Football Analytics

Ioannis Ntzoufras

Athens University of Economics & Business
Disclaimer

- I am not Sotirakopoulos or WikiSports.

- I have a family: Two daughters (and a wife).
  ⇒ no actual time to watch sports and their developments as a fan.

- So I am not a database of sports events.

- But I know about methods and some results I am involved in.

- **Beware:** I am a normal academic scientist
  ⇒ I do research “useless” & “non understandable” topics that usually Academics do.
Why I am here

• Because I know Nikolas

• Sports Analytics is my/our academic hobby
  – Started with Dimitris Karlis as phd students back in 1998

• First paper in 2000 with Dimitris for fun in “Student”
Analysis of sports data by using bivariate Poisson models

Dimitris Karlis

Athens University of Economics and Business, Greece

and Ioannis Ntzoufras

University of the Aegean, Chios, Greece

[Received November 2001. Final revision April 2003]

Summary. Models based on the bivariate Poisson distribution are used for modelling sports data. Independent Poisson distributions are usually adopted to model the number of goals of two competing teams. We replace the independence assumption by considering a bivariate Poisson model and its extensions. The models proposed allow for correlation between the two scores, which is a plausible assumption in sports with two opposing teams competing against each other. The effect of introducing even slight correlation is discussed. Using just a bivariate Poisson distribution can improve model fit and prediction of the number of draws in football games. The model is extended by considering an inflation factor for diagonal terms in the bivariate joint distribution. This inflation improves in precision the estimation of draws and, at the same time, allows for overdispersed, relative to the simple Poisson distribution, marginal distributions. The properties of the models proposed as well as interpretation and estimation procedures are provided. An illustration of the models is presented by using data sets from football and water-polo.

Keywords: Bivariate Poisson regression; Difference of Poisson variates; Inflated distributions; Soccer

Google Scholar (30/1/2019):
Analysis of sports data by using bivariate Poisson models
D Karlis, I Ntzoufras
Journal of the Royal Statistical Society: Series D (The Statistician) 52 (3…
AUEB Sports Analytics Group (founded in 2015)

Hosted in the Computational & Bayesian Statistics Lab of AUEB

- Two Faculty Members
- 5 Collaborating researchers
- 5 International Professors as external/occasional collaborators
- 2 PhD Students
Our events

A Series of annual workshops:


AUEB Sports Analytics Group organizes an annual conference dedicated to all topics where mathematics and sport meet. AUEB Sports Analytics Workshop 2018 is hosted by Athens University of Economics and Business (Greece) and organized by the Department of Statistics from Monday 26th of November to Tuesday 27th of November 2018. It will be the 3rd conference in Greece that brings together professionals and academics with a common interest in applying cutting-edge quantitative methods on Sports.

**Topics include:**
- Mathematical and physical models in sports
- Performance measures and models
- Optimisation of sports performance
- Statistics and probability models
- Match outcome models
- Competitive strategy
- Game theoretical models
- Optimal tournament design and scheduling,
- Decision support systems
- Econometrics in sport
- Analysis of sporting technologies
- Computationally intensive methods
- Financial valuation in sport

**Announcement**

A limited number of contributed talks could be accepted.
Deadline for abstract submission:
11 November 2018
Submit your abstract [here](#)
In order to edit your abstract, you have to create a free account in [https://easychair.org/](https://easychair.org/)
Math Sport International

7th in Series
previously hosted in
• Manchester,
• Groningen,
• Manchester,
• Leuven,
• Loughborough &
• Padova

MathSport International organizes biennial conferences dedicated to all topics where mathematics and sport meet. MathSport International 2019 is hosted by Athens University of Economics and Business (Greece) and organized by the Department of Statistics from Monday 1st of July to Wednesday 3rd of July 2019. It will be the 7th conference in Europe that brings together Maths and Sport. A social event is foreseen for the evening of Sunday 30th of June.

Topics include:
• Mathematical and physical models in sports
• Performance measures and models
• Optimisation of sports performance
• Statistics and probability models
• Match outcome models
• Competitive strategy
• Sports Quantitative marketing
• Game theoretical models
• Optimal tournament design and scheduling,
• Decision support systems
• Analysis of rules and adjudication
• Econometrics in sport
• Analysis of sporting technologies

Keynote Speakers
• Luke Born (Simon Fraser University & Strategy and Analytics of Sacramento Kings)
• Simon Jenkins (University of Winchester)
• Ioannis Kosmidis (University of Warwick)
• Stephanie Kovalchik (Victoria University)
• Raymond Stefani (California State University)
Introduction

Football/Soccer is the best sport for implementing Science/Statistics/Analytics

• Low number of events (so difficult to predict)
• High uncertainty (so difficult to predict)
• Very popular (because it is difficult to predict?)
• Very profitable (because it is difficult to predict?)
• High Financial Risk of investment (because passion becomes more important than numbers and science) – Professional Teams are usually acting as win-maximizers and not profit-maximizers
Main Topics Quantitative analysis of Football/Sports

- Prediction
- Player Evaluation & Performance analytics
- Physical Metrics of Players in training
- Inline game metrics with wearables
- Scheduling
- Sports Economics & Competitive Balance
- Other (Passing Network Analytics, Referee effects, Red card effect, Home effect, Corruption Analytics, Analysis of substitution times)
Prediction

- Offline (before the game)
- Inline (within the game)
Offline Prediction

Modeling of

• Game Scores
  – Poisson based models and extensions
  – Modeling the difference using the Skellam model

• Final outcome of a game (Win/Draw/Loss)
  – Multinomial regression model
  – Bradley Terry Model
Models for Counts

• Simple Poisson Model (Maher, 1982; Lee, 1992; Dixon & Coles, 1997, Karlis and Ntzoufras, 2000)
• Bivariate Poisson Model (Karlis & Ntzoufras, 2003)
• Negative Binomial Model (see e.g. Ntzoufras 2009)
• Skellam Model for the goal difference (Karlis & Ntzoufras, 2009)
• Poisson-log-normal random effects model (not the best for football counts; see e.g. Ntzoufras 2009)
Models for Scores

Such models allow us not only to predict a single football game but also (simulation based results)

- Final League reproduction
- Estimate probabilities of winning a league, winning European tickets, or relegation.
- Estimate final rankings
- Estimate results under different scenarios/assumptions (by changing covariates i.e. conditions of the game)
Offline Prediction

**Poisson Based models**

- Vanilla model: home effect + teams attacking and defensive parameters
- Models with time evolved team parameters (time and form matters!)
- Additional covariates
  - Odds from betting teams (easily accessible – good covariates)
  - Team performance (ingame and before the game)
  - Information about events and formation (team strategy, formation, injuries etc.)
  - Economo-demographic variables (Stability, tradition, Budget, Player Value, Coach Value, Country of origin for European leagues)
  - Prior information (previous games between the teams)
  - Team form (e.g. performance in last 5 games)
Offline Prediction

The simple (vanilla) Poisson model

The model is expressed by

\[ Y_{ij} \sim \text{Poisson}(\lambda_{ik}) \quad \text{for } j = 1, 2 \]

\[
\log(\lambda_{i1}) = \mu + \text{home} + a_{\text{HT}_i} + d_{\text{AT}_i}
\]

\[
\log(\lambda_{i2}) = \mu + a_{\text{AT}_i} + d_{\text{HT}_i} \quad \text{for } i = 1, 2, \ldots, n,
\]

where \( n \) = number of games, \( \mu \) = constant parameter; home = home effect; \( \text{HT}_i \) and \( \text{AT}_i \) = home and away teams in \( i \) game; \( a_k \) and \( d_k \) = attacking and defensive effects–abilities of \( k \) team for \( k = 1, 2, \ldots, K \); and \( K \) = number of teams in the data (here \( K = 20 \)).

In full (balanced) round-robin leagues, the parameters can be easily calculated by considering averaged of scored/conceded goals for each team.
Data for the simple (vanilla) model

• **Observations**
  – $2 \times \text{Number of games (N)}$
  – Each game will occupy two lines/observations (one for home team and one for away team)

• **Response Variable**: Goals scored by each team in each game

• **Covariates**
  – **Home effect**: Binary for home and away teams (1 for home teams and zero otherwise)
  – **Scoring team**: Categorical factor for the team scoring the number of goals (the corresponding coefficient will estimate the attacking ability of each team)
  – **Team accepting goals**: Categorical factor for the team receiving the number of goals (the corresponding coefficient will estimate the defensive ability of each team).
Important Assumptions

• Dependence/Independence of Goals of a game
• Time dependent attacking and defending parameters
• What about draw inflation?
• What about Over-dispersion?
• Shall we focus on modeling scores or outcomes (win/draw/loss)?

Checking the performance of the predictions

• Checking model fit and prediction using in-sample and out-of-sample measures
What can we do with such models?

We now can calculate via simulation

- The probability of a specific score
- The probability of a score difference
- The probability of win/loss/draw
- Calculate the probability of winning the league or each position
- Reproduce the league under the model
What if analysis for 2017-18

Super league 2017-2018

<table>
<thead>
<tr>
<th>#</th>
<th>ΟΜΑΔΑ</th>
<th>ΑΓΩΝΕΣ</th>
<th>ΒΑΘΜΟΙ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Α.Ε.Κ.</td>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td>Π.Α.Ο.Κ.</td>
<td>30</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>ΟΛΥΜΠΙΑΚΟΣ</td>
<td>30</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>ΑΤΡΟΜΗΤΟΣ ΑΘ.</td>
<td>30</td>
<td>56</td>
</tr>
<tr>
<td>5</td>
<td>ΑΣΤΕΡΑΣ ΤΡΙΠ.</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>Α.Ο. ΞΑΝΘΗ</td>
<td>30</td>
<td>45</td>
</tr>
</tbody>
</table>

- PAOK -3 points for the game PAOK-AEK (11/3/2018) and lost this game
- PAOK also lost the game PAOK-OLYMPIAKOS (25/2/2918) without playing
What if analysis for 2017-18

Super league 2017-2018

<table>
<thead>
<tr>
<th>#</th>
<th>ΟΜΑΔΑ</th>
<th>ΑΓΟΝΕΣ</th>
<th>ΒΑΘΜΟΙ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Α.Ε.Κ.</td>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td>Π.Α.Ο.Κ.</td>
<td>30</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>ΟΛΥΜΠΙΑΚΟΣ</td>
<td>30</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>ΑΤΡΟΜΗΤΟΣ ΑΘ.</td>
<td>30</td>
<td>56</td>
</tr>
<tr>
<td>5</td>
<td>ΑΣΤΕΡΑΣ ΤΡΙΠ.</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>Α.Ο. ΞΑΝΘΗ</td>
<td>30</td>
<td>45</td>
</tr>
</tbody>
</table>

PAOK – OLYMPIAKOS

PAOK Wins 57.5%
Draw 16%
Olympiakos Wins 26.5%
### Super league 2017-2018

<table>
<thead>
<tr>
<th>#</th>
<th>ΟΜΑΔΑ</th>
<th>ΑΓΩΝΕΣ</th>
<th>ΒΑΘΜΟΙ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Α.Ε.Κ.</td>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td>Π.Α.Ο.Κ.</td>
<td>30</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>ΟΛΥΜΠΙΑΚΟΣ</td>
<td>30</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>ΑΤΡΟΜΗΤΟΣ ΑΘ.</td>
<td>30</td>
<td>56</td>
</tr>
<tr>
<td>5</td>
<td>ΑΣΤΕΡΑΣ ΤΡΙΠ.</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>Α.Ο. ΞΑΝΘΗ</td>
<td>30</td>
<td>45</td>
</tr>
</tbody>
</table>

**PAOK – AEK**

- **PAOK Wins**: 44.5%
- **Draw**: 22%
- **AEK Wins**: 33.5%
Super league 2017-2018

Final Result
- PAOK Champion 60%
- Tie PAOK & AEK 25%
- AEK Champion 15%
Who is going to be the champion?

<table>
<thead>
<tr>
<th>Θέση</th>
<th>Ομάδα</th>
<th>A</th>
<th>N</th>
<th>l</th>
<th>H</th>
<th>ΓΥ</th>
<th>ΓΚ</th>
<th>+/-</th>
<th>ΒΑΘ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ΠΑΟΚ</td>
<td>19</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>37</td>
<td>8</td>
<td>29</td>
<td>51</td>
</tr>
<tr>
<td>2</td>
<td>Ολυμπιακός</td>
<td>19</td>
<td>14</td>
<td>3</td>
<td>2</td>
<td>35</td>
<td>9</td>
<td>26</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>ΑΕΚ</td>
<td>19</td>
<td>12</td>
<td>4</td>
<td>3</td>
<td>33</td>
<td>8</td>
<td>25</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>Ατρόμητος</td>
<td>19</td>
<td>9</td>
<td>6</td>
<td>4</td>
<td>28</td>
<td>19</td>
<td>9</td>
<td>33</td>
</tr>
<tr>
<td>5</td>
<td>Άρης</td>
<td>19</td>
<td>9</td>
<td>2</td>
<td>8</td>
<td>21</td>
<td>19</td>
<td>2</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>Παναθηναϊκός</td>
<td>19</td>
<td>9</td>
<td>7</td>
<td>3</td>
<td>27</td>
<td>16</td>
<td>11</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>Παναιτωλικός</td>
<td>19</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>23</td>
<td>25</td>
<td>-2</td>
<td>26</td>
</tr>
<tr>
<td>8</td>
<td>Σάνθη</td>
<td>19</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>15</td>
<td>15</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>Πανιώνιος</td>
<td>19</td>
<td>6</td>
<td>4</td>
<td>9</td>
<td>18</td>
<td>29</td>
<td>-11</td>
<td>22</td>
</tr>
<tr>
<td>10</td>
<td>Αστέρας Τρίπολης</td>
<td>19</td>
<td>6</td>
<td>3</td>
<td>10</td>
<td>16</td>
<td>22</td>
<td>-6</td>
<td>21</td>
</tr>
<tr>
<td>11</td>
<td>Λωμία</td>
<td>19</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td>14</td>
<td>25</td>
<td>-11</td>
<td>20</td>
</tr>
<tr>
<td>12</td>
<td>Λόρισα</td>
<td>19</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td>16</td>
<td>23</td>
<td>-7</td>
<td>20</td>
</tr>
<tr>
<td>13</td>
<td>ΟΦΗ</td>
<td>19</td>
<td>3</td>
<td>8</td>
<td>8</td>
<td>18</td>
<td>28</td>
<td>-10</td>
<td>17</td>
</tr>
<tr>
<td>14</td>
<td>ΠΑΣ Πάνινα</td>
<td>19</td>
<td>5</td>
<td>2</td>
<td>12</td>
<td>14</td>
<td>29</td>
<td>-15</td>
<td>17</td>
</tr>
<tr>
<td>15</td>
<td>Λεβαδειακός</td>
<td>19</td>
<td>3</td>
<td>4</td>
<td>12</td>
<td>11</td>
<td>27</td>
<td>-16</td>
<td>13</td>
</tr>
<tr>
<td>16</td>
<td>Απόλλων Σμύρνης</td>
<td>19</td>
<td>1</td>
<td>3</td>
<td>15</td>
<td>7</td>
<td>31</td>
<td>-24</td>
<td>6</td>
</tr>
</tbody>
</table>
Who is going to be the champion?

PAOK
Olympiakos
AEK
Atromitos
Panathinaikos
Aris
Panetolikos
Xanthi
Asteras Tripolis
Panionios
Lamia
Larisa
OFI Crete
Giannina
Levadeiakos
Apollon
Who is going to be the champion?

PAOK
Olympiakos
AEK
Atromitos
Panathinaikos
Aris
Panetolikos
Xanthi
Asteras Tripolis
Panionios
Lamia
Larisa
OFI Crete
Giannina
Levadeiakos
Apollon

2.5% 50% 97.5%

PAOK 70 77 82
Olympiakos 60 68 74
AEK 55 62 68
Atromitos 44 51 58
Panathinaikos 39 46 54
Aris 36 43 50
Who is going to be the champion?

Champion 2018-19?

- PAOK: 95.8%
- Tie: 2.8%
- Olympiakos: 95.8%
Who is going to be the champion?

Champion 2018-19?

- **PAOK**: 95.8%
- **Tie**: 2.8%
- **Olympiakos**: 2%
Who is going to be the champion?

Champion 2018-19 If Olympiakos Wins?

- **PAOK**: 85% (-10%)
- **Tie**: 10% (+7%)
- **Olympiakos**:
Who is going to be the champion?

<table>
<thead>
<tr>
<th>Team</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt;</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Olympiakos</td>
<td>79.4%</td>
<td>11%</td>
</tr>
<tr>
<td>AEK</td>
<td>11%</td>
<td>81%</td>
</tr>
</tbody>
</table>
Who is going to be the champion?

- Aris: 68.6%
- Panathinaikos: 38.9%
- Atromitos: 25%
Prediction within the game

Modeling of

- Time to event (goal)
  - Survival analysis based models
    - Dixon & Robinson (1998, RSSD)
    - Nevo and Ritov (2013, JQAS)
    - Boshnakov, Kharrat, McHale (2017, Int. J. Forecasting)
    - Work in progress by our team

- Model the probability of event for short intervals (every 1 or 5 minutes)
  - Using Binomial mixed models for repeated measures
EURO 2012 FINAL

Spain – Italy = 4-0

Survival Inline Plot
(based on a Bayesian Model using posterior medians of the expected arrival times)

From our work in progress with I. Leriou & D. Karlis
Player Evaluation

**Aim**

- Estimate the contribution of players in a team
- Rank, identify and reward best players
- Scouting – Early Identification of talents
- Estimate the future performance/value of a Player
- Help the manager to decide the best formation
Methods

• Simple approach with binary indicators
• Random effects
• Analysis based on Game Performance Indicators
• Expected Goals (xG) and Expected Assists (xA)
• Player Economic/Marketing Value and performance
Methods (2)

• Simple approach with indicators
  – Build a model with indicators whether a player was in the field
  – Binary indicators for players
  – Difficult to build a dataset. Each game should be splitted in multiple lines according to substitution times

• Analysis based on Game Performance Indicators
  – Build a model to identify the importance of each event in the game (goals, shots, steals, passes, speed, stamina, area covered etc.)
  – Use model indicators to build an index of players
  – McHale, Scarf & Folker (2012, Interfaces) building different indexes based on different response measures
Methods (3)

- Random effects
  - Use random effects to identify individual contribution
  - Goal Scoring: McHale & Szczepanski (2014, JRSSA)
  - Passing Skills: Szczepanski & McHale (2016, JRSSA)
- Player Economic/Marketing Value and performance
  - Evaluating the efficiency of the association football transfer market using regression based player ratings (pre-print only)
Methods (4)
McHale, Scarf & Folker (2012, Interfaces)
building different indexes based on different response measures
Index ingredients:
• Subindex 1: Modelling Match Outcome (model based with outcome probability)
• Subindex 2: Points-Sharing Index (time played by each players and points)
• Subindex 3: Appearance Index (time played by each players)
• Subindex 4: Goal-Scoring Index
• Subindex 5: Assists Index
• Subindex 6: Clean-Sheets Index
Player Evaluation

Methods (5)

Expected Goals (xG)
- We model every shot
- Response measure: is the probability of a shot resulting in a goal
- The sum of these probabilities will give the xG of a player and a team
- Similar for assists (xA)
- References:
  - [https://www.optasports.com/services/analytics/advanced-metrics/](https://www.optasports.com/services/analytics/advanced-metrics/)
  - [https://understat.com/](https://understat.com/)
Expected Goals (xG): https://understat.com/
Xgoals also in Greek media

χGoals: Η επανάσταση που άλλαξε το ποδόσφαιρο

Τι είναι τα xGoals και γιατί έχουν αλλάξει το ποδόσφαιρο; Το Sport24.gr σας παρουσιάζει τα expected goals, που έχουν βελτιώσει τον τρόπο που βλέπουμε το ποδόσφαιρο, το πώς παίζουν οι ομάδες και αξιολογούνται οι ποδοσφαιριστές. Ο θέμας Καίσαρης αναλύει την επανάσταση που επέτυχε να βάξει το ποδόσφαιρο σε σωστές βάσεις και μας επιτρέπει καλύτερη ανάλυση του παιχνιδιού.


Mega analysis: Τα xGoals ανοίγουν τα X-files της Super League

Ποια ομάδα φτιάχνει τις καλύτερες τελικές της Super League και ποια δέχεται τις πιο επικίνδυνες; Ποιοι εξεχωρίζουν στα μάρκα τίτλου και γιατί είναι πίσω ο Ολυμπιακός; Σεκάστε τα γκαλ, το θέμα είναι τα xGoals, που χρησιμοποιούνται για πρώτη φορά στο ελληνικό πρωτάθλημα. Το Sport24.gr και ο θέμας Καίσαρης σας δίνουν όλες τις απαντήσεις, για όλες τις ομάδες, σε μια UNIQUE ανάλυση.

Xgoals also in Greek media

### Xgoals in American Soccer Analysis website

https://www.americansocceranalysis.com/

<table>
<thead>
<tr>
<th>Shooter/Team Model</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Z-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.172</td>
<td>0.170</td>
<td>24.589</td>
<td>0.000</td>
</tr>
<tr>
<td>Distance (log-yds)</td>
<td>-2.353</td>
<td>0.047</td>
<td>-50.056</td>
<td>0.000</td>
</tr>
<tr>
<td>Goal Mouth Available (quadratic-yds)</td>
<td>-0.026</td>
<td>0.007</td>
<td>-3.785</td>
<td>0.000</td>
</tr>
<tr>
<td>Goal Mouth Available (yds)</td>
<td>0.069</td>
<td>0.019</td>
<td>3.716</td>
<td>0.000</td>
</tr>
<tr>
<td>Headed (binary)</td>
<td>-0.648</td>
<td>0.066</td>
<td>-9.746</td>
<td>0.000</td>
</tr>
<tr>
<td>Cross (binary)</td>
<td>-0.380</td>
<td>0.061</td>
<td>-6.206</td>
<td>0.000</td>
</tr>
<tr>
<td>Through ball (binary)</td>
<td>0.909</td>
<td>0.074</td>
<td>12.292</td>
<td>0.000</td>
</tr>
<tr>
<td>Corner (binary)</td>
<td>-0.622</td>
<td>0.064</td>
<td>-9.753</td>
<td>0.000</td>
</tr>
<tr>
<td>Free kick (binary)</td>
<td>0.539</td>
<td>0.117</td>
<td>4.592</td>
<td>0.000</td>
</tr>
<tr>
<td>Indirect Free kick (binary)</td>
<td>-0.192</td>
<td>0.080</td>
<td>-2.393</td>
<td>0.017</td>
</tr>
<tr>
<td>Fastbreak (binary)</td>
<td>0.680</td>
<td>0.106</td>
<td>6.397</td>
<td>0.000</td>
</tr>
<tr>
<td>Penalty (binary)</td>
<td>2.735</td>
<td>0.134</td>
<td>20.336</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Example of the Simple approach with indicators

- 351 matches of the La Liga Season 2015/2016
- 954 goals (555 goals were scored by home teams, 399 conceded)
- 110 scored by Real Madrid, 34 conceded
- M.Sc. Thesis at AUEB by A. Mourtopallas
Impact of players

Category:
- a: Bad at both
- b: Better in attack
- c: Better in defence
- d: Good at both

Players:
- Nacho Fernandez
- Alvaro Arbeloa
- Danilo
- Karim Benzema
- Cristiano Ronaldo
- Luka Modric
- Raphael Varane
- Marcelo
- Isco
- Casemiro
- Keylor Navas
- Mateo Kovacic
- Daniel Carvajal
- Lucas Vazquez
- Sergio Ramos
- Francisco Casilla
- Gareth Bale
- Pepe
- James Rodriguez
- Borja Mayoral
Impact of midfielders

- Luka Modric
- Toni Kroos
- Isco
- Casemiro
- Mateo Kovacic
- James Rodriguez

Salary:
- 150
- 125
- 100
- 75
- 50

Starter:
- No
- Yes
Conclusions

Cristiano Ronaldo is the key player of the team

Tony Kroos’ impact is higher than we may presume

Nacho Fernandez improved since last season (very high def contribution)

Lucas Vasquez is a very promising player (contributed positively in both attack and defensive dimensions with low salary)

Gareth Bale performed less than expected (overprized)

Pepe ⇒ low defensive contribution – high salary (overprized?)
Metrics for physical improvement and training

**Aim**

- Improve the physical condition of athletes
- Focus on specific skills and measure them
- Avoid injuries
- Improves the team by optimizing allocated training time
Inline game metrics with wearables

The aim is to measure
- Movement of players in the game
- Speed and coverage
- Physical condition
- Physical and tactics performance

It helps
- Evaluate the performance of players and teams within a game
- The manager to decide formation and substitutions
League and Contest Scheduling

**AIM**

- Fair scheduling
- Eliminate bias due to the sequence of games
- Strengthen competitiveness (related with next slides)
- Incorporate constraints (incl. other sports, safety issues, other events, tv requirements etc.)

**HOW?**

- Using Operational Research and optimization methods
- Hybrid search methods
- Validate using simulation methods from Statistical models
Competitive Balance

• A balanced league increases the interest of the fans and improves the athletic product

• The notion of a balanced league is not yet very well defined
  – Equal Strength between all teams? or
  – Equal Strength between best teams (or the teams with the highest number of fans?)
What league do we want to see?

- All fans like the fact that a weaker team occasionally wins a game or a league
- May neutral fans follow the weakest team e.g. Greece in Euro 2004

But

- They do not like their team to lose
- They like or they are willing to pay an expensive ticket to see a final with high ranked and expensive teams e.g. Bayern-Barcelona
Competitive balance in Greek League

Joint work with V. Manasis

Moving Averages of lag five for $DN_1$ (Champion) from 1959-2008

Value of Competitive Balance

Season


1983-84 Παναθηναϊκός
1984-85 Π.Α.Ο.Κ.
1985-86 Παναθηναϊκός
1986-87 Ολυμπιακός
1987-88 Λάρισα
1988-89 Α.Ε.Κ.
1989-90 Παναθηναϊκός
1990-91 Παναθηναϊκός

Competitive balance in Greek League

Joint work with V. Manasis

Greece

England

Germany
What about the last decade in the Greek League?

Source: SAW2019 presentation of V. Manasis
ManU won 13 out of 17 leagues for the period 1992-2009 and it was not ranked in lower position than 3rd.

3 cases in England ⇒ promoted team ⇒ won the championship:
- Ipswich (1961)
- Nottingham (1997)
- Leicester (2015-16 – not in the Figure)

Joint work with V. Manasis
### Premier League after 13 games of the 2015/16 season (when Leicester won)

<table>
<thead>
<tr>
<th>TEAM</th>
<th>P</th>
<th>W</th>
<th>D</th>
<th>L</th>
<th>GD</th>
<th>PTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leicester City</td>
<td>13</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>Manchester United</td>
<td>13</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>Manchester City</td>
<td>13</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td>Arsenal</td>
<td>13</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>12</td>
<td>26</td>
</tr>
<tr>
<td>Tottenham Hotspur</td>
<td>13</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>West Ham United</td>
<td>13</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td>Everton</td>
<td>13</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>Southampton</td>
<td>13</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>Liverpool</td>
<td>13</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>Crystal Palace</td>
<td>13</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>Stoke City</td>
<td>13</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>West Bromwich Albion</td>
<td>13</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>Watford</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Swansea City</td>
<td>13</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>Chelsea</td>
<td>13</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>Norwich City</td>
<td>13</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Newcastle United</td>
<td>13</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>Sunderland</td>
<td>13</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>Bournemouth</td>
<td>13</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>Aston Villa</td>
<td>13</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>14</td>
<td>5</td>
</tr>
</tbody>
</table>
Premier League after 13 games of the 2015/16 season

“Where anybody can beat anybody”

WHERE ANYONE... CAN BEAT ANYONE
What about the last decade in Premier League?

Source: SAW2019 presentation of V. Manasis

Credits: Manasis (2018) Sports Analytics Workshop Presentation

<table>
<thead>
<tr>
<th>Year</th>
<th>Club</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006–07</td>
<td>Manchester United (16)</td>
</tr>
<tr>
<td>2009–10</td>
<td>Chelsea (4)</td>
</tr>
<tr>
<td>2010–11</td>
<td>Manchester United (19)</td>
</tr>
<tr>
<td>2011–12</td>
<td>Manchester City (3)</td>
</tr>
<tr>
<td>2012–13</td>
<td>Manchester United (20)</td>
</tr>
<tr>
<td>2015–16</td>
<td>Leicester City</td>
</tr>
<tr>
<td>2016–17</td>
<td>Chelsea (6)</td>
</tr>
</tbody>
</table>
Moving Averages of lag five for DN_1 (Champion) from 1959-2008

One case => promoted team => won the championship: Kaiserslauten in 1998
How to design Knockout Tournaments?

• Do we support the stronger or the weakest teams?

We do not wish to see

• many strong teams to be disqualified early
• Two weak or not popular teams in the final

We do wish to see

• Some strong teams to be disqualified early
• Some weak teams to qualify further against all odds
Sports Economics & Competitive balance

In round-robin contests (National leagues)?
• Do we support the stronger or weakest teams?
• Small or large leagues?
• Playoffs?
• Give more money to strong teams (reward) or to weak teams (balance)?
• What about promotion/relegation rules (refreshes the interest or just recycles bad teams?)

We do wish to see
• A large enough group of teams to be close and compete for the championship
• A large enough group of teams to be close and compete for European tickets

We do not wish to see
• A team having big margin of points from all the rest (so the champion is known early)
• Teams with low number of points so they are not competitive (early relegation)
• Teams with economic problems
For UEFA Champions League

- Does it need improvement?
- Not metrics to measure balance
- Big discussion of how to reward teams and share income
- Closed or Open League?
- How many teams from each National League/Country
- The current income share and reward system destroys the balance in National teams in second ranked leagues like Greece.
Home sweet home

• Home effect/advantage is well established in
  • «Καφενείον» discussions
  • Data based Studies
• Pollard (1986)
  ⇒ relatively stable in English League from 1888!
  ⇒ ~64% of the points from home teams for 1970-1981
FunStats & Facts

Home effect estimates (% points won) for 1998-2004

Home effect in Super league
+24% (2017-18)
+53% (2018-19)

This is Balkans!
Home effect +10%

Figure 1. Map of Europe showing home advantage in the national league of each country.
FunStats & Facts

Home sweet home

• Home effect is stronger in 2\textsuperscript{nd} division leagues compared to 1\textsuperscript{st} division

• Home effect is lower/smaller in derbies (reported in various studies)

Referee bias contributes to home advantage in English Premiership football

Figure 1. Mean home advantage in terms of goal differential for each of the 50 referees included in the analysis (diamonds) after controlling for team ability and crowd size compared to the league-wide average home advantage (dashed line). Error bars represent standard errors.

Home effect (goal difference) in Super league
+0.43 (2017-18)
+0.39 (2018-19)
FunStats & Facts – The Red Card Effect
Cerveny, Ours, Tuijl (2018). Empirical Economics

“Ten do it better” myth
The effect of black uniforms

- Significant in American Football (NFL) and Hockey (NHL)
- Not in Association Football? (At least not in Turkey!)
FunStats & Facts - Jersey Color Effect!

- **Red jersey** judged more harshly than the rest for tackles from behind
- “We revealed that tackles from behind were judged more harshly against players dressed in red than against those dressed in blue, green and yellow”
- (the effect is merely significant – p.value=0.07)
- Significant difference vs. blue
FunStats & Facts - **Jersey Color Effect!**

- **RED Color** is the right color for a **goalkeeper** to catch a penalty!!!
- “Players facing red-clad goalkeepers scored on fewer penalty kicks than those facing either blue- or green-clad goalkeepers, but no differences in expectancy of success emerged.”
FunStats & Facts - Jersey Color Effect!

Journal of Sports Sciences

ISSN: 0264-0414 (Print) 1466-447X (Online) Journal homepage: https://shapeamerica.tandfonline.com/loi/rjsp20

Red shirt colour is associated with long-term team success in English football

Martin J. Attrill, Karen A. Gresty, Russell A. Hill & Robert A. Barton

To cite this article: Martin J. Attrill, Karen A. Gresty, Russell A. Hill & Robert A. Barton (2008) Red shirt colour is associated with long-term team success in English football, Journal of Sports Sciences, 26:6, 577-582, DOI: 10.1080/02640410701736244
The home advantage over the first 20 seasons of the English Premier League: Effects of shirt colour, team ability and time trends

Mark S. Allen\textsuperscript{a*} and Marc V. Jones\textsuperscript{b}

\textsuperscript{a}Department of Applied Science, London South Bank University, London, UK; \textsuperscript{b}Centre for Sport, Health and Exercise Research, Staffordshire University, Stoke-on-Trent, UK

(Received 3 July 2012; final version received 22 October 2012)
Wearing red helps you win

Tim Radford, science editor
Thu 19 May 2005 11,45 BST

Red is the tint for winners. When all else is equal, a sporting strip of scarlet is enough to tip the balance, British scientists report in Nature today.

Almost on the eve of an FA Cup final clash between two teams that both normally sport a red strip, Russell Hill and Robert Barton of the University of Durham have identified a new variable for sporting tipsters and a new challenge for the athletics authorities: red seems to confer an advantage.

"Our results suggest that the evolutionary psychology of aggressive competition is likely to be a fertile field for further investigation," they report. "The implication for regulations governing sporting attire may also be important."

Redness indicates anger, testosterone and male aggression in humans, mandrills and sticklebacks. In experiments, red leg bands have helped ringed birds win a higher place in the pecking order. Red plays a big role in signalling superiority throughout the animal world.

The two scientists decided to investigate the role of red in human contests. They ignored Team Ferrari, with its special tint, and Manchester United and Arsenal's blood-red combat kits, and focused on the sports where the colours are randomly assigned. They examined the outcomes in boxing, tae kwon do, Graeco-Roman and freestyle wrestling, the contact sports of the 2004 Olympics, where contestants were randomly given either red or blue outfits. If colour had nothing to do with it, then the number of red and blue winners should be evenly matched.
FunStats & Facts - **Jersey Color Effect!**

- Man. UND
- Arsenal
- Liverpool
Color Psychology in Football: The Effect of Shirt Color on a Team’s Performance in the Dutch Eredivisie

Rosenbaum and Rubin (1983). The found result in this paper is that red colored teams have an advantage in earning points per game and scoring goals relative to getting goals against. These results can have very important implications for club policy makers who want to change the club colors or for people who want to start up a new football team. According to the results, they should choose the color red as the major color for their home shirts.

- **RED** is still important
- Found in Dutch football
- Not in German League
- Also in other sports (handball, Australian football).
- Be careful! Strongly Confounding with Jerseys of top teams
- What about Greek Football? Is Red Jersey Important?
Which league is the wealthiest?

![Chart showing revenue of the biggest (Big Five*) European soccer leagues from 1996/97 to 2018/19 (in million euros)]

Source: Deloitte
© Statista 2018

Additional Information:
Europe; Deloitte, 1996/97 to 2016/17
Who was the best player for 2018 (Jan 2019)?

Lars Magnus Hvattum
Professor of Quantitative Logistics
Molde University College

Molde University College
Specialized University in Logistics
Who was the best player for 2018 (Jan 2019)?

https://gaming.youtube.com/watch?v=jLfACAC4V-I&feature=share
Who was the best player for 2018 (Jan 2019)?

<table>
<thead>
<tr>
<th></th>
<th>Player</th>
<th>Games</th>
<th>Role</th>
<th>Club</th>
<th>Rating</th>
<th>Goals</th>
<th>Assists</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lionel Messi</td>
<td>31</td>
<td>F</td>
<td>Barcelona</td>
<td>53278</td>
<td>0.365</td>
<td>0.308</td>
</tr>
<tr>
<td>2</td>
<td>Neymar</td>
<td>26</td>
<td>F</td>
<td>Paris Saint-Germain</td>
<td>20897</td>
<td>0.313</td>
<td>0.305</td>
</tr>
<tr>
<td>3</td>
<td>Thomas Müller</td>
<td>29</td>
<td>F</td>
<td>Bayern München</td>
<td>36226</td>
<td>0.319</td>
<td>0.298</td>
</tr>
<tr>
<td>4</td>
<td>Sadio Mané</td>
<td>26</td>
<td>F</td>
<td>Liverpool</td>
<td>22828</td>
<td>0.304</td>
<td>0.297</td>
</tr>
<tr>
<td>5</td>
<td>Mohamed Salah</td>
<td>26</td>
<td>F</td>
<td>Liverpool</td>
<td>21439</td>
<td>0.283</td>
<td>0.278</td>
</tr>
<tr>
<td>6</td>
<td>David Alaba</td>
<td>26</td>
<td>DM</td>
<td>Bayern München</td>
<td>27614</td>
<td>0.282</td>
<td>0.277</td>
</tr>
<tr>
<td>7</td>
<td>Bernardo Silva</td>
<td>24</td>
<td>MF</td>
<td>Manchester City</td>
<td>16382</td>
<td>0.278</td>
<td>0.273</td>
</tr>
<tr>
<td>8</td>
<td>Robert Lewandowski</td>
<td>30</td>
<td>F</td>
<td>Bayern München</td>
<td>38539</td>
<td>0.303</td>
<td>0.266</td>
</tr>
<tr>
<td>9</td>
<td>Luis Suárez</td>
<td>31</td>
<td>F</td>
<td>Barcelona</td>
<td>39495</td>
<td>0.325</td>
<td>0.264</td>
</tr>
<tr>
<td>10</td>
<td>Thiago Alcántara</td>
<td>27</td>
<td>M</td>
<td>Bayern München</td>
<td>17819</td>
<td>0.272</td>
<td></td>
</tr>
</tbody>
</table>

https://gaming.youtube.com/watch?v=jLfACAC4V-I&feature=share
Who was the best player for 2018 (Jan 2019)?

https://gaming.youtube.com/watch?v=jLfACAC4V-I&feature=share
Who was the best player for 2018 (Jan 2019)?

https://gaming.youtube.com/watch?v=jLfACAC4V-I&feature=share
What about Ronaldo?

- Position 23 from 9 (June 2018)
- Only player over 33 at top30
- Peak rating $\uparrow$
- Current rating $\downarrow$ due to
  - Age
  - Transfer

https://gaming.youtube.com/watch?v=jLfACAC4V-I&feature=share
Is the Greek Super League the worst?

Mean values per country for the most comprehensive Special Dynamic Concentration index
### FunStats & Facts

**Is the Greek Super League the worst?**

#### Average stadium utilization at professional football matches in Europe between 2010 and 2017, by league

<table>
<thead>
<tr>
<th>League</th>
<th>Average Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premier League</td>
<td>94.95%</td>
</tr>
<tr>
<td>Deutsche Fussball Liga GmbH</td>
<td>91.27%</td>
</tr>
<tr>
<td>Eredivisie CV</td>
<td>88.23%</td>
</tr>
<tr>
<td>Ligue de Football Professionnel</td>
<td>70.67%</td>
</tr>
<tr>
<td>LaLiga</td>
<td>Liga De Fútbol Profesional</td>
</tr>
<tr>
<td>Norsk Toppfotball</td>
<td>63.33%</td>
</tr>
<tr>
<td>Scottish Professional Football League</td>
<td>61.89%</td>
</tr>
<tr>
<td>Belgium Pro League</td>
<td>58.22%</td>
</tr>
</tbody>
</table>
FunStats & Facts

Celebrating can be Dangerous

More than 1 in 20 soccer injuries are caused by celebrating goals on the pitch.

In the Turkish league a study found that almost 6% of the injuries were caused by goal celebrations.

American Journal of Sports Medicine, 2005, pp. 1237–1240
Picture from FACTSLIDES: https://www.factslides.com/s-Soccer
Concluding remarks

To conclude with

- **Prediction** is important for fans (in terms of betting) ⇒ increases profits of bet companies and interest for the sport product (in macro perspective).

- **Inline prediction** is important for fans (in terms of betting) ⇒ increases profits of bet companies and interest for the sport product (Media – TV, Radio, Internet).
Concluding remarks

• **Player performance and evaluation** ⇒ Of main interest for: the fans (Player Ranking), Teams (Scouting, Future Performance and Value), Companies (Sponsoring), Players (A lot of money from all previous), Coaches/Managers (Selection of better players)

• **Physical Measurements** (Training and Games): It is related with player evaluation. Main value to help managers/coaches to improve their teams. In macro perspective also the teams financial position is also improving.

• **Scheduling and Competitive Balance**: More Fair and Balanced contests lead to better product and more profit.
NO Matter How Many Goals You Save People Always Remember The One You Miss.

That's all Folks!

THANK YOU