Modeling NFL Field Goal Attempt Outcomes in a Bayesian Framework using Informative Missingness

1 Abstract

We account for informative missingness in Bernoulli data, where the number of trials is determined in part by the probability of success. Our example dataset is play-by-play data from the National Football League (NFL) from the years 2008 through 2010. Noting that whether a kick is attempted is a decision made in part based on the probability the kicker will make the kick, we encode this information in our model in a way that is straightforward and sensible for this type of embedded data. We show the potential benefits of our model through simulation. We then apply this model to the data and compare the results to the results of a model using only the observed kick data, noting the unique features of each model. Finally, we analyze the results of our model to develop rankings of kickers under three ranking systems.

2 Introduction

Field goal percentage is the most oft-quoted measure of kicker performance. If all field goals were equally difficult, this would be the best way to objectively evaluate performance. However, because various conditions can make some field goals more difficult than others, it is possible for a more skilled kicker to have a worse field goal percentage than a less skilled one. In fact, one can envision a Simpson's paradox scenario in which one kicker has another "dominated," i.e. has a higher probability of making any particular kick, and yet may have a lower long-run percentage of kicks made. While this issue may apply to any performance metric taken over heterogeneous difficulty levels, it is exascerbated in cases where the difficulty level is expected to be highest for the highest performers. This is the case for NFL field goal attempts, where better kickers may be asked to take more kicks in difficult circumstances (high situational pressure, far away, into the wind, etc.). In such cases, modelling this heterogeneity provides an opportunity for fair comparisons to be made among kickers.

While many have attempted to account for heterogeneity in field goal difficulty, there are a number of unique features to the analysis we provide. Bilder and Loughin [1998] treat kickers as interchangeable in their modelling process, citing the fact that kickers sometimes are not resigned when they want more money. This statement is of course true of all performance related occupations, and in fact a worthwhile question for an NFL team to ask itself might be, “Do we expect this kicker to perform well enough compared to other kickers so that investing more money in him is worthwhile?” So while for Bilder and Laughlin’s modelling purposes it may have been useful to ignore kicker-to-kicker differences, it is crucial to take differences between kickers into account if our goal is to evaluate performance. Bilder and Laughlin also note that differences between kickers are most evident at the farthest distances, where modelling is unreliable due to the sparsity of attempts for many kickers. This is another instance of how the unique features of our model allow for stronger differentiation of kicker ability. By using the choice to attempt or not attempt a field goal as indirect evidence of whether a kick will be made or not, we can distinguish between kickers with the requisite ability to make long field goals and those lacking such ability. Because the ability to make long field goals provides an obvious tactical advantage,, any reasonable kicker rating should incorporate this skill. One might consider using an empirical strategy in which probabilities for making field goals are considered by examining empirical probabilities of making “similar” kicks. Once again, the sparsity issue becomes apparent when partitioning the data by kicker. Our model uses information gained about these probabilities from all kickers, but adjusts estimation based on individual kicker performance as well, thus requiring few actual attempts by an individual before reasonable predictions can be made. Obviously these predictions become more accurate with more data, but a lack of actual attempts at a particular distance does not preclude making inferences about the outcome of kicks attempted at that distance.

3 Data

Our data come from play-by-play information kept over an 11 year period by Arm Chair Analysis.com Erny [2011]. The website contains a Game Index dataset, a Roster dataset, and a Play by

Play dataset. The Game Index dataset contains 135 variables for each of 2921 games. Each row corresponds to an individual game and each column corresponds to a variable with values for each game. One such variable is "GAME ID," which assigns each NFL game for the years 2000 through 2010 a unique number between 1 and 2921. "SEAS" denotes the season that each particular game took place in with an integer between 2000 and 2010. "V" amd "H" provide two- or three-letter tags for which of the 32 NFL teams were the visiting and home teams, respectively, for a particular game. For example, a "V" value of "SF" denotes that the San Francisco 49ers were the visiting team for the GAME ID corresponding to that row. More specifics on individual variables are explained in the "Game Index Table" section of the Data Reference pdf that accompanies the zip folder Arm Chair Analysis.com provides.

The Roster dataset provides a list of active players on each team for each of the games. Each of 76,758 rows corresponds to a player/game combination and each column corresponds to a variable about that player/game combination. The "NAME" variable lists a player's name, usually in the form (First Initial.Last Name), with more letters of the first name used as necessary for the purpose of distinguishing between similarly named individuals on the same team. The "GAME ID" and "SEAS" variables are defined as in the Game Index dataset, but exist only for GAME ID 2121 through 2921 and corresponding SEAS 2008-2010. The variable "TEAM" lists the team that an individual played on for that GAME ID, using the same team identification system as the Game Index dataset used for "H" and "V." The variable "POS" lists a player's position, with a "POS" of K denoting that a particular player is a kicker. More specifics on individual variables are explained in the "Rosters Table" section of the Data Reference pdf that accompanies the zip folder that Arm Chair Analysis.com provides.

The Play by Play dataset provides information on each of the 473,621 plays that took place during the 2000 through 2010 NFL seasons. Each row corresponds to one such play, and each column corresponds to a variable about that play. The variables "GAME ID" and "SEAS" are defined as in the "Game Index" dataset. The "PLAY ID" variable gives a unique identification number between 1 and 473,621 for each play. The variable "GOOD" takes on values "Y","N", and " ", depending on whether a field goal or extra point was made, missed, or not attempted, respectively, on a particular play. The Play by Play datset lists 21,704 field goals and extra points as made and 2,289 field goals and extra points as missed. The variable “FKICKER” lists, using the same abbreviations as the Rosters dataset, the name of the kicker if a field goal or extra point was attempted on that play. Otherwise this variable is blank. The variable “OFF” lists the team with possession of the ball on a particular play using the same abbreviations as “TEAM” in the Rosters dataset. “LEN” lists the approximate length of a play in seconds. “QTR”,“MIN”, and “SEC” describe the quarter (values 1-5, where 5 denotes overtime), and minutes and seconds left in that quarter, respectively. “DWN” lists what down a play took place on (values 1-4), and “YFOG” lists the yards that the team with possession was from its own goal at the start of that play. More specifics on individual variables are explained in the “Play by Play Table” section of the Data Reference pdf that accompanies the zip folder Arm Chair Analysis.com provides.

4 Models

4.1 Independent Logistic Regression (ILR)

Let m_k be the number of field goals attempted by kicker k, and let K be the number of kickers that have attmepted a field goal. For the ith field goal attempted by kicker k, let

$$Y_{ik} = \begin{cases} 0 & \text{if field goal is missed, and} \\ 1 & \text{if field goal is made} \end{cases}$$

We assume $Y_{ik} \sim Ber(p_{ik})$ where p_{ik} is the probability of kicker k making the field goal on 4th down i, and all Y_{ik} random variables are independent. We model p_{ik} in the following way:

$$\text{logit}(p_{ik}) = x'_{ik}\beta_k \tag{1}$$

where $\text{logit}(p_{ik}) = \log\left(\frac{p_{ik}}{1-p_{ik}}\right)$, $x'_{ik} \in \mathbb{R}^S$ are covariates related to the probability of making a field goal such as distance, and β_k is a vector of the coefficients expressing how the probability of kicker k making a field goal is affected by those covariates. We calculate the β_k's that provide a line of best fit independently for each $k \in 1 : K$.

4.2 Hierarchical Logistic Regression

Instead of independently modelling the βk's for each kicker, it is possible to model these effects as draws from a normal distribution, i.e.

$$\beta_k \overset{iid}{\sim} N_S(\mu_\beta, \Sigma_\beta)$$

This is an example of Hierarchical Logistic Regression (HLR). Through this modeling choice, every β_k estimate is influenced by every observed field goal attempt. One advantage of making this assumption is it may greatly reduce the uncertainty in each individual kicker's parameter estimates by incorporating attempts from other kickers. For example, after two successful field goals at short distances, IRL would estimate that a kicker would make all field goals from any distance. Using HLR ensures that parameter estimates are closer to average values, while still allowing for variation in estimation from kicker to kicker.

4.3 Informative Missingness (IM) Model

Let n_k be the number of 4th downs encountered by kicker k, and let K be the number of kickers who have attempted a field goal. For the ith 4th down play encountered by kicker k, let

$$W_{ik} = \begin{cases} 0 & \text{if field goal is not attempted, and} \\ 1 & \text{if field goal is attempted} \end{cases}$$

We assume $W_{ik} \sim Ber(q_{ik})$ for $i = 1, ..., n_k, k = 1, ..., K$ where q_{ik} is the probability of kicker k attempting a field goal on 4th down i, and all W_{ik} random variables are independent. Similarly, let

$$Y_{ik} = \begin{cases} 0 & \text{if field goal is missed, and} \\ 1 & \text{if field goal is made} \end{cases}$$

We assume $(Y_{ik}|W_{ik}=1) \sim Ber(p_{ik})$ where p_{ik} is the probability of kicker k making the field goal on 4th down i given that a field goal was attempted, Y_{ik} is missing for $W_{ik}=0$, and all Y_{ik} random variables are independent. We model p_{ik} as in equation 1. Furthermore, we model

$$\begin{aligned}\text{logit}(q_{ik}) &= z'_{ik}\alpha_k + \gamma \cdot \text{logit}(p_{ik}) \\ &= z'_{ik}\alpha_k + \gamma x'_{ik}\beta_k \end{aligned} \tag{2}$$

where $z'_{ik} \in \mathbb{R}^R$ are covariates related to the probability of attempting a field goal such as how making that field goal would affect the empirical probability of a team winning, and α_k is a vector of the coefficients expressing how the probability of kicker k attempting a field goal is affected by these covariates. In practice, these would most likely be determined by the strengths and weaknesses of the competing teams and game-specific effects. Through this model setup, we have allowed the commonsense notion that instances where field goals are not attempted should influence our belief on whether such kicks would have been made. Rather than making independent predictions for each kicker based on a potentially small number of 4th downs and kick attempts, we once again wish to incorporate information from other kickers as well in making our predictions. To do this, we model the α 's and β 's such that,

$$\beta_k \overset{iid}{\sim} N_S(\mu_\beta, \Sigma_\beta) \qquad \alpha_k \overset{iid}{\sim} N_R(\mu_\alpha, \Sigma_\alpha)$$

$$\begin{bmatrix} \alpha_k \\ \beta_k \end{bmatrix} \sim N_{R+S}\left(\begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix}, \begin{bmatrix} \Sigma_\alpha & \Sigma_{\alpha\beta} \\ \Sigma_{\beta\alpha} & \Sigma_\beta \end{bmatrix} \right)$$

Finally, we set prior distributions on γ, μ_β, Σ_β, μ_α, and Σ_α. For our problem, with a large amount of data and no prior data or expectations on these parameters, we want to make our priors diffuse and uninformative. Thus we choose $\gamma \sim N(\gamma_0, \sigma^2_{\gamma_0})$, $\mu_\beta \sim N_S(\beta_0, \sigma^2_{\beta_0} I)$, $\mu_\alpha \sim N_R(\alpha_0, \sigma^2_{\alpha_0} I)$, $\Sigma_\beta \sim IW(I, S+1)$ and $\Sigma_\alpha \sim IW(I, R+1)$, where IW is the Inverse Wishart distribution. More on IW parametrization is explained by Mardia et al Mardia and Bibby [1979]

5 Results

5.1 Simulation

We performed a simulation of 25 kickers with 100 kick opportunities each to compare the IM and HLR models. We generate $\beta_k \sim N_2\left(\begin{pmatrix} 6.4 \\ -0.13 \end{pmatrix}, \begin{pmatrix} 0.33 & 0 \\ 0 & .014 \end{pmatrix}\right)$, where $x'_{ik} = \begin{pmatrix} 1 \\ U(18, 78) \end{pmatrix}$. These covariates and their associated $\beta's$ were chosen to be indicative of an intercept and "distance" in the actual dataset. We independently generate $\alpha_k \sim N_2\left(\begin{pmatrix} -1.1 \\ .35 \end{pmatrix}, \begin{pmatrix} .23 & 0 \\ 0 & .072 \end{pmatrix}\right)$, where $z'_{ik} = \begin{pmatrix} 1 \\ (2 * Beta(40, 30)) - 1 \end{pmatrix}$. These covariates and their associated α's were chosen to be indicative of an intercept and "ΔWP" (explained in section 5.2) in the actual dataset. Finally we let $\gamma = 1$ and generate data from these parameters. We fit our model on the generated data and obtain results. We also fit a model without equation (2) and using only the observed values of Y and the HLR methodology. Comparing these models demonstrates the additional information gained by incorporating the informative missingness into our model.

We created 95% credible intervals for each β parameter under our two models and compare these intervals. Figure 4 shows the estimates and credible intervals (red for informative missingness model, blue for model based on observed data) for the $\beta_{k,1}$ parameters corresponding to kicker specific intercepts(left) and slopes (right), ordered by the size of the true parameter.

Looking at a Mean Squared Error (MSE), here calculated as

$$\text{(Mean of Posterior - True Parameter)}^2 + \text{Variance of Posterior}$$

it appears that under the simulated conditions the larger model is preferable under this criterion. Figure 2 shows the reduced (HLR) model and IM model MSEs.

The full model has smaller median MSE for both the slopes and intercepts. Preliminary results under various simulation conditions have revealed that model comparison between the IM Model and the HLR model is highly dependent on the size of the simulation and the underlying model by which parameters are generated.

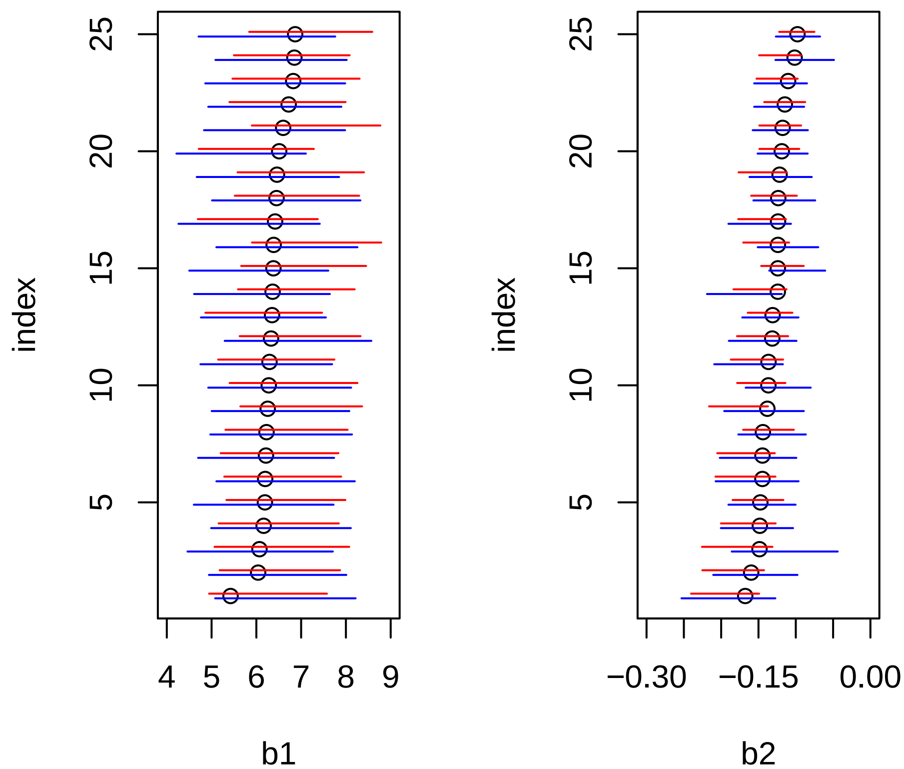

Figure 1: Intercepts and Slopes for HLR and IM Models

5.2 Covariate Selection for NFL Data

Since a vast majority of the heterogeneity in kick difficulties can be explained by distance, we chose to use a simple model with an intercept and distance as our only covariate on equation 1. We defined the distance variable by calculating (117-YFOG), as this value is the distance a field goal would need to be attempted from on any play. For this model to be useful, we needed to somewhat accurately capture the probability that a team would attempt a field goal. Otherwise, when a team chose to not attempt a field goal for various tactical reasons besides a kicker's ability to make the kick, it would negatively impact our predicted probability that a kicker could make such a kick. We used Advanced NFL Stat's WP calculator **?** to create the variable ΔWP

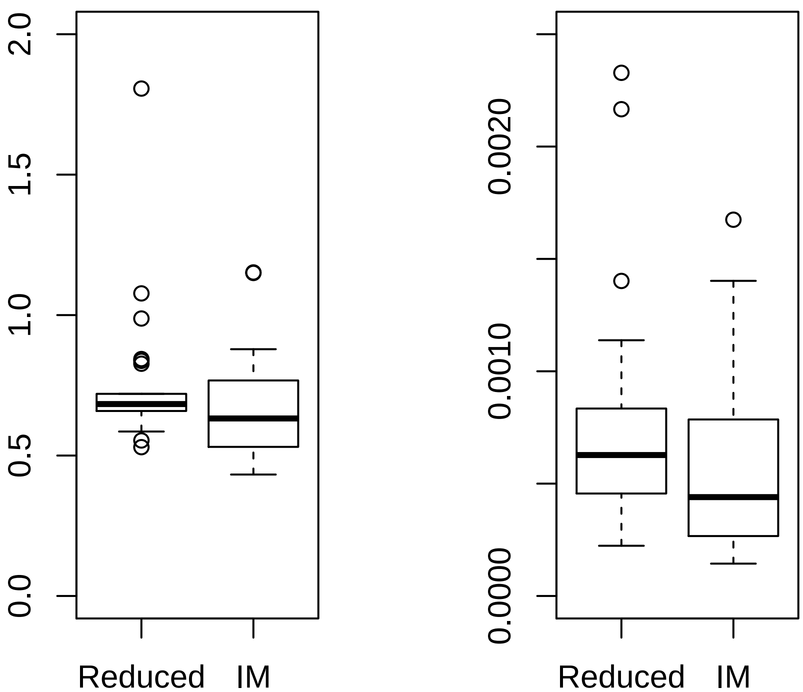

Figure 2: MSE in Intercepts and Slopes

as an aid in modeling the probability that a field goal is attempted. This variable is defined as (Win Probability if a field goal is made on play i and the other team receives the ball on first down at their own Win Probability before play i). These Win Probabilities were calculated using Advanced NFL Stat's WP Calculator. Ideally, we would like to use a calculation involving

$$P(\text{Winning before play i}|\text{Field Goal not attempted on play i})$$

rather than $P(\text{Winning before play i})$, but that estimate is not readily available from the WP Calculator. While additional covariates may be required to accurately define the probability an attempt being taken and made, these variables at least help to model the heterogeneity in makes and attempts.

5.3 Model Fit with Chosen Covariates

We ran iterations from reasonable starting points with a burnin of 20,000 and found an initial estimate of posterior likelihood for the parameters of the IM model for the 46 kickers with 20 or more kick opportunities. We then used 4 chains of initial values that were maximums and minimums of that initial posterior estimate. After a burn-in period of 60,000 iterations for each of four (highly dispersed) chains, we had no evidence of a lack of convergence. Among our 204 parameters of interest (4 kicker specific effects for each of 46 kickers, 4 mean estimates, and a 4x4 covariance matrix) our largest Rhat was 1.004. At this point we could consider our iterations as draws from the posterior distribution. We took 400,000 such samples, thinning to use only 100,000 of them. The parameter estimates for each of the 46 kickers examined are shown below.

5.4 Parameter Estimates

Below are parameter estimates for all 46 kickers, along with the standard deviations of those estimates. These estimates are ordered by the slope of the best fit line of the logit(probability of success) as distance increases by 1 under the IM model.

	Name	Intercept	Int S.D.	Slope	Slope S.D.
1	D.Buehler	4.578	0.547	-0.0874	0.0143
2	L.Tynes	4.979	0.489	-0.0974	0.0133
3	J.Hanson	5.806	0.526	-0.0984	0.0123
4	M.Nugent	4.837	0.557	-0.0988	0.0158
5	R.Bironas	6.03	0.507	-0.1016	0.0121
6	G.Hartley	5.793	0.531	-0.1021	0.0142
7	S.Janikowski	6.114	0.497	-0.1035	0.0115
8	D.Rayner	6.064	0.691	-0.107	0.0191
9	J.Nedney	5.948	0.512	-0.1083	0.0129

10	J.Scobee	5.696	0.518	-0.1084	0.0127
11	K.Brown	5.743	0.504	-0.1085	0.0136
12	M.Prater	6.024	0.496	-0.1094	0.0121
13	R.Succop	5.75	0.522	-0.1097	0.0137
14	C.Barth	5.888	0.544	-0.111	0.014
15	A.Vinatieri	5.925	0.504	-0.1112	0.0133
16	R.Gould	6.112	0.511	-0.1126	0.0128
17	R.Longwell	6.327	0.514	-0.1136	0.0129
18	G.Gano	5.883	0.602	-0.1141	0.0165
19	J.Elam	6.053	0.555	-0.1158	0.0157
20	O.Mare	6.13	0.499	-0.1164	0.0132
21	S.Gostkowski	6.073	0.5	-0.117	0.0133
22	N.Rackers	6.417	0.523	-0.117	0.0131
23	M.Gramatica	6.114	0.751	-0.1172	0.023
24	J.Feely	6.351	0.513	-0.1198	0.0129
25	J.Kasay	6.538	0.534	-0.12	0.013
26	J.Brown	6.52	0.523	-0.1201	0.0124
27	M.Bryant	6.332	0.53	-0.1227	0.0142
28	S.Suisham	6.215	0.523	-0.1239	0.0136
29	J.Carney	6.342	0.537	-0.1282	0.0152
30	J.Reed	6.676	0.537	-0.1302	0.0135
31	P.Dawson	6.71	0.558	-0.1309	0.0141
32	C.Stitser	6.049	0.707	-0.1315	0.0263
33	S.Hauschka	6.132	0.692	-0.1317	0.0224
34	N.Kaeding	6.654	0.538	-0.1318	0.0138
35	B.Cundiff	6.603	0.579	-0.1323	0.0158
36	D.Carpenter	6.958	0.575	-0.1338	0.0138
37	R.Lindell	6.792	0.551	-0.1349	0.0139

38	N.Folk	6.62	0.547	-0.1362	0.0141
39	N.Novak	6.238	0.722	-0.1375	0.0251
40	M.Crosby	7.019	0.575	-0.142	0.0141
41	D.Akers	7.338	0.587	-0.1466	0.0144
42	S.Graham	7.018	0.592	-0.1531	0.0166
43	M.Stover	7.372	0.684	-0.1566	0.0186
44	S.Andrus	6.482	0.775	-0.1605	0.0278
45	B.Coutu	5.905	0.829	-0.3835	0.0963
46	R.Lloyd	5.845	0.836	-0.4301	0.0948

While one can gain a feel for the relative kicker abilities at short lengths by looking at the intercepts provided, and the relative rate of dropoff as distance increases by looking at the slopes provided, it is useful to translate these parameter estimates into a more tangible analysis. Provided are estimates for each kicker's estimated percentage of makes at 30, 40, 50, and 60 yards based on the model parameter estimates of posterior medians. The kickers are ordered by estimated probability of making a 60 yard field goal.

	Name	30	40	50	60
1	R.Bironas	95.2	87.7	72.1	48.3
2	S.Janikowski	95.3	87.8	71.9	47.6
3	J.Hanson	94.5	86.6	70.8	47.5
4	G.Hartley	93.9	84.7	66.5	41.7
5	D.Rayner	94.6	85.6	67.1	41.2
6	R.Longwell	94.9	85.6	65.6	38
7	M.Prater	93.9	83.9	63.5	36.8
8	J.Nedney	93.7	83.4	63	36.6
9	N.Rackers	94.8	85	63.8	35.4

10	R.Gould	93.9	83.3	61.9	34.5
11	J.Kasay	95	85	63.1	34
12	D.Buehler	87.6	74.7	55.1	33.9
13	J.Brown	94.9	84.8	62.6	33.5
14	A.Vinatieri	93	81.4	59.1	32.2
15	K.Brown	92.3	80.3	57.9	31.7
16	C.Barth	92.8	81	58.3	31.6
17	J.Scobee	92	79.6	56.9	30.9
18	R.Succop	92.1	79.6	56.7	30.4
19	J.Feely	94	82.6	59	30.3
20	O.Mare	93.3	81.4	57.7	29.9
21	L.Tynes	88.7	74.7	52.8	29.7
22	J.Elam	92.9	80.5	56.5	29
23	M.Gramatica	93.1	80.6	56.3	28.5
24	S.Gostkowski	92.8	80.1	55.6	28
25	G.Gano	92.1	78.9	54.5	27.7
26	M.Bryant	93.4	80.6	54.9	26.3
27	D.Carpenter	95	83.3	56.6	25.5
28	M.Nugent	86.7	70.8	47.4	25.1
29	J.Reed	94.1	81.3	54.2	24.3
30	P.Dawson	94.2	81.4	54.1	24.2
31	S.Suisham	92.4	77.9	50.5	22.8
32	N.Kaeding	93.7	79.9	51.6	22.2
33	R.Lindell	94	80.2	51.2	21.4
34	B.Cundiff	93.3	78.8	49.7	20.8
35	J.Carney	92.4	77.1	48.3	20.6
36	D.Akers	95	81.4	50.2	18.9
37	M.Crosby	94	79.3	48	18.3

38	N.Folk	92.7	76.4	45.3	17.5
39	S.Hauschka	89.9	70.4	38.9	14.6
40	C.Stitser	89.1	68.8	37.2	13.7
41	N.Novak	89.2	67.7	34.6	11.8
42	M.Stover	93.5	75.2	38.7	11.7
43	S.Graham	91.9	70.9	34.5	10.2
44	S.Andrus	84.1	51.6	17.6	4.1
45	B.Coutu	0.4	0	0	0
46	R.Lloyd	0.1	0	0	0

Rob Bironas and Sebastian Janikowski, two kickers noted for their extremely strong legs, are now predicted to have the highest probability of making a 60 yard field goal. Two backup kickers have unreasonably low estimations for their probabilities of making a field goal. We treated all decisions to not attempt a field goal with a particular kicker equally, which may have been an unreasonable assumption. This is discussed further in the discussion section. Going from these estimates to a reasonable ranking of kickers is obviously dependent upon the distribution of field goals upon which we wish to base their performance. We examine three possible ranking systems in section 5.6.

5.5 Benefits of Kicker Comparison Using IM Model

To demonstrate some of the unique features of the data, it is helpful to examine a simple independent logistic regression model with distance as the only covariate. The resulting curve along with the proportion of kicks made at each distance between 19 and 62 yards leaguewide from 2000-2010 are shown in Figure 3.

From the graph it is evident that the model slightly underestimates the probability of making most kicks and overestimates the probability of making very long kicks. The overestimation is more noteworthy when one considers that the empirical results for that particular yardline are attempts taken by the kickers with the strongest legs rather than by kickers at random. This overestimation

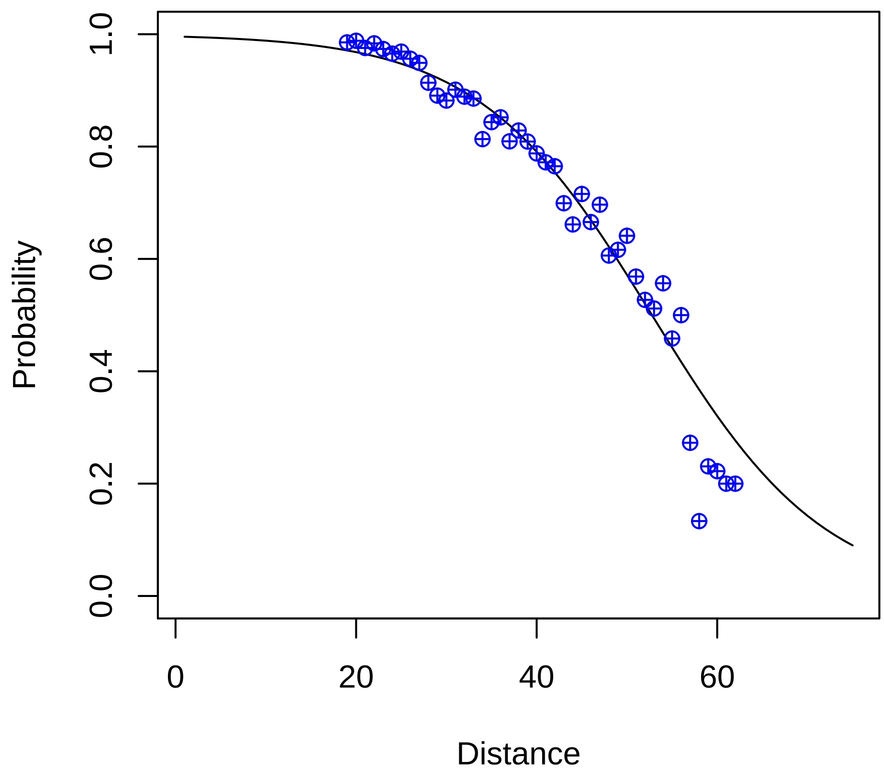

Figure 3: Simple Logistic Regression on Distance

will prove to create difficulties when fitting kicker specific model effects. Intuition tells us that the kickers who take more long field goal attempts are given these opportunities because they are better at making such kicks than kickers who take no such kicks. However, our fitted curves will suggest just the opposite. For kickers with no kicks in these long ranges, this missing data is treated as random, and they are often estimated to make a relatively high percentage of these longer kicks. This holds true whether the kicker specific effects are treated as random (HLR) or fixed (ILR).

One example of this phenomenon can be seen by independently fitting logistic curves as above for kickers Sebastian Janikowski and Lawrence Tynes (Figure 4). Sebastian Janikowski is widely regarded as the best long range kicker in the NFL over the three seasons we examined. In that period he made a 61 yard field goal and 8 of his 83 made field goals were longer than Tynes' longest made field goal of 53 yards. Due to the fact that Tynes has little data over 50 yards, the logistic model fits a slope and intercept parameter that is weighted heavily by shorter distances. As we have seen that the average dropoff in a kicker's probability of making a field goal is different in the 5 yard period between 25 and 30 yards than in the 5 yard period between 50 and 55 yards, Tynes is (wrongly) predicted to be an outstanding long range kicker. Janikowski, on the other hand, has actually taken many kicks at these longer distances. Thus the logistic model attempts to fit the data in this region, resulting in predictions that (wrongly) evaluate him as an only slightly better than average long range kicker. While modeling the kicker specific effects as random would create less extreme differences in the fits of these two kickers, the fact remains that Lawrence Tynes will be estimated to make a higher proportion of his field goals at long distances than Sebastian Janikowski due to the nonrandom missingness.

The IM Model provides more sensible inferences in these cases, as Figure 5 indicates. Figure 5 contains the new estimated curves for Janikowski and Tynes across distance under the Informative Missingness Model, along with the curve of an average kicker, constructed using μ_β median posterior values under the IM model. Note that we now have intuitive results that are harmonious with our expectations. Sebastian Janikowski is estimated to make a much higher than average proportion of long field goals and Lawrence Tynes is predicted to make a slightly below average proportion of long field goals. The estimated probability of an average kicker making a field goal at long range still appears to be higher than expected, but this is no longer an issue in being able to compare kickers against one another to develop a ranking.

5.6 Rankings and Kicker Comparison

We can use our model to create a ranking of kickers. To do so, we need to examine the expected number of field goals made over a particular distribution of plays. An interesting question is whether to use all kicks that were attempted as the distribution (AKA), all kicks that could have been attempted if the kicker took them all (AKP), or a kick distribution consisting of all kicks but weighted by the probability that the kick would be taken by an average kicker(MWK). AKA may

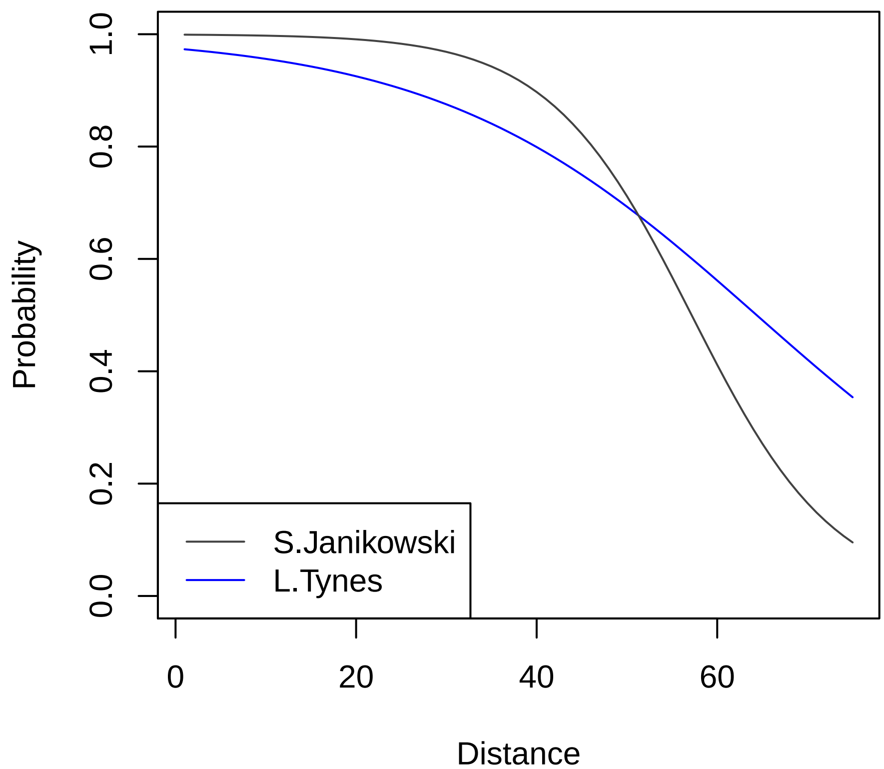

Figure 4: Logistic Fits for Lawrence Tynes and Sebastian Janikowski

undervalue a kicker with the ability to make long field goals due to the fact that he could have taken many more of the possible kicks. AKP may overvalue the long range kicker, since a team would nearly always choose to punt rather than attempt a field goal that has a low probability of going in and a high probability of yielding very good field position to the other team. MWK is using the parameter estimates for μ_β and μ_β to create the weighted distribution. Specifically, the weights are

$$\mu_{\alpha_0} + \mu_{\alpha_1} * \text{DeltaWP} + \gamma * \text{logit}(\mu_{\beta_0} + \mu_{\beta_1} * \text{Distance}$$

For AKA and AKP, we sum the probability of a make over all kicks in the distribution (kick attempts or kick opportunities). For MWK, we sum the probability of a make times that kick's

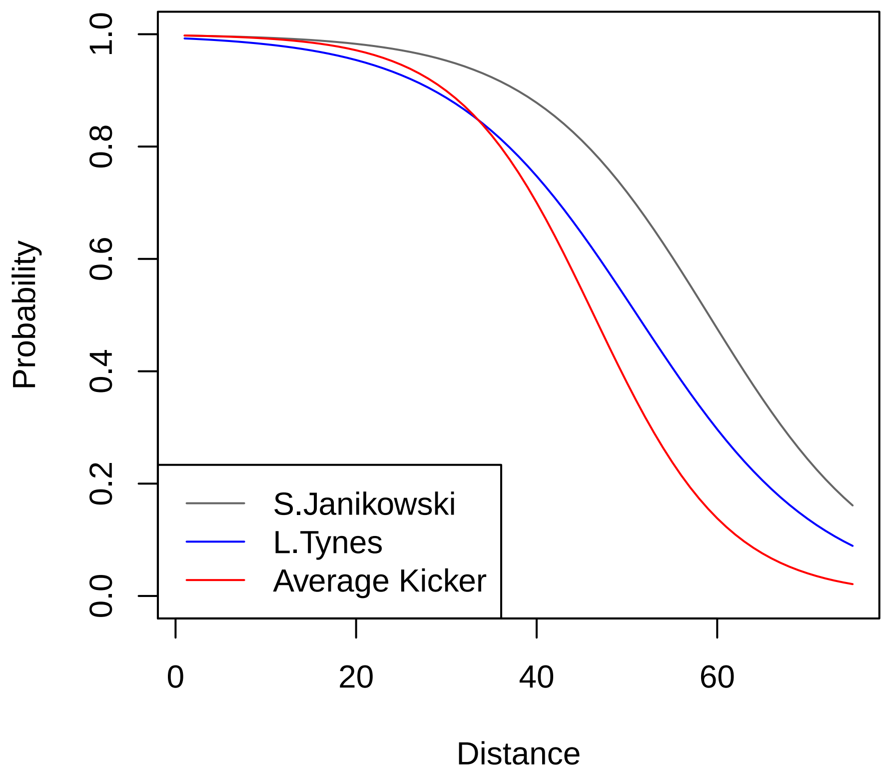

Figure 5: IM Model Fits for Lawrence Tynes and Sebastian Janikowski

associated weight over all kick opportunities. Field goal percentages from all three outlined possible distributions are provided (ordered by MWK):

	Name	MWK	AKA	AKP
1	S.Janikowski	2025	2427	5088
2	R.Bironas	2024	2426	5120
3	J.Hanson	2009	2403	5072
4	D.Rayner	1992	2377	4789
5	R.Longwell	1990	2373	4660
6	N.Rackers	1981	2358	4552

7	J.Kasay	1980	2356	4500
8	G.Hartley	1980	2358	4792
9	J.Brown	1977	2351	4478
10	M.Prater	1967	2337	4583
11	J.Nedney	1962	2328	4566
12	R.Gould	1960	2324	4484
13	D.Carpenter	1954	2310	4166
14	J.Feely	1949	2306	4315
15	A.Vinatieri	1934	2285	4358
16	O.Mare	1933	2281	4272
17	P.Dawson	1931	2273	4075
18	J.Reed	1930	2272	4079
19	C.Barth	1928	2274	4324
20	D.Akers	1928	2262	3902
21	M.Gramatica	1923	2265	4207
22	M.Bryant	1923	2262	4130
23	J.Elam	1922	2263	4219
24	K.Brown	1919	2261	4311
25	S.Gostkowski	1917	2254	4175
26	R.Lindell	1916	2246	3953
27	N.Kaeding	1913	2243	3972
28	R.Succop	1911	2247	4250
29	J.Scobee	1910	2247	4267
30	M.Crosby	1905	2223	3828
31	G.Gano	1902	2231	4136
32	B.Cundiff	1900	2219	3900
33	S.Suisham	1889	2206	3942
34	J.Carney	1880	2188	3848

35	N.Folk	1872	2170	3731
36	M.Stover	1860	2137	3512
37	L.Tynes	1842	2147	4104
38	D.Buehler	1835	2144	4263
39	S.Graham	1816	2063	3361
40	S.Hauschka	1803	2058	3483
41	M.Nugent	1788	2064	3837
42	C.Stitser	1785	2028	3414
43	N.Novak	1776	2007	3327
44	S.Andrus	1623	1736	2702
45	B.Coutu	104	61	121
46	R.Lloyd	44	24	51

With notably long range kickers Sebastian Janikowski and Rob Bironas near the top of the list, it is clear that kickers in this ranking system are no longer being penalized for being trusted to attempt long field goals. Due to the (still) seemingly elevated overall probability of success on long field goals, the actual expected values may not be trustworthy. However, this ranking system provides a fair comparison between kickers, modeling their probability of making a field goal in a way that yields more reasonable relative kicker effects.

6 Discussion

While it is certain that further work could be done in finding new covariates and properly utilizing the covariates present to create a better model of field goal behavior, we have demonstrated through this process that utilizing a Bayesian hierarchical model such as the one we propose leads to more sensible inferences than models dealing only with the observed kick data. While the specific details may vary, the basic framework of this model could be applied to a myriad of problems where the data contain informative missingness. This modelling and ranking framework could be used for any instance in which the number and/or difficulty of trials is in some way correlated with the response variable. An example might be trying to rank surgeons based on their performance, when the top surgeons may take on the most challenging surgeries. We performed linear logistic regression, but we could rather easily extend this approach for performing linear regression on a different type of numeric response variable, or for performing logistic regression with nonlinear covariates. For our particular dataset, one could envision a scenario in which distance had a nonlinear effect on the logit of the probability that they attempted a field goal.

As with any observational study, there are potentially lurking variables present that would aid in explaining the probability a field goal is attempted and/or made. We only made use of a few covariates, and incorporating additional information on field conditions would almost undoubtedly aid in predicting the probability of making a kick. Furthermore, it is obvious that modelling the probability of attempting a field goal is a crucial aspect of our hierarchical model. We could incorporate other covariates to explain the variation in our attempt model as well, an endeavor that would no doubt lead to better estimation. To accomplish the selection of further covariates from a list of reasonable possibilities, one could perform forward selection using DIC as a criterion. Covariates available through the Game Index dataset but unused in our analysis included WSPD (wind speed in miles per hour at the start of the game), COND (a factor variable with 18 choices related to weather conditions), and SURF (a factor variable with 9 choices including grass and several types of turf). Another improvement that could be made to our model is to incorporate what choice a team made over attempting a field goal with a specific kicker, rather than just whether they made that choice. By far the two worst performing kickers according to our model were a secondary kicker on their roster. That might suggest that they were not as good, but the choice not to kick a short field goal using one of these individuals does not reflect on his ability to make such kicks in the same way that it would if he were the only available kicker on his team.

Further work could be done through simulation in assessing the effect of the relative size of parameters on the information gained through incorporating the informative missingness into our model. Logic and preliminary simulation suggest that the IM model outperforms the HLR for:

- small values of the relative frequency of attempts (i.e. for small values for $logit(q)$)
- small amounts of variability in the $\alpha's$
- large values of γ

Obviously, there are a number of crucial differences between the simulated data and our actual data. Most importantly, we generated the data in our simulation to come from a model of the form of the full model we are using. So all we have truly shown through simulation is the benefit in using the proposed model over a simpler logistic regression model if the proposed model is correct. While this may be the case, there is reason to believe that the benefit of using the proposed model over this simpler model on the real data may provide even greater benefit. While the simpler model may not be making use of all the available data in the simulation, creating more variation in estimation, this model is at least long-term unbiased. However, as we saw that the logistic curve overestimated the probability of a make at long distances, kickers who took no such kicks will have inflated estimates on their success probabilities compared to those who actually took such kicks (and did worse than the model anticipates). There is still much work to be done in the field of incorporating indirect evidence to model problems in which informative missingness is present, but our analysis shows that this procedure can provide solutions to the counterintuitive inferences made when random missingness assumptions are violated for logistic regression.

References

C.R. Bilder and T.M. Loughin. "It's Good!" an analysis of the probability of success for placekicks. *CHANCE-BERLIN-*, 11:20–24, 1998.

Dennis Erny. Armchair Analysis.Com play by play dataset. 2011.

Kent J.T. Mardia, Kanti V. and J.M. Bibby. Multivariate Analysis. 1979.