The Effectiveness of Using Out Probabilities to Measure Fielding in Baseball

Abstract

This paper determines the extent to which objectively measuring fielding in baseball using out probabilities provides an improvement over measuring fielding subjectively by scoring errors. Generalized linear mixed models show that objectively measuring fielding using out probabilities provides a 6.4% improvement in describing run prevention when compared to errors. The models use barrels, strikeouts, and walks as fixed effects, and obtain linear weights similar to established values, although the models find strikeouts to be slightly less valuable and errors more costly on defense. Furthermore, the models consider different stadiums as random effects and obtain generally similar results to park effects recently published by Statcast.

Introduction

In baseball, errors traditionally measure fielding. Errors are subjective, and do not measure how fast a fielder gets to a batted ball (a slow player who does not reach the ball to make a play cannot get an error).

Objective statistics are intended to improve measuring fielding. For example, "Outs Above Average (OAA)" (n.d.) measures the number of plays fielders make and their difficulties using out probabilities . Out probability is the likelihood of a hit or out based on a batted ball's exit speed and launch angle off the bat.

This report compares two models for predicting runs allowed. One uses errors to measure fielding, and the other uses out probabilities.

Methods

Game data from 2017–2020 was obtained from Retrosheet's Game Logs (2020). The response is *Runs* allowed, and predictors are Strikeouts (*K*); *Walks*; and Errors (*E*). *Walks* include intentional and unintentional walks, and hit-by-pitches.

Two additional predictors were obtained from Statcast, and merged into the dataset.[1] First, a "Barrel" (n.d.) is a batted ball with an ideal exit speed and launch angle, and 80% of Barrels result in hits (58% of Barrels are Home Runs). Second, a fielding metric (*Fld*) adds up how well fielders turned balls in play into outs based on out probabilities. For example, if a defense made an out on a 25% out probability play, they receive 0.75 points. If the defense did not make an out, 0.25 points were subtracted from the *Fld* score. The average *Fld* score per game is 0.016, ranging from -9.2 to 6.8, with a 1.9 standard deviation.

A correlation table is:

[1] Statcast data was scraped using the baseballr package developed by Petti (2021).

	Runs	HR	K	E	Barrels	Fld	Walks
Runs	1.0000000	0.60357089	-0.142410098	0.22090419	0.50702673	-0.4791406	0.383013864
HR	0.6035709	1.00000000	-0.051174551	0.02015249	0.68772349	-0.0778602	0.103860934
K	-0.1424101	-0.05117455	1.000000000	-0.02554938	-0.07449600	-0.1135270	0.005435108
E	0.2209042	0.02015249	-0.025549383	1.00000000	0.02482873	-0.2748278	0.095860225
Barrels	0.5070267	0.68772349	-0.074496005	0.02482873	1.00000000	-0.0142910	0.117315047
Fld	-0.4791406	-0.07786020	-0.113527007	-0.27482782	-0.01429100	1.0000000	-0.145263207
Walks	0.3830139	0.10386093	0.005435108	0.09586022	0.11731505	-0.1452632	1.000000000

The highest correlation is between *Runs* and Home Runs (HR). However, instead of Home Runs, *Barrels*—which indicate hit quality before the result is known—were used in the models. Home Runs can be random depending on the stadium. Many runs are scored in Colorado due to elevation (batted balls travel farther in high altitude), and some stadiums have more home runs due to smaller outfield dimensions (Monagan, 2014). Therefore, the stadium where the game was played is considered a random effect in mixed models.

In summary, the response, predictor variables, and random effect used in the models are:

Model 1 (using Errors to measure fielding):

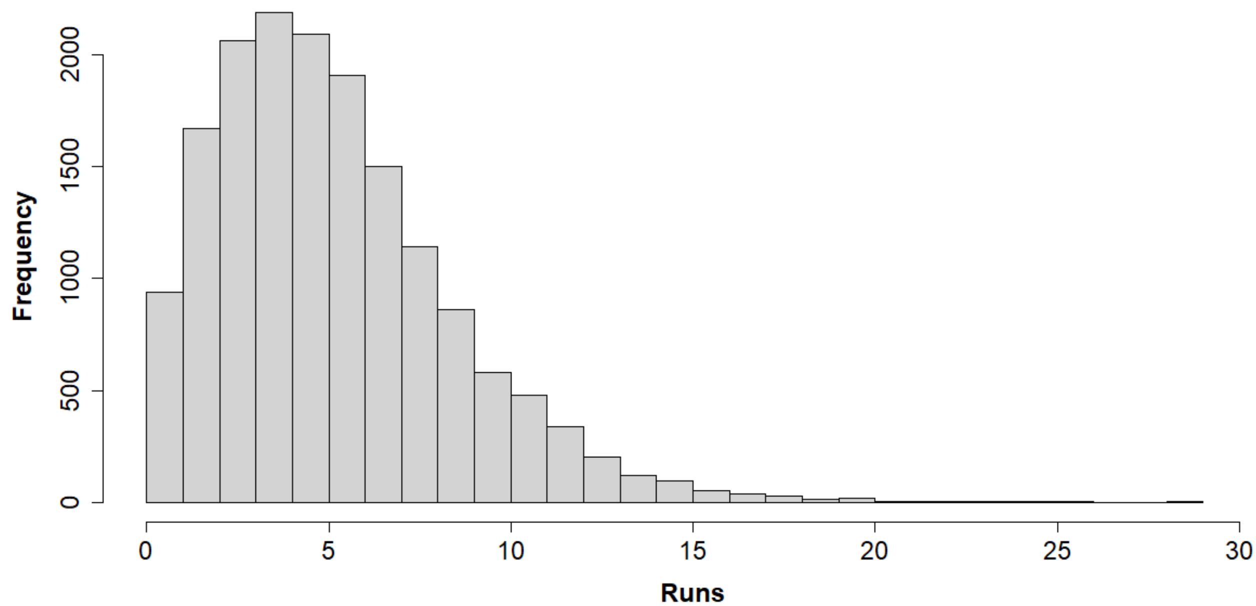
Variable	Type
Runs Allowed	Response
Barrels Allowed	Fixed Effect
Walks Allowed (includes Hit-by-Pitch & Intentional Walks)	Fixed Effect
Strikeouts	Fixed Effect
Errors	Fixed Effect
Stadium	Random Effect

Model 2 (using Out Probabilities to measure fielding):

Variable	Type
Runs Allowed	Response
Barrels Allowed	Fixed Effect
Walks Allowed (includes Hit-by-Pitch & Intentional Walks)	Fixed Effect
Strikeouts	Fixed Effect
Fielding measured by Out Probabilities (abbreviated *Fld*)	Fixed Effect
Stadium	Random Effect

Essentially, Barrels, Walks, and Strikeouts measure pitching performance analogous to Fielding Independent Pitching (FIP). Then, the fielding statistic (errors in the first model and *Fld* using out probabilities in the second model) and stadium effects are intended to account for the remaining variation in runs allowed.

The dataset contains 16,338 observations (8,169 games with two entries for each team). *Runs* allowed per game is counting data that is right-skewed:



The mean is 4.64 and the variance is 10.6. Because the response is counting data that is not normal and has overdispersion, negative binomial generalized linear mixed models (GLMMs) were fit using the *glmmTMB* package in R.

Results

The two models have the following output:

<u>Errors</u>

```
 Family: nbinom2  ( log )
Formula:          Runs ~ Walks + K + E + Barrels + (1 | ParkID)
Data: Baseball

     AIC      BIC   logLik deviance df.resid
 73363.8  73417.7 -36674.9  73349.8    16331

Random effects:

Conditional model:
 Groups Name        Variance Std.Dev.
 ParkID (Intercept) 0.003102 0.05569
Number of obs: 16338, groups:  ParkID, 32

Overdispersion parameter for nbinom2 family (): 12.9

Conditional model:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.875964   0.019244   45.52   <2e-16 ***
Walks        0.089171   0.001765   50.53   <2e-16 ***
K           -0.025766   0.001470  -17.53   <2e-16 ***
E            0.148383   0.005170   28.70   <2e-16 ***
Barrels      0.203583   0.002883   70.62   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

<u>Fld</u>

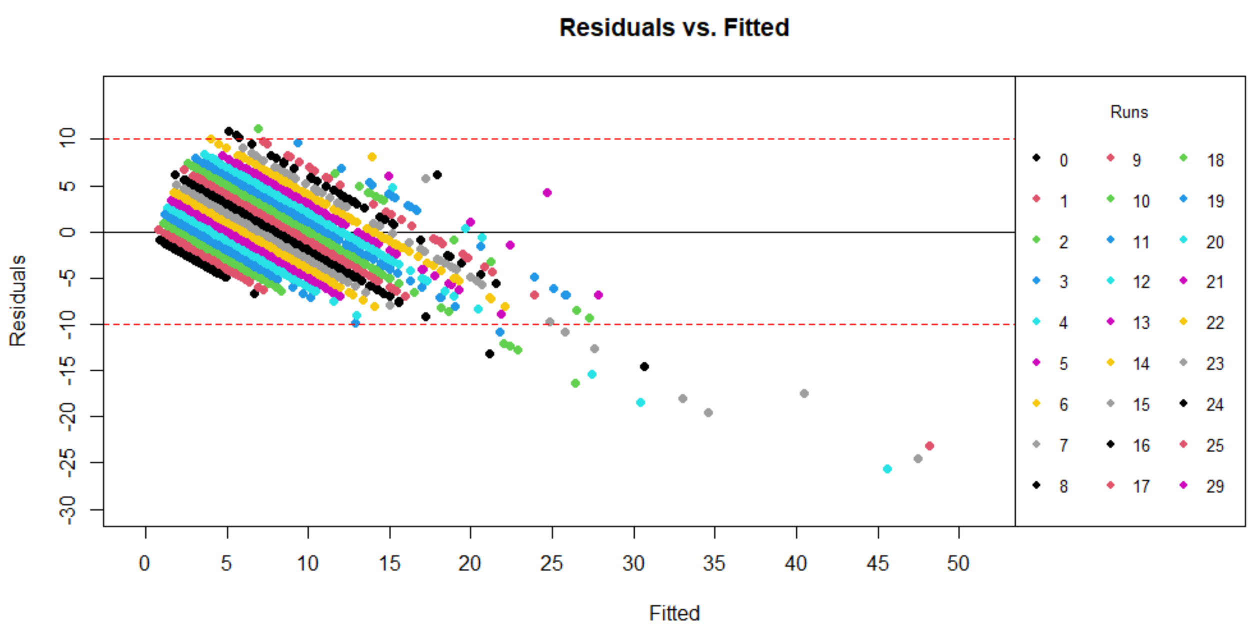
```
 Family: nbinom2  ( log )
Formula:
Runs ~ Walks + K + Fld + Barrels + (1 | ParkID)
Data: Baseball

     AIC      BIC   logLik deviance df.resid
 68649.9  68703.9 -34318.0  68635.9    16331

Random effects:

Conditional model:
 Groups Name        Variance Std.Dev.
 ParkID (Intercept) 0.001438 0.03792
Number of obs: 16338, groups:  ParkID, 32

Overdispersion parameter for nbinom2 family ():  141

Conditional model:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.105307   0.015579   70.95   <2e-16 ***
Walks        0.073076   0.001492   48.97   <2e-16 ***
K           -0.036149   0.001278  -28.28   <2e-16 ***
Fld         -0.152524   0.001896  -80.45   <2e-16 ***
Barrels      0.191758   0.002392   80.16   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The model using *Fld* as the fielding statistic reduces the deviance from 73,350 to 68,636. Both models have similar fixed effects coefficients and all of the fixed effects are highly significant. In the model using *E*, the standard deviation of the variance component for the stadium random effect is 0.056, and in the *Fld* model, the standard deviation is 0.038.

For the *Fld* model, exponentiated Best Linear Unbiased Predictors (BLUPs) are ranked as:

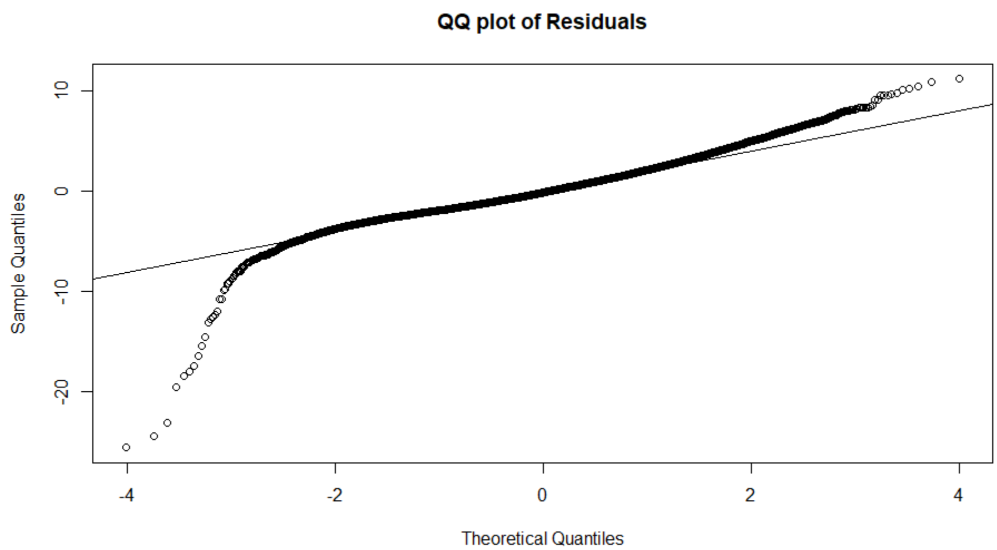
	Team	Stadium	exp_BLUP
1	Colorado Rockies	Coors Field	1.0807406
2	Cincinnati Reds	Great American Ballpark	1.0622354
3	Texas Rangers	Rangers Ballpark in Arlington	1.0546105
4	New York Yankees	Yankee Stadium II	1.0498611
5	Houston Astros	Minute Maid Park	1.0433716
6	Baltimore Orioles	Oriole Park at Camden Yards	1.0341307
7	Los Angeles Angels	Angel Stadium of Anaheim	1.0311173
8	Toronto Blue Jays	Sahlen Field	1.0182414
9	Chicago White Sox	Guaranteed Rate Field	1.0158292
10	Seattle Mariners	Safeco Field	1.0115278
11	San Diego Padres	PETCO Park	1.0065320
12	Arizona Diamondbacks	Chase Field	1.0038191
13	Atlanta Braves	SunTrust Park	1.0013063
14	Texas Rangers	Globe Life Field in Arlington	1.0002678
15	Milwaukee Brewers	Miller Park	0.9970425
16	Toronto Blue Jays	Rogers Centre	0.9946793
17	Cleveland Indians	Progressive Field	0.9944326
18	Minnesota Twins	Target Field	0.9925494
19	Oakland As	Oakland-Alameda County Coliseum	0.9920892
20	Philadelphia Phillies	Citizens Bank Park	0.9920677
21	Detroit Tigers	Comerica Park	0.9907220
22	Kansas City Royals	Kauffman Stadium	0.9902595
23	Chicago Cubs	Wrigley Field	0.9855759
24	New York Mets	Citi Field	0.9825991
25	Boston Red Sox	Fenway Park	0.9788995
26	Tampa Bay Rays	Tropicana Field	0.9756067
27	Los Angeles Dodgers	Dodger Stadium	0.9675260
28	Pittsburgh Pirates	PNC Park	0.9664935
29	St Louis Cardinals	Busch Stadium III	0.9633704
30	Washington Nationals	Nationals Park	0.9624212
31	San Francisco Giants	AT&T Park	0.9451326
32	Miami Marlins	Marlins Park	0.9358373

A diagnostic plot for the *Fld* model comparing fitted values to residuals is:



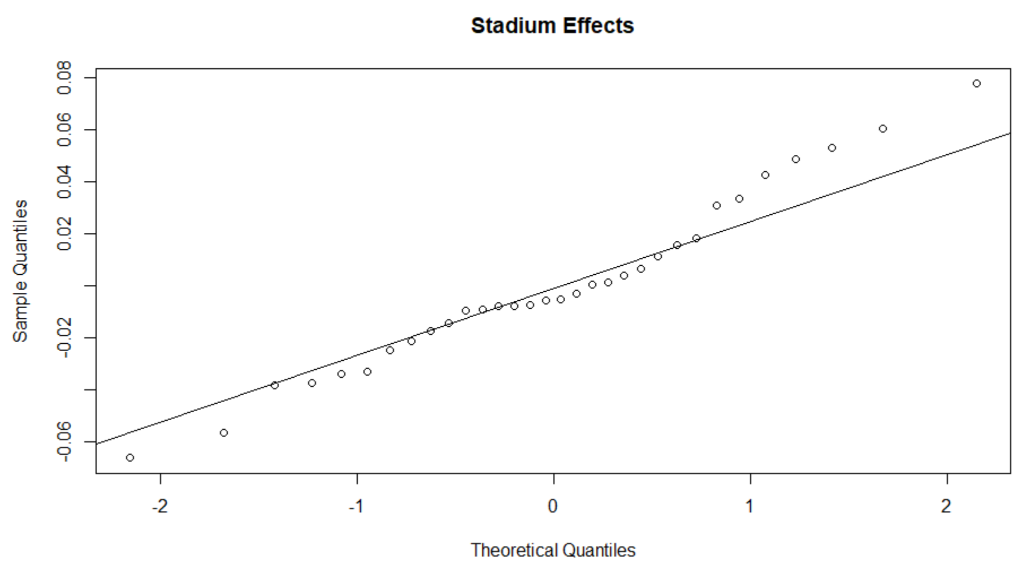
The fitted values equal the actual number of runs allowed on the black horizontal line where the residual is zero. The fitted values are almost always within ±10 runs of the actual runs allowed (as indicated by the red dotted lines) for every game where teams allowed 0–9 runs. At 10 runs or more, when the residuals are highest, the residuals are almost always negative meaning the model overestimated the number of runs in high scoring games.

The residuals Q-Q plot is:



The residuals generally stay on the straight line, although there is heavy-tailed behavior, with the most extreme negative residuals due to the model overestimating the runs allowed.

A Q-Q plot of the stadium random effect shows for the most part, a normal distribution:



Discussion

The model using *Fld* instead of *E* reduces the deviance by 6.4% (from 73,350 to 68,636). Therefore, objectively measuring fielding using Out Probabilities provides a modest improvement in explaining *Runs* allowed, compared to Errors.

The fixed effects coefficients are interpreted in relation to the average *Runs* allowed—4.64. For example, the coefficient for *Walks* in the *Fld* GLMM is 0.0731. Accordingly, an additional walk increases *Runs* by a factor of $e^{0.0731}$, or 1.076. This equals 0.352 more runs than average (i.e., (1.076 x 4.64) - 4.64 = 0.352).

Calculating the run values of all of the fixed effects this way provides values comparable to linear weights as reported by Tangotiger (2018):

Fixed Effect	*E* GLMM	*Fld* GLMM	Tangotiger (2017–2018, except Barrels)
Walks (includes Hit-by-Pitch)	0.433	0.352	0.30–0.36
Strikeouts	-0.165	-0.188	-0.288 to -0.283
Errors	0.742		0.46–0.55
Barrels	1.048	0.981	0.93[2]

The GLMMs find strikeouts as slightly less valuable for the defense, and errors to cause more runs than the reported values from the linear model.

The standard deviations of the variance components of the random effect in both models (0.056 and 0.038) are larger than the coefficients for strikeouts (-0.026 and -0.036). In Colorado, teams score the most runs—8% more per game than the average stadium.

The BLUPs ranking roughly corresponds with park effect rankings measured by Statcast and recently released (Petriello, 2021). The stadiums with the three highest BLUPs are also in the top ten in Statcast's "Park Factors Leaderboard" (n.d.) for 2017–2019 (Colorado #1; Texas #2; Cincinnati #10). There is also correspondence at the bottom of both lists (San Francisco and Miami are the last two in each). The largest discrepancy is Nationals Park (#30 GLMM vs. #3 Statcast).

Conclusion

When using GLMMs with stadiums as random effects, objectively measuring fielding using Out Probabilities provides a 6.4% improvement in describing run prevention compared to subjectively measuring fielding by Errors.

[2] Tangotiger (2017) reports a scaled run value for barrels of 1.428. The unscaled value is calculated as described in Weinberg (2016) to be 0.93 (i.e., 1.428/1.2-0.26=0.93), where 1.2 is a scaling coefficient and -0.26 is the value of an out.

References

"Barrel". (n.d.). MLB. http://www.mlb.com/glossary/statcast/barrel

"Game Logs". (2020, November 30). Retrosheet. http://www.retrosheet.org/gamelogs

"Outs Above Average (OAA)". (n.d.). MLB. http://www.mlb.com/glossary/statcast/outs-above-average

Monagan, M. (2014, April 28). *Baseball's many physical dimensions.* Cut4 by MLB.com. https://www.mlb.com/cut4/the-dimensions-of-every-big-league-ballpark-in-this-cool-infographic/c-73642276

"Park Factors Leaderboard". (n.d.). Savant. http://baseballsavant.mlb.com/leaderboard/statcast-park-factors?type=year&year=2019&batSide=&stat=index_wOBA&condition=All&rolling=

Petriello, M. (2021, April 29). *A new way to dissect baseball's park factors.* MLB. http://www.mlb.com/news/park-factors-measured-by-statcast

Petti, B. (2021, February 19). *Baseballr 0.8.4.* http://billpetti.github.io/baseballr

Tangotiger [tangotiger]. (2018, September 7). Linear Weights Run Values year by year since 2014 [Tweet]. Twitter. https://twitter.com/tangotiger/status/1038232219595235328

Tangotiger. (2017, January 16). Re: Statcast Lab: Barrels [Blog comment]. http://www.tangotiger.com/index.php/site/comments/statcast-lab-barrels#30

Weinberg, N. (2016, April 11). *The beginner's guide to deriving wOBA*. Fangraphs. http://library.fangraphs.com/the-beginners-guide-to-deriving-woba