

An Additive Model for Swing Manipulation

Steven Silverman

February 20, 2017

1 Introduction

The primary job of a pitcher is to get outs and prevent runs. The best way to achieve this goal is to prevent the batter from putting the ball in play. Two ways to do so are preventing swings at pitches that are likely to be called strikes, while *inducing* swings at pitches that are *unlikely* to be called strikes.

Absent granular batted-ball data, we can use called strike probability, as defined by TruMedia’s model, as a proxy for some nebulous notion of “contact quality”: that is, a pitch in the very middle of the strike zone is likely to be struck, and struck well, if swung at, while pitches way outside or in the dirt are likely to be tipped or missed entirely. As such, pitchers who manage to get batters to swing at such pitches will be better off, as will pitchers who prevent swings on middle-middle meatballs.

In this report, I define a metric for how well pitchers do just that, then construct a model to predict it using the other pitch-tracking data provided by TruMedia. Ideally, the model will provide insight into how certain pitchers maintain their effectiveness, as well as pointing towards specific areas players could work on to improve.

2 Methods

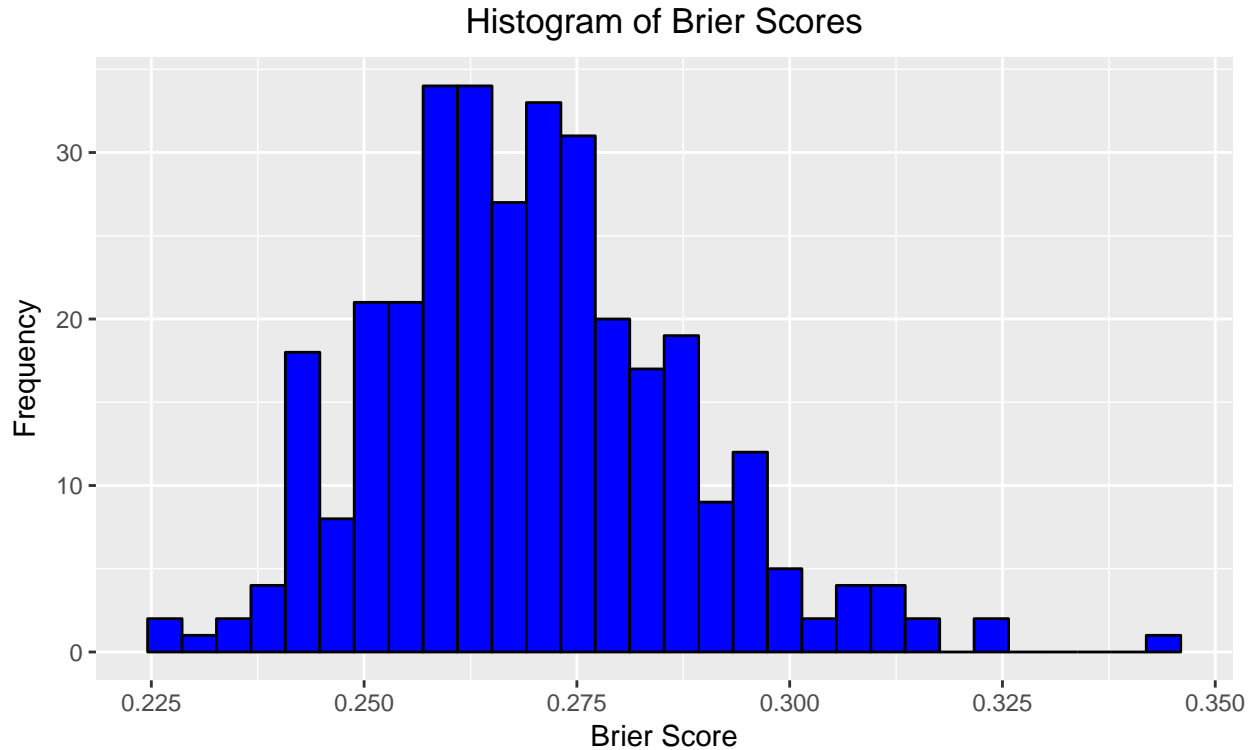
To measure how well pitchers manage swings, I used a scoring rule known as the Brier score. It is designed to measure how accurate probabilistic forecasts are, so in this instance TruMedia’s called strike probability will be the “prediction,” while the swing/no swing will be the outcome. For binary outcome variables like this one, the Brier score can be thought of as the mean squared error of the predictions. So, for example, predicting a called strike at 70% probability for a pitch where the batter swung would result in a Brier score of $(1 - 0.7)^2 = 0.09$ for that pitch.

Below are hitter and pitcher leaderboards for this metric the combination of the 2014 and 2015 seasons (pitchers on the left, hitters on the right). For pitchers, higher numbers are better (indicating more deception); for hitters, lower numbers are better (indicating a better eye). I used a cutoff of 1500 pitches, which corresponds to around 400 plate appearances/batters faced.

Table 1: Brier Score Leaderboards

Pitcher	Pitches	Brier Score	Batter	Pitches seen	Brier Score
Andrew Miller	2075	0.345	Brandon Belt	3557	0.202
Mark Melancon	2081	0.322	George Springer	3291	0.202
Luke Gregerson	1950	0.322	Freddie Freeman	4697	0.203
Greg Holland	1886	0.317	Chris Iannetta	2851	0.210
Will Smith	2191	0.314	Dexter Fowler	5072	0.211
Dellin Betances	2765	0.313	Andrew McCutchen	5453	0.213
Zach Putnam	1706	0.312	Yunel Escobar	3889	0.215
Brett Cecil	1745	0.311	Ryan Raburn	1684	0.217
Jose Fernandez	1779	0.311	Melvin Upton	3097	0.217
Sergio Romo	1837	0.307	John Jaso	2183	0.218

These leaderboards largely reinforce what we know about pitchers' filthiness and batters' eyes: Andrew Miller has possibly the best slider in the majors, Mark Melancon has a lethal cutter-curveball combination, and so on; meanwhile, Brandon Belt and Freddie Freeman were both in the top six in zone-swing percentage last year and draw a lot of walks. They also give a good feel for the magnitude of the scores in general. To get a better idea of the overall distribution, here's a histogram for the 333 pitchers who met the pitches-thrown cutoff:



It's a bit skewed right, meaning there are few excellent scores but no truly terrible ones. However, if you remove even just the top two players (Miller and Melancon), the distribution is quite close to symmetric.

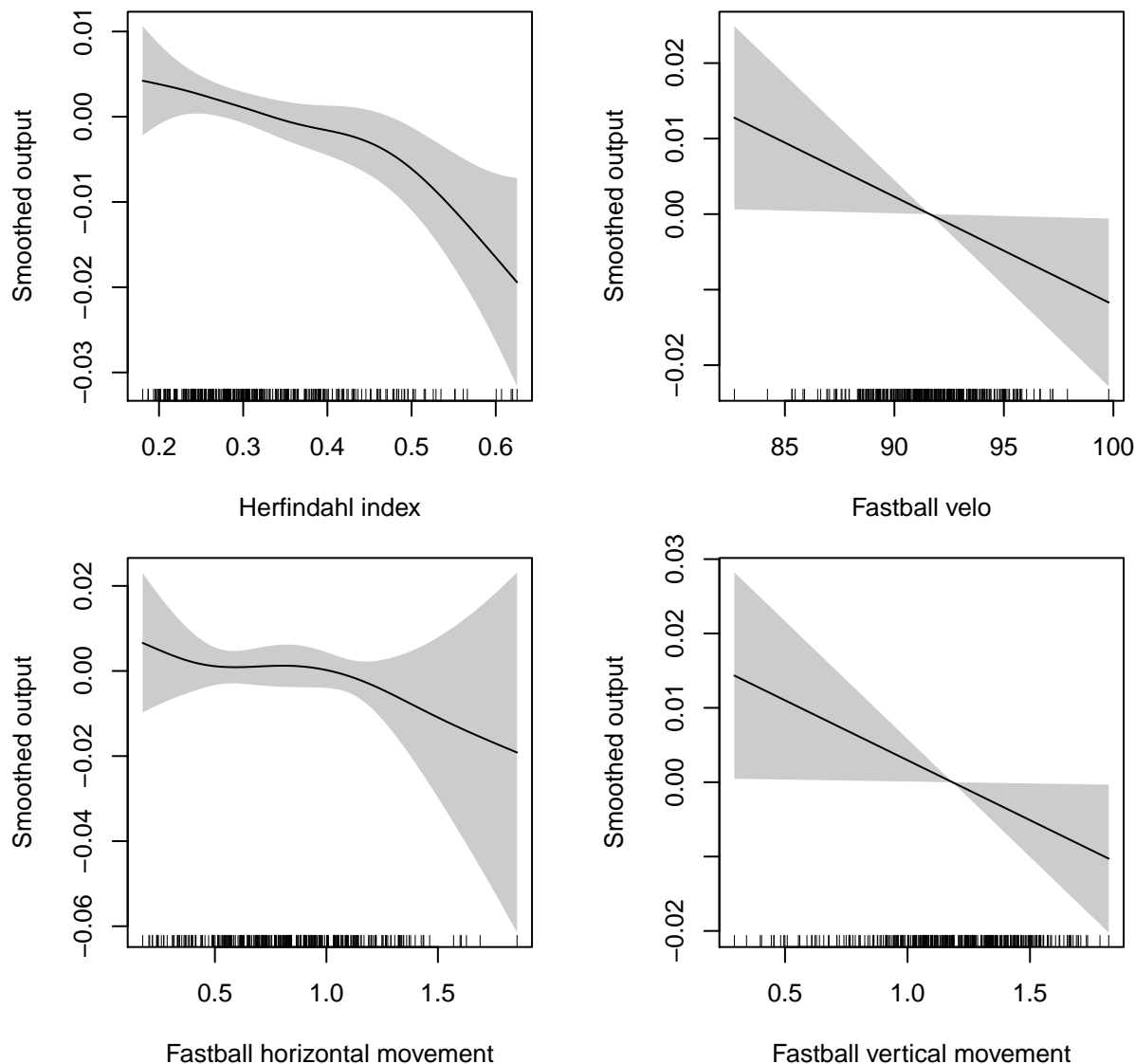
Next, I constructed a generalized additive model (GAM) to predict the Brier scores. I chose a GAM because it's not too computationally intensive to fit, allows for nonlinearity and other flexibility, and provides an easy way to model interactions between variables. I had a list of several different predictor variables, along with several combinations of the same, and I fit GAMs on the training data for all combinations (subject to a couple variables always being present). I used the generalized cross-validation score returned by the `gam()` function of the `mgcv` package in R to select the best model. The variables included in it are the following:

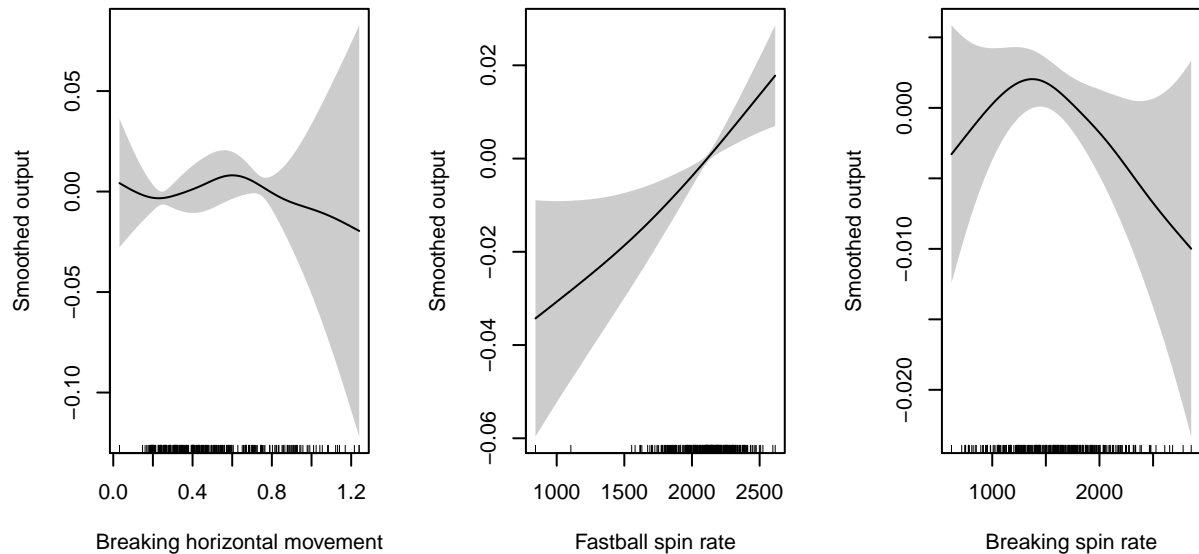
- Average fastball spin rate (using Brooks Baseball's definition of which pitches are considered "fastballs")
- Average breaking pitch spin rate (again using Brooks Baseball's definition)
- Average initial fastball velocity (measured by TruMedia at 50 feet from home plate)
- Average (magnitude of) horizontal fastball movement
- Vertical fastball movement, with gravity removed. This was not measured directly, but since position and velocity were available, I used the kinematics equations to calculate it. That requires some questionable assumptions, like constant acceleration during flight, but it's a sufficient proxy
- Average (magnitude of) horizontal breaking pitch movement
- Herfindahl index of pitch arsenal. This was designed to measure competition within an industry, but here measures how varied a pitcher's selection is. The usage percentages for each pitch are squared,

and then all the values are summed to get a number between 0 and 1, with 1 being only ever throwing a single pitch, and numbers closer to 0 indicating more variation.

- The interaction between fastball velocity and velocity on offspeed pitches
- The interactions among average fastball, breaking, and offspeed location (both horizontal and vertical)

With GAMs, after fitting the model one can examine the partial response functions (essentially, how the model changes for a single variable if all others are constant). Below are the partial responses for the individual variables. The shading shows two-standard-error confidence intervals for the smoothed values; the tick marks on the x-axis correspond to the training values.

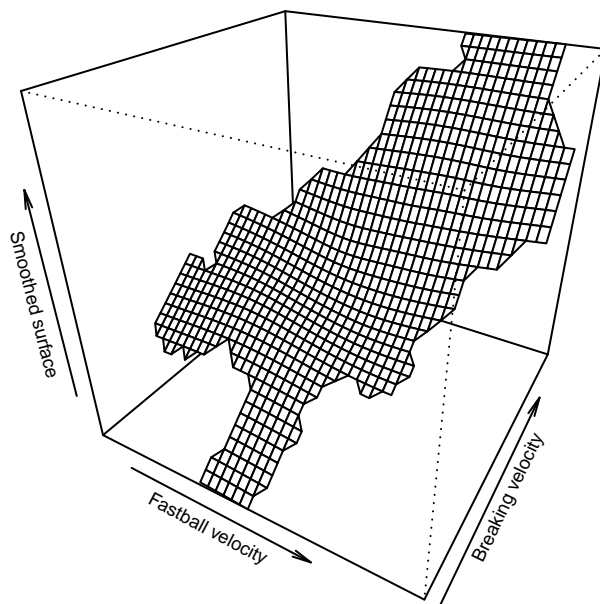




The most interesting takeaway from the partial responses are that getting more vertical movement on fastballs is negatively associated with Brier score, but having a higher spin rate is *positively* associated. It's not clear why two variables that are pretty well-correlated ($r = 0.55$) would have such dramatically different effects. The error bars on the vertical movement are fairly large, though, and what with the measurement error introduced by how I calculated that variable, it's possible the discrepancy is just a modeling artifact.

Also, pitchers with lower Herfindahl indexes had higher (better) Brier scores; one possible explanation is that elite pitchers who only throw two pitches must have two excellent pitches to survive in the majors, while having more variety allows each individual pitch to not be as outstanding.

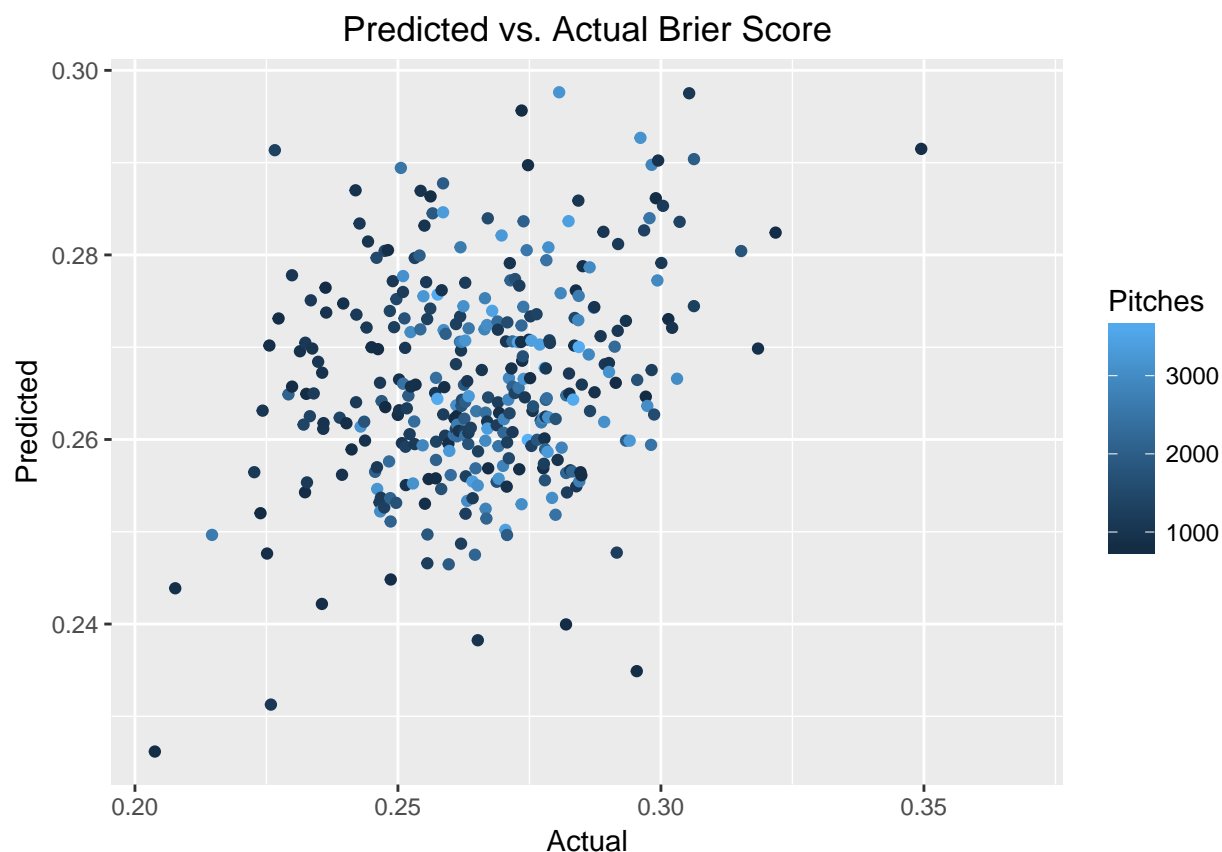
We can also create a three-dimensional plot for the interaction of fastball velocity and breaking velocity. The resulting surface shows how the two terms relate to each other and the predicted values.



In general, increased velocity for both fastballs and breaking pitches are positively associated with Brier score when taken together, with the effect being more pronounced for breaking pitches.

3 Results

Since the model was trained on 2014 and 2015 data, and we have 2016 available as well, we can use that as a test set to evaluate how well it predicts. The following is a plot of predicted versus actual Brier scores for the 2016 season, among pitchers with at least 750 pitches thrown:



There's clearly a positive association, and the correlation between the two sets is $r = 0.27$. More relevant, however, is the average prediction error. I chose to use mean absolute error, which just averages the magnitudes of the errors: the MAE for this model is 0.016. For a predicted value that has a theoretical range of 0 to 1 and a practical range of about 0.2 to 0.35, this is not bad at all, as it definitely serves to classify players into rough groupings of skill levels. (As expected, number of pitches thrown does not have a strong association with either score variable.)

4 Future Improvements

As usual for a hackathon, this project has a lot of room for extensions and improvements. Chiefly, I'd like to attempt some more modeling techniques which are more computationally expensive, like kernel regression or mixed models. It's tough with only a laptop and a couple weeks to really process the data properly. Also, it would be interesting to see how much the Brier score really affects performance for both pitchers and hitters: how well does it correlate with statistics like, say, wOBA (or wOBA against), or the offensive components of value statistics like WAR?

The current model does not control for catchers or umpires (unless TruMedia's called strike probability model already takes that into account), so adding that would help in future iterations. I'd also like to explore using different scoring rules to see if they're more predictive of actual performance, since the Brier score is one of many options. If I could merge this data with Statcast data like exit velocity, then the scores could become more granular: for example, allowing 80 MPH contact on a middle-middle pitch is less bad than allowing similar contact on a ball a foot outside. The same goes for launch angle.

Lastly, I would like to examine which individual player-pitches, rather than pitchers themselves, do the best. The metric devised here could serve as a sort of "nasty factor" showing how much hitters get fooled by a particular pitch. That's mainly a curiosity which fans would enjoy, but it could certainly have some predictive power for run prevention as well.