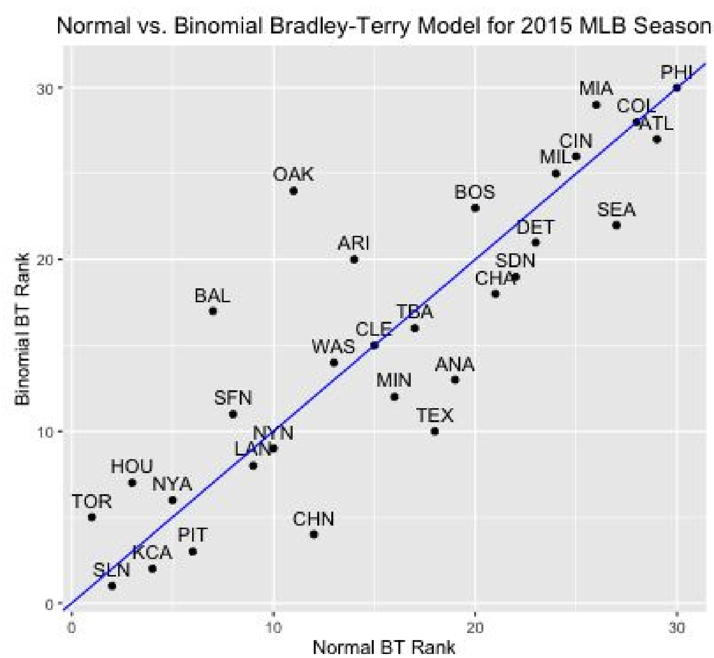
Applying the Bradley-Terry Model to the 2015 MLB Season

With a 162-game sample in Major League Baseball, the best teams separate themselves from the pack more often than in other sports. Thus, oftentimes, it is enough to look at the final regular season standings to essentially rank how good teams are. However, there is one caveat: in baseball, teams play an unbalanced schedule. Teams play 19 games against each of the four teams in their division – just under 50 percent of all of their regular season games. In order to develop the best ranking of teams across all divisions, we need to account for the competition each team faces. (Just compare the 2015 NL Central, which had three teams at at least 97 wins, to the 2015 AL West, whose winner had just 88 wins.) The Bradley-Terry model is effective in this situation because it adjusts each team's game-by-game results by how good its opponent is.

To assess the strength of each team in the league, we fitted two Bradley-Terry models to data set of 2015 MLB regular season play. The league data, downloaded using the Retrosheet library in R, included home and away team scores for each game. Using this feature space, we fitted a regularized Bradley-Terry model and a binomial Bradley-Terry model to rank each team in the league (see Appendix 1). Some teams are consistently ranked in the top of the league. For example, the St. Louis Cardinals, winners of the aforementioned NL Central, are ranked first overall by the normalized B-T model and second by the binomial B-T model. However, because the models take different approaches to evaluating success, we see that this is not always the case.



Take for example, the Oakland A's. The A's are given a ranked 11th by the normal B-T model, but 25th by the binomial model – a difference of 14 spots, the most of any team in the majors. This rank-discrepancy can be explained with some intuition for the model's construction. The response variable for the regularized binomial model is a binary one (1 if the home team wins, 0 if the home team loses), while the response for the regularized normal model is the score differential. Thus, the normal model rewards teams who win by large score differentials. The A's had the second-highest average margin of victory over the training set with 4.1 runs, behind only the Toronto Blue Jays with 4.4 runs. However, they were 57-74 over the time period covered by the training data, good for the seventh-worst record in baseball over that time. Thus, their high score-margin gave them a higher ranking in the regularized B-T model, but their below-average win-loss record pushed them down in the binomial model.

Conversely, the Texas Rangers were ranked No. 10 overall by binomial model and No. 19 overall by the normal model – the largest difference on the other side of the $y = x$ line in Figure 1 (above). While the A's were rewarded by the normal model for their high margin of victory, the Rangers were penalized for their large score differential in losses. Over their 68 losses over the time span that the training data covered, the average run differential in the Rangers' losses was 3.9 runs, the third-highest in the majors.

When evaluating the overall performance of the normal and binomial Bradley-Terry models, we examined its number of correct predictions on a separate data set. To do this, we used a simple validation set approach by partitioning the first 80 percent of our data for training purposes and the remaining 20 percent for testing the model. The following table provides our results:

	Normal BT	**Binomial BT**
Training Set	57.15%	57.87%
Test Set	54.93%	56.17%

Since the model was built to interpolate the training data, it makes sense that the model is better fit on the training data than the test data.

One possible improvement to the model could be a more robust scoring system for home field advantage, rather than a placing the same value, 1, for every home team. This scoring system could take into account historical averages for home-field wins over the past decade. Since every field provides its own advantages (and disadvantages), it is worth differentiating each team's park conditions and fans to better fit the data.

Nonetheless, both models pass the eye test, as a majority of the actual playoff teams from the 2016 season are among the ones are that have beta values in the top-10 of both the normal and binomial Bradley-Terry models.

Appendix A. Ranks and beta values of all teams for both the regularized normal and regularized binomial Bradley-Terry models.

Team	Normal BT		Binomial BT	
	Beta	Rank	Beta	Rank
ANA	-0.035	19	0.006	13
ARI	0.054	14	-0.049	20
ATL	-0.745	29	-0.214	27
BAL	0.249	7	-0.030	17
BOS	-0.114	20	-0.077	23
CHA	-0.220	21	-0.033	18
CHN	0.103	12	0.164	4
CIN	-0.377	25	-0.192	26
CLE	0.032	15	-0.008	15
COL	-0.598	28	-0.230	28
DET	-0.283	23	-0.059	21
HOU	0.506	3	0.113	7
KCA	0.468	4	0.273	2
LAN	0.233	9	0.107	8
MIA	-0.384	26	-0.240	29
MIL	-0.369	24	-0.161	25
MIN	0.017	16	0.052	12
NYA	0.441	5	0.131	6
NYN	0.183	10	0.096	9
OAK	0.112	11	-0.132	24
PHI	-0.842	30	-0.249	30
PIT	0.372	6	0.256	3
SDN	-0.277	22	-0.046	19
SEA	-0.390	27	-0.071	22
SFN	0.247	8	0.062	11
SLN	0.617	2	0.320	1
TBA	-0.026	17	-0.009	16
TEX	-0.033	18	0.067	10
TOR	0.986	1	0.157	5
WAS	0.061	13	-0.007	14