7

We are going to use a mixed hierarchical linear regression model in this project. The response variable is the fan duel points received by the player. The predictor variables are the 5 week running mean of the points scored, the 6 week running mean of points against positions allowed by the opposing team, the position that a player plays, the indicator of whether it is a home or away game and the rank of a player based on their recent performance.

8

In this mixed linear regression model, we assume the performance of a player is normally distributed around a true mean value. The true value has an intercept of alpha, which is the 5 week running mean of the fanduel points earned.

It also has a fixed effect term which is the defense beta. It measures the defense effect of the opposing team adjusted for the position. The defense effects of the same position are pooled from the same position specific distribution.

The last two terms of the true value are the home and away effects on a player. They are some sort of random effects, which are pooled based on the same rank and position.

9

Here is a diagram of the DAG model. We model the defense effect, as how well a particular team’s defense has performed against the player’s position.

We pool the effect based on the position of the player. That is, the defense effect, beta\_p, is normally distributed with mean delta\_p and a constant diffuse variance. The hyperparameter delta\_p is distributed with a non-informative prior

10

We model that each position and rank pair has the same home and away effect. Those effects, beta\_home and beta\_away, for a specific rank are pooled together from the same rank specific normal distribution with eta\_r and rho\_r, and a constant variance. The eta\_r and rho\_r are hyperparameters distributed with a non-informative prior.

11

We use proper prior to approximate non-informative improper prior in JAGS. Sigma is an inv-gamma with parameters: .0001. The rest of the hyperparameters: delta, eta, and rho are all using normal distribution of 0 mean and 10000 square variance. We have checked from the result that those variance values are not restrictive.

12

The BUG model is a relative straight forward translation from the DAG model. Note that care has been taken to make sure the X.defense, X.home and X.away matches the indexing for the for loops.

13

We use JAGS and R to run this model. As initialization, we set up 4 chains with different over-dispersed starting points to effectively expose the parameter space.

We use the data from 2016 for training the model. 2016 is the most recent data set with a full season.

We set aside the last week of 2016 data for testing the model prediction accuracy

For convergence, we used trace plots and gelman statistics to check.

We run enough iterations to get an effective sample size of 400 or higher for all parameters.

14

Here is a sample of trace plots for convergence diagnostic and the gelman statistics.

As we can see, the trace plots are well distributed and the gelman statistics all have a upper bound less than 1.01

15

Here is a summary of the effective sample size of all parameters. Note that some beta away parameters have a slow mix in. Therefore, we use thinning to achieve the same desired effective sample size with a more reasonable number of iterations

16

Here are some of the summary results for the model parameters. Here, the beta.defense[1] corresponds to the PK position defense effect, and the beta.defense[2] corresponds to the QB position defense effect.