To: Dr. Paul Sabin

From: Frank Daly

Date: 2 December 2021

Subject: Mixed Models for Predicting NFL Spreads

Intro

The goal for this project is to accurately predict NFL spreads to be used in a betting model. It is excessively difficult to be successful in sports betting, especially in football. Therefore, we are hoping that if we cannot beat the spreads, we will at the very least be able to accurately identify spreads. This problem can be seen as a classification problem because we need to classify the possibilities with our results. These possibilities are do we cover, not cover, or push. Cover means we can predict the spread and be confident in a team's ability to win or lose in accordance with the margin. As for not covering, this means our bet did not hit and we were not able to win or lose within the given margin. The last possibility is if we push, which means simple means we tied on the bet. To accomplish this task, we will first develop a mixed model to generate power rankings for all NFL teams. The power rankings will yield us offensive, defensive, and special teams effects which we will use to predict spreads between any two teams. By the end we hope that we can have a useful model that allows us to test our predictions with that of Vegas.

Lit Review

Our first literature source includes the article "Expected Points and EPA Explained" by ESPN Analyst Alok Pattani. This article was a good refresher for what "expected points" and “expected points added” (EPA) are. In the article, Pattani explains that expected points was a statistic created to capture important features that may be lost in looking at total points (Pattani). This statistic posed a solution to evaluating teams where high scoring games do not properly indicate the way a team may have performed. The example that Pattani uses is when the Jets and Raven met in 2011 and despite 6 touchdowns combined, only one was scored by the offense (Pattani). Therefore, the introduction of expected points solves cases like this. Expected points is measured while accounting for factors such as down and distance, position on the field, home field advantage, and time remaining. Pattani further elaborates by explaining how EPA followed EP and EPA is a difference in EP at the start of the play versus the end of the play (Pattani). This source was helpful in reviewing the important of EPA and how it has been used as an advanced measure in football.

Another informative source was “An Introduction to Linear-Mixed Effects Modeling in R,” where we became knowledge of the statistics behind the model and the detailed aspects that we need to pay attention to. Mixed models are good substitute when we are met with limitations with multiple regressions and ANOVAs. Some of the major pros of mixed models is that they can handle missing data better and will not lower our likelihood like multiple regressions and ANOVAs would (Brown). Understanding what mixed models are, we come to learn that these are models that have both fixed and random effects. This allows us to keep specific factors consistent, while other factors and observations with higher variability can be properly accounted for through random effects. Brown also mentions to be weary of whether the model contains by-participant or by-item randomized slopes. She goes on to state that because we can add a random effect does not mean we should, and we should take note if a model is not converging because of an unnecessary effect (Brown). This article was beneficial in helping us understand how mixed models in general work.

Data

Our data is broken up into two parts. We have the metadata which consists of the play-by-play data for a given NFL season. This play-by-play data is then used to generate the other part of our data which is our effects model. This effects model will then be used to predict the spread for each game by accounting for the average number of plays in a given NFL game.

We must also understand certain statistics and factors specific to the NFL and our data. The first statistic we will introduce is "expected points added" (EPA). Before EPA came "expected points" which is just an average of points you would expect a team to score on a given play. There are certain factors like field location, down and distance, and maybe even team strengths that go into calculating expected points. As for EPA it is an extension of expected points where it is calculated as a difference in the teams expected points at the end of the play and the team's expected points at the start of the play. This can be seen as a state space statistic because it is looking at a measure in one state versus another state. A major discovery in football data and analysis was the standard of models accounting for EPA per play. Therefore, our mixed model will set EPA/play as our response variable. The next factor of our data that should be understood is our fixed effect, home field advantage. Throughout various sports, sports data analytics has determined home field advantage to be a major factor in predicting a team's probability to win a zero-sum game. Some reasons why home field advantage in football is a critical factor includes teams traveling across the country to play, an environment the team is familiar, and most importantly a home crowd that makes player to coach communication difficult on away teams. Therefore, home field advantage will be the fixed effect in our model. The two criteria of our data that is most important is understanding that EPA/play is the response variable and home field advantage is the fixed effect

Methodology

Our goal is to develop a model that accurately predicts the spread of NFL games. To accomplish this, we will essentially be solving a classification problem. Here we will need to classify the prediction in one of three ways: we cover, we do not cover, or we push. Typically, most spreads are not whole numbers, so we are more concerned with covering or not covering rather than pushing. But before we can classify the predication, we need to define a model to produce the predictions. For this we will develop a mixed model.

We will create three mixed models, two of which we are expecting to me more influential than the third. The three models are power rankings for all 32 NFL teams in terms of their offensive, defensive, and special team effects. Therefore, we will have a mixed model that yields the top offensive teams and likewise for defensive and special teams. The challenge with this project is finding a way to accurate use the different power rankings effectively together to develop an accurate prediction. We are expecting the offensive and defensive models to be more important to our analysis than the special teams one so we will not use special teams, but this could be problematic because according to the first result on Google, roughly 20% of NFL plays are on special teams. Ultimately, we will stick to just offense and defense for now.

As mentioned briefly in the Data section of this report, our model will consist of a response variable (EPA/play), a fixed effect (home field advantage), and random effects (offensive effect, defensive effect, and special teams effect). Using these factors within our mixed model, we can take the effects we produce and create a spread from it. To do this, we will multiply the given effects by the average number of plays per game. The reason for this is our model is predicting EPA/play for an entire season, therefore, to get a spread for a single game we need to account for this by using the average number of plays per game.

Results

The first set of results we gathered was our summaries for our offensive, defensive, and special teams effects. These results can be found in the Appendix in *Tables 1-3.* These tables give us estimates on each teams effect as well as the standard errors and t values for each team. The estimates from these tables will in turn be used to generate the offensive, defensive, and special teams effects.

The next results from our data are the offensive, defensive, and special team effects plots. These plots can be found in the Appendix labeled as *Plots 1-3.* There are a couple things to note in these plots that we will need to come back to and further investigate. Foremost, the most glaring feature is that the offensive and special team effects are identical. What is even more concerning is that all but one team (the Bills) are listed as positive effects. We checked this with our previous tables and we do see that the offense and special team effects are identical, but I do not have only one negative effect. I believe this to be an issue with plotting the data as opposed to the data and the model we created. Furthermore, another concerning visual from these plots is that the Buffalo Bills are listed as the only team with all negative effects. This is weird because we will see later that the Bills are found to be a top 3 offensive and defensive team.

As we move into the next set of results we have are the power rankings as well as the spread we created from these rankings. These results can be found in *Tables 4-5.* Here we see that the top three offensive teams from our results include Packers, Bills, and Cardinals. If we compare these results to that of OddsShark, they have the top three offensive teams as the Cowboys, the Bucs, and the Colts. As for defensive power rankings that we generated we found the top three defesive teams to be the Patriots, Cowboys, and the Bills. As for OddsShark, their top three are the Bills, the Patriots, and the Broncos. If we look at the overall rankings, we see that our top three teams are Patriots, Bills, and Cardinals whereas OddsShark has Patriots, Bills, and Bucaneers.

In order to generate a spread we attempted two approaches. One where we multiply the effects based on average points per game and one where it was average plays per game. For our first approach, I need to multiply our effects by the average number of points scored in a game. For this, we created two different models. One model we multiply the effects by the league average of total points score, the other model we multiply the effects by home and away team averages. We found the league average total points per game to be 45.89119, the away team average to be 22.7772, and the home team average to be 23.11399. According to the pro football reference the average number of points scored by all teams to be 22.9. If we add the average of our home and away team averages we also get 22.9 and if we multiply that by 2 we also get a match in our points per game. Our model was split into offensive and defensive ratings so we used each of these models to create a spread, one for defense, one for offense. Therefore, we took the offensive effect that we generated from our mixed model and multiplied it by the average number of points scored in a football game. The top highest spreads generated from our offensive model were the Bills at 5.62, the Packers 5.38, and the Cardinals at 4.96. As for the defensive model the top three highest spreads were the Patriots

8.92, the Cowboys 5.62, and the Bills 5.03. The results from this approach were thrown away as this is not the preferred method.

In our preferred approach we did almost the exact same, except this time our effects were multiplied by the average number of plays per game. This is the go to method that is favored by sports analyst. In this method we used the league average of 63 points per game and discovered that our top three teams for offense were Packers, Bills, and Cardinals. Likewise, the top three teams for defense were Patriots, Cowboys, and Bills. We notice that our magnitudes on both ends is larger than that of using average points per game. Even though our top teams for offense and defense effects stay the same, we note that in this model the Bills and Packers are switched for offense. If we look at our worse offensive teams we see that the magnitude is increased and in our first model the worst team offensively was the Jaguars, whereas in this model it is the Texans.

One potential issue with our model is that we develop our spreads using the average number of plays per game. This can be problematic because it does not group this average by team, it is a league average. Because of this we may have biased results favored towards teams closer to the average than above or below it. Certain teams may have a higher average number of plays per game compared to other teams. A good example of this would be looking at a team like the Chiefs versus a team like the Bears. The Chiefs are an offensive powerhouse and rely heavily on the passing attack. For this reason, they might have a much higher average number of plays per game than the Bears. We know from previous knowledge that in football, the offensive EPA standard deviation is much larger than defensive. Therefore, teams with strong offenses want more possessions whereas teams with weak offenses want less possessions with higher time of possession.

Another potential issue is that our data does not account for injuries. We know that injuries are a major part of the game and a team’s performance can be dependent upon injuries. Some examples of notable injuries include the starting quarterback being hurt, starting running back being hurt, or a string of injuries on a defense or offense. It is safe to say that a team without their starting QB is going to perform worse compared to if they did have their starting QB. Therefore, not accounting for injuries can be seen as a bias in our data.

Continuing, our model does not account for a prior or historical data. In our case, using previous spreads could produce better and more accurate results. NFL data follows a somewhat normal distribution, therefore by accounting for league spreads in years past we can fit a Bayesian Mixed Model using a prior that would likely improve our results. One issue we may have to look out for that is unique with this year is not using a prior during COVID and not using a prior from major rule changes. We saw from previous ELO models that the 2018 season yielded higher team effects than any other year and that the covid season yielded the lowest. Therefore, if we wanted to account for prior spreads, we should be weary about using these seasons.

Conclusion

After analysis of our results, we can conclude that there is both good and bad with our model. Looking at what our model does well is that foremost we are able to accurately predict the favorite in a game. Out of the test trials we ran for Week 13 and Week 14, we accurately predicted every favorite that Vegas had. Despite the worrisome we had with our power rankings as indicated by Buffalo, our model is still useful to some extent because we can accurately predict the favorite. Another thing our model does well is that for one possession games it provides close estimates. When we had tested with Week 13 and Week 14, the largest error we saw no more than 1.5 points compared to Vegas' Line. For the most point our model would overestimate by 0.5-1 points compared to Vegas. As for what did not go well with our model, is that we seem to have weird results in our effect plots where offensive and special teams are identical and the Bills are rated poorly. We believe this to be an error with plotting and not the actual dataset but is unknown at this time. Our model may also be slightly biased because of the number of plays in a given game. Our model biases would be towards teams that have higher plays per game than the average we used as well as the flip side to teams that have lower average plays per game than the average we used. Overall, our model needs some fine tuning but proves to be a good starting point as we can accurately predict favorites and develop spreads close to that of Vegas for one possession games.

Works Cited

Brown, Violet A. "An Introduction to Linear Mixed-Effects Modeling in R - Violet A. Brown, 2021." *SAGE Journals*, 8 Mar. 2021, https://journals.sagepub.com/doi/full/10.1177/2515245920960351.

Pattani, Alok. "Expected Points and EPA Explained." *ESPN*, ESPN Internet Ventures, 15 Sept. 2012, https://www.espn.com/nfl/story/_/id/8379024/nfl-explaining-expected-points-metric.

Appendix

Table 1: Offensive Summary

Team	Estimate	Std. Error	t value
(Intercept)	-0.11231	0.03635	-3.089
home_tea	0.15826	0.01853	8.539
home_tea	0.14552	0.01836	7.924
home_tea	-0.01664	0.01863	-0.893
home_tea	0.01553	0.01776	0.874
home_tea	0.16687	0.0181	9.218
home_tea	0.07331	0.01823	4.02
home_tea	0.11566	0.01811	6.387
home_tea	0.09738	0.01825	5.337
home_tea	0.02985	0.01866	1.6
home_tea	0.14344	0.01823	7.867
home_tea	0.03679	0.01841	1.998
home_tea	0.15093	0.01777	8.494
home_tea	0.10336	0.01758	5.879
home_tea	0.15515	0.0174	8.915
home_tea	0.04581	0.01806	2.537
home_tea	0.11744	0.01766	6.651
home_tea	0.12413	0.01827	6.793
home_tea	0.12206	0.01798	6.787
home_tea	0.04542	0.01781	2.551
home_tea	0.07377	0.01816	4.062
home_tea	0.09702	0.01799	5.394
home_tea	0.03825	0.01847	2.07
home_tea	0.05524	0.01852	2.982
home_tea	0.15314	0.01834	8.35
home_tea	0.19229	0.01876	10.248
home_tea	0.15935	0.01817	8.768
home_tea	0.09941	0.01749	5.683
home_tea	0.06289	0.01722	3.651
home_tea	0.05884	0.01895	3.105
home_tea	0.07539	0.01772	4.255
home_tea	0.09884	0.01852	5.337

Table 2: Defensive Summary

Team	Estimate	Std. Error	t value
(Intercept)	-0.11231	0.03635	-3.089
home_tea	0.15826	0.01853	8.539
home_tea	0.14552	0.01836	7.924
home_tea	-0.01664	0.01863	-0.893
home_tea	0.01553	0.01776	0.874
home_tea	0.16687	0.0181	9.218
home_tea	0.07331	0.01823	4.02
home_tea	0.11566	0.01811	6.387
home_tea	0.09738	0.01825	5.337
home_tea	0.02985	0.01866	1.6
home_tea	0.14344	0.01823	7.867
home_tea	0.03679	0.01841	1.998
home_tea	0.15093	0.01777	8.494
home_tea	0.10336	0.01758	5.879
home_tea	0.15515	0.0174	8.915
home_tea	0.04581	0.01806	2.537
home_tea	0.11744	0.01766	6.651
home_tea	0.12413	0.01827	6.793
home_tea	0.12206	0.01798	6.787
home_tea	0.04542	0.01781	2.551
home_tea	0.07377	0.01816	4.062
home_tea	0.09702	0.01799	5.394
home_tea	0.03825	0.01847	2.07
home_tea	0.05524	0.01852	2.982
home_tea	0.15314	0.01834	8.35
home_tea	0.19229	0.01876	10.248
home_tea	0.15935	0.01817	8.768
home_tea	0.09941	0.01749	5.683
home_tea	0.06289	0.01722	3.651
home_tea	0.05884	0.01895	3.105
home_tea	0.07539	0.01772	4.255
home_tea	0.09884	0.01852	5.337

Table 3: Special Teams Summary

Team	Estimate	Std. Error	t value
(Intercept)	-0.01496	0.03642	-0.411
home_tea	-0.03249	0.01866	-1.741
home_tea	0.073399	0.018186	4.036
home_tea	-0.01706	0.018503	-0.922
home_tea	-0.09354	0.01812	-5.162
home_tea	-0.00388	0.017957	-0.216
home_tea	-0.01393	0.018235	-0.764
home_tea	-0.02379	0.018107	-1.314
home_tea	0.061553	0.018542	3.32
home_tea	-0.08389	0.018894	-4.44
home_tea	-0.04773	0.018202	-2.622
home_tea	-0.03965	0.01852	-2.141
home_tea	-0.08418	0.017985	-4.681
home_tea	0.039346	0.01787	2.202
home_tea	-0.05087	0.017638	-2.884
home_tea	0.042917	0.018123	2.368
home_tea	0.050261	0.017843	2.817
home_tea	0.066101	0.018498	3.573
home_tea	-0.01935	0.018254	-1.06
home_tea	-0.09979	0.01788	-5.581
home_tea	0.001133	0.018287	0.062
home_tea	0.126864	0.018234	6.957
home_tea	-0.00893	0.018552	-0.481
home_tea	-0.09097	0.01867	-4.872
home_tea	-0.01665	0.018534	-0.898
home_tea	0.166833	0.019052	8.757
home_tea	-0.01866	0.018034	-1.035
home_tea	0.00859	0.017906	0.48
home_tea	0.039831	0.017314	2.301
home_tea	0.020024	0.019091	1.049
home_tea	0.031982	0.01797	1.78
home_tea	-0.01548	0.018563	-0.834

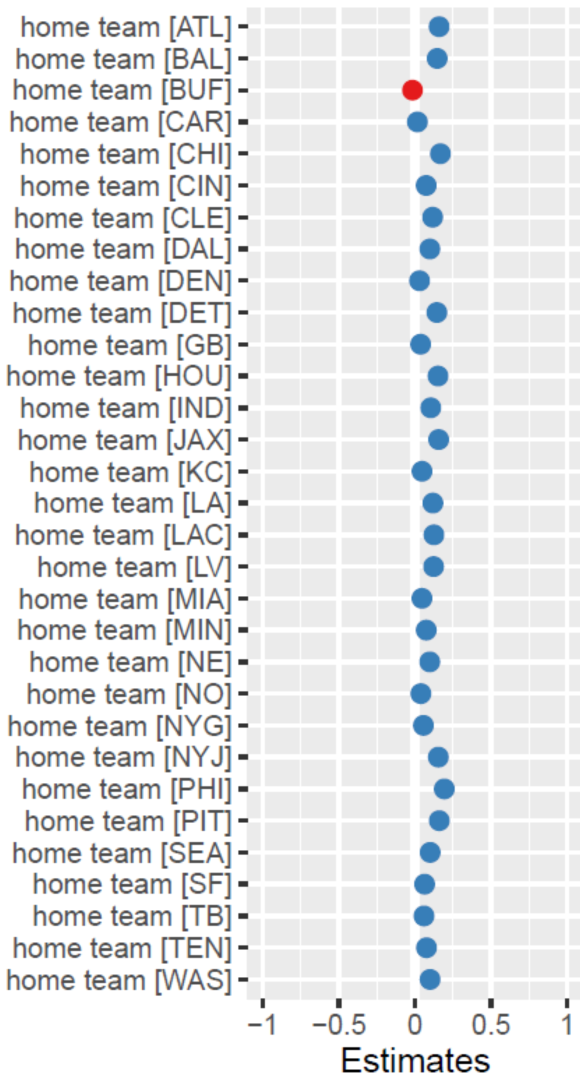
Table 4: Offensive Power Rankings and Spread

	team	off_eff	team_valu	team_rank	spread
1	ARI	0.108116	0.108116	3	6.811338
2	ATL	-0.10889	-0.10889	28	-6.86013
3	BAL	-0.01052	-0.01052	19	-0.66258
4	BUF	0.114747	0.114747	2	7.229045
5	CAR	-0.04946	-0.04946	25	-3.11606
6	CHI	-0.12844	-0.12844	31	-8.09158
7	CIN	-0.00468	-0.00468	18	-0.29466
8	CLE	0.006548	0.006548	15	0.41252
9	DAL	0.050995	0.050995	9	3.2127
10	DEN	0.054992	0.054992	8	3.464513
11	DET	-0.11521	-0.11521	29	-7.25828
12	GB	0.117229	0.117229	1	7.385401
13	HOU	-0.1779	-0.1779	32	-11.2075
14	IND	0.08421	0.08421	6	5.305224
15	JAX	-0.12567	-0.12567	30	-7.91696
16	KC	0.103668	0.103668	5	6.531093
17	LA	0.064635	0.064635	7	4.072013
18	LAC	0.01127	0.01127	14	0.710004
19	LV	-0.01493	-0.01493	21	-0.9403
20	MIA	-0.01446	-0.01446	20	-0.91088
21	MIN	0.050519	0.050519	11	3.182686
22	NE	0.04784	0.04784	12	3.013942
23	NO	-0.0461	-0.0461	24	-2.90429
24	NYG	-0.02173	-0.02173	23	-1.369
25	NYJ	-0.10279	-0.10279	27	-6.47548
26	PHI	0.003554	0.003554	16	0.223888
27	PIT	-0.0521	-0.0521	26	-3.28251
28	SEA	-0.01591	-0.01591	22	-1.00262
29	SF	0.050865	0.050865	10	3.20448
30	TB	0.105931	0.105931	4	6.673677
31	TEN	-0.00248	-0.00248	17	-0.15626
32	WAS	0.016135	0.016135	13	1.016516

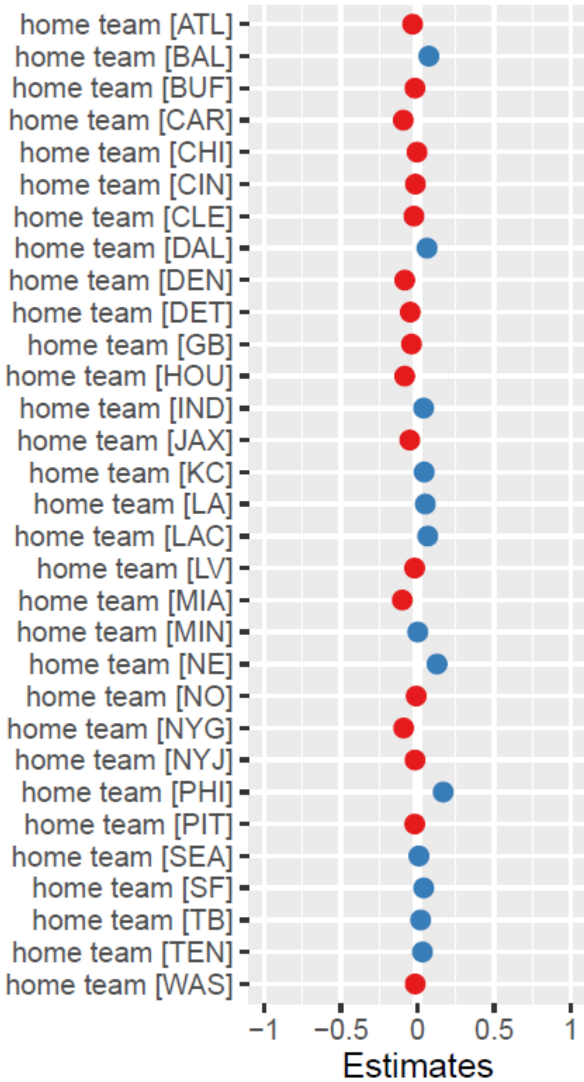
Table 5: Defensive Power Rankings and Spread

	team	def_eff	team_valu	team_rank	spread
1	ARI	0.106316	0.106316	4	6.69793
2	ATL	-0.08826	-0.08826	29	-5.56029
3	BAL	0.005566	0.005566	16	0.350657
4	BUF	0.109456	0.109456	3	6.895736
5	CAR	0.010727	0.010727	14	0.675803
6	CHI	-0.0494	-0.0494	23	-3.11203
7	CIN	0.053195	0.053195	7	3.35126
8	CLE	-0.01059	-0.01059	18	-0.66716
9	DAL	0.121984	0.121984	2	7.684983
10	DEN	-0.03877	-0.03877	21	-2.44279
11	DET	-0.07571	-0.07571	27	-4.76999
12	GB	-0.02649	-0.02649	20	-1.66908
13	HOU	-0.07969	-0.07969	28	-5.02038
14	IND	0.04774	0.04774	9	3.007617
15	JAX	-0.13375	-0.13375	32	-8.42634
16	KC	0.012996	0.012996	13	0.818758
17	LA	0.020856	0.020856	12	1.313929
18	LAC	0.027916	0.027916	11	1.758696
19	LV	-0.056	-0.056	26	-3.52811
20	MIA	-0.04885	-0.04885	22	-3.07776
21	MIN	-0.01199	-0.01199	19	-0.75545
22	NE	0.187541	0.187541	1	11.81506
23	NO	0.085212	0.085212	5	5.368369
24	NYG	-0.0516	-0.0516	24	-3.25067
25	NYJ	-0.12938	-0.12938	31	-8.15122
26	PHI	0.0512	0.0512	8	3.225571
27	PIT	-0.05536	-0.05536	25	-3.48757
28	SEA	0.002237	0.002237	17	0.140931
29	SF	0.008012	0.008012	15	0.504763
30	TB	0.075815	0.075815	6	4.776359
31	TEN	0.034548	0.034548	10	2.176535
32	WAS	-0.10546	-0.10546	30	-6.64411

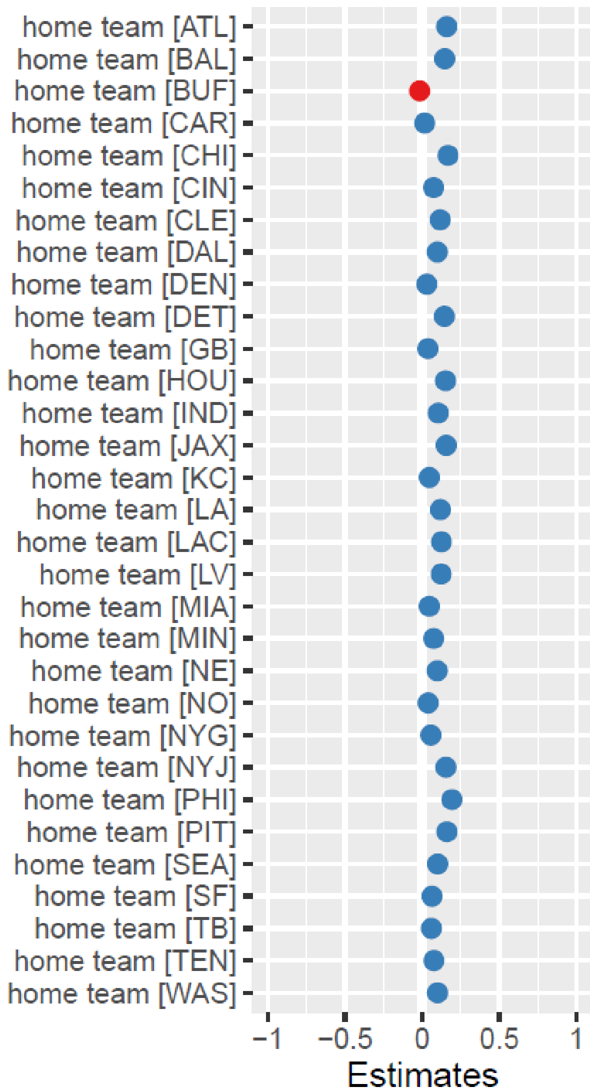
Plot 1: Offensive Effects



Plot 2: Defensive Effects



Plot 3: Special Team Effects


home team [ATL]
home team [BAL]
home team [BUF]
home team [CAR]
home team [CHI]
home team [CIN]
home team [CLE]
home team [DAL]
home team [DEN]
home team [DET]
home team [GB]
home team [HOU]
home team [IND]
home team [JAX]
home team [KC]
home team [LA]
home team [LAC]
home team [LV]
home team [MIA]
home team [MIN]
home team [NE]
home team [NO]
home team [NYG]
home team [NYJ]
home team [PHI]
home team [PIT]
home team [SEA]
home team [SF]
home team [TB]
home team [TEN]
home team [WAS]
−1
−0.5
0
0.5
1
Estimates