10-2011

A Hierarchical Bayesian Variable Selection Approach to Major League Baseball Hitting Metrics

Blakeley B. McShane
University of Pennsylvania

Alexander Braunstein

James M. Piette III
University of Pennsylvania

Shane T. Jensen
University of Pennsylvania

Recommended Citation

McShane, B. B., Braunstein, A., Piette, J. M., & Jensen, S. T. (2011). A Hierarchical Bayesian Variable Selection Approach to Major League Baseball Hitting Metrics. *Journal of Quantitative Analysis in Sports, 7* (4), http://dx.doi.org/10.2202/1559-0410.1323

A Hierarchical Bayesian Variable Selection Approach to Major League Baseball Hitting Metrics

Abstract
Numerous statistics have been proposed to measure offensive ability in Major League Baseball. While some of these measures may offer moderate predictive power in certain situations, it is unclear which simple offensive metrics are the most reliable or consistent. We address this issue by using a hierarchical Bayesian variable selection model to determine which offensive metrics are most predictive within players across time. Our sophisticated methodology allows for full estimation of the posterior distributions for our parameters and automatically adjusts for multiple testing, providing a distinct advantage over alternative approaches. We implement our model on a set of fifty different offensive metrics and discuss our results in the context of comparison to other variable selection techniques. We find that a large number of metrics demonstrate signal. However, these metrics are (i) highly correlated with one another, (ii) can be reduced to about five without much loss of information, and (iii) these five relate to traditional notions of performance (e.g., plate discipline, power, and ability to make contact).

Keywords
baseball, hierarchical, Bayesian, mixture, model, random effects, variable selection

Disciplines
Statistics and Probability

Journal of Quantitative Analysis in Sports

Volume 7, *Issue* 4 2011 *Article* 2

A Hierarchical Bayesian Variable Selection Approach to Major League Baseball Hitting Metrics

Blakeley B. McShane, *Northwestern University*
Alexander Braunstein, *Chomp, Inc.*
James Piette, *University of Pennsylvania*
Shane T. Jensen, *University of Pennsylvania*

Recommended Citation:
McShane, Blakeley B.; Braunstein, Alexander; Piette, James; and Jensen, Shane T. (2011) "A Hierarchical Bayesian Variable Selection Approach to Major League Baseball Hitting Metrics," *Journal of Quantitative Analysis in Sports*: Vol. 7: Iss. 4, Article 2.

A Hierarchical Bayesian Variable Selection Approach to Major League Baseball Hitting Metrics

Blakeley B. McShane, Alexander Braunstein, James Piette, and Shane T. Jensen

Abstract

Numerous statistics have been proposed to measure offensive ability in Major League Baseball. While some of these measures may offer moderate predictive power in certain situations, it is unclear which simple offensive metrics are the most reliable or consistent. We address this issue by using a hierarchical Bayesian variable selection model to determine which offensive metrics are most predictive within players across time. Our sophisticated methodology allows for full estimation of the posterior distributions for our parameters and automatically adjusts for multiple testing, providing a distinct advantage over alternative approaches. We implement our model on a set of fifty different offensive metrics and discuss our results in the context of comparison to other variable selection techniques. We find that a large number of metrics demonstrate signal. However, these metrics are (i) highly correlated with one another, (ii) can be reduced to about five without much loss of information, and (iii) these five relate to traditional notions of performance (e.g., plate discipline, power, and ability to make contact).

KEYWORDS: baseball, hierarchical, Bayesian, mixture, model, random effects, variable selection

Author Notes: Blakeley B. McShane, Kellogg School of Management, Northwestern University. Alexander Braunstein, Chomp, Inc. James Piette, University of Pennsylvania. Shane T. Jensen, University of Pennsylvania.

1 Introduction

I don't understand. All of a sudden, it's not just BA and Runs Scored, it's OBA. And what is with O-P-S? - Harold Reynolds

The past decade has witnessed a dramatic increase in interest in baseball statistics, as evidenced by the popularity of the books *Moneyball* (Lewis, 2003) and *Curve Ball* (Albert and Bennett, 2003). Beyond recent public attention, the quantitative analysis of baseball continues to be active area of sophisticated research (e.g., James (2008), Kahrl et al. (2009)). Traditional statistics such as the batting average (AVG) are constantly being supplemented by more complicated modern metrics, such as the power hitting measure isolated power (ISO; Puerzer (2003)) or the base-running measure speed (SPD; James (1987)). The goal of each measure remains the same: estimation of the true ability of a player on some relevant dimension against a background of inherent randomness in outcomes. This paper will provide a statistical framework for evaluating the reliability of different offensive metrics where reliability is defined by consistency or predictive performance.

There has been substantial previous research into measures of offensive performance in baseball. Silver (2003) investigates the randomness of interseason batting average (AVG) and finds significant mean reversion among players with unusually high batting averages in individual seasons. Studeman (2007b) used several players to investigate relationships between infield fly balls, line drives, and hits. Null (2009) uses a sophisticated nested Dirichlet distribution to jointly model fourteen batter measures and finds that statistical performance is mean reverting. Baumer (2008) uses algebraic relationships to demonstrate the superiority of on-base percentage (OBP) over batting average (AVG).

Considerable interest surrounds the batting average on balls in play (BABIP). Studeman (2007a) considers four defense-independent pitching statistics (DIPS) for individual batters: walk rate, strikeout rate, home-run rate, and BABIP. These four measures form a sequence where each event is removed from the denominator of the next event (e.g., a player cannot strike out if he walks, he cannot hit a home run if he walks or strikes out, etc.). Studeman (2007b) finds that the first three measures are quite consistent whereas BABIP is quite noisy. This BABIP measure has been modified in many subsequent works. Lederer (2009) considers BABIP and groundball outs in the 2007 and 2008 seasons and concludes that handedness and position (a proxy for speed) are useful for predicting the two measures. Brown (2008) builds on this analysis by finding five factors that are predictive of BABIP and groundball outs: the ratio of pulled groundballs to opposite field groundballs, the percentage of grounders hit to center field, speed (SPD), bunt hits per plate appearance, and the ratio of home runs to fly balls.

Fair (2008) analyzes the effects of age on various offensive metrics for hitters. Kaplan (2006) decomposes several offensive statistics into both player and team level variation, and finds that player-level variation accounts for the large majority of observed variation.

Our own contribution focuses on the following question: which offensive metrics are consistent measures of some aspect of player ability? We use a hierarchical Bayesian variable selection model to partition metrics into those with predictive power versus those that are overwhelmed by noise. Scott and Berger (2006) use a similar variable selection approach to perform large-scale analysis of biological data. They provide a detailed exploration of the control of multiple testing that is provided by their hierarchical Bayesian framework, which is an advantage shared by our approach.

We implement our model on fifty offensive metrics using MCMC methods and present results for several parameters related to the within-player consistency of these offensive measures. For external validation, we compare the results of our posterior inference to those generated by various special cases of our model as well as to another popular variable selection approach, the Lasso (Tibshirani, 1996).

A large number of the fifty metrics demonstrate some degree of signal and there is considerable overlap with the results of the Lasso. We identify five metrics which stand out. These five metrics can account for much of the variation in the other forty-five. Furthermore, they are related to traditional notions of performance (e.g., plate discipline, speed, power, ability to make contact).

2 Methodology

Our goal is a model that can evaluate offensive metrics on their ability to predict the future performance of an individual player based on his past performance. A good metric is one that provides a consistent measure for that individual, so that his past performance is indicative of his future performance. A poor metric has little predictive power: one would be just as well served (or possibly better served) predicting future performance by the overall league average rather than taking into account past individual performance. We formalize this principle with a Bayesian variable selection model for separating out players that are consistently distinct from the overall population on each offensive measure. In addition to providing individual-specific inferences, our model also provides global measures of the signal in each offensive measure.

Our data comes from the Appelman (2009) database. We have fifty available offense metrics which are outlined in Appendix A. The data contains 8,596 player-seasons from 1,575 unique players spanning the 1974-2008 seasons (data for ten of

DOI: 10.2202/1559-0410.1323

the fifty offensive metrics were not available before the 2002 season so for those metrics we fit our model on 1,935 player-seasons from 585 unique players[1]).

2.1 Hierarchical Bayesian Variable Selection Model

For a particular offensive metric, we let y_{ij} denote the metric value for player i during season j. We model each metric independently, and, in particular, the player-seasons y_{ij} for each player's performance on a given metric are modeled as following a normal distribution with underlying individual player mean $(\mu + \alpha_i)$ and individual player-season variance $w_{ij} \cdot \sigma^2$,

$$y_{ij} \sim \text{Normal}(\mu + \alpha_i\,, w_{ij} \cdot \sigma^2). \tag{1}$$

The parameter μ denotes the overall population mean (i.e., the Major League Baseball mean) for the given offensive metric and the α_i denote the player-specific differences from the population mean μ.

The weight term w_{ij} addresses the fact that the variance of a season-level offensive metric for player i in season j should depend on player i's number of opportunities in season j. For metrics which are rates (e.g., on-base percentage (OBP), batting average (AVG)), the player-seasons with more opportunities should have a lower variance while, for metrics which are totals (e.g., homeruns (HR), hits (H)), the variance should be higher. In order to achieve this behavior, we set $w_{ij} = \bar{u}/u_{ij}$ for rates and $w_{ij} = u_{ij}/\bar{u}$ for totals where u_{ij} denotes the weight function for player i in season j and $\bar{u}$ represents the mean weight over all player-seasons. The raw weights u_{ij} used for each offensive metric are given in Appendix A.

With this formulation, the parameter σ^2 represents the global variance of the offensive metric for player-seasons with an average number of opportunities. The global parameters μ and σ^2 are unknown and are given the following prior distributions,

$$\mu \sim \text{Normal}(0, K^2) \qquad \sigma^2 \sim \text{Inverse} - \text{Gamma}(\alpha_0, \beta_0). \tag{2}$$

We tried several settings for the hyperparameters K^2, α_0, and β_0 to insure our posterior inferences were not sensitive to the values chosen and settled on $K^2 = 10000$, $\alpha_0 = .01$, $\beta_0 = .01$ as non-informative choices for these prior distributions.

We also need to address our unknown player-specific parameters α_i. We could employ a conventional Bayesian random effects model which utilizes a Normal prior distribution shared by all α_i parameters. Instead, we propose a more sophisticated model for the unknown individual α_i's; our strategy builds on Bayesian

[1]These metrics are BUH, BUH/H, FB/BIP, GB/BIP, GB/FB, HR/FB, IFFB/FB, IFH, IFH/H, and LD/BIP.

variable selection methodologies (George and McCulloch, 1997) and allows differentiation between players who are consistently different from the population mean versus players who are not.

We formulate our sample of players as a mixture of (i) "zeroed" players for whom $\alpha_i = 0$ versus (ii) "non-zeroed" players for whom $\alpha_i \neq 0$. We use the binary variable γ_i to denote the unknown group membership of each player i (i.e., $\gamma_i = 0 \Leftrightarrow \alpha_i = 0$; $\gamma_i = 1 \Leftrightarrow \alpha_i \neq 0$). We denote by p_1 the unknown proportion of players that are in the non-zeroed group ($\gamma_i = 1$) and use the prior distribution $\alpha_i \sim \text{Normal}(0, \tau^2)$ for them. For the players in the zeroed group, we have a point-mass at $\alpha_i = 0$. The variance parameter τ^2 represents the differences among individual players who themselves differ from the overall league mean. When τ^2 is large (particularly in relation to σ^2), this means that there can be a potentially wide gulf between zeroed and non-zeroed players.

George and McCulloch (1997) demonstrate that using a pure point-mass for a mixture component complicates model implementation. They suggest approximating the point-mass with a second normal distribution that has a much smaller variance, $v_0 \cdot \tau^2$, where v_0 is a hyperparameter set to be quite small. In our model implementation, we set $v_0 = 0.01$, meaning that the zeroed component has 1/100th of the variance of the non-zeroed component. Thus, our mixture model on the player-specific parameters is

$$\alpha_i \sim \begin{cases} \text{Normal}(0\,,\,\tau^2) & \text{if } \gamma_i = 1 \\ \text{Normal}(0\,,\,v_0 \cdot \tau^2) & \text{if } \gamma_i = 0. \end{cases} \tag{3}$$

We illustrate this mixture in Figure 1.

The last two parameters of our model are τ^2 and p_1. We give the following prior distribution to τ^2,

$$\tau^2 \sim \text{Inverse} - \text{Gamma}(\psi_0, \delta_0). \tag{4}$$

Gelman (2006) cautions that the inverse-Gamma family, when used as a prior on the group-level variance (i.e., the variance τ^2 of the player coefficients α_i in our setting), can sometimes be surprisingly informative even when ψ_0 and δ_0 are set to low values. Instead, he suggests a uniform prior on τ,

$$p(\tau) \propto 1 \quad \Rightarrow \quad p(\tau^2) \propto 1/\tau \tag{5}$$

which we implement by setting $\psi = -1/2$ and $\delta_0 = 0$ in Equation 4 thereby preserving conjugacy.

DOI: 10.2202/1559-0410.1323

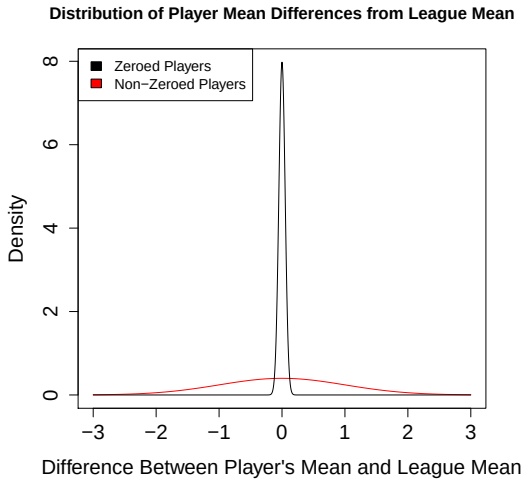

Figure 1: Illustration of mixture for player-specific parameters α_i. The black curve approximates a point mass at zero with a normal component that has a very small variance relative to the normal component for the non-zeroed players.

Finally, we allow the mixing proportion parameter p_1 to be unknown with prior distribution

$$p_1 \sim \text{Uniform}(0,1). \tag{6}$$

As discussed by Scott and Berger (2006), allowing p_1 to be estimated by the data provides an automatic control for multiple comparisons, an important advantage of our Bayesian methodology. Alternative approaches such as standard regression testing of individual means would require an additional adjustment for the large number of tests (i.e., 1,575 players) being performed.

The mixing proportion p_1 is also an important model parameter for evaluating the overall reliability of an offensive metric as it gives the probability that a randomly chosen player shows consistent differences from the population mean. Therefore, metrics with high signal should have a high p_1.

2.2 MCMC Implementation

Let $\mathbf{y}$ be the vector of all player-seasons y_{ij} for a given offensive metric. Similarly, let $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ denote the vectors of all α_i's and all γ_i's respectively. The use of conjugate prior distributions outlined in Section 2 allows us to implement our model with

a Gibbs sampler (Geman and Geman (1984)) where each step has a nice analytic form. Specifically, we iteratively sample from the following conditional distributions of each set of parameters given the current values of the other parameters.

1. Sampling μ from $p(\mu|\boldsymbol{\alpha},\sigma^2,\mathbf{y})$:
Letting i index players and j seasons within a player, the conditional distribution for μ is

$$\mu|\boldsymbol{\alpha},\sigma^2,\mathbf{y} \sim \text{Normal}\left(\frac{\sum_{i,j}\frac{y_{ij}-\alpha_i}{w_{ij}\cdot\sigma^2}}{\sum_{i,j}\frac{1}{w_{ij}\cdot\sigma^2}+\frac{1}{K^2}}, \frac{1}{\sum_{i,j}\frac{1}{w_{ij}\cdot\sigma^2}+\frac{1}{K^2}}\right).$$

2. Sampling $\boldsymbol{\alpha}$ from $p(\boldsymbol{\alpha}|\mu,\boldsymbol{\gamma},\sigma^2,\tau^2,\mathbf{y})$:
Again letting i index players and j seasons within a player, the conditional distribution for each α_i is

$$\alpha_i|\mu,\gamma_i,\sigma^2,\tau^2,\mathbf{y} \sim \text{Normal}\left(\frac{\sum_{j}\frac{y_{ij}-\mu}{w_{ij}\cdot\sigma^2}}{\sum_{j}\frac{1}{w_{ij}\cdot\sigma^2}+\frac{1}{\tau_i^2}}, \frac{1}{\sum_{j}\frac{1}{w_{ij}\cdot\sigma^2}+\frac{1}{\tau_i^2}}\right)$$

where $\tau_i^2=\tau^2$ if $\gamma_i=1$ or $\tau_i^2=v_0\cdot\tau^2$ if $\gamma_i=0$.

3. Sampling σ^2 from $p(\sigma^2|\mu,\boldsymbol{\alpha},\mathbf{y})$:
Letting N be the total number of observed player-seasons, the conditional distribution for σ^2 is

$$\sigma^2|\mu,\boldsymbol{\alpha},\mathbf{y} \sim \text{Inv}-\text{Gamma}\left(\alpha_0+\frac{N}{2}, \beta_0+\sum_{i,j}\frac{(y_{ij}-\alpha_i-\mu)^2}{2\cdot w_{i,j}}\right).$$

4. Sampling τ^2 from $p(\tau^2|\boldsymbol{\alpha})$:
Letting m be the number of players, the conditional distribution for τ^2 is

$$\tau^2|\boldsymbol{\alpha} \sim \text{Inv}-\text{Gamma}\left(\psi_0+\frac{m}{2}, \delta_0+\sum_i\frac{\alpha_i^2}{2\cdot v_i}\right)$$

where $v_i=1$ when $\gamma_i=1$ and $v_i=v_0$ if $\gamma_i=0$.

DOI: 10.2202/1559-0410.1323

5. Sampling $\boldsymbol{\gamma}$ from $p(\boldsymbol{\gamma}|\boldsymbol{\alpha}, \tau^2, p_1)$:
Again letting i index players, the conditional distribution of each γ_i is a Bernoulli draw with probability

$$p(\gamma_i = 1|\alpha_i, \tau^2, p_1) = \frac{p_1 \cdot \exp\left(-\frac{\alpha_i^2}{2\tau^2}\right)}{\frac{(1-p_1)}{\sqrt{v_0}} \cdot \exp\left(-\frac{\alpha_i^2}{2v_0\tau^2}\right) + p_1 \cdot \exp\left(-\frac{\alpha_i^2}{2\tau^2}\right)}.$$

6. Sampling p_1 from $p(p_1|\boldsymbol{\gamma})$:
Finally, the mixing proportion p_1 has the conditional distribution

$$p_1|\boldsymbol{\gamma} \sim \text{Beta}\left(1 + \sum_i \gamma_i\,,\, 1 + \sum_i (1-\gamma_i)\right).$$

Sampling Scheme:
We independently run our model and Gibbs sampler for each of our fifty offensive metrics. In particular, we run each Gibbs sampler for 60,000 iterations and discard the first 10,000 iterations as burn-in. The remainder of the chain is thinned to retain every 50th iteration in order to eliminate autocorrelation of the sampled values. We present our results from our estimated posterior distributions in Section 3.

2.3 Submodels, Identifiability, and Signal Assessment

In Section 4.1, we consider three submodels (i.e., special cases of the main model outlined in Sections 2.1 and 2.2) for the purpose of validation. While one often employs model selection criteria such as the Deviance Information Criterion (Spiegelhalter et al., 2002) or Bayes Factors (Kass and Raftery, 1995) for this purpose, such criteria prove intractable here (the discrete γ_i and improper prior on τ^2 rule out DIC and Bayes Factors respectively). Instead, we opt to compare models using a holdout sample (see Section 4.1 for details).

The first special case of our model is the standard Bayesian random effects mentioned above. Unlike our mixture prior on the α_i, the Bayesian random effects model places a single Normal prior an all α_i. While this is the standard way to think of the random effects model, there is another way to think of it that is more natural in the context of our mixture model: the Bayesian random effects model is the special case of our model when p_1 is fixed at one. It is a partial pooling model where each player's mean $\mu + \alpha_i$ is estimated as a weighted average of his observed average and overall league average.

The other two special cases of our model are also subcases of the Bayesian random effects model: the "no pooling" model and the "complete pooling" model (Gelman and Hill, 2006). The no pooling model estimates each player's mean $\mu + \alpha_i$ by his observed mean. It is thus the special case of our model that sets p_1 to one and τ^2 to ∞. The complete pooling model, on the other hand, estimates each player to have the same mean. That is, α_i is set to zero for all i. There are two ways to obtain the complete pooling model from ours: (i) fix p_1 at zero or (ii) fix p_1 at one and τ^2 at zero.

The fact that the complete pooling model can be derived from our model in two ways seems to present a problem: our model is not identifiable. In fact, since we have suggested that p_1 is an important model parameter for evaluating the overall reliability of an offensive metric (because it gives the probability that a randomly chosen player shows consistent differences from the population mean), the problem appears even more grave because the complete pooling model can be obtained from the main model by setting p_1 to its extreme values of zero or one. The former would suggest a metric lacks signal whereas the latter suggests it has high signal.

In fact, this problem goes beyond identifiability. The likelihoods of the main model and various submodels will be very similar when τ^2 is small relative to σ^2, regardless of the value of p_1. Thus, they will result in nearly identical estimates of player performance.

Hence, having a large fraction of players differ from the overall league mean (i.e., having a high p_1) is a *necessary* but *not sufficient* condition for a metric to be high signal. If a metric is going to be useful to managers, these differences have to be meaningful, that is practically significant and not merely statistically significant. This will be the case when τ^2 is large relative to σ^2. Hence, to evaluate metrics, we look at two quantities derived from our model: (i) $\widehat{p}_1$, the posterior mean of the p_1 parameter, and (ii) $\widehat{r}$, the posterior mean of $\tau^2/(\tau^2+\sigma^2)$. The former gives the fraction of players who differ from the league mean and the latter gives the fraction of the variance in the response that is due to individual player differences as opposed to chance. When both of these quantities are high, a metric contains a large amount of signal.

3 Results

Before implementing our Bayesian variable selection model on the fifty offense metrics outlined in Appendix A, we examined the distribution of the data observed for each one in order to assess the normality assumption of Equation 1. For the majority of these metrics (36/50), the assumption proved reasonable. However,

a smaller subset of metrics (14/50) exhibit substantial skewness[2]. Examples are triples (3B) and stolen bases (SB) where the vast majority of players have very small values but there also exists a long right tail consisting of a small number of players with much larger values. The large proportion of zero values also makes many of these metrics less amenable to transformation. We proceeded to implement our model on all fifty measures, but in the results that follow we will differentiate between those measures that fit the normality assumption versus those that do not.

3.1 Evaluating Signal in Each Offensive Measure

As discussed in Section 2.1, there are two aspects of our posterior results which are relevant for evaluating the overall signal in an offensive metric: the fraction of players who differ from the league mean (estimated by $\widehat{p}_1$) and the fraction of the variance in the response that is due to individual player differences as opposed to chance (estimated by $\widehat{r}$). When both of these are large, individual mean estimates will have substantially greater predictive power than the overall league mean.

In Figure 2, we plot $\widehat{p}_1$ against $\widehat{r}$ for our fifty offensive metrics (note, the values plotted for this and all similar figures can be found in Appendix B). Metrics colored in red were the majority that were reasonably approximated by a normal distribution whereas metrics colored in black were not. Metrics which appear in the upper right portion of the plot are those that our model identifies as demonstrating high signal.

Several facts stand out from this figure. First, our model tends to identify the non-normally distributed metrics given in black as being low signal. Outlying datapoints which violate the model assumptions do not automatically "fool" the model into thinking the metric is high signal. Second, while a large number of the normal metrics given in red have a large $\widehat{p}_1$, there is substantial variance in $\widehat{r}$. Third, the distribution of the $\widehat{r}$ has no sharp breaks and, consequently, there is no natural boundary between metrics which have high and low signal. That said, nine metrics appear to have the highest signal: K/PA, GB/BIP, FB/BIP, HR/FB, SPD, HR/PA, BB/PA, ISO, and K. All nine of these have $\widehat{p}_1$ near one and $\widehat{r}$ among the highest of all metrics. For the remainder of the discussion, we restrict ourselves K/PA, SPD, ISO, BB/PA, and GB/BIP[3].

[2]These metrics are 3B, 3B/PA, BUH, BUH/H, CS, CS/OB, HBP, HDP/PA, IBB, IBB/PA, SB, SB/OB, SBPA, and SH.

[3]We do this because (i) K/PA and K, (ii) HR/FB, HR/PA, and ISO, and (iii) GB/BIP and FB/BIP measure more or less the same thing. The first set measures strikeouts, the second power hitting, and the third the tendency to hit grounders as opposed to fly balls.

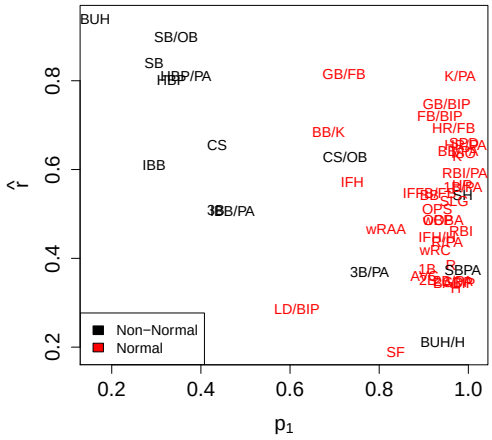

Figure 2: Plot of $\widehat{r}$ against $\widehat{p}_1$: $\widehat{p}_1$ is the posterior mean of the p_1 parameter and estimates the fraction of players who differ from the league mean whereas $\widehat{r}$ is the posterior mean of $\tau^2/(\tau^2+\sigma^2)$ and estimates the fraction of the variance in the response that is due to individual player differences as opposed to chance. The values plotted here can be found in Appendix B.

This set of five "best" metrics spans several different aspects of offensive ability: K/PA and BB/PA are all related to plate discipline, SPD represents speed, ISO measures hitting power, and GB/BIP captures the tendency to hit ground balls. Furthermore, they have some support in the literature. Studeman (2007a) finds that strikeout rate (K/PA) and walk rate (BB/PA) are very consistent. We also have confirmation on the low end: like Studeman (2007a), we find that BABIP is a low signal metric.

A final interesting fact is that ratio-based metrics seem to have lower signal than their two constituent parts. For instance, while both GB/BIP and FB/BIP demonstrate high signal in Figure 2, their ratio GB/FB demonstrates much less signal. A similar fact holds for K/PA and BB/PA on the one hand and BB/K on the other.

3.2 Examining Individual Players

In this section, we examine individual players by focusing on four of the five high signal metrics: ISO, BB rate, SPD and K rate. Each of these metrics measures a

DOI: 10.2202/1559-0410.1323

ISO - Isolated Power			**BB/PA - Walk Rate**		
Player	**Mean** $(\mu+\alpha_i)$		**Player**	**Mean** $(\mu+\alpha_i)$	
	Estimate	**SD**		**Estimate**	**SD**
Mark McGwire	0.320	0.010	Barry Bonds	0.204	0.004
Barry Bonds	0.304	0.008	Gene Tenace	0.186	0.007
Ryan Howard	0.293	0.016	Jimmy Wynn	0.183	0.010
Jim Thome	0.287	0.009	Ken Phelps	0.176	0.011
Albert Pujols	0.281	0.011	Jack Cust	0.176	0.012
Population Mean $\hat{\mu}=0.142$			Population Mean $\hat{\mu}=0.087$		
SPD - Speed			**K/PA - Strikeout Rate**		
Player	**Mean** $(\mu+\alpha_i)$		**Player**	**Mean** $(\mu+\alpha_i)$	
	Estimate	**SD**		**Estimate**	**SD**
Vince Coleman	8.55	0.30	Jack Cust	0.388	0.018
Jose Reyes	8.22	0.40	Russell Branyan	0.376	0.021
Carl Crawford	8.14	0.36	Melvin Nieves	0.371	0.020
Willie Wilson	8.13	0.25	Rob Deer	0.351	0.010
Omar Moreno	7.89	0.31	Mark Reynolds	0.347	0.018
Population Mean $\hat{\mu}=4.11$			Population Mean $\hat{\mu}=0.166$		

Table 1: Top players for four high signal metrics. For each player, we provide the posterior estimate and posterior standard deviation for their individual mean $(\mu+\alpha_i)$. The estimated $\widehat{\gamma_i}$ was equal to 1.00 for each of these cases. The posterior estimate of the population mean μ is also provided for comparison.

different aspect of offensive ability: ISO relates to hitting power, BB rate relates to plate discipline, SPD relates to speed, and K rate relates to the ability to make contact. We further explore our results by focusing on the top individual players for each of these measures, as estimated by our model. In Table 1, we show the top five players in terms of their estimated individual means $(\mu+\alpha_i)$ for each of these metrics.

For the isolated power (ISO) metric, each of the top five players are well-known hitters that have led the league in home runs at least once during their careers. Even more striking is the magnitude of their estimated individual means $(\mu+\alpha_i)$: they are more than double the population mean $\hat{\mu}=0.142$. Barry Bonds appears in the top five baseball players for both ISO and BB rate, and, more generally, there is a fairly strong correspondence between these two metrics beyond the results shown in Table 1. This finding suggests that there is correlation between the skills that determine a batters plate discipline and the skills that lead to hitting for

power. Other well-known power hitters ranking high on BB rate (but outside of the top five) are Jim Thome, Mark McGwire, Frank Thomas, and Adam Dunn. Nevertheless, Bonds stands out dramatically with a walk rate that is almost 2% higher than the next highest player–about thirteen extra walks per season. This difference seems especially substantial when taking into account the small standard deviation (0.4%) of his estimated mean.

Jack Cust appears in the top five baseball players for both BB rate and K rate. This is especially interesting since having a high BB rate is beneficial whereas having a high K rate is detrimental. However, it is not particularly surprising: players with good plate discipline will frequently be in high count situations that can also lead to strike outs. Cust is especially well-known for having a "three-outcome" (i.e. walk, strikeout, or home run) approach at the plate. Moving beyond the top five players, other power hitters such as Ryan Howard, Adam Dunn, and Jim Thome also exhibit high K rates. The top players on Bill James' speed metric SPD are a much different set of players than those highlighted by the other three metrics. The highest estimated individual mean is held by former Rookie of the Year Vince Coleman who led the National League in stolen bases from 1985 to 1990.

A general theme of all four metrics examined in Table 1 is that there is consistency *within players*, as indicated by the relatively small standard deviations, but clear evidence of substantial heterogeneity *between players* since the top players are estimated to have such a large deviation from the population mean. These two factors are an ideal combination for a high signal offensive metric and are precisely what is measured by having a high $\widehat{r}$.

4 External Validation and Principal Components

In this section, we perform various analyses of our data and model relative to external methods. First, we compare our model to the submodels outlined in Section 2.3. Second, we compare our results to an alternative variable selection approach based upon the Lasso. Finally, we explore the correlation between offensive metrics with a principal component analysis.

4.1 Comparison to Submodels

We begin our comparison of our general model to its three submodels by recalling the definition of high signal metric versus a low signal metric. A high signal metric is one that provides a consistent measure for an individual so that a player's past performance is indicative of his future performance. A low signal one has little predictive power: one would be just as well served predicting future performance

DOI: 10.2202/1559-0410.1323

by the overall league average rather than taking into account past individual performance. This implies a simple way to directly assess this question: we can hold out a portion of our data and compare the predictions of our model to the mean-only complete pooling model outlined in Section 2.3. In particular, we hold out the 2008 season values for each player, estimate the two models (i.e., our mixture model and the mean-only complete pooling model) using the values from seasons prior to 2008, and then use two models to forecast the 2008 season values.

For each posterior draw j from our mixture model, we obtain a league mean μ^j, a set of player deviations from the league mean $\{\alpha_i^j\}_i$, and a variance σ^{2j}. We can thus get a predicted value for player i on posterior draw j by simulating $y_i^{j\star} \sim Normal(\mu^j + \alpha_i^j, w_{i,2008} \cdot \sigma^{2j})$ from the likelihood. Finally, we can average these over all of our posterior draws to form our prediction $\hat{y}_{i,2008}^{MM} = \sum_{j=1}^{J} y_i^{j\star}/J$ for player i (where MM refers to our mixture model prior on the α_i).

Similarly, for each posterior draw j from the complete pooling model, we obtain a league mean $\tilde{\mu}^j$ and a variance $\tilde{\sigma}^{2j}$. We can thus get a predicted value for player i on posterior draw j by simulating $\tilde{y}_i^{j\star} \sim Normal(\tilde{\mu}^j, w_{i,2008} \cdot \tilde{\sigma}^{2j})$ from the likelihood. Finally, we can average these over all of our posterior draws to form our prediction $\hat{y}_{i,2008}^{CP} = \sum_{j=1}^{J} \tilde{y}_i^{j\star}/J$ for player i (where CP denotes complete pooling).

Using the heldout $y_{i,2008}$, we can calculate the root mean square error (RMSE) of our model's predictions, $RMSE^{MM} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{i,2008} - y_{i,2008}^{MM})^2}$ where i indexes the N players who played in 2008[4]. We can do likewise for the complete pooling model. For high signal metrics, our model should have a substantially lower RMSE.

Evidence for this is shown in Figure 3 which plots our model's $\widehat{p}_1$ and $\widehat{r}$ against the RMSE of our mixture model relative to the RMSE of the mean-only complete pooling model (i.e., $Relative\ RMSE = RMSE^{MM}/RMSE^{CP}$; small values of this quantity denote good performance by our model). There is dramatic correlation between Relative RMSE and $\widehat{r}$. Metrics with a low $\widehat{r}$ are ones for which our model performs similarly to the mean-only model–precisely the definition of a low signal metric. Concomitantly, our model dramatically outperforms the mean-only model for those metrics we previously identified as high signal in Section 3.1 (i.e., those with high $\widehat{p}_1$ and high $\widehat{r}$). For example, for strikeout rate, our model has an RMSE that is less than half of that of the mean-only complete pooling model

The low correlation between $\widehat{p}_1$ and Relative RMSE is not surprising given (i) the low correlation of $\widehat{p}_1$ and $\widehat{r}$ and (ii) the high correlation between $\widehat{r}$ and Rela-

[4]Since our aim is to estimate the consistency of a player with respect to a metric over time, we excluded from the RMSE calculation all players for whom 2008 was the first year of play. There is no notion of consistency for these players since it is their one and only year in the dataset. The analysis would not be much changed by including them, however, since both models predict the league mean for such players.

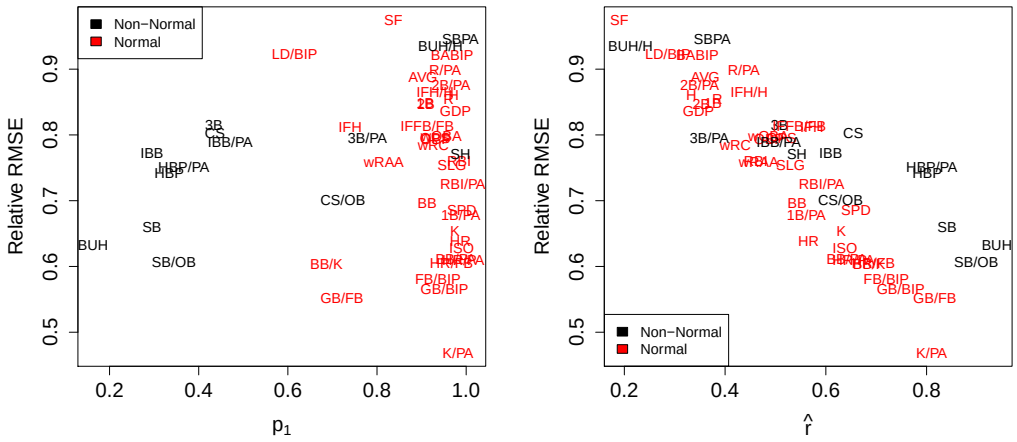

Figure 3: Left: Plot of $\widehat{p}_1$ against the RMSE of the full model relative to the complete pooling model. Right: Plot of $\widehat{r}$ against the RMSE of the full model relative to the complete pooling model. The values plotted here can be derived from those given in Appendix B.

tive RMSE. However, we do not view this low correlation as problematic and rather view it as an extra dimension on which to screen metrics. In particular, we view the $\widehat{r}$ / Relative RMSE dimension as serving to screen metrics on how much individual players vary on them versus the inherent season to season variation in them (i.e., τ^2 versus σ^2). On the other hand, the additional dimension provided by $\widehat{p}_1$ facilitates further differentiation among metrics. For instance, among metrics that have high $\widehat{r}$ / low Relative RMSE, $\widehat{p}_1$ differentiates (i) normal metrics from the skewed metrics and, within the normal metrics, (ii) ratio metrics such as BB/K and GB/FB from non-ratio metrics.

We are encouraged that (i) those metrics flagged by our model to have high signal (i.e., those with high $\widehat{p}_1$ and $\widehat{r}$) substantially outperform the mean-only model on holdout prediction and (ii) those flagged to have low signal perform similarly to the mean-only model on holdout prediction. It means that the metrics identified as high signal fit the natural definition of it: metrics that players perform consistently with respect to such that predictions based on individual information trump the league mean as a prediction. Moreover, it suggests that, when the mean-only model is "correct", our model identifies it as the appropriate special case and estimates it.

That said, one wonders whether allowing the point mass mixture at the league mean (and thereby forcing some fraction of players to be estimated at the

DOI: 10.2202/1559-0410.1323

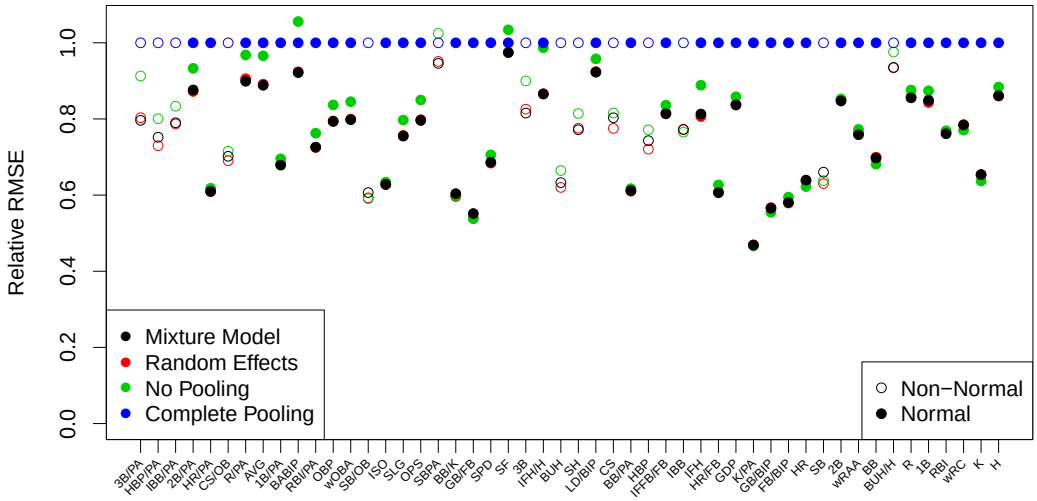

Figure 4: Holdout RMSEs of the main mixture model and three submodels discussed in Section 2.3 relative to the holdout RMSE of the mean-only complete pooling model. The mixture model is given in black, the random effects model in red, the no pooling model in green, and the complete pooling model in blue. The x-axis is ordered by size of the complete pooling model RMSE. Normal measures are represented by filled circles and non-normal ones by open circles. The values plotted here can be derived from those given in Appendix B.

league mean) is too restrictive. Perhaps some of the metrics identified as low signal do indeed contain signal and that player-specific information would trump the league mean as a predictor. In order to assess this, we fit both the Bayesian random effects model (p_1 fixed at one) and the no pooling model (p_1 fixed at one and τ^2 fixed at ∞), both of which use player-specific information for all players. We then drew predictions from the posterior distribution following exactly the same process outlined above for the mixture model and calculated RMSEs on the holdout sample.

We plot the holdout RMSEs of all four models relative to the mean-only model in Figure 4. As can be seen, our mixture model in black never performs meaningfully worse than any of the other three models (i.e., it is always at or near the bottom for each metric). Hence, we can conclude that our procedure of fitting a data-determined fraction of players to the league mean does not cause us to miss out on any signal in the metrics or falsely conclude that a metric is low signal.

This plot shows a further benefit as well. It shows that our model performs about as well or better than the random effects and no pooling model for all metrics. Not only does this provide validation for our model, but it also shows that we are

not sacrificing predictive performance in order to gain the insight provided by the model outputs $\widehat{p}_1$ and $\widehat{r}$. That is, we pay no price for interpretability.

4.2 Comparison to the Lasso

The Lasso (Tibshirani (1996)) is a penalized least squares regression that uses an L_1 penalty on the estimated regression coefficients,

$$\hat{\beta}^{Lasso} = \underset{\hat{\beta}}{\operatorname{argmin}} \left[\sum_{i,j} (y_{ij} - X_i\hat{\beta})^2 + \lambda \sum_i |\hat{\beta}_i| \right], \quad \lambda \geq 0. \tag{7}$$

The Lasso enforces sparsity on the covariate space by forcing some coefficients to zero and can therefore be used for variable selection. A more intuitive reformulation of the Lasso is as a minimization of $\sum_{i,j}(y_{ij} - X_i\hat{\beta})^2$ subject to $\frac{\sum_i |\hat{\beta}_i|}{\sum_i |\hat{\beta}_i^{OLS}|} \leq f$, where $\hat{\beta}_i^{OLS}$ is the coefficient from variable i in the ordinary least squares solution. The free parameter f is known as the Lasso "fraction" because it is the ratio of the L_1 sizes of the Lasso solution and the OLS solution. Furthermore, f directly corresponds to λ in Equation 7 and ranges between zero (corresponding to fitting only an overall mean or $\lambda = \infty$ in Equation 7) and one (corresponding to the ordinary least squares regression solution or $\lambda = 0$ in Equation 7).

We apply the Lasso to our problem by centering each offensive metric and then fitting the regression model consisting only of indicators for each player. Each component of the $\hat{\beta}^{Lasso}$ vector corresponds to the individual mean of a given player, and we are interested in which of these individual means are fitted to be different from zero. To select a value of the free parameter f, we implemented multiple five-fold cross validation by randomly subdividing all player-seasons into five groups. That is, we fix f and fit the Lasso on four of the groups and predict fifth group of player-seasons, which has been held out-of-sample. We then repeat this four times, holding each of the groups as out of sample and fitting on the other four. Finally, we repeat this procedure ten times, yielding fifty out-of-sample RMSEs for a given value of f. After conducting this procedure along a fine grid of possible f values ranging between 0 and 1, we selected the f with the lowest cross-validated average RMSE. We then fit the Lasso model to the full dataset using this value of f.

The outcome of interest from this Lasso regression is Lasso%, the percentage of players that are fitted with non-zero coefficients by the Lasso (i.e., the percent of $\hat{\beta}_i^{Lasso} \neq 0$). This measure represents a global measure of signal for each metric, and thus serves as an alternative to our model-based measures of $\widehat{p}_1$ and $\widehat{r}$. We compare our model-based measures to the Lasso% measure in Figure 5. We begin with the right plot which shows that $\widehat{r}$ and the Lasso% are highly correlated, both for the

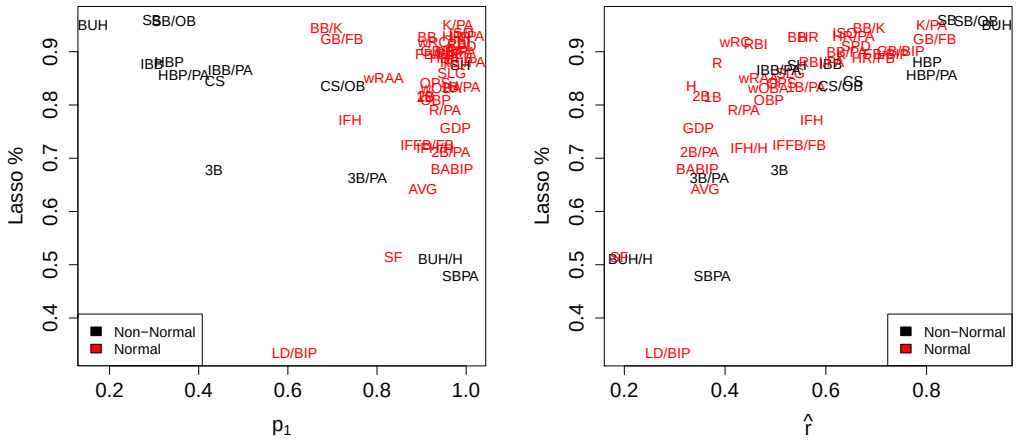

Figure 5: Left: Plot of $\widehat{p}_1$ against the percentage of players with non-zero means selected by the Lasso. Right: Plot of $\widehat{r}$ against the percentage of players with non-zero means selected by the Lasso. The values plotted here can be found in Appendix B.

normal metrics in red and the skewed ones in black. Thus, there is a broad correspondence between one measure of signal from our own model and the external Lasso model's measure of signal.

More interesting is the comparison of $\widehat{p}_1$ and Lasso% presented in the left plot of Figure 5. These two measures are intimately related as both attempt to estimate the fraction of players who differ from the league mean. Consequently, we see agreement between Lasso% and $\widehat{p}_1$ for many measures, especially the red measures that fit the normal model. These high signal measures with large $\widehat{p}_1$ also tend to have a large percentage of non-zero coefficients. The main difference between the two methods is with the black metrics that have skewed (non-normal) data distributions. These measures tend to have a high Lasso% but a low $\widehat{p}_1$, meaning that a Lasso-based analysis would attribute much more signal to these metrics than our mixture model-based analysis. Neither our model nor the Lasso is meant for the highly skewed data of these black metrics. The fact that the our model is more cautious about these metrics than the Lasso suggests an advantage to our approach.

4.3 Principal Components Analysis

Among our metrics with a high $\widehat{p}_1$ in Figure 2, there was a smooth distribution of $\widehat{r}$ with no clean breaks. Ideally, there would be a more stark divide in the performance of these metrics, allowing us to focus on only a small subset of metrics as a complete summary of offensive performance. However, this is a difficult task in large part

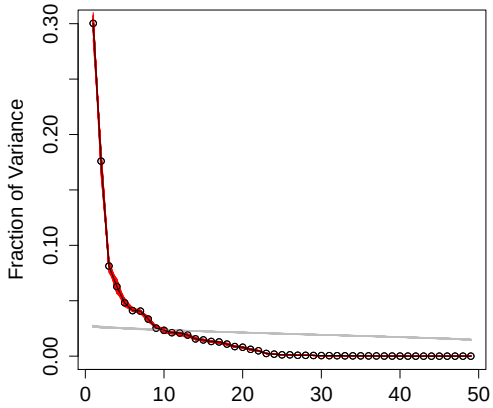

Figure 6: Plot of the variance explained by each principal component. We create a grey null band by randomly permuting the values within each column to demonstrate the strong significance of our results. In addition, we demonstrate the variability in our own principal components by creating bootstrap samples of player-seasons and calculating the variance of the bootstrap principal components in red.

because of the high correlation between many of these metrics (e.g., consider OPS which is the sum of OBP and SLG). We performed a more systematic assessment of the correlation between metrics using a Principal Components Analysis. PCA projects the data onto an orthogonal space such that each orthogonal component describes a decreasing amount of variance.

Note that one of our fifty metrics, SBPA, was not included in this analysis due to a high number of player-seasons which had a denominator (SB+CS) equal to zero. Furthermore, so we could include all metrics, we work with the reduced set of 1,935 player-seasons which begin in 2002. The results from our principal components analysis on the remaining forty-nine metrics are shown in Figure 6. We see that among the forty-nine metrics represented in Figure 6, only about eight principal components have variance exceeding the null bands, which suggests that there are only about eight unique (orthogonal) metrics among the entire set of metrics. Furthermore, these eight principle components account for about 80% of the total variance.

In Section 3.1, we identified a set of nine metrics as high signal metrics and further reduced this to five (i.e., K/PA, BB/PA, SPD, ISO, and GB/BIP) due to collinearity considerations. One wonders how will these five metrics explain the other forty-nine given the results of the principal components analysis. Do they account for a large fraction or a small fraction? One can assess this question by

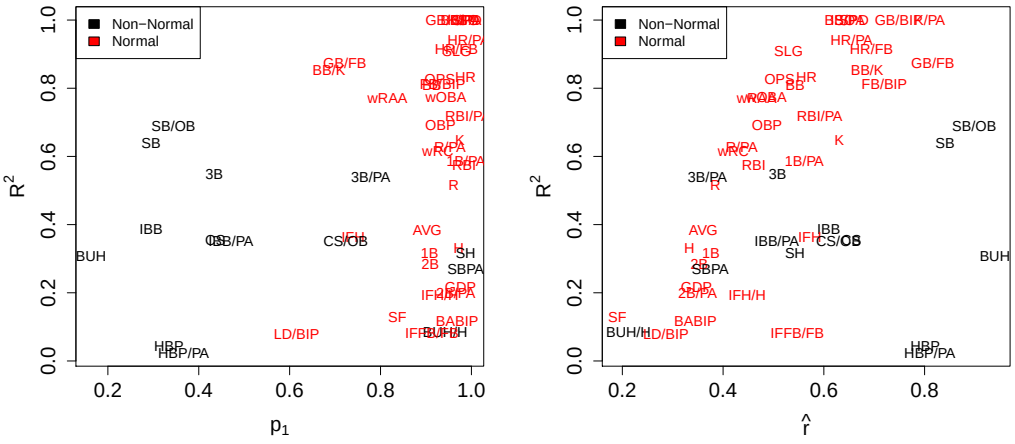

Figure 7: Left: Plot of $\widehat{p}_1$ against R^2. Right: Plot of $\widehat{r}$ against R^2. R^2 is calculated from the regression of each metric on K/PA, BB/PA, SPD, ISO, and GB/BIP. The values plotted here can be found in Appendix B.

regressing each of the forty-nine metrics on the five selected metrics and taking one minus the sum of squares of the residuals divided by the sum of squares of the forty-nine metrics. This corresponds exactly to the usual multivariate R^2 calculation and to the fraction of variance explained in principal components analysis (in particular, if we scale the forty-nine variables by their standard deviations before regressing on the five selected variables it corresponds to principal components on the correlation matrix; otherwise, it corresponds to principal components on the covariance matrix). We find that our five selected metrics explain about 55% of the total variance, quite favorable compared to the 80% explained by the first eight principal components especially when one considers that our metrics are directly interpretable unlike principal components.

More interesting is to discover which of the forty-nine metrics our set of five (i.e., K/PA, BB/PA, SPD, ISO, and GB/BIP) can predict well and which ones it cannot. We plot the R^2 from the regression of each of the metrics on our five metrics in Figure 7. There is not much relation between $\widehat{p}_1$ and R^2 but there is a strong correlation between $\widehat{r}$ and R^2, particular for the normally distributed metrics. Looking more specifically at what metrics our five metrics can serve as surrogates for (i.e., provide a high R^2 for), we see (in addition to the five metrics themselves) power hitting metrics like SLG, HR/PA, HR/FB, HR, and OPS; plate discipline metrics like BB, BB/K; hit location metrics like FB/BIP and GB/FB; and metrics which relate to a player's run contribution to his team like wRAA and wOBA. Further, metrics which our five cannot predict are ones that seem to be largely unpredictable

in the sense that they largely depend defensive players. Such metrics include hit by pitch total (HBP) and rate (HBP/PA), sacrifice flies (SF), BABIP, and bunt hit rate (BUH/H).

In sum, our set of fifty offensive metrics are highly correlated. However, the set of five selected by our model provide a substantial reduction in the dimensionality. Managers can look at these five an obtain much of the information contained in the full set of fifty.

5 Discussion

We have introduced a hierarchical Bayesian variable selection model, which allows us to determine how well hitting metrics provide sound predictions across time and players. Our model does not require adjustment for multiple testing across players and imposes shrinkage of player-specific parameters towards the population mean. For fifty different offensive metrics, the full posterior distributions of our model parameters are estimated with a Gibbs sampling implementation. We evaluate each of these metrics with several proxies for reliability or consistency, such as the proportion of players found to differ from the population mean ($\widehat{p}_1$) as well as the variation in individual player performance relative to variation from season to season ($\widehat{r}$).

We identified five metrics which stand out has having high signal. Beyond that, there is a continuum of metrics which all demonstrate high $\widehat{p}_1$ but slightly less $\widehat{r}$ (see Figure 2). That is, players do stand out from the league mean for these metrics but the differences across players are less and less. The existence of this slow tailing off in $\widehat{r}$ is largely a function of the high correlation among the set of fifty metrics. Our principal components analysis suggests that only a small subset of metrics are substantively different from one another and that our set of five accounts for a substantial amount of this difference.

A direction of future research would be the creation of a reduced set of consistent metrics that were less highly dependent but still directly interpretable. Our sophisticated hierarchical model could be extended further to share information between metrics instead of the separate metric-by-metric analysis that we have performed. The relatively high correlation between some metrics could be used to cluster metrics together and reduce dimensionality. This approach would have the advantage of pooling across related measures and increasing effective degrees of freedom.

DOI: 10.2202/1559-0410.1323

A Offensive Measures

Our fifty offensive measures are subdivided into three categories for ease of presentation. Terms that are not defined in this appendix are AB (at bats), BIP (balls in play), OB (total number of times on base), PA (plate appearances), and PA* (plate appearances minus sacrifice hits). As noted in the main text, the weights $w_{i,j}$ actually used in the analysis are a function of the raw weights $u_{i,j}$. In particular, we set $w_{ij} = \bar{u}/u_{ij}$ for rates and $w_{ij} = u_{ij}/\bar{u}$ for totals where u_{ij} denotes the raw weight function for player i in season j and $\bar{u}$ represents the mean weight over all player-seasons.

1. Simple Hitting Totals and Rates

Metric y_{ij}	Raw Weight u_{ij}	Description	Metric y_{ij}	Raw Weight u_{ij}	Description
1B	PA	singles	1B/PA	PA	single rate
2B	PA	doubles	2B/PA	PA	double rate
3B	PA	triples	3B/PA	PA	triple rate
HR	PA	home runs	HR/PA	PA	home run rate
R	PA	runs	R/PA	PA	run rate
RBI	PA	runs batted in	RBI/PA	PA	runs batted in rate
BB	PA	base on balls (walk)	BB/PA	PA	walk rate
IBB	PA	intentional walk	IBB/PA	PA	intentional walk rate
K	PA	strike outs	K/PA	PA	strike out rate
HBP	PA	hit by pitch	HBP/PA	PA	hit by pitch rate
BUH	H	bunt hits	BUH/H	H	bunt hit proportion
H	PA	hits	GDP	PA	ground into double play
SF	PA	sacrifice fly	SH	PA	sacrifice hit

2. More Complicated Hitting Totals and Rates

Metric y_{ij}	Raw Weight u_{ij}	Description
OBP	PA*	on base percentage (OB/PA*)
AVG	AB	batting average (H/AB)
SLG	AB	slugging percentage
OPS	AB $\times$ PA*	OPB + SLG
ISO	AB	isolated power (SLG-AVG)
BB/K	PA	walk to strikeout ratio
HR/FB	PA	home run to fly ball ratio
GB/FB	BIP	ground ball to fly ball ratio
BABIP	BIP	batting average for balls in play
LD/BIP	BIP	line drive rate
GB/BIP	BIP	ground ball rate
FB/BIP	BIP	fly ball rate
IFFB/FB	FB	infield fly ball proportion
IFH	GB	in field hit
IFH/H	GB	in field hit proportion
wOBA	PA*	weighted on base average
wRC	PA	runs created based on wOBA
wRAA	PA	runs above average based on wOBA

3. Baserunning Totals and Rates

Metric y_{ij}	Raw Weight u_{ij}	Description	Metric y_{ij}	Raw Weight u_{ij}	Description
SB	OB	stolen bases	SB/OB	OB	stolen base rate
CS	OB	caught stealing	CS/OB	OB	caught stealing rate
SBPA	SB + CS	stolen bases per attempt i.e., SB/(SB+CS)	SPD	PA	Bill James' speed metric

B Model Statistics and Holdout RMSEs

Metric	Model Statistics				Holdout RMSEs			
	$\widehat{p}_1$	$\widehat{r}$	Lasso%	R^2	Mixture Model	Random Effects	No pooling	Complete pooling
1B	0.908	0.376	0.815	0.317	22.4	22.2	23.0	26.3
2B	0.910	0.353	0.817	0.285	9.65	9.68	9.71	11.4
3B	0.434	0.509	0.678	0.549	2.02	2.05	2.23	2.48
AVG	0.903	0.361	0.642	0.386	0.0249	0.0249	0.0270	0.0280
BABIP	0.969	0.345	0.680	0.119	0.0282	0.0283	0.0323	0.0306
BB	0.913	0.543	0.928	0.811	16.3	16.3	15.9	23.4
BB/K	0.687	0.686	0.945	0.855	0.182	0.18 0	0.18 0	0.301
BB/PA	0.977	0.642	0.900	1.000	2.07	2.06	2.09	3.38
BUH	0.163	0.940	0.950	0.307	1.78	1.74	1.87	2.81
BUH/H	0.943	0.213	0.511	0.086	23.2	23.1	24.2	24.8
CS	0.436	0.655	0.844	0.354	2.60	2.51	2.64	3.24
FB/BIP	0.937	0.720	0.896	0.812	3.89	3.88	3.99	6.70
GB/FB	0.721	0.816	0.925	0.876	0.257	0.251	0.251	0.467
GB/BIP	0.952	0.749	0.903	1.000	3.74	3.75	3.67	6.61
GDP	0.976	0.347	0.756	0.217	4.97	4.98	5.09	5.94
H	0.972	0.333	0.836	0.332	34.3	34.3	35.2	39.9
HBP	0.334	0.801	0.881	0.044	3.09	3.00	3.21	4.16
HR	0.987	0.565	0.929	0.833	6.80	6.78	6.62	10.6
HR/FB	0.967	0.695	0.889	0.914	3.49	3.50	3.61	5.76
IBB	0.295	0.610	0.877	0.387	3.67	3.67	3.64	4.75
IFFB/FB	0.913	0.548	0.725	0.083	3.61	3.61	3.71	4.43
IFH	0.740	0.572	0.773	0.364	4.59	4.55	5.02	5.64
IFH/H	0.930	0.448	0.720	0.195	2.22	2.22	2.53	2.57
ISO	0.990	0.638	0.935	1.000	0.0373	0.0373	0.0377	0.0594
K/PA	0.982	0.810	0.950	1.000	3.09	3.10	3.08	6.60
LD/BIP	0.615	0.287	0.335	0.080	2.66	2.67	2.77	2.89
OBP	0.932	0.487	0.810	0.692	0.0268	0.0269	0.0283	0.0338
OPS	0.930	0.512	0.842	0.828	0.0748	0.0750	0.0798	0.0939
R	0.961	0.384	0.879	0.516	21.9	22.0	22.4	25.6
RBI	0.984	0.461	0.916	0.576	21.4	21.5	21.6	28.1
SB	0.295	0.840	0.960	0.640	7.24	6.92	7.00	11.0
SF	0.837	0.190	0.516	0.130	2.40	2.40	2.55	2.47
SH	0.987	0.543	0.876	0.317	2.20	2.21	2.32	2.85
SLG	0.968	0.530	0.861	0.908	0.0526	0.0527	0.0555	0.0696
K	0.975	0.631	0.895	0.647	24.8	24.7	24.1	37.9
SPD	0.990	0.660	0.911	1.000	1.17	1.17	1.21	1.71
wOBA	0.944	0.486	0.832	0.776	0.0289	0.0290	0.0306	0.0363
wRAA	0.815	0.467	0.851	0.771	12.8	12.8	13.0	16.8
wRC	0.926	0.420	0.918	0.615	23.2	23.2	22.8	29.6
1B/PA	0.988	0.561	0.834	0.586	0.0207	0.0208	0.0212	0.0305
2B/PA	0.966	0.350	0.711	0.199	0.0115	0.0114	0.0122	0.0131
3B/PA	0.779	0.369	0.663	0.541	0.00341	0.00344	0.00391	0.00428
HR/PA	0.994	0.655	0.930	0.943	0.00964	0.00963	0.00977	0.0158
R/PA	0.953	0.437	0.792	0.628	0.0211	0.0213	0.0228	0.0235
RBI/PA	0.992	0.592	0.881	0.718	0.0235	0.0234	0.0247	0.0323
IBB/PA	0.471	0.507	0.867	0.354	0.00616	0.00614	0.00650	0.00780
HDP/PA	0.367	0.810	0.857	0.024	0.00560	0.00543	0.00596	0.00744
SB/OB	0.344	0.899	0.957	0.690	0.0351	0.0343	0.0342	0.0579
CS/OB	0.724	0.629	0.837	0.352	0.0133	0.0130	0.0135	0.0189
SBPA	0.987	0.375	0.479	0.271	0.258	0.259	0.279	0.272

DOI: 10.2202/1559-0410.1323

References

Albert, J. and J. Bennett (2003): *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game*, Copernicus Books.

Appelman, D. (2009): "Major league baseball database," *Fangraphs.com*, http://www.fangraphs.com

Baumer, B. (2008): "Why on-base percentage is a better indicator of future performance than batting average: An algebraic proof." *Journal of Quantitative Analysis in Sports*, 4, Article 3.

Brown, L. D. (2008): "In-season prediction of batting averages: A field-test of simple empirical bayes and bayes methodologies," *Annals of Applied Statistics*, 2, 113–152.

Fair, R. (2008): "Estimated age effects in baseball." *Journal of Quantitative Analysis in Sports*, 4, Article 1.

Gelman, A. (2006): "Prior distributions for variance parameters in hierarchical models," *Bayesian Analysis*, 1, 515–533.

Gelman, A. and J. Hill (2006): *Data Analysis Using Regression and Multilevel/Hierarchical Models*, New York, NY: Cambridge University Press.

Geman, S. and D. Geman (1984): "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6, 721–741.

George, E. I. and R. E. McCulloch (1997): "Approaches for bayesian variable selection," *Statistica Sinica*, 7, 339–373.

James, B. (1987): *The Bill James Baseball Abstract 1987*, Ballantine Books.

James, B. (2008): *Bill James Handbook 2009*, ACTA Publications.

Kahrl, C., S. Goldman, and N. Silver (2009): *Baseball Prospectus 2009: The Essential Guide to the 2009 Baseball Season*, Plume.

Kaplan, D. (2006): "A variance decomposition of individual offensive baseball performance," *Journal of Quantitative Analysis in Sports*, 2, Article 2.

Kass, R. E. and A. E. Raftery (1995): "Bayes factors," *Journal of the American Statistical Association*, 90, 773–795.

Lederer, R. (2009): "Babip: Slicing and dicing groundball out rates babip: Slicing and dicing groundball out rates babip: Slicing and dicing groundball out rates," *http://baseballanalysts.com/archives/2009/01/babip-slicing-a.php*, January 27, 2009.

Lewis, M. (2003): *Moneyball: The art of winning an unfair game*, W.W. Norton & Co.

Null, B. (2009): "Modeling baseball player ability with a nested dirichlet distribution." *Journal of Quantitative Analysis in Sports*, 5, Article 5.

Puerzer, R. J. (2003): "Engineering baseball: Branch rickey's innovative approach to baseball management," *NINE: A Journal of Baseball History and Culture*, 12, 72–87.

Scott, J. and J. Berger (2006): "An exploration of aspects of bayesian multiple testing," *Journal of Statistical Planning and Inference*, 136, 2144–2162.

Silver, N. (2003): "Lies, damned lies, randomness: Catch the fever!" *Baseball Prospectus*, May 14, 2003., URL `https://baseballprospectus.com/article.php?articleid=1897`

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002): "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583–639.

Studeman, D. (2007a): "A quick look at four hitting rates," *The Hardball Times*, URL `http://www.hardballtimes.com/main/article/a-quick-look-at-four-hitting-rates/`.

Studeman, D. (2007b): "Should jose reyes hit more ground balls?" *The Hardball Times*, URL `http://www.hardballtimes.com/main/article/should-jose-reyes-hit-more-groundballs/`.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, B*, 58, 267–288.

DOI: 10.2202/1559-0410.1323