterior distribution for our age-specific average model integrates over these augmented variables,

$$p(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma^2} | \boldsymbol{S}, \boldsymbol{X}) \quad \propto \quad \int p(\boldsymbol{S}|Z) \cdot p(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{X}) \cdot p(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\sigma^2}) \cdot p(\boldsymbol{\mu}, \boldsymbol{\sigma^2}) d\boldsymbol{Z}, \quad (5)$$

where $\boldsymbol{\mu}$ is the collection of $\boldsymbol{\mu}_a$ at all ages a in Age. The individual components of the posterior distribution described are:

$$\begin{aligned}
p(\boldsymbol{S}|\boldsymbol{Z}) &= \prod_{i=1}^{m} \prod_{a \in Age_i} \prod_{j=1}^{n_{ia}} \left(I(S_{iaj} = 1, Z_{iaj} \geq 0) + I(S_{iaj} = 0, Z_{iaj} \leq 0) \right), \\
p(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{X}) &\propto \prod_{i=1}^{m} \prod_{a \in Age_i} \prod_{j=1}^{n_{ia}} \exp\left(-\frac{1}{2} (Z_{iaj} - \boldsymbol{X}_{iaj} \cdot \boldsymbol{\beta}_{i,a_{it}})^2 \right), \\
p(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\sigma^2}) &\propto \prod_{k=0}^{4} (\sigma_k^2)^{-\frac{T}{2}} \cdot \prod_{i=1}^{m} \prod_{a \in Age_i} \exp\left(-\frac{1}{2\sigma_k^2} (\beta_{ia_{it}k} - \mu_{ak})^2 \right), \\
p(\boldsymbol{\mu}, \boldsymbol{\sigma^2}) &\propto \prod_{k=0}^{4} (\sigma_k)^{-1},
\end{aligned}$$

where Age_i denotes the set of ages of player i in their observed seasons.

We obtain samples $\boldsymbol{\beta}, \boldsymbol{\mu}$, and $\boldsymbol{\sigma^2}$ from the posterior distribution (5) by iteratively sampling from:

1. $p(Z_{iaj}|\boldsymbol{\beta}, S_{iaj}, \boldsymbol{X}_{iaj})$ for each $i = 1, \ldots, m$, $a \in Age_i$ and $j = 1, \ldots, n_{ia}$
2. $p(\boldsymbol{\beta}_{i,a_{it}}|\boldsymbol{Z}_{ia}, \boldsymbol{X}_{ia}, \boldsymbol{\mu}_a, \boldsymbol{\sigma^2})$ for each $i = 1, \ldots, m$ and $a_{it} \in Age_i$
3. $p(\boldsymbol{\mu}_a|\boldsymbol{\beta}_{\cdot a}, \boldsymbol{\sigma^2})$ for each age a
4. $p(\boldsymbol{\sigma^2}|\boldsymbol{\beta}, \boldsymbol{\mu})$

where $\boldsymbol{\beta}_{\cdot a}$ is the collection of age-specific parameters from all players at age a. Step 1 is a sample from a truncated Normal distribution for each player i, their age a in season t and BIP j,

$$p(Z_{iaj}|\boldsymbol{\beta}_{i,a_{it}}, \boldsymbol{S}, \boldsymbol{X}) \propto \exp\left(-\frac{1}{2}(Z_{iaj} - \boldsymbol{X}_{iaj} \cdot \boldsymbol{\beta}_{ia})^2\right) \cdot \left[\mathrm{I}(S_{iaj} = 1, Z_{iaj} \geq 0) + \mathrm{I}(S_{iaj} = 0, Z_{iaj} \leq 0)\right].$$

Step 2 samples the age-specific coefficients $\boldsymbol{\beta}_{i,a_{it}}$ for each age $a \in Age_i$,

$$\boldsymbol{\beta}_{i,a_{it}} \sim \mathrm{N}\left(((\boldsymbol{\sigma^2})^{-1} + \boldsymbol{X}'_{ia}\boldsymbol{X}_{ia})^{-1} \cdot ((\boldsymbol{\sigma^2})^{-1}\boldsymbol{\mu}_a + \boldsymbol{X}'_{ia}\boldsymbol{Z}_{ia}), ((\boldsymbol{\sigma^2})^{-1} + \boldsymbol{X}'_{ia}\boldsymbol{X}_{ia})^{-1}\right).$$

Step 3 samples the age-specific population averages $\boldsymbol{\mu}_a$,

$$\mu_{ak} \sim \text{Normal}\left(\overline{\beta_{ak}}, \frac{\sigma_k^2}{m_a}\right),$$

where $\overline{\beta_{ak}} = \sum_i \beta_{iak}/m_a$ and m_a is the number of players with observed seasons at age a. Step 4 samples the variance parameters $\boldsymbol{\sigma^2}$ by setting $\sigma_k^2 = A_k^{-1}$ where

$$A_k \sim \text{Gamma}\left(\frac{T-1}{2}, \frac{\sum_i \sum_{a \in Age_i} (\beta_{iak} - \mu_{ak})^2}{2}\right).$$

The Gibbs sampler for Model 2 was run for 100000 iterations from multiple starting points with convergence occurring within the first 20000 iterations. Those first 20000 iterations were discarded as burn-in and the remaining samples were thinned (taking only every 100th sample) to remove any autocorrelation between samples.

D Model 3 Details and MCMC Implementation

In this section, we outline our Gibbs sampling (Geman and Geman, 1984) algorithm for Model 3, Autoregressive Age Model for Fielding Ability, presented in Section 3.3. This algorithm is substantially more complicated than our previous models due to the state-space formulation.

We implement this model by again augmenting our data with random variables $\boldsymbol{Z}$ just as in Appendix A. The posterior distribution for our model integrates over these augmented variables

$$p(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\sigma^2}, \boldsymbol{\tau^2}|\boldsymbol{S}, \boldsymbol{X}) \propto \int p(\boldsymbol{S}|\boldsymbol{Z}) \cdot p(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{X}) \cdot p(\boldsymbol{\beta}|\boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\sigma^2}, \boldsymbol{\tau^2}) \cdot p(\boldsymbol{\phi}, \boldsymbol{\sigma^2}, \boldsymbol{\alpha}, \boldsymbol{\tau^2})\, d\boldsymbol{Z} \quad (6)$$

where:

$$
\begin{aligned}
p(\boldsymbol{S}|\boldsymbol{Z}) &= \prod_{i=1}^{m}\prod_{t=1}^{T_i}\prod_{j=1}^{n_{it}} \left(I(S_{itj}=1, Z_{itj}\geq 0) + I(S_{itj}=0, Z_{itj}\leq 0)\right), \\
p(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{X}) &\propto \prod_{i=1}^{m}\prod_{t=1}^{T_i}\prod_{j=1}^{n_{it}} \exp\left(-\frac{1}{2}(Z_{itj} - \boldsymbol{X}_{itj}\cdot\boldsymbol{\beta}_{it})^2\right), \\
p(\boldsymbol{\beta}|\boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\sigma^2}, \boldsymbol{\tau^2}) &\propto \prod_{k=0}^{4}(\sigma_k^2)^{-\frac{\sum_i (T_i-1)}{2}} \prod_{i=1}^{m}\prod_{t=2}^{T_i} \exp\left(-\frac{1}{2\sigma_k^2}(\beta_{ia_{it}k} - \phi_{a_{it}k}\beta_{ia_{it-1}k})^2\right) \\
&\quad \prod_{k=0}^{4}(\tau_k^2)^{-\frac{m}{2}} \prod_{i=1}^{m} \exp\left(-\frac{1}{2\tau_k^2}(\beta_{ia_{i,1}k} - \phi_{a_{it}k}\alpha_{ik})^2\right) \\
p(\boldsymbol{\phi}, \boldsymbol{\sigma^2}, \boldsymbol{\alpha}, \boldsymbol{\tau^2}) &\propto \prod_{k=0}^{4}(\tau_k\cdot\sigma_k)^{-1}.
\end{aligned}
$$

We obtain samples from the posterior distribution (6) by iteratively sampling from the following conditional distributions,

1. $p(\boldsymbol{Z}_{i,a_{it},j} \,|\, \boldsymbol{\beta}_{i,a_{it}}, \boldsymbol{S}_{i,a_{it},j}, \boldsymbol{X}_{i,a_{it},j})$ for $i = 1,\ldots,m,\ \ t = 1,\ldots,T_i$ and $j = 1,\ldots,n_{it}$
2. $p(\alpha_i \,|\, \boldsymbol{\beta}_i, \boldsymbol{\tau^2}, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, \boldsymbol{Z}, \boldsymbol{X})$ for each $i = 1,\ldots,m$
3. $p(\boldsymbol{\tau^2} \,|\, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, \boldsymbol{Z}, \boldsymbol{X})$
4. $p(\boldsymbol{\phi}_a \,|\, \boldsymbol{\beta}_a, \boldsymbol{\alpha}_a, \boldsymbol{\sigma^2})$ for each $a \in Age$
5. $p(\boldsymbol{\sigma^2} \,|\, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\phi})$
6. $p(\boldsymbol{\beta}_{i,a_{it}} | \boldsymbol{\alpha}, \boldsymbol{\tau^2}, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, \boldsymbol{Z}, \boldsymbol{X})$ for each $i = 1,\ldots,m$ and $t = 1,\ldots,T_i$

In step 1, we sample latent variables $Z_{i,a_{it},j}$ in the same manner as our previous models:

$$
\begin{aligned}
p(Z_{i,a_{it},j}|\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{X}) &\propto \exp\left(-\frac{1}{2}(Z_{i,a_{it},j} - \boldsymbol{X}_{i,a_{it},j}\cdot\boldsymbol{\beta}_{i,a_{it}})^2\right) \\
&\quad \cdot \left[\mathrm{I}(S_{i,a_{it},j}=1, Z_{i,a_{it},j}\geq 0) + \mathrm{I}(S_{i,a_{it},j}=0, Z_{i,a_{it},j}\leq 0)\right].
\end{aligned}
$$

For steps 2 and 3, there were several position $\times$ BIP-type combinations that did not have adequate sample sizes to ensure stable of the full posterior distribution of the $\boldsymbol{\alpha}$ and $\boldsymbol{\tau^2}$ parameters. Instead, we fixed the $\boldsymbol{\alpha}$ and $\boldsymbol{\tau^2}$ parameters at their maximum likelihood values.

Step 4 samples the autoregression coefficients $\boldsymbol{\phi}_a$, where $\boldsymbol{\phi}_a$ is a diagonal matrix with diagonal terms $\boldsymbol{\phi}_a[k,k] = \phi_a$ for $k = 1, \ldots, 5$. The conditional posterior distribution for ϕ_a is

$$\phi_a | \boldsymbol{\beta}_a, \boldsymbol{\alpha}_a, \sigma^2 \ \sim \ \text{Normal}\left(\frac{\sum_{k=0}^{4} E_{ak}}{\sum_{k=0}^{4} F_{ak}}, \frac{1}{\sum_{k=0}^{4} F_{ak}} \right),$$

where:

$$E_{ak} = \frac{\sum\limits_{i \in A_a} \beta_{i(a-1)k} \beta_{iak} + \sum\limits_{i \in B_a} \alpha_{ik} \beta_{iak}}{\sigma_k^2} \qquad \text{and} \qquad F_{ak} = \frac{\sum\limits_{i \in A_a} \beta_{i(a-1)k}^2 + \sum\limits_{i \in B_a} \alpha_{ik}^2}{\sigma_k^2}.$$

The set A_a is all players who played a non-debut season at age a and the set B_a is all players whose age was a during their debut season.

In step 5, we sample $\boldsymbol{\sigma^2}$, the diagonal variance matrix of the error term in the state evolution of each player's age-specific parameters $\beta_{i,a_{it}}$. We sample each diagonal element σ_k^2 from its conditional posterior distribution by setting $\sigma_k^2 = C_k^{-1}$ where

$$C_k \sim \text{Gamma}\left(\frac{T-1}{2}, \frac{1}{2} \sum_a \left[\sum_i (\beta_{iak} - \phi_a \beta_{i(a-1)k})^2 + \sum_i (\beta_{iak} - \phi_a \alpha_{ik})^2 \right] \right)$$

The final step of the Gibbs sampler uses a forward-filtering, backwards-sampling (Carter and Kohn, 1994) scheme to sample the player-specific parameters $\boldsymbol{\beta}_i = (\boldsymbol{\beta}_{i,a_{i1}}, \ldots, \boldsymbol{\beta}_{i,a_{i,T_i}})$ for each player i.

$$p(\boldsymbol{\beta}_i | \boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\theta}) \ = \ p(\boldsymbol{\beta}_{i,a_{i,T_i}} | \boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\theta}) \prod_{t=1}^{T_i - 1} p(\boldsymbol{\beta}_{i,a_{it}} | \boldsymbol{\beta}_{i,a_{i,t+1}}, \boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\theta})$$

where $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\tau^2}, \boldsymbol{\sigma^2})$ collects the other model parameters. During a forward filtering pass, we use the Kalman filter (Kalman, 1960) to calculate the conditional means $\text{E}(\boldsymbol{\beta}_{i,a_{it}} | \boldsymbol{\beta}_{i,a_{i,t+1}}, \boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\theta})$ and conditional variances $\text{Var}(\boldsymbol{\beta}_{i,a_{it}} | \boldsymbol{\beta}_{i,a_{i,t+1}}, \boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\theta})$ for these Gaussian densities. These means and variances are used during a backwards sampling step, where each $\boldsymbol{\beta}_{i,a_{it}}$ is sampled conditional on the previously sampled $\boldsymbol{\beta}_{i,a_{i,t+1}}$. The Gibbs sampler for Model 3 was run for 25000 iterations from multiple starting points with convergence occurring within the first 12500 iterations. Those first 12500 iterations were discarded as burn-in and the remaining samples were thinned (taking only every 100th sample) to remove any autocorrelation between samples.

References

Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 422, 669–679.

Appelman, D. and Lichtman, M. (2011). Fangraphs: www.fangraphs.com.

Carey, B. (2010). New camera system takes the guesswork out of baseball stats. *Popular Science.* **Feb. 12, 2010**.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81**, 3, 541–553.

Dewan, J. (2006). *The Fielding Bible*. ACTA Sports.

Dewan, J. (2009). Baseball info solutions: www.baseballinfosolutions.com.

Dutton, C. and Bendix, P. (2008). Batters and BABIP. *The Hardball Times* **Dec. 02, 2008**.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2003). *Bayesian Data Analysis, 2nd Edition*. Chapman and Hall/CRC.

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.

James, B. (2001). *The New Bill James Historical Baseball Abstract, Revised Edition*. Free Press.

Jensen, S. T., Shirley, K. E., and Wyner, A. J. (2009). Bayesball: A Bayesian hierarchical model for evaluating fielding in major league baseball. *The Annals of Applied Statistics* **3**, 2, 491–520.

Kalist, D. E. and Spurr, S. J. (2006). Baseball Errors. *Journal of Quantitative Analysis in Sports* **2**, 4.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transaction of the ASME—Journal of Basic Engineering* **82**, 35–45.

Kaplan, D. (2008). Univariate and Multivariate Autoregressive Time Series Models of Offensive Baseball Performance: 1901-2005. *Journal of Quantitative Analysis in Sports* **4**, 3.

Lichtman, M. (2003). Ultimate Zone Rating (UZR). *Baseball Think Factory.* **Mar. 3, 2003**.

Lichtman, M. (2010). The FanGraphs UZR Primer. *FanGraphs.* **May 19, 2010**.

Null, B. (2009). Modeling Baseball Player Ability with a Nested Dirichlet Distribution. *Journal of Quantitative Analysis in Sports* **5**, 2.

Pinto, D. (2003). A Probabilistic Model of Range. *Baseball Musings.* **Sep. 19, 2003**.

Plaschke, B. (1993). Hit or Error? A Question of Judgement. *Los Angeles Times* **Jul. 11, 1993**.

Reich, B. J., Hodges, J. S., Carlin, B. P., and Reich, A. M. (2006). A Spatial Analysis of Basketball Shot Chart Data. *The American Statistician* **60**, 1, 3–12.

Schwarz, A. (2006). Finally, an error-free way to measure fielding. *New York Times* **Apr. 2, 2006**.

Tango, T., Lichtman, M., Dolphin, A., and Palmer, P. (2007). *The Book: Playing the Percentages in Baseball*. Potomac Books Inc.

Zimmerman, J. (2009). UZR's (and Most Other Defensive Metrics) Limitation in Year to Year Analysis. *Beyond the Boxscore* **Aug. 15, 2009**.