Measuring the Efficacy of League Formats in Ranking Football Teams

Jan Lasek[1]
(joint work with Marek Gągolewski)

[1]Institute of Computer Science PAN & deepsense.ai

Computational Intelligence Seminar

Faculty of Mathematics and Information Science
Warsaw University of Technology

CV abstract

Education:

- **2013 - 2015 – Institute of Computer Science PAN:** Interdisciplinary PhD studies "Information Technologies: Research and their Interdisciplinary Applications" – Data Mining, Machine Learning, Statistics (working under supervision of dr. hab. inż. Marek Gągolewski).
- **2010 - 2013 – University of Warsaw:** Master studies at Faculty of Economic Sciences.
- **2011 - 2012 – Vrije Universiteit Amsterdam:** Master studies in mathematics within joint master programme with Faculty of Mathematics, University of Warsaw.
- **2007 - 2010 – University of Warsaw:** Double degree programme in Economics and Mathematics.

Current position:

- **Since 2015 – deepsense.ai (CodiLime):** Data scientist

Publication list I

Lasek, J. and Gągolewski, M. (2015). Predictive efficacy of a new association football league format in Polish Ekstraklasa. In *Proc. Machine Learning and Data Mining for Sports Analytics Workshop at ECML/PKDD*.

Lasek, J. and Gągolewski, M. (2015). Estimation of tournament metrics for association football league formats. In *Selected problems in information technologies (Proc. ITRIA'15)*, volume 2, pages 67–78. Institute of Computer Science, Polish Academy of Sciences.

Lasek, J., Szlávik, Z., Gągolewski, M., and Bhulai, S. (2016). How to improve a team's position in the FIFA ranking? A simulation study. *Journal of Applied Statistics*, 43(7):1349–1368.

Publication list II

Lasek, J. and Gągolewski, M. (2015). The winning solution to the AAIA'15 Data Mining Competition: Tagging Firefighter Activities at a Fire Scene. In *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 375–380.

Brefeld, U., Lasek, J., and Mair, S. (2017). Probabilistic player movement models and zones of control. *Submitted for review to Machine Learning Journal*.

Lasek, J., Szlávik, Z., and Bhulai, S. (2013). The predictive power of ranking systems in association football. *International Journal of Applied Pattern Recognition*, 1:27–46.

Gągolewski, M. and Lasek, J. (2015). Learning experts' preferences from informetric data. In Alonso, J., Bustince, H., and Reformat, M., editors, *Proc. IFSA/EUSFLAT'15*, pages 484–491. Atlantis Press.

Publication list III

Gągolewski, M. and Lasek, J. (2015). The use of fuzzy relations in the assessment of information resources producers' performance. In *Proc. 7th IEEE International Conference Intelligent Systems IS'2014, Vol. 2: Tools, Architectures, Systems, Applications*, volume 323 of *Advances in Intelligent Systems and Computing*, pages 289–300. Springer.

Lasek, M. and Lasek, J. (2016). Are stock markets driven more by sentiments than efficiency? *Journal of Engineering, Project, and Production Management*, 6(1):53–62.

Bogucki, R., Lasek, J., and J. K. Milczek, M. T. (2016). Early warning system for seismic events in coal mines using machine learning. In *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 213–220.

Lasek, J. (2016). Euro 2016 predictions using team rating systems. In *Proc. Machine Learning and Data Mining for Sports Analytics Workshop at ECML/PKDD*.

Selected research

Optimizing a team's position in the official FIFA ranking

FIFA World Ranking

67	Bolivia	483
68	El Salvador	481
69	Poland	474
70	Republic of Ireland	473
71	Trinidad and Tobago	470

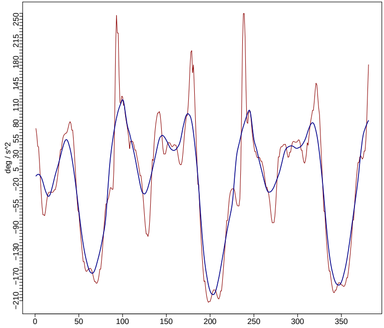
→

3	Argentina	**1399** (1399.02)
4	Switzerland	**1329** (1329.44)
5	Poland	**1319** (1318.66)
6	Portugal	**1267** (1267.06)
7	Chile	**1250** (1249.81)

Lasek, J., Szlávik, Z., Gągolewski, M., and Bhulai, S. (2016). How to improve a team's position in the FIFA ranking? A simulation study. *Journal of Applied Statistics*, 43(7):1349–1368.

Selected research

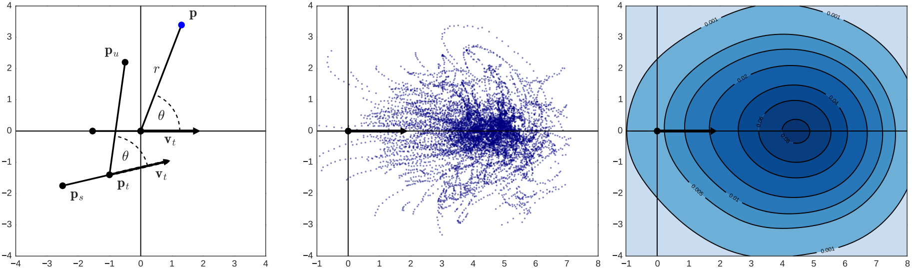
Tagging firefighters activities at a fire scene



Lasek, J. and Gągolewski, M. The winning solution to the AAIA'15 Data Mining Competition: Tagging Firefighter Activities at a Fire Scene. In *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 375–380.

Selected research

Estimating probabilistic movement models from positional data



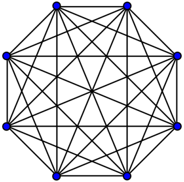
Brefeld, U., Lasek, J., and Mair, S. (2017). Probabilistic player movement models and zones of control. *Submitted for review to Machine Learning Journal.*

Measuring the Efficacy of League Formats in Ranking Football Teams

Motivation

There are multiple league formats that have been introduced. There is a need for an objective comparison between them.

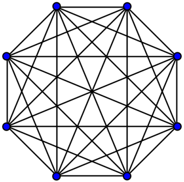
Here we focus on **predictive abilities** of different **league formats** based on **how accurately the ranking they produce matches the true ranking** based on team latent strength (ratings).



Motivation

There are multiple league formats that have been introduced. There is a need for an objective comparison between them.

Here we focus on **predictive abilities** of different **league formats** based on **how accurately the ranking they produce matches the true ranking** based on team latent strength (ratings).



The problem was studied by, e.g., Appleton (1995), McGarry and Schutz (1997), Scarf et al. (2009), and Ryvkin (2010) and is extended here for popular league formats.

Motivation

Recently Ekstraklasa underwent two changes:

- In the 2013/2014 season: 2nd competition stage was introduced (with points division)
- In the 2017/2018 season: dividing points in the 2nd stage was abandoned

Motivation

Recently Ekstraklasa underwent two changes:

- In the 2013/2014 season: 2nd competition stage was introduced (with points division)
- In the 2017/2018 season: dividing points in the 2nd stage was abandoned

Which **competition format** does produce **the most accurate team rankings**?

Many faces of the competition

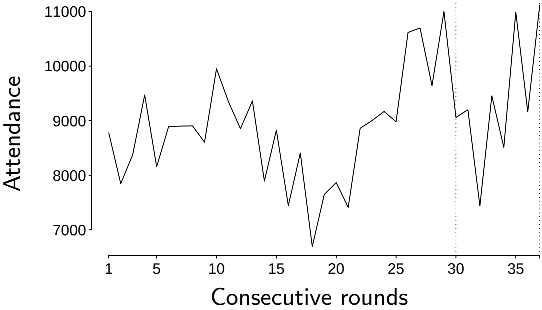


Figure : Average attendance in consecutive rounds of Polish Ekstraklasa in the 2013/2014 - 2016/2017 seasons.

Approach outline

Simulation study of league format efficacy

To compare different formats, we run simulations based on **latent team strength models** and compare them with the **final league standings**.

1. Review of team rating systems
2. Overview of league formats
3. Evaluation measures
4. Results

Review of team rating systems

Rating (ranking) systems play an important role in decision support.

Review of team rating systems

Rating (ranking) systems play an important role in decision support.

In particular, in sports, they are used in many contexts:

Review of team rating systems

Rating (ranking) systems play an important role in decision support.

In particular, in sports, they are used in many contexts:

1. They constitute a basic approach toward match outcome prediction

$$\textit{Data} \rightarrow \textit{Ratings} \rightarrow \textit{Predictions}$$

Review of team rating systems

Rating (ranking) systems play an important role in decision support.

In particular, in sports, they are used in many contexts:

1. They constitute a basic approach toward match outcome prediction

$$\textit{Data} \rightarrow \textit{Ratings} \rightarrow \textit{Predictions}$$

2. They can be used for tournament seedings

Review of team rating systems

Rating (ranking) systems play an important role in decision support.

In particular, in sports, they are used in many contexts:

1. They constitute a basic approach toward match outcome prediction

$$\textit{Data} \rightarrow \textit{Ratings} \rightarrow \textit{Predictions}$$

2. They can be used for tournament seedings
3. Ratings provide an objective measure of competitor strength

Review of team rating systems

Rating (ranking) systems play an important role in decision support.

In particular, in sports, they are used in many contexts:

1. They constitute a basic approach toward match outcome prediction

$$\textit{Data} \rightarrow \textit{Ratings} \rightarrow \textit{Predictions}$$

2. They can be used for tournament seedings
3. Ratings provide an objective measure of competitor strength
4. In football, the official FIFA ranking is used to grant work permits

Review of team rating systems

Many rating systems were proposed so far:

- The ordinal logistic regression model (Koning, 2000)
- Exponentially weighted moving average ratings by Cattelan et al. (2013)
- Poisson model by Maher (1982) or Dixon and Coles (1997)
- Bayesian Poisson model by Rue and Salvesen (2000)
- Elo rating system for association football by Hvattum and Arntzen (2010)
- Bivariate Posson model by Karlis and Ntzoufras (2003), or Koopman and Lit (2015)
- pi-ratings by Constantinou and Fenton (2013)
- **The attack-defence correlated Poisson model – our contribution**

Review of team rating systems

Many rating systems were proposed so far:

- The ordinal logistic regression model (Koning, 2000)
- Exponentially weighted moving average ratings by Cattelan et al. (2013)
- Poisson model by Maher (1982) or Dixon and Coles (1997)
- Bayesian Poisson model by Rue and Salvesen (2000)
- Elo rating system for association football by Hvattum and Arntzen (2010)
- Bivariate Posson model by Karlis and Ntzoufras (2003), or Koopman and Lit (2015)
- pi-ratings by Constantinou and Fenton (2013)
- **The attack-defence correlated Poisson model – our contribution**

We discuss the correlated Poisson model below (submitted for review to Statistical Modelling Journal).

Basic Poisson model

The assumption here is that the goals scored by a team can be modelled as a Poisson distributed variable.

Given the attacking and defensive skills (model's parameters) of teams i and j, a_i, a_j and d_i, d_j, respectively, the rates of Poisson variables for a home team i and visiting team j, λ and μ respectively, are modelled as:

$$\lambda = c + h + a_i - d_j,$$

$$\mu = c + a_j - d_i.$$

Basic Poisson model

In this model, the probability of a score x to y is a product of two individual Poisson variables with rates λ and μ respectively and equal to

$$\frac{\lambda^x \cdot e^{-\lambda}}{x!} \cdot \frac{\mu^y \cdot e^{-\mu}}{y!}.$$

Basic Poisson model

In this model, the probability of a score x to y is a product of two individual Poisson variables with rates λ and μ respectively and equal to

$$\frac{\lambda^x \cdot e^{-\lambda}}{x!} \cdot \frac{\mu^y \cdot e^{-\mu}}{y!}.$$

the model's parameters are estimated using the maximum likelihood method with regularization:

$$L(\mathbf{r}, h, c|\mathcal{M}) = \sum_{m \in \mathcal{M}} \log \mathbb{P}(g_i^{(m)}|\mathbf{r}, h, c) + \log \mathbb{P}(g_j^{(m)}|\mathbf{r}, h, c) - \frac{\lambda}{2}\|\mathbf{r}\|_2^2,$$

where

- $\mathbf{r} = (a_1, d_1, a_2, d_2, \ldots, a_n, d_n)$ are the attacking and defensive ratings for all n teams and
- $\mathcal{M}$ is the dataset of all matches m.

Extended Poisson model

Note that the Poisson model does not take into account correlation between

- goals scored by the opposing teams
- attacking and defensive team strengths

The first problem was addressed by e.g., Maher (1982), Dixon and Coles (1997), Karlis and Ntzoufras (2003), or Koopman and Lit (2015).

Here we focus on the second problem of accounting for correlations between attacking and defensive ratings on the team level.

Extended Poisson model

Empirically, we can observe correlation between attacking and defensive ratings.

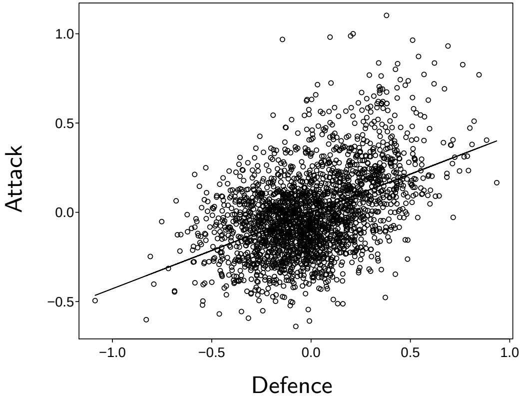


Figure : Attacking (x-axis) against defensive capabilities (y-axis) for a group of teams with a linear trend line. Correlation between the two ratings is ca. 0.467.

Extended Poisson model

Correlation between ratings also allows for realistic simulation of strength paths throughout the season in a dynamic version of the model.

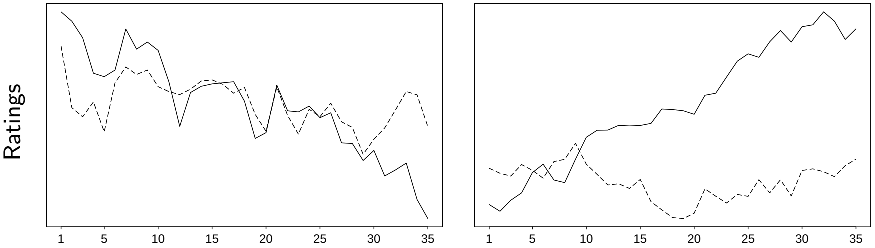


Figure : Difference in simulation for the correlated (left) and uncorrelated (right) attacking and defensive ratings illustrated for a single team throughout a season with 35 rounds.

Extended Poisson model

Correlation is included by extending the penalty operator

$$L(\mathbf{r}, h, c|\mathcal{M}) = \sum_{m \in \mathcal{M}} \log \mathbb{P}(g_i^{(m)}|\mathbf{r}, h, c) + \log \mathbb{P}(g_j^{(m)}|\mathbf{r}, h, c) \\ - \lambda \cdot \left(\frac{1}{2} \|\mathbf{r}\|_2^2 - \rho \cdot \langle \mathbf{a}, \mathbf{d} \rangle \right), \quad (1)$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{d} = (d_1, d_2, \ldots, d_n)$.

Extended Poisson model

Correlation is included by extending the penalty operator

$$L(\mathbf{r}, h, c|\mathcal{M}) = \sum_{m\in\mathcal{M}} \log \mathbb{P}(g_i^{(m)}|\mathbf{r}, h, c) + \log \mathbb{P}(g_j^{(m)}|\mathbf{r}, h, c) - \lambda \cdot \left(\frac{1}{2}\|\mathbf{r}\|_2^2 - \rho \cdot \langle \mathbf{a}, \mathbf{d} \rangle\right), \quad (1)$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{d} = (d_1, d_2, \ldots, d_n)$.

In general, basic regularization operators (L_1, L_2, or combinations thereof) were used by, e.g., Groll et al. (2015) or Gagolewski and Lasek (2015b). A more refined operator was used for example by in Elo++ model by Sismanis (2010).

Extended Poisson model

For a single team with a ratings pair (a, d) the penalty term can be rewritten to

$$\exp\left(-\lambda \cdot \left(\frac{1}{2}a^2 + \frac{1}{2}d^2 - \rho a d\right)\right) = \exp\left(-\frac{1}{1-\rho^2} \cdot \frac{a^2 + d^2 - 2\rho a d}{2\sigma^2}\right)$$

with $\sigma = \left(\lambda(1-\rho^2)\right)^{-\frac{1}{2}}$. This can be recognized as the (not normalized) bivariate Gaussian density with mean 0, variance σ^2 in both dimensions and correlation ρ between them.

$$\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}\right)$$

Extended Poisson model

In general, for all teams, the regularization component can be viewed as $2n$-dimensional Gaussian distribution with mean zero and correlation matrix $\mathbf{\Sigma} = [\Sigma_{ij}] \in \mathbb{R}^{2n \times 2n}$ where

$$\Sigma_{ij} = \begin{cases} \sigma^2 & \text{for } i = j, \\ \rho\sigma^2 & \text{for } |i - j| = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

$$\begin{bmatrix} \sigma^2 & \rho\sigma^2 & 0 & 0 & \dots & 0 & 0 \\ \rho\sigma^2 & \sigma^2 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \sigma^2 & \rho\sigma^2 & \dots & 0 & 0 \\ 0 & 0 & \rho\sigma^2 & \sigma^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \sigma^2 & \rho\sigma^2 \\ 0 & 0 & 0 & 0 & \dots & \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

Extended Poisson model

The penalty as a function of model parameters can be rewritten using the defined correlation matrix and equals

$$F_\lambda(\mathbf{a}, \mathbf{d}) = \lambda \cdot \left(\frac{1}{2} \|\mathbf{r}\|_2^2 - \rho \cdot \langle \mathbf{a}, \mathbf{d} \rangle \right) = \frac{1}{\det(\boldsymbol{\Sigma})} \cdot \mathbf{r} \boldsymbol{\Sigma}^{-1} \mathbf{r}^T.$$

As for the maximum likelihood optimization problem to be well-posed, $F_\lambda(\mathbf{a}, \mathbf{d})$ needs to be bounded from below.

Extended Poisson model

The penalty as a function of model parameters can be rewritten using the defined correlation matrix and equals

$$F_\lambda(\mathbf{a}, \mathbf{d}) = \lambda \cdot \left(\frac{1}{2} \|\mathbf{r}\|_2^2 - \rho \cdot \langle \mathbf{a}, \mathbf{d} \rangle \right) = \frac{1}{\det(\mathbf{\Sigma})} \cdot \mathbf{r} \mathbf{\Sigma}^{-1} \mathbf{r}^T.$$

As for the maximum likelihood optimization problem to be well-posed, $F_\lambda(\mathbf{a}, \mathbf{d})$ needs to be bounded from below.

This means that the matrix $\mathbf{\Sigma}$ needs to be positive semidefinite ($\rho^2 \leqslant 1$).

Extended Poisson model

Comparison of prediction results for the correlated and the base model:

- The model was verified for 120 seasons – for 5 major European leagues times 24 seasons
- The fraction of seasons in which the correlated model produced better results was computed

Extended Poisson model

Evaluation

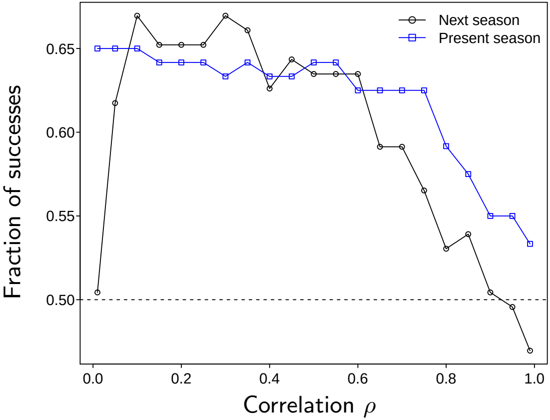


Figure : Fraction of test set season evaluations (y-axis) in which the correlated Poisson model achieves lower error rate than its uncorrelated counterpart. The x-axis presents correlation parameter ρ value.

Tuning the model

To run simulations, we need to determine models' parameters. This is done by maximizing the predictive power of rating systems.

- The 2015/2016 and 2016/2017 seasons were used as a training and a test set, respectively
- The results were compared according to logloss and accuracy of predictions

Tuning the model

Table : Logloss of predictions for the 2016/2017 season.

	Germany	**Poland**	**Scotland**
Poisson	1.010	1.023	0.944
Bookmaker odds	1.000	0.980	0.912

Table : Accuracy of predictions for the 2016/2017 season.

	Germany	**Poland**	**Scotland**
Poisson	49.4%	51.7%	55.9%
Bookmaker odds	53.3%	53.4%	55.9%

Toward dynamic rating systems

Exponentiated L_2 loss can be interpreted as multidimensional Gaussian density function. Using this observation we sample initial team ratings independently from the normal distribution. Moreover, the ratings are perturbed from round to round.

Toward dynamic rating systems

Exponentiated L_2 loss can be interpreted as multidimensional Gaussian density function. Using this observation we sample initial team ratings independently from the normal distribution. Moreover, the ratings are perturbed from round to round.

More precisely, for each team i:

1. $r_i^{(1)} \sim \mathcal{N}(0, \sigma)$
2. $r_i^{(k+1)} \sim \mathcal{N}(r_i^{(k)}, \sigma_i)$ for a team-specific drift parameter σ_i

Toward dynamic rating systems

Exponentiated L_2 loss can be interpreted as multidimensional Gaussian density function. Using this observation we sample initial team ratings independently from the normal distribution. Moreover, the ratings are perturbed from round to round.

More precisely, for each team i:

1. $r_i^{(1)} \sim \mathcal{N}(0, \sigma)$
2. $r_i^{(k+1)} \sim \mathcal{N}(r_i^{(k)}, \sigma_i)$ for a team-specific drift parameter σ_i

In our study, for each team i, we set $\sigma_i = \omega_i^{-1}$ for ω_i being a sample from the Gamma distribution $\Gamma(\alpha, 1)$ with the density function:

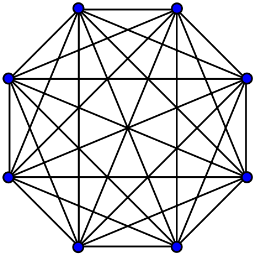
$$g(x|\alpha, 1) = \frac{x^{\alpha-1}}{\Gamma(\alpha)} \cdot \exp(-x)$$

where $\alpha > 0$ is a shape parameter and $x > 0$.

League formats

A *round-robin tournament* (RR) is the building block for domestic championships. For even number of n teams it constitutes

- $n - 1$ rounds with $\frac{1}{2}n$ matches played in each round
- $\binom{n}{2}$ matches in total



A kRR tournament means playing a single RR tournament k times.

League formats

- "+" denotes that a given format employs a two-stage league
- "$\frac{1}{2}$" that the points after the first stage are divided by two

Format	Example country	Rounds 12	Rounds 16	Matches 12	Matches 16
$3RR + (1RR/1RR)$	Scotland	38	52	228	416
$\frac{1}{2} \cdot 3RR + (1RR/1RR)$	–	38	52	228	416
$3RR$	Finland	33	45	198	360
$2RR + (2RR/2RR)$	Israel	32	44	192	352
$\frac{1}{2} \cdot 2RR + (2RR/2RR)$	Romania	32	44	192	352
$2RR + (1RR/1RR)$	Poland	27	37	162	296
$\frac{1}{2} \cdot 2RR + (1RR/1RR)$	Poland (in past)	27	37	162	296
$2RR$	Spain	22	30	132	240
$1RR$	–	11	15	66	120

Experiment set up

Definition of team strength and evaluation metrics

The “true” team ranking is obtained by averaging team strength throughout the season:

$$\bar{r}_i = \frac{1}{K} \sum_{k=1}^{K} a_i^{(k)} + d_i^{(k)}.$$

Experiment set up

Definition of team strength and evaluation metrics

The "true" team ranking is obtained by averaging team strength throughout the season:

$$\bar{r}_i = \frac{1}{K}\sum_{k=1}^{K} a_i^{(k)} + d_i^{(k)}.$$

The simulated and the final league ranks are compared according to several criteria:

- Kendall's τ correlation
- Spearman's Footrule distance and
- the fraction of the best team wins

Results

Average z-scores of tournament metrics across all simulation settings.

Format	No. of matches	Kendall's τ	Spearman's Footrule	The best team wins
$3RR + (1RR/1RR)$	228	0.964	-0.979	0.943
$\frac{1}{2} \cdot 3RR + (1RR/1RR)$	228	0.840	-0.851	0.729
$3RR$	198	0.636	-0.644	0.481
$2RR + (2RR/2RR)$	192	0.308	-0.297	0.558
$\frac{1}{2} \cdot 2RR + (2RR/2RR)$	192	0.199	-0.185	0.336
$2RR + (1RR/1RR)$	162	-0.007	0.015	0.049
$\frac{1}{2} \cdot 2RR + (1RR/1RR)$	162	-0.111	0.120	-0.151
$2RR$	132	-0.471	0.477	-0.603
$1RR$	66	-2.358	2.346	-2.342

Results

Influence of the number of matches

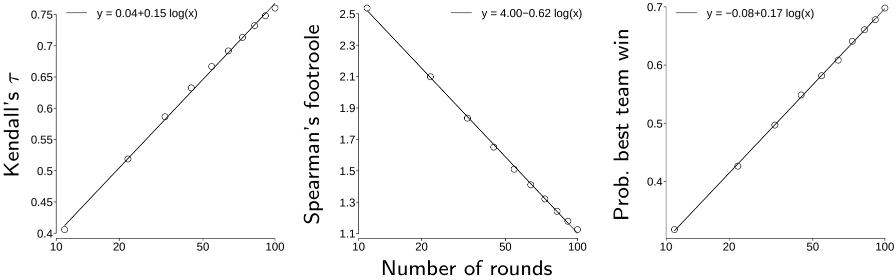


Figure : The three metrics considered (from the left): Kendall's τ correlation, Spearman's Footrule distance and the fraction of the best team wins (y-axis) as the function of the number of rounds (x-axis, on the logarithmic scale) in *kRR* tournament, $k = 1, 2, \ldots, 10$ for $n = 12$ teams.

Enhancing the 3RR + (1RR/1RR) format

By changing the number of points awarded for a win

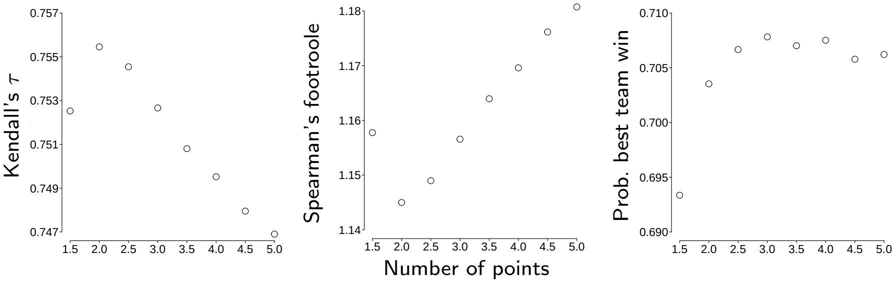


Figure : The three metrics considered (from the left): Kendall's τ correlation, Spearman's Footrule distance and the fraction of the best team wins (y-axis) as the function of the number of points awarded for a win (x-axis).

Further work

The study sets up other interesting research questions to answer.

- Different pay-offs for a win may influence a team's attitude. How to account for this?
- Investigate analytically the empirical logarithmic law for different metrics as a function of rounds played.
- Compare the accuracy of the extended Poisson model according to other criteria.

Further work - related areas

Other related research problems that are the focus of my PhD project:

- Developing player level-based rating systems
- Enhancing the official FIFA ranking

Further work - related areas

Other related research problems that are the focus of my PhD project:

- Developing player level-based rating systems
- Enhancing the official FIFA ranking

The proposed thesis title is: ***New data-driven rating systems in association football***.

References I

Appleton, D. (1995). May the best man win? *Journal of the Royal Statistical Society: Series C (The Statistician)*, 44(4):529–538.

Cattelan, M., Varin, C., and Firth, D. (2013). Dynamic Bradley-Terry modelling of sports tournaments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(1):135–150.

Constantinou, A. and Fenton, N. (2013). Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries. *Journal of Quantitative Analysis in Sports*, 9(1):37–50.

Dixon, M. J. and Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(2):265–280.

Elo, A. (1978). *The rating of chessplayers, past and present*. New York.

References II

Groll, A., Schauberger, G., and Tutz, G. (2015). Prediction of major international soccer tournaments based on team-specific regularized Poisson regression: An application to the FIFA World Cup 2014. *Journal of Quantitative Analysis in Sports*, 11(2):97–115.

Hvattum, L. M. and Arntzen, H. (2010). Using Elo ratings for match result prediction in association football. *International Journal of Forecasting*, 26(3):460–470.

Karlis, D. and Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):381–393.

Koning, R. H. (2000). Balance in competition in Dutch soccer. *Journal of the Royal Statistical Society: Series C (The Statistician)*, 49(3):419–431.

Koopman, S. J. and Lit, R. (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1):167–186.

References III

Lasek, J. and Gagolewski, M. (2015). Predictive efficacy of a new association football league format in Polish Ekstraklasa. In *Proc. Machine Learning and Data Mining for Sports Analytics Workshop at ECML/PKDD*, pages 67–78.

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118.

McGarry, T. and Schutz, R. W. (1997). Efficacy of traditional sport tournament structures. *The Journal of the Operational Research Society*, 48(1):65–74.

Rue, H. and Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3):399–418.

Ryvkin, D. (2010). The selection efficiency of tournaments. *European Journal of Operational Research*, 206(3):667–675.

References IV

Scarf, P., Yusof, M. M., and Bilbao, M. (2009). A numerical study of designs for sporting contests. *European Journal of Operational Research*, 198(1):190–198.

Sismanis, Y. (2010). How I won the "Chess Ratings - Elo vs the Rest of the World" competition. *CoRR*, abs/1012.4571.

Questions?

Thank you!

Appendix

Ordinal logistic regression model

In this model, match outcomes - H (home team win), D (draw) and A (away team win) - are linked to team ratings via the following equations

$$\begin{aligned}
\mathbb{P}(H) &= \frac{1}{1+e^{c-(r_i-r_j+h)}}, \\
\mathbb{P}(D) &= \frac{1}{1+e^{-c-(r_i-r_j+h)}} - \frac{1}{1+e^{c-(r_i-r_j+h)}}, \\
\mathbb{P}(A) &= 1-\frac{1}{1+e^{-c-(r_i-r_j+h)}},
\end{aligned}$$

where $h > 0$ is a parameter accounting for the home team advantage and $c > 0$ in an intercept which governs the draw margin.

Ordinal logistic regression model

In this model, match outcomes - H (home team win), D (draw) and A (away team win) - are linked to team ratings via the following equations

$$\begin{aligned}
\mathbb{P}(H) &= \frac{1}{1+e^{c-(r_i-r_j+h)}}, \\
\mathbb{P}(D) &= \frac{1}{1+e^{-c-(r_i-r_j+h)}} - \frac{1}{1+e^{c-(r_i-r_j+h)}}, \\
\mathbb{P}(A) &= 1 - \frac{1}{1+e^{-c-(r_i-r_j+h)}},
\end{aligned}$$

where $h > 0$ is a parameter accounting for the home team advantage and $c > 0$ in an intercept which governs the draw margin.

For $c = 0$ the model reduces to the expected match outcome function in the famous Elo rating system (Elo, 1978).

Ordinal logistic regression model

Model is fitted using the maximum likelihood method with L_2 regularization:

$$L(\mathbf{r}, h, c|\mathcal{M}) = \sum_{m \in \mathcal{M}} \log \mathbb{P}(o^{(m)}) - \frac{\lambda}{2}\|\mathbf{r}\|_2^2, \tag{3}$$

where:

- $\mathbb{P}(o^{(m)})$ equal to the probability of the actual outcome of a match m in the set of all matches $\mathcal{M}$ attributed by the model and
- $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ denote the vector of team rating parameters.

Tuning the models

Side note: Measuring the competitive balance

We may also refer to the parameter λ as an alternative a measure of competitive balance of teams in a league (Koning, 2000).

Table : Optimal parameter values (c, h, λ) for three different leagues in the 2016/17 season.

	Germany	**Poland**	**Scotland**
Poisson	(0.085, 0.371, 14.5)	(0.063, 0.37, **27.5**)	(0.124, 0.196, 13)

Toward dynamic rating systems

To determine parameter α we analyse Kendall's τ correlation coefficient between the probability of becoming champion before the start of the season derived from bookmaker odds and final league rankings for three league seasons: 2013/14, 2014/15 and 2015/16.

Table : Kendall's τ correlation between probability of outright winner derived from bookmaker odds and final league position.

	Germany	**Poland**	**Scotland**
2013/14	0.499	0.333	0.788
2014/15	0.569	0.700	0.364
2015/16	0.464	0.346	0.515

Toward dynamic rating systems

Table : Kendall's τ correlation coefficient between the initial and the final team strength for different parameter settings.

$\alpha \setminus \sigma$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**
2	0.094	0.180	0.257	0.321	0.374	**0.424**	0.465
5	0.257	0.445	0.561	0.644	0.701	0.744	0.777
10	0.469	0.674	0.770	0.823	0.857	0.881	0.898
20	0.691	0.835	0.889	0.916	0.932	0.944	0.952
50	0.870	0.935	0.957	0.967	0.974	0.979	0.981
100	0.936	0.968	0.978	0.984	0.987	0.989	0.991
∞	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Results

Average tournament metrics for the parameter setting $(\alpha, \sigma) = (2, 0.6)$.

Format	**No. of matches**	**Kendall's τ**	**Spearman's Footrule**	**The best team wins**
$3RR + (1RR/1RR)$	228	0.753	1.157	0.708
$\frac{1}{2} \cdot 3RR + (1RR/1RR)$	228	0.748	1.177	0.703
$3RR$	198	0.738	1.218	0.680
$2RR + (2RR/2RR)$	192	0.718	1.300	0.696
$\frac{1}{2} \cdot 2RR + (2RR/2RR)$	192	0.714	1.318	0.693
$2RR + (1RR/1RR)$	162	0.709	1.338	0.667
$\frac{1}{2} \cdot 2RR + (1RR/1RR)$	162	0.705	1.354	0.666
$2RR$	132	0.687	1.429	0.623
$1RR$	66	0.584	1.846	0.511