

Management Science Final Project (Spring 2024)

Yen-Chun Chou & Howard Hao-Chun Chuang

Throughout the semester, you have learnt three important techniques in management science – optimization, simulation, and forecasting. We would like to ask each group to solve a predict-then-optimize problem for fantasy basketball (NBA) decision-making. Please see the example *Fantasy_NBA.ipynb* in which two datasets – *boxscores_dataset.csv* and *target_games.csv* are used. For your project, please finish the following tasks.

- 1) Collect your own NBA players' and games' dataset from more recent periods. Do NOT use the same dataset we give you guys.
- 2) Build a linear regression (LR) model to predict players' fantasy scores (Y). Use up-to-date fantasy score formula. Create new predictors/features (X) on top of those already in the example. Note that those predictors must be known available before any new game. Assess the model's prediction performance and discuss effect/importance of predictors.
- 3) Formulate a linear integer programming model to optimize your lineup for fantasy games. The objective is to form a team with the highest fantasy scores. Feel free to add or change constraints (total salary limit, position requirement such as at least one PG in the team, etc.). Show and explain your lineup of fantasy NBA players.
- 4) Collect residuals from LR in 2) and fit a LR model using $LSig = \ln(\text{squared residuals})$ as Y . Predict $LSig$ and assume error $\sim \text{Normal}(0, \text{stdev} = \exp(LSig)^{0.5})$, so you can simulate score errors e_1, e_2, \dots, e_S . Then construct the distribution of fantasy score forecast for a player using Predicted Score + e_j for $j=1, 2, \dots, S$. Use the simulated score distribution in your optimization model to maximize the probability of total fantasy scores \geq a threshold (you define) or any other performance metric you prefer. Any difference in the lineup?
- 5) **(Bonus)** For one or more out-of-sample weeks, compare fantasy scores of your team built from 3) and 4) to the TRUE optimal line-up with highest scores in your data.
- 6) **(Bonus)** This is NOT compulsory so you do not have to do this. That said, you will get extra points if you conduct **variable selection using Lasso (L1) and L0** (see below for L0).

https://colab.research.google.com/github/Gurobi/modeling-examples/blob/master/linear_regression/l0_regression.ipynb

Please do upload the following items onto wm5 by 23:59 on June 22.

- 1) A power point file with a link to your presentation video clip (no more than 12 mins) that clearly explains what you have done and found. Make sure the link to your video works!
- 2) A clearly and succinctly written report with necessary equations, tables, and/or graphics.
- 3) Python code and data (we will check this).