Predicting Fantasy Football with 2-Step Bayesian Hierarchical Model

Duke

Michael Sarkis

Introduction

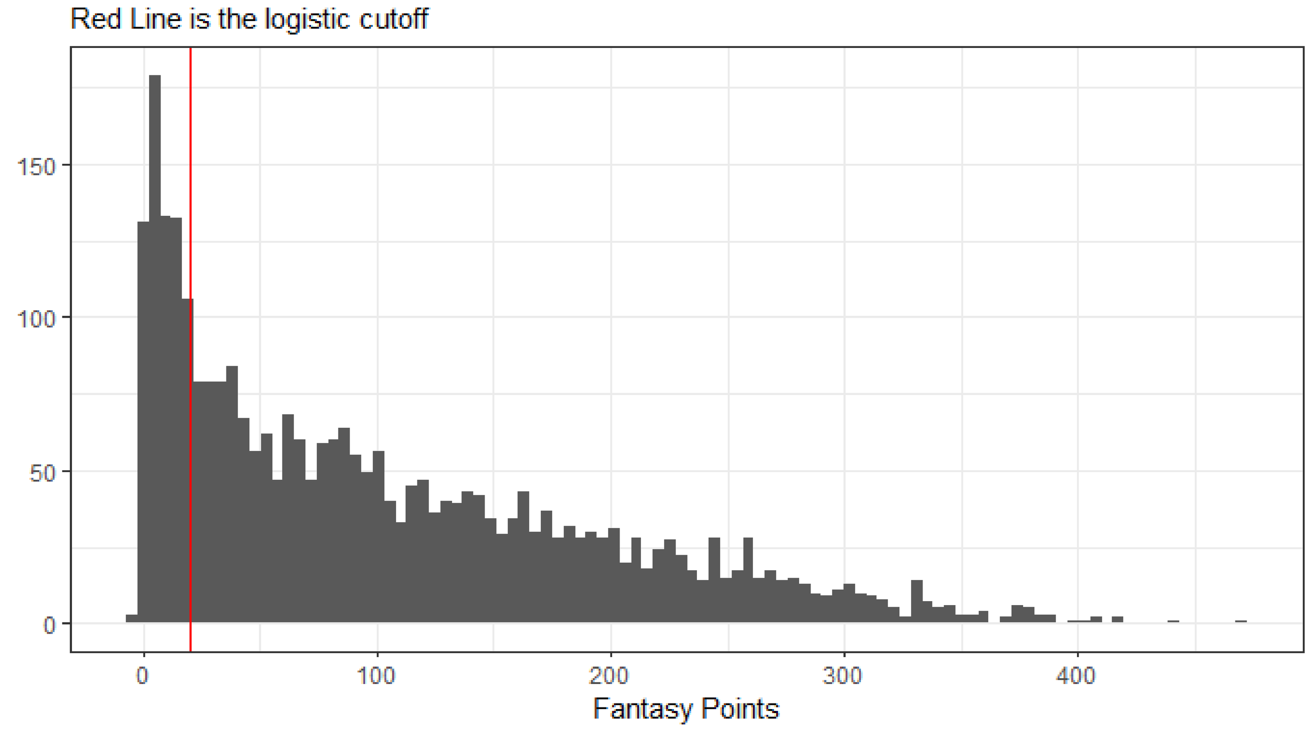
Fantasy football is a popular game with over 46 million players globally that brought in $22 billion dollars in profit for sports companies.[1] The basic concept of the game is player draft a team of NFL players and get points based on how well those players do in real football games with a goal of beating their opponents fantasy team each week. The goal of this project is to predict next year's fantasy point output based on past performance.

Data & Methods

The dataset for this project consists 8 years of NFL game data and fantasy football scoring as well as salary information. Just under 3000 player seasons were used in this model. The response, shown below, of next year's fantasy score is right skewed with a large portion of values near 0.

Next Year Fantasy Point Distribution

Red Line is the logistic cutoff



1. Hierarchical Logistic Regression model
   - To determine if a player will score "zero" fantasy points next season or move onto be predicted by the next model
   - The cutoff for "zero" is <15 points
   - Age, previous fantasy points, games started, average per year salary (APY) were considered for effects
   - Position and Team were considered for grouping

2. Bayesian Hierarchical model
   - Fit on the non-zero observations from the logistic model
   - Age, previous fantasy points, games started, average per year salary (APY) were considered for effects
   - Position and Team were considered for grouping
   - "Zero" players predicted with 15

Results

1. Hierarchical Logistic Regression

$$IsZero_{ip}|x_{ip} \sim Bernoulli(\pi_{ip}),\ i = 1, \dots, n;\ \ p = 1, .., 4$$

$$\log\left(\frac{\pi_{ip}}{1-\pi_{ip}}\right) = \beta_0 + b_{1p} + \beta_1 PFP + \beta_2 APY + \sum_a \beta_{3a} I(Age = a)$$

$\beta_0 =$ Intercept

$b_{1p} =$ Random Effect of Position; $b_{1p} \sim N(0, \sigma^2)$

$\beta_1 =$ Fixed Effect for Previous Year's Fantasy Points

$\beta_2 =$ Fixed Effect for Average Salary Per Year

$\beta_{3a} =$ Fixed Effect for Age Categories

2. Bayesian Hierarchical Model

$$(Y_{ipt}|IsZero_{ip} = 0) = \mu + b_{1p} + b_{2t} + \beta_1 PFP + \beta_2 APY + \sum_a \beta_{3a} I(Age = a) + \epsilon_{ipt}$$

$$(Y_{ipt}|IsZero_{ip} = 1) = 15$$

$\mu =$ Overall Mean;

$b_{1p} =$ Random Effect of Position; $b_{1p} \sim N(0, \tau_p^2)$;

$b_{2t} =$ Random Effect of Team; $b_{2t} \sim N(0, \tau_t^2)$

$\epsilon_{ipt} \sim N(0, \sigma^2) \perp b_{1p}, b_{2t}$

Priors

$\mu \sim N(100, 40)$

$\tau_p^2 \sim IG(2, 10)$

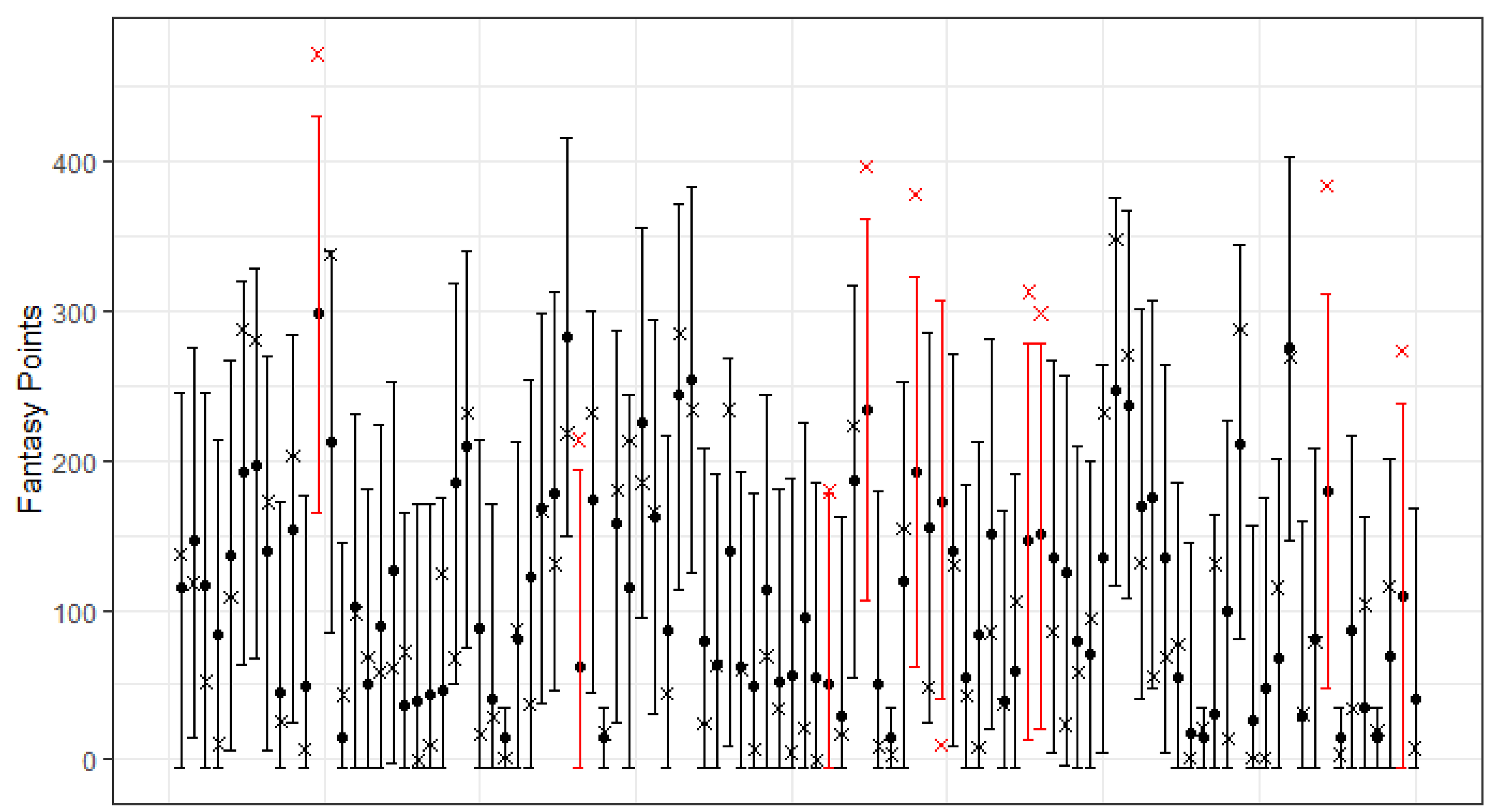
$\tau_t^2 \sim IG(2, 5)$

$\sigma^2 \sim IG(2.5, 150)$

93.5% Prediction Interval Accuracy

Prediction Intervals for Data

x Shows true value



Results Continued

Logistic Regression Accuracy		
	Predicted 0	**Predicted Non-0**
True 0	2.72%	15.46%
True Non-0	2.15%	79.67%

Full Prediction RMSE		
Two-Step	**One-Step**	**Naive**
64.60	64.62	72.62

Conclusions

Fantasy football data is very noisy and difficult to predict due to the random nature of football and the prevalence of injuries. The hierarchical model is a useful approach that helps to quantify the team and positional effects, which appear to have meaningful effects. The logistic stage of the model was lacking and did not appear to improve the RMSE by much if at all, but could hopefully be improved in the future to better improve the projections.

The final model's predictions were a significant improvement over the naïve approach, but still somewhat inaccurate. For future analysis a different approach to the large number of zeros and low values in the response data could be used. Additionally, using game specific data (snap counts, etc.) or schedule strength data may be helpful.

References

1. Fantasy Sports & Gaming Association - https://thefsga.org/fsga-faqs/

2. Sports Reference LLC. "NFL Fantasy Rankings" Pro-Football-Reference.com - Pro Football Statistics and History. https://www.pro-football-reference.com/. 8/28/21

3. Over The Cap. "Contract History" overthecap.com/contract-history/