

# Swinging, Fast and Slow: Interpreting variation in baseball swing tracking metrics

Scott Powers<sup>1\*</sup> and Ronald Yurko<sup>2\*</sup>

<sup>1</sup>Department of Sport Management, Rice University

<sup>2</sup>Department of Statistics & Data Science, Carnegie Mellon University

July 1, 2025

## Abstract

In 2024, Major League Baseball released new bat tracking data, reporting swing-by-swing bat speed and swing length measured at the point of contact. While exciting, the data present challenges for their interpretation. The timing of the batter’s swing relative to the pitch determines the point of measurement relative to the full swing path. The relationship between swing metrics and swing outcomes is confounded by the batter’s pitch recognition. We introduce a framework for interpreting bat tracking data in which we first estimate the batter’s intention conditional on ball-strike count and pitch location using a Bayesian hierarchical skew-normal model with random intercept and random slopes for batter. This yields batter-specific effects of count on swing metrics, which we leverage via instrumental variables regression to estimate causal effects of bat speed and swing length on contact and power outcomes. Finally, we evaluate the tradeoff between contact and power due to bat speed by modeling a plate appearance as a Markov chain. We conclude that batters can reduce their strikeout rate by reducing bat speed as strikes increase, but the tradeoff in reduced power approximately counteracts the benefit to the average batter.

## 1 Introduction

Baseball has a long history as fertile ground for statistical analysis of rich datasets. For Major League Baseball (MLB), detailed records are publicly available at the play-by-play level going back to 1912 (Retrosheet, 2025). Beginning in 2007, MLB made optical pitch tracking data freely available online (Fast, 2010), enabling unprecedented insights into player performance (Swartz et al., 2017). In 2016, MLB made waves again when they updated their ball tracking technology to incorporate Doppler radar and released batted ball tracking data to the public (Arthur, 2016). In May 2024, MLB released their most exciting new dataset since batted ball tracking in 2016 by making bat tracking data available at the pitch-by-pitch level (Petriello, 2024).

The new bat tracking metrics include *bat speed*, *fast-swing rate*, *squared-up rate*, *blasts* and *swing length*. Of these, only bat speed and swing length represent new information: The other metrics (fast-swing rate, squared-up rate and blasts) are functions of bat speed and other ball tracking data that were already publicly available. Both bat speed and swing length are measured at the moment of bat-ball contact (or at the point nearest to contact, in the case of a swing-and-miss). Bat speed is the linear speed of the “sweet spot” of

---

\*Both authors contributed equally to this work.

the bat, approximately six inches from the end; and swing length is the distance traveled by the end of the bat from the start of the swing until contact. Critically, bat speed and swing length are available for each swing, not just aggregated at the player level. This opens the door to pitch-level analysis within player, not just analysis of differences between player averages.

An open research question in baseball is the effect of batter aggressiveness. Conventional coaching wisdom would suggest that the harder a player swings, the less likely they are to make solid contact, but the harder they are likely to hit the ball if they do make solid contact. What is the effect of bat speed on contact? Figure 1 illustrates a simple analysis, and the result is counterintuitive. We observe that swings which result in solid contact actually have slightly faster-than-average bat speeds (relative to the batter’s own average). This shows that bat speed is positively correlated with making solid contact. Taken at face value, it would seem to suggest that batters should always swing as hard as possible because it does not reduce contact quality and may even increase it.

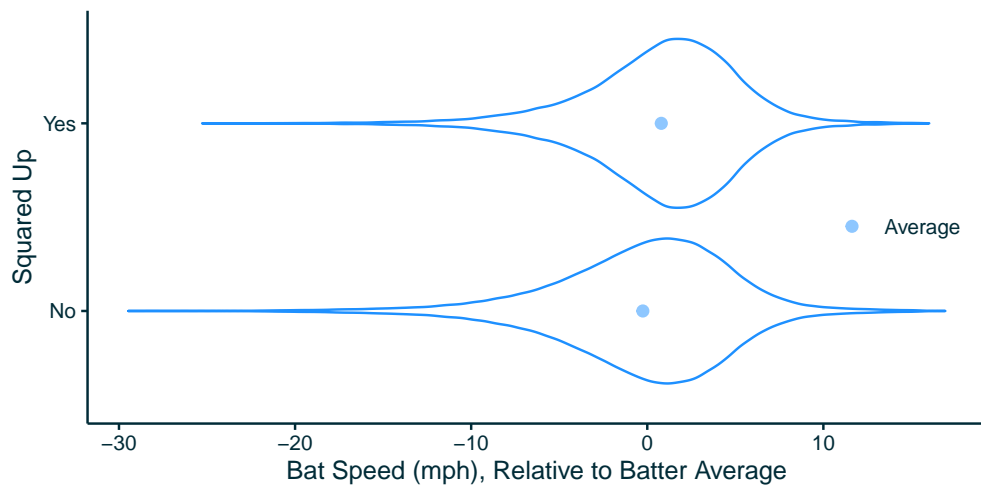


Figure 1: *Distribution of bat speed relative to batter average by contact quality (squared up or not), across all swings in the dataset. The x-axis represents the difference between the bat speed and the batter’s average bat speed. Following MLB’s definition, a swing is considered “squared up” if the batted ball’s exit velocity is at least 80% of the theoretical maximum (given pitch speed and bat speed), a proxy for solid contact.*

Like the famous problem of the bat and ball whose total cost is \$1.10 (Kahneman, 2011), this is a problem for which slower, more deliberate thinking is helpful. The dilemma is due to how the data are measured. The measurement occurs at the point of contact, the timing of which is closely related to the outcome of the swing. Figure 2 illustrates two hypothetical swings with identical swing mechanics, but because one swing is early and the other is late, the measurements are taken at different points in the swing. Thus, the timing of the swing affects the measured bat speed and swing length. From this perspective, bat speed and swing length are themselves outcomes of the swing, not entirely under the batter’s control.

We hypothesize that the counterintuitive results shown in Figure 1 are explained by an unobserved confounder: the extent to which the batter correctly identifies the pitch type early in its trajectory (by lucky guess or by other means—see, e.g., Elmore and Matthews (2022)). For example, when the batter prepares for a high fastball and this guess is correct, they are likely to complete their swing with full intent. When the batter preparing for a high fastball is met with a curveball, they need to make a mid-swing adjustment, reducing their bat speed in the process. The latter scenario is less likely to result in contact than the former.

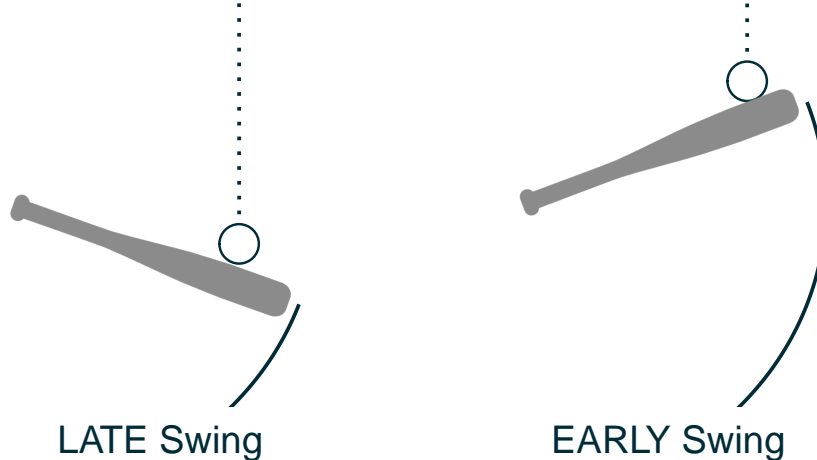


Figure 2: A diagram illustrating the effect of swing timing on swing length measurement due to the fact that swing length is measured at the point of contact. Imagining two swings with the exact same mechanics, the left image shows the swing length measurement if the swing is late, and the right image shows the swing length measurement if the swing is early.

Our approach to this problem is to model the *intent* of the batter rather than taking observed bat speed and swing length at face value. We control for contact quality, and we model the distribution of swing metrics for each batter *when they are successful* to better identify what they are trying to do. We estimate separate Bayesian hierarchical skew-normal models for intended bat speed and intended swing length, and we demonstrate heterogeneity between batters in how they modulate their intended swing metrics as the number of strikes increases. We perform instrumental variables regression to estimate the causal effect of these intended swing metrics on swing outcomes. Finally, we model a plate appearance as a Markov chain to evaluate the tradeoff between contact and power when modulating bat speed and swing length.

## 1.1 Related Work

Shortly after MLB released the new bat tracking data, there were several analyses published in online outlets. Clemens (2024) explored correlations on the player level between the new bat tracking metrics and existing performance metrics, finding that the new metrics did not explain much variation in overall performance. This was followed by more careful analyses at the swing level. Woodward (2024) estimated a generalized additive model (GAM) to predict bat speed (as a percentage of the batter’s peak bat speed) based on pitch location and other variables; and then manually created rules based on the residuals of this model to label certain individual swings as “guess” swings or “catch-and-redirect” swings in an effort to identify batter intent. Orr (2024) pointed out that swings resulting in pulled fly balls tend to have greater swing lengths because the contact point tends to be further in front of home plate; and then fit a GAM to predict swing length from pitch location and bat speed in an effort to isolate contact depth.

In the academic literature, we have not found any previous studies analyzing newly available bat tracking data or examining the relationship between swing aggressiveness and batting outcome in baseball. Nevins et al. (2019) modeled the relationship between bat speed and batted ball speed for fastpitch softball batters. Orishimo et al. (2024) investigated correlations between kinetic/kinematic features and bat speed, finding that ground force reactions (vertical, anterior/posterior, and resultant) under the lead foot were most

correlated with bat speed. In perhaps the most intriguing and closely related study to the present work, Nakashima et al. (2025) collected full-path bat tracking data from 145 swings made by 29 collegiate batters in a laboratory setting and calculated for each swing the acceptable range of timing error for contacting a pitch with the “sweet spot” of the bat. In a different sport, Burton and Powers (2015) developed a model for the relationship between volleyball serving aggressiveness and service outcomes and calculated optimal aggressiveness based on a server’s rates of aces and errors.

Recent research on statistics in sports has used some of the methodological pieces in the present work. Judge (2024) presented a skew-normal likelihood for the distribution of a batter’s exit velocity in baseball, which motivated our approach in Section 3.1. Putman and Tolhurst (2025) used instrumental variables regression (as we do in Section 3.2) to model the causal effect of pass/run play calling on offensive play outcomes in American football.

## 2 Data

Our dataset comprises 685,143 pitches thrown during the 2024 MLB regular season. This publicly available dataset was downloaded from Baseball Savant (MLB Advanced Media, 2024) using the R package `sabRmetrics` (Powers, 2025). For each pitch, we observe the identities and handedness (left/right) of the batter and pitcher involved and the ball-strike count prior to the pitch. We also observe pitch tracking information, most notably the pitch type (e.g. fastball, curveball, changeup) and the  $(x, z)$  location where the pitch crosses the front of home plate. The pitch tracking information is sufficient to recreate an estimate of the pitch’s full trajectory as a three-dimensional quadratic function of time  $(x(t), y(t), z(t))$  and to recreate the corresponding pitch metrics such as release point, release speed, and vertical/horizontal break. Each pitch has a categorical outcome, which can be one of: Hit by Pitch (HBP), Called Ball, Called Strike, Swinging Strike, Foul Ball, or Fair Ball.

For pitches on which the batter swings, we observe bat tracking information describing the bat speed and the swing length (detailed in Section 2.1). For pitches that are hit into play (i.e. the Fair Ball outcome), we observe the initial exit speed and vertical exit angle (relative to the ground) of the ball as it leaves the bat. We also observe each fair batted ball’s manually charted hit coordinates (on the field), which serves as a proxy for the horizontal exit angle of the ball off the bat.

### 2.1 Bat Speed and Swing Length Measurement

Bat tracking metrics were made possible by improvements to MLB’s Hawk-Eye Statcast system in 2023 which featured five high frame rate (300 frames per second) cameras dedicated to tracking movements of the pitcher and the batter (Goldbeck, 2023). Through computer vision, this system tracks not only player center of mass but full limb orientation (including the bat) throughout the pitching and batting movements, generating a rich kinematic time series of data on each pitch.

The raw kinematic time series data are not publicly available, but MLB has begun to release some metrics derived from them, starting with bat speed and swing length on every pitch. Bat speed is defined as the linear speed of the “sweet spot” of the bat (roughly six inches from the end) at the point of contact with the ball, and swing length is defined as the distance traveled by the end of the bat from the start of the swing until the point of contact (Petriello, 2024). In the event of a swing and miss, both metrics are measured at the point nearest to contact. These metrics represent a tiny fraction of the kinematic time series data, and the process by which they are measured complicates their interpretation on a swing-by-swing basis.

## 2.2 Data Cleaning

The initial dataset includes attempted bunts (which are labeled) and attempted *check-swings* (i.e., partial swings, not labeled). We aim to limit our analysis to full swings. When reporting aggregate metrics like average bat speed, MLB filters out each batter’s bottom 10% of swings by bat speed, labeling the remaining swings as *competitive*. We take a more inclusive approach by filtering out only swings with bat speed below 50 mph (which filters out all bunt attempts). Based on video review, this cutoff does a good job of distinguishing between full and partial swings while only removing  $\approx 2.5\%$  of swings.

We reproduce MLB’s derived metric for *squared-up* contact as:

$$\text{batted ball speed} > 80\% \times \{1.23 \cdot (\text{bat speed}) + 0.23 \cdot (\text{pitch speed})\}. \quad (1)$$

The right-hand side of Equation (1) is a rough approximation for 80% of the theoretical maximum batted ball speed given bat speed and pitch speed at contact, with additional assumptions about the mass of bat and ball (Nathan, 2000). We observe batted ball speeds as high as 110% of this rough theoretical maximum.

## 2.3 Building Blocks

The present work relies on some models that are not novel, but nevertheless warrant explanation for an audience not steeped in baseball analytics literature.

### 2.3.1 Linear Weights

Linear weights (Thorn and Palmer, 1984) are the standard metric for the run value of a plate appearance outcome (strikeout, walk, single, etc.). They are based on the concept of base-out run expectancy, which models the progression of an inning as a Markov chain between base-out states, meaning that the state of the Markov chain is defined by which bases are occupied by runners and by the number of outs. For each state, the run expectancy is defined as the expected number of runs scored from that state until the terminal state (end-of-inning, three outs). The transition probability function is generally calculated using the league-wide empirical frequency of transitions between base-out states across a large, multi-year sample of data.

The linear weight of a plate appearance outcome is defined as the average change in base-out run expectancy that occurs on that outcome. For example, the linear weight of a walk is the average change in base-out run expectancy across all walks in the data. Some walks may result in greater changes in run expectancy (e.g. with bases loaded and zero outs) or lesser changes in run expectancy (e.g. with bases empty and two outs), but the linear weight does not depend on the context in which the outcome occurs.

For the present work, we followed the above methodology to calculate linear weights using all regular season MLB data from 2022 through 2024.

### 2.3.2 Hit Outcome Model

A hit outcome model is a regression model which predicts the outcome of a fair ball based on batted ball tracking data. For example, most hard-hit line drives result in hits, but a batter might hit a hard line drive that results in an out because it is unluckily directed right where a fielder is positioned.

We use the gradient boosting hit outcome model from Powers and Iglesias (2023). For this model, the response variable is the change in base-out run expectancy resulting from the batted ball, and the features are batted ball exit speed, vertical/launch angle and horizontal/spray angle. It is trained on all MLB regular season batted balls from 2022 through 2024. The hyperparameters are reported in Table 1. We call the predictions from this model *expected linear weight* (xLW).

Model	Description	XGBoost Hyperparameters			
		nrounds	eta	max_depth	min_child_weight
	Hit outcome model	1000	0.05	9	100
	<i>Pitch outcome model</i>				
{A}	Probability of swing	1500	0.05	9	10
{B}	Probability of HBP given no swing	400	0.05	6	10
{C}	Probability of contact given swing	1000	0.01	6	100
{D}	Probability of strike given called pitch	2000	0.01	9	10
{E}	Probability of fair ball given contact	1500	0.01	9	100
{F}	Expected xLW given fair ball	1000	0.01	6	100
	Hyperparameter search space	(0, 2000]	{0.01, 0.05, 0.3}	{3, 6, 9}	{10, 30, 100}

Table 1: *Descriptions and hyperparameters for hit outcome model and pitch outcome model components. Gradient boosting hyperparameters were tuned using 5-fold cross-validation, except for  $\gamma$  (0), `subsample` (0.65), and `colsample_by_tree` (0.7), which were fixed in advance.*

### 2.3.3 Pitch Outcome Model

We utilize the pitch outcome model from Powers and Iglesias (2023). This is a collection of six component models, as illustrated in Figure 3. Each component is a gradient boosting model with five contextual features (balls, strikes, batter side, strike zone top, and strike zone bottom) and nine features from the trajectory of the pitch (2D location at plate, 3D velocity vector at plate, 3D acceleration vector at plate, 1D extension at release). Five of the component models predict probabilities of binary outcomes, and one component model predicts the expectation of a continuous outcome.

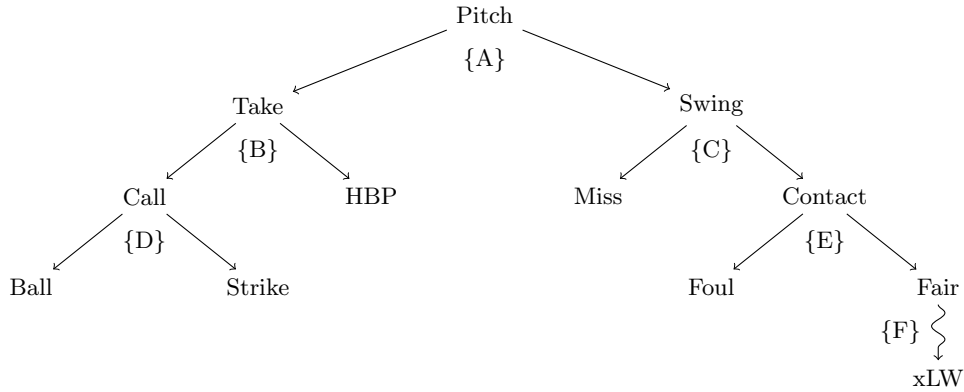


Figure 3: *A pitch outcome tree diagram for the components of the pitch outcome model. Each split in the tree is a binary split with a corresponding binary gradient boosting model, except for the final split, corresponding to component model {F}, for which the outcome is continuous.*

Table 1 provides details of the pitch outcome component models. We trained the pitch outcome model using the R package `predpitchscore` (Powers and Iglesias, 2025), which relies on the eXtreme Gradient Boosting (XGBoost) algorithm (Chen and Guestrin, 2016). Models were trained on MLB regular season pitches from 2022 through 2024 with 5-fold cross-validation for hyperparameter tuning.

### 3 Methods

#### 3.1 Intention Model

Acknowledging that swing timing biases the measurement of bat speed and swing length (see Figure 1), we first filter our dataset to consider only swings that result in squared-up contact against pitchers’ primary fastballs. In effect, this is a form of “matching” to mitigate the confounding bias of timing and pitch recognition on the observed bat speed and swing length measurements. We assume that this filtered dataset of 32,094 high-quality contact swings against primary fastballs captures the batter’s *intention*. Thus, we consider models fit to this filtered dataset as a way of capturing variation in the batter’s *intended* swing length and bat speed.

We hypothesize that two covariates explain real (non-artifactual) differences in swing mechanics: ball-strike count and pitch location. To the extent that a batter’s swing tracking metrics co-vary with count, we describe this as their *approach*. To the extent that a batter’s swing tracking metrics co-vary with pitch location, we describe this as swing *adaptation*. Additionally, based on visualizations of batter swing length and bat speed distributions (such as examples in Figure 4) we are also interested in capturing variation in the shape of the intended distributions for batters.

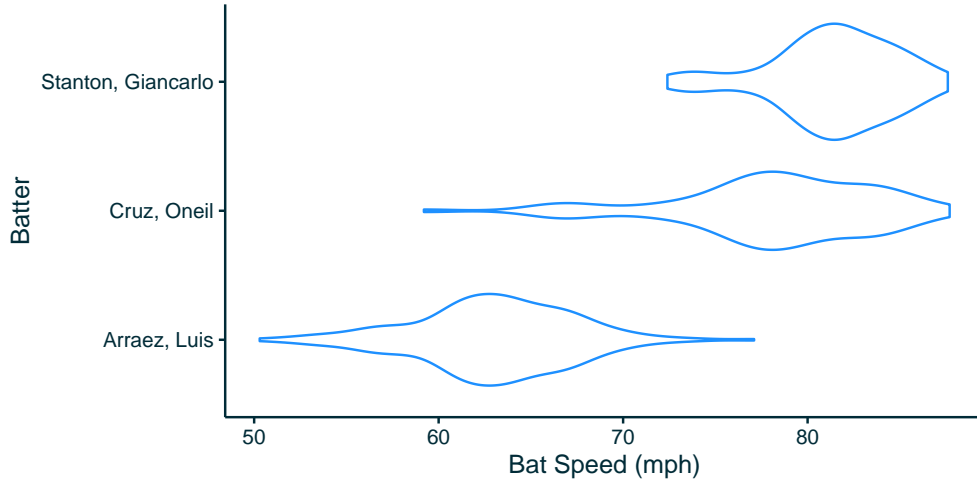


Figure 4: *Distribution of bat speed for three different example batters, considering only swings which resulted in squared-up contact against pitchers’ primary fastballs.*

To this end, we fit the following skew-normal (SK) multilevel model:

$$\begin{aligned}
 (\text{swing length})_i &\sim \text{SK}(\mu_i, \sigma, \alpha_i) \\
 \mu_i &= \mu_0 + \gamma_{p_i} + \gamma_{b_i} + \beta^B \cdot (\text{balls})_i + (\beta^S + \gamma_{b_i}^S) \cdot (\text{strikes})_i \\
 &\quad + (\beta^X + \gamma_{b_i}^X) \cdot (\text{pitch loc x})_i + (\beta^Z + \gamma_{b_i}^Z) \cdot (\text{pitch loc z})_i \\
 \alpha_i &= \alpha_0 + \nu_{b_i}
 \end{aligned} \tag{2}$$

where  $b_i$  denotes the batter on swing  $i$  against pitcher  $p_i$ . We use the same specification for both swing length and bat speed intention models.

In detail, the  $\gamma$  parameters are random intercepts and random slopes describing the mean of the swing

distribution such that

$$\begin{aligned}\gamma_{p_i} &\sim \mathcal{N}(0, \sigma_p^2) \\ \boldsymbol{\gamma}_{b_i} &\sim \mathcal{N}(\mathbf{0}, \Sigma)\end{aligned}\tag{3}$$

where  $\boldsymbol{\gamma}_{b_i}$  is the vector of batter random intercept and slopes  $(\gamma_{b_i}, \gamma_{b_i}^S, \gamma_{b_i}^X, \gamma_{b_i}^Z)$  which follow a multivariate normal distribution centered around the zero vector  $\mathbf{0}$  with covariance matrix  $\Sigma$ . This  $\Sigma$  contains the relevant random effect variances for the batter-level terms,  $(\sigma_b, \sigma_b^S, \sigma_b^X, \sigma_b^Z)$ , as well as their respective correlations. This allows us to quantify variation between batters and pitchers at the intercept level, as well as capturing variation in batter approach and swing adaptation.

The  $\nu_{b_i}$  parameters are random intercepts enabling us to capture variation between batters in the shape  $\alpha$  of the swing length and bat speed distributions, where

$$\nu_{b_i} \sim \mathcal{N}(0, \tau_b^2).\tag{4}$$

These models tell us: What are the bat speed and swing length (by batter, count and pitch location) when the timing is good? We interpret the prediction from this model on each pitch as the intended bat speed and swing length.

We fit both bat speed and swing length models in a Bayesian framework via the R package **brms** (Bürkner, 2021), which provides an interface for Bayesian modeling with Stan (Carpenter et al., 2017). We use weakly informative priors for the parameters, with vague half- $t_3$  priors (i.e., a Student’s  $t$  distribution centered at zero with 3 degrees of freedom but truncated to positive values) for the standard deviation parameters (Gelman, 2006) and the LKJ prior for the correlations between batter-level effects (Lewandowski et al., 2009). Our Bayesian approach naturally provides uncertainty quantification for the model parameters via their posterior distributions, which are estimated using MCMC sampling. For model fitting, we use four parallel chains with 6,000 (3,000 burn-in) and 4,000 (2,000 burn-in) samples for the swing length and bat speed models, respectively. We observe evidence of convergence of the MCMC algorithm based on trace plots and  $\hat{R}$  statistics close to 1 (Brooks and Gelman, 1998; Gelman and Rubin, 1992), along with no indication of problematic effective sample sizes (Gelman et al., 2020).

For comparison, we also fit simpler Gaussian versions of the swing length and bat speed models. These models follow a specification similar to that of the SK models in Equation (2), but with the shape parameter removed (i.e., the Gaussian distribution is a SK with shape 0). We use an 80/20 train/test split to compare the out-of-sample performance of the SK and Gaussian fits against each other. We compute the expected log pointwise predictive density (ELPD) on the 20% test data, which involves computing the log-likelihood of each test observation for each posterior sample. In addition to providing us with standard errors about the difference in model performance, we choose ELPD as our evaluation metric because of our interest in comparing the skewness of the SK posterior predictive distributions relative to the symmetric Gaussian.

### 3.2 Causal Model

The intention model from Section 3.1 sets us up to estimate a causal model for the effect of bat speed and swing length on swing outcomes via instrumental variables regression (Bollen, 2012). Relying on baseball domain knowledge, we hypothesize that the relationship between swing tracking metrics and outcomes is confounded by a latent, unobserved variable: whether the batter correctly identifies the pitch type early in its trajectory (by lucky guess or by other means) or needs to make a mid-swing adjustment. When the batter fails to correctly identify the pitch type early, he is less likely to make good contact, and he also adjusts his swing, reducing the bat speed and affecting the swing length. Critically, this confounder can vary



significantly from pitch to pitch as batters frequently limit their focus to one or two pitch types because it is too difficult to prepare for any pitch type (Gray, 2002).

A good instrumental variable is one which has an effect on the treatment (in this case, swing tracking metrics) and is known to be uncorrelated with the confounder (Bollen, 2012). As the instrument varies, we observe the effect of the treatment with the effect of the confounder stripped away. Relying on the framework from Section 3.1, we choose the batter’s count-intended bat speed and swing length as our instrumental variables. Count-intended swing length is the player-specific contribution of count to swing length, as estimated by (2), and count-intended bat speed is defined analogously.

Count-intended swing metrics affect observed swing metrics as modeled in Section 3.1, and we expect that they are approximately uncorrelated with early pitch recognition. The count is known before the pitch and does not depend on the pitch itself. Precise independence may not be a valid assumption, for example, if batters who tend to struggle more with early pitch recognition in two-strike counts tend also to modify their two-strike swing intentions a certain way. But this choice of instrumental variables mitigates any confounding effect of early pitch recognition from pitch to pitch.

We consider the effect of bat speed and swing length on three components of the pitch outcome model (see Figure 3): {C} probability of contact conditioned on swing; {E} probability of fair ball conditioned on contact; and {F} expected xLW conditioned on fair ball (xLW is explained in Section 2.3.2). For each of these components, we estimate a logistic or linear regression of the outcome on the batter’s count-intended bat speed and swing length, with an offset for the prediction from the pitch outcome model described in Section 2.3.3. This offset controls for the difficulty of each pitch when estimating the effects of count-based bat speed and swing length adjustments.

Using  $i \in \{1, \dots, n\}$  to index swings, the random variables  $Y_i^{\text{con}} \in \{0, 1\}$ ,  $Y_i^{\text{fair}} \in \{0, 1\}$  and  $Y_i^{\text{xLW}} \in \mathbb{R}_+$  represent whether swing  $i$  results in contact; whether swing  $i$  results in a fair ball; and the xLW of the batter ball on swing  $i$  (undefined if  $Y_i^{\text{fair}} \in \{0, 1\} = 0$ ), respectively. The corresponding predictions from the pitch outcome model are denoted by  $\hat{p}_i^{\text{con}} \in (0, 1)$ ,  $\hat{p}_i^{\text{fair}} \in (0, 1)$  and  $\hat{y}_i^{\text{xLW}} \in \mathbb{R}$ . We use  $b_i$  to denote the batter on swing  $i$ , and  $\hat{\gamma}_{b_i}^{\text{BS}}$  and  $\hat{\gamma}_{b_i}^{\text{SL}}$  are the estimated batter random slopes for strikes in the bat speed intention model and the swing length intention model (2), respectively.

$$\log \left( \frac{\mathbb{P}(Y_i^{\text{con}} = 1)}{1 - \mathbb{P}(Y_i^{\text{con}} = 1)} \right) = \log \left( \frac{\hat{p}_i^{\text{con}}}{1 - \hat{p}_i^{\text{con}}} \right) + \alpha^{\text{con}} + \beta_{\text{BS}}^{\text{con}} \cdot \hat{\gamma}_{b_i}^{\text{BS}} + \beta_{\text{SL}}^{\text{con}} \cdot \hat{\gamma}_{b_i}^{\text{SL}}, \quad (5)$$

$$\log \left( \frac{\mathbb{P}(Y_i^{\text{fair}} = 1 \mid Y_i^{\text{con}} = 1)}{1 - \mathbb{P}(Y_i^{\text{fair}} = 1 \mid Y_i^{\text{con}} = 1)} \right) = \log \left( \frac{\hat{p}_i^{\text{fair}}}{1 - \hat{p}_i^{\text{fair}}} \right) + \alpha^{\text{fair}} + \beta_{\text{BS}}^{\text{fair}} \cdot \hat{\gamma}_{b_i}^{\text{BS}} + \beta_{\text{SL}}^{\text{fair}} \cdot \hat{\gamma}_{b_i}^{\text{SL}}, \quad (6)$$

$$Y_i^{\text{xLW}} \mid \{Y_i^{\text{fair}} = 1\} \sim \text{Normal}(\hat{y}_i^{\text{xLW}} + \alpha^{\text{xLW}} + \beta_{\text{BS}}^{\text{xLW}} \cdot \hat{\gamma}_{b_i}^{\text{BS}} + \beta_{\text{SL}}^{\text{xLW}} \cdot \hat{\gamma}_{b_i}^{\text{SL}}, \sigma^2). \quad (7)$$

In each of these models, we interpret  $\alpha$  as an intercept to account for the fact that the mean outcome may not be the same in this dataset as in the dataset on which the pitch outcome model was trained. We interpret  $\beta_{\text{BS}}$  and  $\beta_{\text{SL}}$  as the effects of bat speed and swing length adjustments on the outcomes. The parameter  $\sigma^2$  in the xLW model is a nuisance parameter.

The Gaussian distribution assumed by (7) is likely not true. A substantial portion of batted balls are routine outs ( $\text{xLW} \approx -0.27$ , the minimum) while some batted balls are likely home runs ( $\text{xLW} > 1$ ). The unconditional distribution of xLW is right-skewed with skewness 1.55. This violation of the Gaussian assumption does not cause problems for the point estimates of the regression coefficients in (7), but it does cause our methodology to under-estimate their standard errors.

### 3.3 Run Value Model

The models from Section 3.2 estimate the effect of swing adjustments on contact and power, but they don't help us make value judgments when these effects move in opposite directions, e.g. sacrificing power for increased contact. In baseball, the units of value are runs, which translate directly to wins (James, 1983). The standard metric for estimating the run value of batting performance is linear weight, which assigns a run value to each type of batter outcome (strikeout, walk, home run, etc.) according to its average change in base-out run expectancy (Thorn and Palmer, 1984). In this section, we describe a method for converting the effects on contact and power into effects on linear weight.

We model the progression of a plate appearance as a Markov reward process whose nonterminal states are all possible ball-strike counts  $\mathcal{S}_N \equiv \{0, 1, 2, 3\} \times \{0, 1, 2\}$  and whose terminal states are  $\mathcal{S}_T \equiv \{\text{Strikeout, Walk, Hit by Pitch, Fair Ball}\}$ . From Section 3.2, we have the means by which to estimate the probability of each possible outcome of a pitch given its trajectory and the batter's approach, assuming an otherwise average batter. These probabilities induce transition probabilities between the states  $\mathcal{S} \equiv \mathcal{S}_N \cup \mathcal{S}_T$  of the Markov reward process because the pitch outcome (HBP, Called Ball, Called Strike, Swinging Strike, Foul Ball or Fair Ball) probability is the product of conditional probabilities from models {A}–{E}, and the state transition is a deterministic function of the pitch outcome. We use  $\hat{q}(s'|s, \vec{x}, \vec{\gamma})$  to denote the estimated transition probability from  $s \in \mathcal{S}$  to  $s' \in \mathcal{S}$  induced by the estimated pitch outcome probabilities given pitch trajectory features  $\vec{x} \in \mathbb{R}^9$  and batter approach  $\vec{\gamma} \equiv (\gamma^{\text{BS}}, \gamma^{\text{SL}}) \in \mathbb{R}^2$ .

Our goal is to estimate the expected run value of a plate appearance given any batter approach  $\gamma$ . We assume that the trajectory of each pitch in the plate appearance is drawn from the leaguewide empirical distribution of pitch trajectories in the 2024 MLB regular season, conditioned on the ball-strike count. Indexing these pitches by  $i = 1, \dots, n$ , we use  $s_i \in \mathcal{S}_N$  to denote the ball-strike count of pitch  $i$ , and we use  $\vec{x}_i \in \mathbb{R}^9$  to denote the trajectory features of pitch  $i$ . Under the foregoing assumption, the transition probability function given batter approach  $\vec{\gamma} \in \mathbb{R}^2$  is

$$\tilde{q}(s'|s, \vec{\gamma}) = \frac{\sum_{i=1}^n \mathbb{I}\{s_i = s\} \hat{q}(s'|s, \vec{x}_i, \vec{\gamma})}{\sum_{i=1}^n \mathbb{I}\{s_i = s\}} \text{ for } s \in \mathcal{S}_N, s' \in \mathcal{S}.$$

For three terminal states, we used fixed linear weights calculated using 2022-2024 MLB data (as detailed in Section 2.3.1):  $\text{LW}(\text{Strikeout}) = -0.27$ ,  $\text{LW}(\text{Walk}) = +0.33$ , and  $\text{LW}(\text{Hit by Pitch}) = +0.36$ . The linear weight of the final terminal state, Fair Ball, depends on the count and the batter approach through the pitch outcome model predictions. The earlier assumption that the trajectory of each pitch is drawn from the leaguewide distribution of trajectories, conditioned on count, implies that

$$\text{LW}(\text{Fair Ball} \mid s, \vec{\gamma}) = \frac{\sum_{i=1}^n \mathbb{I}\{s_i = s\} \hat{E}[Y_i^{\text{xLW}} \mid s_i, \vec{x}_i, \vec{\gamma}]}{\sum_{i=1}^n \mathbb{I}\{s_i = s\}}.$$

Finally, we define the reward function of the Markov reward process as

$$r(s'|s, \vec{\gamma}) = \begin{cases} \text{LW}(s) & \text{if } s' \in \{\text{Strikeout, Walk, Hit By Pitch}\} \\ \text{LW}(\text{Fair Ball} \mid s, \vec{\gamma}) & \text{if } s' = \text{Fair Ball} \end{cases}.$$

We calculate the value of each state by iteratively solving the Bellman equation (Bellman, 1957); and the corresponding value of the initial state  $s = (0, 0)$  gives us the estimated run value of the approach  $\vec{\gamma}$ .

## 4 Results

### 4.1 Intention Model

The out-of-sample ELPD results are shown in Table 2 for both swing length and bat speed models. For both settings, the SK model displayed an ELPD of around three standard errors better than the Gaussian model. The remaining results in this manuscript are based on the SK intention models.

	$\Delta\text{ELPD}$	SE	# SEs
Swing length	34.13	11.72	2.91
Bat speed	87.60	27.88	3.14

Table 2: *Model comparison via expected log predictive density (ELPD). Shown are the difference in ELPD between the skew-normal (SK) and Gaussian models ( $\Delta\text{ELPD}$ ), the standard error of ELPD difference (SE), and the number of standards errors by which the SK model outperforms the Gaussian model.*

Table 3 summarizes the fixed effects of the bat speed and swing length intention models. We observe intuitive relationships for both models with regard to the relationships between pitch-level context and each observed swing measurement. As the number of balls increase ( $\beta^B$ ), on average we observe faster and longer swings. The opposite behavior is observed as the number of strikes increases ( $\beta^S$ ), with slower and shorter swings on average. Figure 5 displays the joint distribution of the posterior mean for the batter random slopes for strikes ( $\hat{B}^S + \hat{\gamma}_b^S$ ) in both intention models. Regardless of the individual batter effect, we expect a decrease in bat speed and swing length per strike. A similar relationship is observed for pitches with higher vertical locations ( $\beta^Z$ ). The two models diverge with regard to the horizontal location of pitches ( $\beta^X$ ), as pitches that are thrown further away from the batter result in longer but slower swings on average. With regard to the shape parameter of the SK distribution, the intercepts ( $\alpha_0$ ) for both bat speed and swing length are negative, which indicate left-skewed distributions.

Parameter	Bat Speed			Swing Length		
	Mean	Lower	Upper	Mean	Lower	Upper
$\mu_0$	76.43	76.06	76.80	8.25	8.20	8.29
$\beta^B$	0.55	0.51	0.59	0.05	0.04	0.05
$\beta^S$	-1.13	-1.20	-1.06	-0.14	-0.14	-0.13
$\beta^X$	-0.52	-0.66	-0.39	0.16	0.14	0.17
$\beta^Z$	-1.87	-2.00	-1.75	-0.46	-0.47	-0.44
$\alpha_0$	-1.85	-1.96	-1.75	-1.46	-1.56	-1.36

Table 3: *Posterior mean and 95% credible interval lower/upper bounds for the fixed effects in the bat speed and swing length intention models.*

Next, Table 4 summarizes the posterior distributions of the random effect variance terms. For both bat speed and swing length intention models, we observe that the highest variation is attributed to differences between batters ( $\sigma_b$ ). Interestingly, the variation between batters in the shape random effects ( $\tau_b$ ) is the second highest for swing length but is smaller than the variation in the pitch location random slopes ( $\sigma_b^X$  and  $\sigma_b^Z$ ) for the bat speed model. The estimated terms demonstrate evidence of variability between batters at the intercept and slope level.

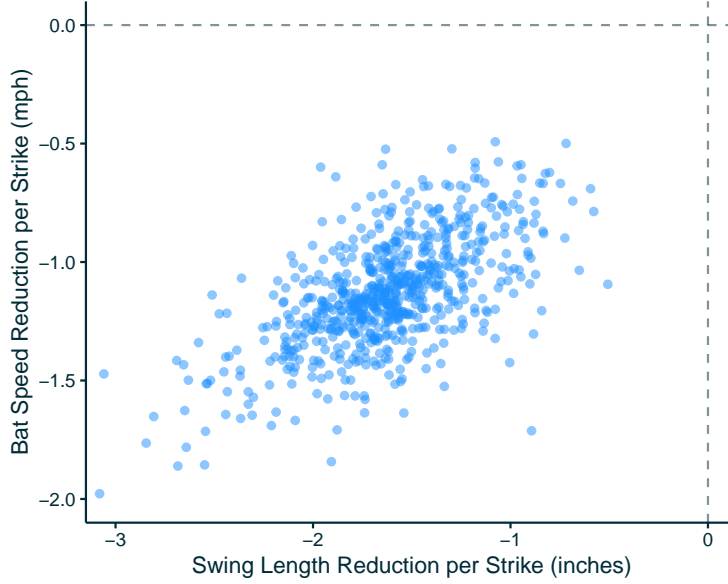


Figure 5: *Estimated batter random slopes for strikes in the intention model. Each point is a batter  $b$ , with the  $x$ -value representing  $\hat{\beta}^S + \hat{\gamma}_b^S$  from the model (2) and the  $y$ -value representing the same quantity from the analogous model for intended bat speed. This quantity is interpretable as the expected change in swing length (or bat speed) corresponding to a one-strike increase in the ball-strike count.*

Parameter	Bat Speed			Swing Length		
	Mean	Lower	Upper	Mean	Lower	Upper
$\sigma_p$	0.47	0.41	0.53	0.12	0.11	0.13
$\sigma_b$	3.35	3.04	3.67	0.43	0.40	0.47
$\sigma_b^S$	0.42	0.35	0.50	0.05	0.05	0.06
$\sigma_b^X$	1.04	0.90	1.18	0.11	0.10	0.13
$\sigma_b^Z$	1.05	0.94	1.17	0.13	0.11	0.14
$\tau_b$	0.67	0.52	0.85	0.34	0.08	0.57

Table 4: *Posterior mean and 95% credible interval lower/upper bounds for the standard deviation of the random effect terms in the bat speed and swing length intention models.*

Figure 6 provides an overview of the batter-level  $\mu_i$  random effects in the SK intention models. The variation we observe between batters aligns with intuition, as batters known for taking longer/faster swings (e.g., Oneil Cruz, Giancarlo Stanton, Javier Baez) display higher random intercepts ( $\gamma_b$ ) in comparison with batters known for shorter/slower swings (e.g., Luis Arraez). We observe positive correlations between the random slopes of bat speed and swing length, providing information on the variation between batters with respect to their approach ( $\gamma_b^S$ ) and adaptation ( $\gamma_b^X$ ,  $\gamma_b^Z$ ). For example, Anthony Rizzo’s approach appears to be more conservative, with the biggest decrease in bat speed and swing length for each additional strike relative to the average batter slope. We also observe batters that deviate from the consistent pattern between bat speed and swing length, such as Shohei Ohtani with regard to his horizontal pitch location random slope appearing to be higher than average for swing length despite a lower-than-average value for bat speed.

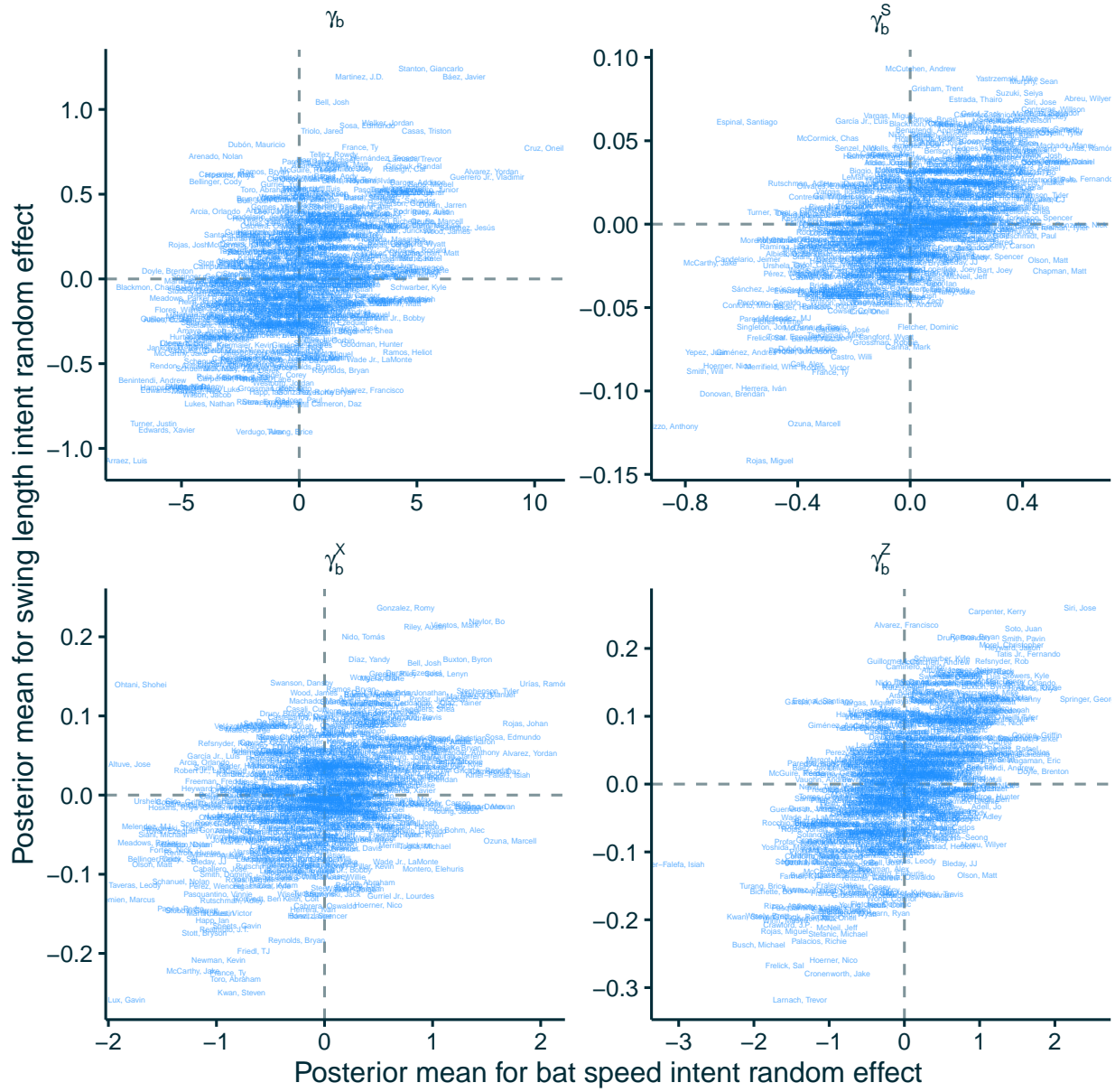


Figure 6: Joint distribution of the posterior means for the batter-level  $\mu_i$  random effects in the SK intention models for both bat speed and swing length. Each batter is displayed by their name, and only batters with at least 25 squared-up swings are displayed.

In contrast, Figure 7 displays the joint distribution of the posterior means for the batter random intercepts in modeling the shape parameter  $\alpha$  of the SK distribution. Unlike the mean-level random effects, we do not observe correlation between the bat speed and swing length shape-level random intercepts. The variation captured by these random intercepts provides some understanding of the differences in the intended swing distribution for players. Luis Arraez displays above-average random intercepts for both bat speed and swing length, indicating that his intention distributions are less skewed left than those of other batters.

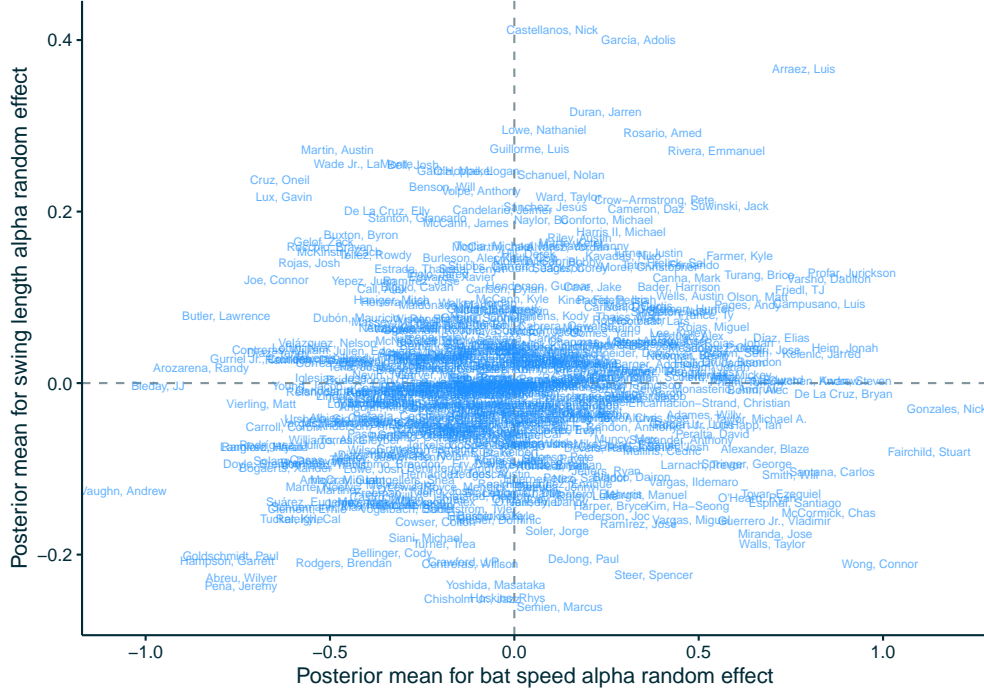


Figure 7: Joint distribution of the posterior means for the batter-level  $\alpha_i$  random effects in the SK intention models for both bat speed and swing length. Each batter is displayed by their name, and only batters with at least 25 squared-up swings are displayed.

## 4.2 Causal Model

The previous section demonstrates that there are differences between batters in how they modulate their swing length and bat speed as the number of strikes in the count increases. This section addresses the consequences of those differences between batters. As described in Section 3.2, we estimate three separate regression models with instrumental variables for contact probability given swing (5), fair ball probability given contact (6) and expected xLW given fair ball (7). The estimated effects of bat speed approach and swing length approach from these models are reported in Table 5.

	Contact Model (5)	Fair/Foul Model (6)	xLW Model (7)
Bat Speed Approach (mph)	$-0.176 \pm 0.016$	$-0.074 \pm 0.014$	$0.022 \pm 0.003$
Swing Length Approach (inches)	$-0.050 \pm 0.010$	$0.003 \pm 0.009$	$0.000 \pm 0.002$

Table 5: Estimated regression coefficients and corresponding standard errors from the causal models (5)–(7). Swing Length Approach is defined as the batter’s change in swing length per strike added to the count, as estimated by (2). Bat Speed Approach is defined as the batter’s change in bat speed per strike added to the count, as estimated by the analogous model for intended bat speed.

From the contact model, the negative coefficient  $-0.176$  for bat speed approach shows that batters who reduce their bat speed more per strike (i.e. more negative bat speed approach) make contact more often in one-strike and two-strike counts. Similarly, the negative coefficient for swing length approach shows that batters who reduce their swing length more per strike make contact more often in these counts. These results match our intuition from baseball domain knowledge: More conservative swings are more likely to

make contact, a more plausible conclusion than the counterintuitive results from the naive analysis shown in Figure 1. Bat speed approach and swing length approach have similar between-batter variance, so in comparing the coefficients we observed that bat speed approach has a much greater effect (more than 3x) on contact than does swing length approach. The fair ball probability model shows a similar effect direction for bat speed approach and no effect for swing length approach.

From the xLW model (i.e. the “power” model), the positive coefficient 0.023 for bat speed approach shows that batters who reduce their bat speed more per strike (i.e. more negative bat speed approach) exhibit less power in one-strike and two-strike counts when they do make contact. This demonstrates evidence for the intuitive tradeoff involved when taking more conservative swings: greater contact but less power. Swing length approach does not seem to have an effect on power.

The magnitudes of the coefficients in Table 5 are difficult to interpret because two of the models are logistic regressions, so the coefficients represent additive effects in the log-odds scale. For example, a pitch with almost 0% contact probability will still have almost 0% contact probability, regardless of approach. The effect of approach on contact is maximized for pitches near 50% contact probability. Figure 5 conveys the average effect size across all 2-strike pitches. We observe that the approach effects for almost all batters (relative to an average approach) are within  $\pm 4\%$  contact probability and  $\pm 0.04$  xLW.

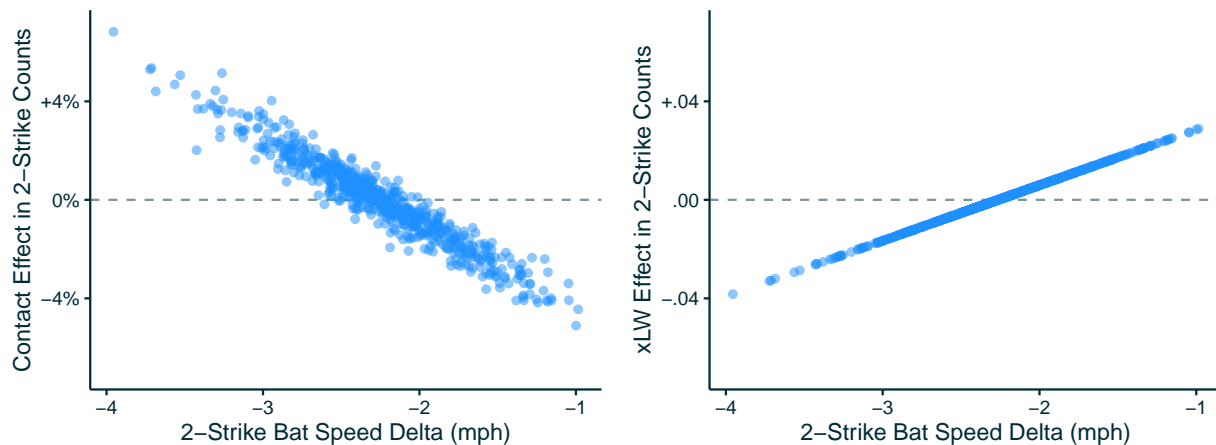


Figure 8: *Estimated causal effect of bat speed on contact (left) and power (right). Each point is a batter; the x-axis shows how much the batter slows their swing in two-strike counts, relative to zero-strike counts; the y-axis shows the estimated combined effect of the batter’s bat speed and swing length adjustments on contact rate (left) and xLW (right) in two-strike counts, from the models (5) and (7), respectively. xLW is expected linear weight, measured in runs per plate appearance (see Section 2.3.2).*

### 4.3 Run Value Model

The previous section demonstrates that there is a contact/power tradeoff involved when batters modulate their swing length and bat speed as the number of strikes in the count increases. This section addresses whether the tradeoff is worthwhile. As described in Section 3.3, we estimate the expected run value of plate appearance outcomes for the average batter if they were to adopt different approaches. Figure 9 plots batter approaches (as previously plotted in Figure 5) now with a color gradient in the background showing the estimated run value of each approach.





Figure 9: *Estimated causal effect of batters’ approaches, measured on the scale of runs per 500 plate appearances. This figure shows the same data as Figure 5 except that the background gradient reports the estimated run value of each approach. If an approach is valued at  $z$  runs, the interpretation is that the average batter with this approach is expected to produce  $z$  runs more than if they adopted an average approach, as measured by linear weights.*

The magnitude of the difference between the highest-value approaches and the lowest-value approaches is small, approximately 4 runs per 500 plate appearances. To put this in context, it amounts to a difference of roughly half a win per season (Slowinski, 2010). Baseball teams are chasing increasingly small competitive advantages, and the cost of half a win is roughly \$4 million in the free agent market (Clemens, 2021). We observe that the highest-value approaches are those which cut down on swing length without cutting down on bat speed. From Table 5, we observe that this is a way to avoid the contact/power tradeoff because cutting down on swing length (all else equal) increases contact without decreasing power.

Table 6 reports the specific approaches and corresponding run values of the top 5 and bottom 5 batters, ranked according to the run value of their approaches *applied to an average batter*. Interestingly, the top three batters (Chapman, Olson and Canha) all spent the first five or more seasons of their MLB careers around the same time with the Oakland Athletics.

## 5 Discussion

In this work, we use newly available bat tracking data to examine the relationship between swing aggression and outcomes in baseball. We recognize that both bat speed and swing length are inherently outcomes of the swing, confounded by whether or not the batter identified the pitch. To address this, we first model the batter’s bat speed and swing length intent via Bayesian hierarchical skew-normal models with evidence of more appropriate fits compared to simpler Gaussian models. Our results demonstrate the variation between batters in multiple aspects, including how they change their swing metrics as the number of strikes increases. We then condition on the batter’s count-intended bat speed and swing length as instrumental variables to estimate the causal effect of these intended swing metrics on swing outcomes. The results are intuitive,



	Batter	Approach		Runs / 500 PA
		Bat Speed (mph)	Swing Length (in.)	
1	Matt Chapman	-0.60	-1.96	1.39
2	Matt Olson	-0.64	-1.88	1.21
3	Mark Canha	-1.14	-2.51	1.13
4	Dominic Fletcher	-1.07	-2.36	1.12
5	Joey Bart	-0.83	-1.95	1.04
688	Jeimer Candelario	-1.71	-1.88	-1.52
689	Chas McCormick	-1.42	-1.00	-1.60
690	Trea Turner	-1.64	-1.54	-1.62
691	Jake McCarthy	-1.84	-1.91	-2.02
692	Santiago Espinal	-1.71	-0.89	-2.65

Table 6: *Top 5 and bottom 5 approaches, ranked by runs per 500 plate appearances. Bat Speed Approach and Swing Length Approach are the batter’s change in bat speed and swing length, respectively, per strike added to the count. We estimate the effect that each approach would have on the average batter’s performance, as measured on the run scale via linear weights.*

indicating that batters who shorten and slow down their swing more per strike are expected to make contact more often in one-strike and two-strike counts relative to other batters. However, we observe a tradeoff as this type of conservative approach at the plate is associated with less power. We quantify this tradeoff in terms of run value, revealing on an interpretable scale that the highest-value approaches cut down on swing length without sacrificing bat speed.

## 5.1 Other Sources of Swing Variation

While Section 4.3 illustrated the tradeoff involved with how batters vary their *approach*, there are of course other sources of variation in bat speed and swing length. For example, in Section 3.1 we define *adaptation* as the extent to which a batter’s swing tracking metrics co-vary with pitch location. As captured by  $\sigma_b^X$  and  $\sigma_b^Z$  in Table 4 and the posterior means of the batter-level random effects in Figure 6, we observe differences between batters in their swing *adaptation* as the pitch location varies horizontally and vertically. To demonstrate the variation, Figure 10 displays the expected swing length from the intention model by pitch location for two sample batters, Nico Hoerner and Juan Soto (assuming a 0-0 count). Hoerner displays a clear relationship with the vertical location with shorter intended swings on average for pitches that are located higher in the zone. In comparison, Soto displays a fairly consistent swing length throughout the zone, with an uptick in length for low and away pitches, versus a shortening of the swing for high and inside pitches. Understanding why batters display such variation in their *adaptation* is left for future work, and could be explored with swing path measurements describing the tilt and attack angle that were released during the 2025 MLB season (Petriello, 2025).

While our definitions for *approach* and *adaptation* are based on the random effects of the intention model, there is additional unexplained variation remaining. Figure 11 displays the standard deviation of the residuals for all swings (not just squared up on primary fastballs) relative to the intention model predictions. We observe clear differences in the remaining variation between batters. For example, Oneil Cruz displays higher residual standard deviations for both bat speed and swing length while Yandy Diaz deviates relatively less from his intention model swing length. We hypothesize that the variation of these residuals is largely

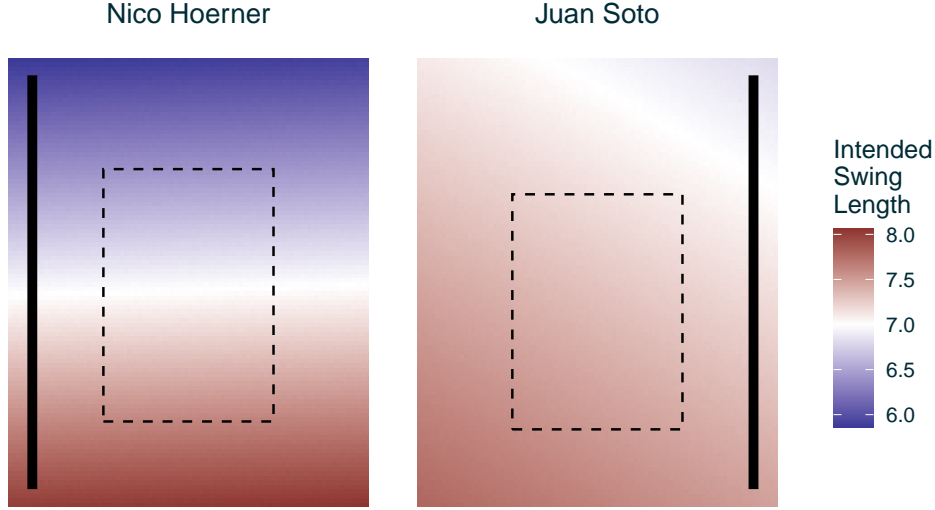


Figure 10: *Expected swing length by pitch location for two sample batters: Nico Hoerner (left) and Juan Soto (right), viewed from behind home plate. The thick vertical bars show the side of the plate on which the batter stands, and the dashed rectangle shows the strike zone, which depends on the height and stance of the batter. The color shows the predicted swing length from model (2) assuming a 0-0 count.*

due to batter *timing*. Further exploration of this is left for future work, which can include exploration of measurements for the “intercept point”, indicating where the bat makes contact with the ball with respect to the position of the batter’s stance and home plate, which was released by MLB during the 2025 season.

## 5.2 Future Work

We limited our analysis of the run value of approach to its effect on a batter of average skill, but it is possible that the results in Section 4.3 could differ depending on the profile of the batter. For example, the value of a specific approach could be different for high-contact, low-power batters versus low-contact, high-power batters. Future work could estimate personalized effects of batter approach based on the batter’s strengths and weaknesses. Additionally, the estimated run value was based on linear weights, which ignore the base-out situation. With a runner on third base and fewer than two outs, the value of putting the ball in play (relative to non-ball-in-play outcomes) increases, so presumably the tradeoff of increased contact for decreased power is more valuable in this situation. Future work could estimate this effect and evaluate the extent to which batters tailor their approach to the base-out situation.

Additional releases of metrics from MLB could help with future work on swing tracking data. Bat speed and swing length, measured at contact, provide only a hint at the full swing path and need to be interpreted carefully. In May 2025, MLB released three new pitch-by-pitch level metrics describing the orientation and movement of the bat at contact: attack angle, attack direction and swing tilt (Petriello, 2025). MLB has hinted that more metrics are in the works, such as contact depth and miss distance (Petriello, 2024). These two metrics would be particularly interesting because contact depth would help contextualize swing length and bat speed, and miss distance would provide significantly more precise data for batter and pitcher evaluation as compared with binary contact/miss outcomes.

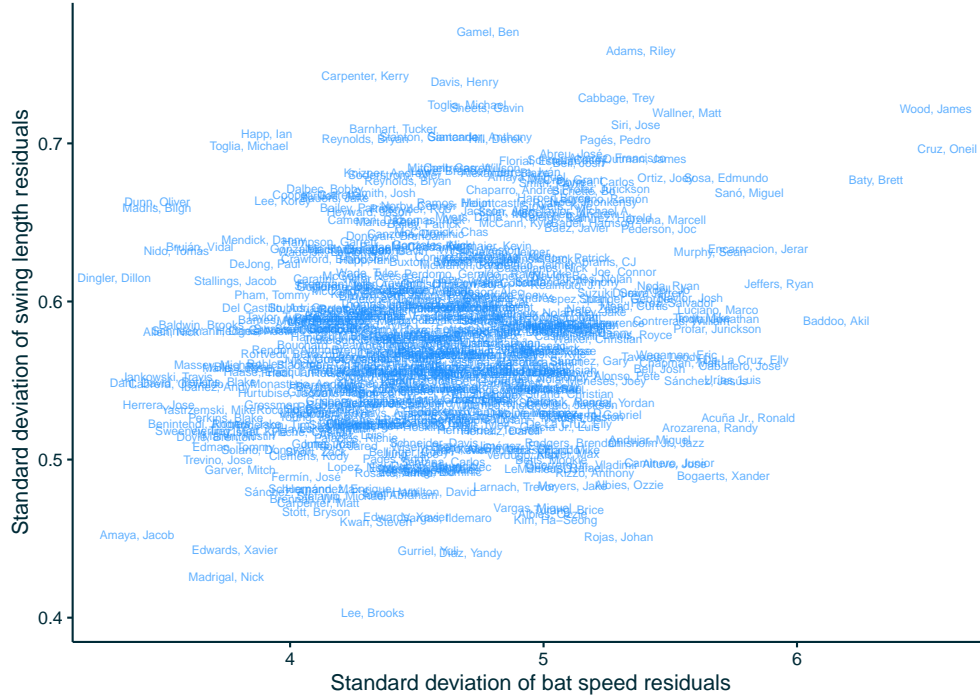


Figure 11: *Joint distribution of the standard deviations of the swing length (y-axis) and bat speed (x-axis) residuals for all swings (not just squared up on primary fastballs) relative to the intention model predictions. Each batter is displayed by their name, and only batters with at least 25 squared-up swings are displayed.*

## Acknowledgments

The authors thank Noah Woodward and Nick Wan for early stage discussions and feedback on this work.

## Code availability

All code related to this paper is available at <https://github.com/saberpowers/swinging-fast-and-slow>.

## References

- Arthur, R. (2016). The New Science of Hitting [April 13, 2016]. *FiveThirtyEight*. <https://fivethirtyeight.com/features/the-new-science-of-hitting/>
- Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.
- Bollen, K. A. (2012). Instrumental Variables in Sociology and the Social Sciences. *Annual Review of Sociology*, 38, 37–72. <https://doi.org/10.1146/annurev-soc-081309-150141>
- Brooks, S. P., & Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434–455. <https://doi.org/10.1080/10618600.1998.10474787>
- Bürkner, P.-C. (2021). Bayesian Item Response Modeling in R with brms and Stan. *Journal of Statistical Software*, 100, 1–54. <https://doi.org/10.18637/jss.v100.i05>

- Burton, T., & Powers, S. (2015). A linear model for estimating optimal service error fraction in volleyball. *Journal of Quantitative Analysis in Sports*, 11(2), 117–129. <https://doi.org/10.1515/jqas-2014-0087>
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76, 1–32. <https://doi.org/10.18637/jss.v076.i01>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Clemens, B. (2021). What Are Teams Paying Per WAR in Free Agency? [December 14, 2021]. *FanGraphs Baseball*. <https://blogs.fangraphs.com/what-are-teams-paying-per-war-in-free-agency/>
- Clemens, B. (2024). What Statcast’s New Bat Tracking Data Does and Doesn’t Tell Us [May 14, 2024]. *FanGraphs Baseball*. <https://blogs.fangraphs.com/what-statcasts-new-bat-tracking-data-does-and-doesnt-tell-us/>
- Elmore, R., & Matthews, G. J. (2022). Bang the Can Slowly: An Investigation into the 2017 Houston Astros. *The American Statistician*, 76(2), 110–116. <https://doi.org/10.1080/00031305.2021.1902391>
- Fast, M. (2010). What the heck is PITCHf/x? *The Hardball Times Baseball Annual*, 153–158. [https://tbt.fangraphs.com/wp-content/uploads/sites/8/images/pitchfx\\_guide.bt.pdf](https://tbt.fangraphs.com/wp-content/uploads/sites/8/images/pitchfx_guide.bt.pdf)
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3), 515–534. <https://doi.org/10.1214/06-BA117A>
- Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>
- Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., & Modrák, M. (2020). Bayesian Workflow. <https://doi.org/10.48550/arXiv.2011.01808>
- Goldbeck, G. (2023). Introducing Statcast 2023: High Frame Rate Bat and Biomechanics Tracking [March 27, 2023]. *MLB Technology Blog*. <https://technology.mlb.com/introducing-statcast-2023-high-frame-rate-bat-and-biomechanics-tracking-3844890264a6>
- Gray, R. (2002). “Markov at the Bat”: A Model of Cognitive Processing in Baseball Batters. *Psychological Science*, 13(6), 542–547. <https://doi.org/10.1111/1467-9280.00495>
- James, B. (1983). *The Bill James Baseball Abstract*. Ballantine Books.
- Judge, J. (2024). Exit Velocity: Modeling Average Maximum Athletic Effort [December 13, 2024]. *Baseball Prospectus*. <https://www.baseballprospectus.com/news/article/95454/prospectus-feature-exit-velocity-modeling-average-maximum-athletic-effort/>
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus, Giroux.
- Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9), 1989–2001. <https://doi.org/10.1016/j.jmva.2009.04.008>
- MLB Advanced Media. (2024). Statcast search. [https://baseballsavant.mlb.com/statcast\\_search](https://baseballsavant.mlb.com/statcast_search)

- Nakashima, H., Horiuchi, G., Kimura, A., & Sakurai, S. (2025). Acceptable range of timing error at bat-ball impact in baseball depends on the bat swing path. *Frontiers in Sports and Active Living*, 7, 1557145. <https://doi.org/10.3389/fspor.2025.1557145>
- Nathan, A. M. (2000). Dynamics of the baseball–bat collision. *American Journal of Physics*, 68(11), 979–990. <https://doi.org/10.1119/1.1286119>
- Nevins, D., Smith, L., & Kensrud, J. (2019). Sensitivity of batted-ball speed to swing speed models. *Proceedings of the Institution of Mechanical Engineers*, 233(3), 416–423. <https://doi.org/10.1177/1754337119835682>
- Orishimo, K. F., Kremenec, I. J., Modica Jr, E., Fukunaga, T., McHugh, M. P., & Bharam, S. (2024). Lower extremity kinematic and kinetic factors associated with bat speed at ball contact during the baseball swing. *Sports Biomechanics*, 23(12), 3406–3417. <https://doi.org/10.1080/14763141.2023.2269418>
- Orr, R. (2024). In Defense of Isaac Paredes [December 27, 2024]. *Baseball Prospectus*. <https://www.baseballprospectus.com/news/article/95688/best-of-bp-in-defense-of-isaac-paredes/>
- Petriello, M. (2024). Everything to know about Statcast’s new bat-tracking data [May 13, 2024]. *MLB.com*. <https://www.mlb.com/news/what-you-need-to-know-about-statcast-bat-tracking>
- Petriello, M. (2025). 4 new swing metrics tell us more than we’ve ever known about contact [May 19, 2025]. *MLB.com*. <https://www.mlb.com/news/new-statcast-swing-metrics-2025>
- Powers, S. (2025). sabRmetrics: Query {statsapi,baseballsavant}.mlb.com and fit fundamental sabermetric models. <https://github.com/saberpowers/sabRmetrics/tree/v1.0.4>
- Powers, S., & Iglesias, V. (2023). Pitch trajectory density estimation for predicting future outcomes. *Saberseminar*. <https://www.youtube.com/watch?v=IgPB4TITass>
- Powers, S., & Iglesias, V. (2025). Predpitchscore: Predictive Pitch Score [R package version 1.0.0]. <https://github.com/saberpowers/predictive-pitch-score>
- Putman, D., & Tolhurst, T. N. (2025). Tackling Endogeneity: Estimating Optimal Pass Rate in the NFL Using Instrumental Variables. <https://doi.org/10.2139/ssrn.5159158>
- Retrosheet. (2025). Retrosheet Event Files. <https://www.retrosheet.org/game.htm>
- Slowinski, P. (2010). Converting Runs to Wins [February 26, 2010]. *FanGraphs Baseball*. <https://library.fangraphs.com/misc/war/converting-runs-to-wins/>
- Swartz, P., Grosskopf, M., Bingham, D., & Swartz, T. B. (2017). The Quality of Pitches in Major League Baseball. *The American Statistician*, 71(2), 148–154. <https://doi.org/10.1080/00031305.2016.1264313>
- Thorn, J., & Palmer, P. (1984). *The Hidden Game of Baseball: A Revolutionary Approach to Baseball and Its Statistics*. Doubleday.
- Woodward, N. (2024). The radar gun doesn’t work here [June 7, 2024]. *The Advance Scout*. <https://theadvancescout.substack.com/p/the-radar-gun-doesnt-work-here>