Bayesian hierarchical models for predicting individual performance in fantasy football (soccer)

Leonardo Egidi\textsuperscript{a} Jonah Sol Gabry\textsuperscript{b}

\textsuperscript{a} Dipartimento di Scienze Statistiche, Università degli Studi di Padova
egidi@stat.unipd.it

\textsuperscript{b} Department of Statistics, Columbia University, New York
jgabry@gmail.com
**Fantasy football** has become a cornerstone among football fans and statistical amateurs. Generally, fantasy games involve

- roster selection at the beginning of the season;
- match-by-match challenges against other participants, with the results determined by the collective performance of the players on the fantasy rosters;
- a lot of free and available data, which allows for statistical analysis.

So far, there is no statistical literature referring to fantasy football models: *we try to fill this gap*, by using **hierarchical Bayesian models** (Gelman and Hill, 2006) for predicting the players’ performances.
For player $i$ in match $t$ the total fantasy rating $y_{it}$ is given by

$$y_{it} = R_{it} + P_{it},$$ \hspace{1cm} (1)

where $R$ is the raw subjective score on a scale from one to ten assigned by some prominent newspaper, and $P$ is the point score, that takes care of specific in-game events.

<table>
<thead>
<tr>
<th>Event</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal</td>
<td>+3</td>
</tr>
<tr>
<td>Assist</td>
<td>+1</td>
</tr>
<tr>
<td>Penalty saved*</td>
<td>+3</td>
</tr>
<tr>
<td>Yellow card</td>
<td>$-0.5$</td>
</tr>
<tr>
<td>Red Card</td>
<td>$-1$</td>
</tr>
<tr>
<td>Goal conceded*</td>
<td>$-1$</td>
</tr>
<tr>
<td>Own Goal</td>
<td>$-2$</td>
</tr>
<tr>
<td>Missed penalty</td>
<td>$-3$</td>
</tr>
</tbody>
</table>

**Table:** Point scores. * = events only applicable to goalkeepers.
Overview of the game

We refer to the Italian fantasy football version *Fantacalcio*. At the beginning of the season, Fantacalcio managers are allocated a limited amount of virtual money with which to buy the players that will comprise their roster.

**Main challenge** There may be **missing values**: in fact, $y_{it}$ will be missing if the player

- does not play in the match;
- does not participate in the match for long enough for being judged by the subjective raw score.

A natural question is: **how modeling the missingness?**
Data All data are from the 2015–2016 season of the Italian Serie A and were collected from the Italian publication La Gazzetta dello Sport.¹

- \( N = 237 \) players, grouped into
- \( J = 4 \) positions (18 goalkeepers, 90 defenders, 78 midfielders, and 51 forwards), and
- \( K = 5 \) team clusters;
- \( T = 38 \) matches.

¹http://www.gazzetta.it.
Predictors and models notation

- $h_{it}$: home/away predictor. $h_{it} = 1$ if player $i$'s team plays match $t$ at its home stadium and $h_{it} = 0$ if the match is played at the opponent's stadium;
- $q_i$: initial standardized price for player $i$;
- $\alpha_i$: individual intercepts corresponding to each player $i = 1, ..., N$;
- $\gamma_{k[i]}$ and $\beta_{k[i],t}$: intercepts for the team-cluster of player $i$ and the team-cluster of the team opposing player $i$ in match $t$, respectively, with $k = 1, ..., K$;
- $\rho_{j[i]}$: the position-specific intercept, with $j = 1, ..., J$;
- $\delta_{j[i]}$: coefficient for the prices;
- $\lambda_{j[i]} \bar{y}_{i,t-1}$: autoregressive term;
- $\zeta_{j[i]} \bar{y}_{i,t-1}$: autoregressive term in the mixture model.
Mixture (MIX) model for the ratings

Assuming that it is very rare for a player to play in every match during a season, we can try to model the overall propensity for missingness. Let $V_{it}$ denote a latent variable

$$V_{it} = \begin{cases} 1, & \text{if player } i \text{ participates in match } t, \\ 0, & \text{otherwise.} \end{cases}$$

If $\pi_{it} = Pr(V_{it} = 1)$, then we can specify a mixture of a Gaussian distribution and a point mass at 0 (Gottardo and Raftery, 2008)

$$p (y_{it} | \eta_{it}, \sigma_y) = \pi_{it} \text{Normal} (y_{it} | \eta_{it}, \sigma_y) + (1 - \pi_{it}) \delta_0, \quad (2)$$

where $\delta_0$ is the Dirac mass at zero and $\eta_{it}$ is the linear predictor:

$$\eta_{it} = \alpha_0 + \alpha_i + \beta_{k[i], t} + \gamma_{k[i]} + \rho_{j[i]} + \delta_{j[i]} q_i + \lambda_{j[i]} \bar{y}_{i,t-1} + \theta h_{it}, \quad (3)$$

and $\sigma_y$ is the standard deviation of the error in predicting the outcome.
Mixture (MIX) model for the ratings

The probability $\pi_{it}$ is modeled using a logit regression,

$$\pi_{it} = \text{logit}^{-1}\left(p_0 + \zeta_j[i] \bar{y}_{i,t-1}\right),$$

which takes into account $\bar{y}_{i,t-1}$, the average rating for player $i$ up to match $t - 1; p_0$ is an intercept for the logit model. The individual-level, position-level, and team-cluster-level parameters are given hierarchical normal priors,

$$\alpha_i \sim \text{Normal}(0, \sigma_{\alpha}), \quad i = 1, \ldots, N \quad (5)$$

$$\gamma_k \sim \text{Normal}(0, \sigma_{\gamma}), \quad k = 1, \ldots, K \quad (6)$$

$$\beta_k \sim \text{Normal}(0, \sigma_{\beta}), \quad k = 1, \ldots, K \quad (7)$$

$$\rho_j \sim \text{Normal}(0, \sigma_{\rho}), \quad j = 1, \ldots, J \quad (8)$$

with weakly informative prior distributions for the remaining parameters and hyperparameters.
Our mixture specification allows for some natural other models extensions

- \( \pi_{it} \sim \logit^{-1} \rightarrow \text{MIX} \)
- \( \pi_{it} = 1, \text{ fixed} \)
  - missing \( y_{it} = 0 \) → Hierarchical autoregressive model (HAr);
  - missing \( y_{it} \sim f \) → Hierarchical autoregressive model with missing model (HAr-Mis);

Remark We want to estimate our models and predict the fantasy rating on a test set. Some interesting issue arise: missingness, model calibration, posterior predictive checks, out-of-sample predictions...

Setup We use the first half of the season as training set and the second half as test set.
MIX and HAr-Mis, that take care of the missingness, produce similar result. (Models fitted via Markov chain Monte Carlo (3000 iter., burn-in=1000)) using RStan Stan Development Team (2016a) and monitored convergence as recommended in Stan Development Team (2016b)).
Posterior predictive checks

Observed vs predicted cumulative ratings
for selected team Napoli

- **HAr**
- **HAr-mis**
- **MIX**
- **- - Observed**

![Graphs showing observed vs predicted cumulative ratings for selected Napoli players.](image)
Calibration for the MIX model
for selected team Napoli
Final aim **Select the best roster**. According to our posterior predictions for the second part of the season, we can create the best roster.

Let us note that the **MIX** is quite competitive; moreover Rudiger (defender, Roma) and Khedira (midfield, Juventus) performed pretty well in the 2016-2017 Serie A season.
Discussion and further work

- We proposed a class of hierarchical Bayesian models for predicting player ratings, in the presence of noisy fantasy football (soccer) data;
- these fantasy ratings may be seen as a crude proxy for players’ performances;
- we took care of the missingness issue;
- after controlling for missingness, the out-of sample predictive fit is good (the selected team appears to be competitive). Still checking for calibration.

Further work

- Dynamic prediction (match after match), adding data for more seasons, adding predictors;
- app for fantasy football managers (working on).

