

Tuesday 7 November 2017

Combining Prediction Markets and Forecasting Contests.

James Reade, Associate Professor, University of Reading

Two popular methods for aggregating individual forecasts are prediction markets, where participants bet on the outcome of future events, and forecasting contests, where participants are ranked according to the accuracy of their forecasts. Can these methods be used in concert to produce more accurate forecasts? We analyze 1.79 million forecasts on oddsportal.com, a social network for sports tipsters. Tipsters are ranked according to the betting return on their tips. We find that an aggregation of these tips predicts sporting outcomes, after controlling for betting/prediction market prices. Rank-order forecasting contests, even without tangible prizes, are useful tools for eliciting crowd forecasts.

A Hierarchical Bayesian Model for Football Scores using the Bookmakers Odds.

Leonardo Egidi, Università degli Studi di Padova

In recent years the challenge of modeling football outcomes has gained increasing attention, in large part due to the potential for making substantial profits in betting markets. According to the current literature, there are two kinds of models designed for predicting football outcomes: so called direct models for the number of goals scored by two competing teams, and indirect models for estimating the probability of the categorical outcome Win/Draw/Loss. In this presentation we discuss an extension to the direct modeling framework with a particular focus on applications to betting. In fact, despite the strong connection existing between football models and betting odds, no authors have used these odds as a part of the model for improving the predictive accuracy on hold-out data and, more generally, the model fit. Roughly speaking, we model the home and the away scores through a hierarchical Bayesian Poisson mixture model by weighting historical data and weekly information on the betting odds provided by several bookmakers. We use the top four European football leagues -Premier League, Bundesliga, Liga and Serie A- and we carry out a variety of graphical posterior predictive checks for checking the fit of the model on in-sample data. Some further simulations show that such a model is useful for developing profitable betting strategies on hold-out data.

References

Dixon, M. J., & 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.

Karlis, D., & 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.

The Carry-Over effect in Round Robin Schedules.

Frits Spijksma, Professor, University of Leuven

Soccer has become a major business involving many stakeholders, such as teams, police, fans, sponsors, and broadcasting companies. Huge amounts of money are being paid for the broadcasting rights, illustrating the economic value of soccer competitions. This fact also emphasizes the relevance of finding a good and fair schedule. We discuss properties that relate to the fairness of a round-robin schedule; among these properties are (i) balancedness of patterns, and (ii) the so-called carry-over value. We describe quantitative measures for these fairness properties, and we apply these to existing round robin schedules.

A Hierarchical Local Search Approach for the NFL Scheduling Problem.

Panagiotis P. Repoussis, Athens University Economics and Business

This work introduces a hierarchical local search framework for solving highly constraint league scheduling problems. Focus is on the National Football League (NFL) that is currently the highest revenue generating professional sports league in the world. The NFL has numerous complex rules in scheduling regular season games and allocating the games to broadcasting networks. Each broadcaster has predefined network windows and preferences depending on the conferences and days of weeks. In addition to the network obligations, the generation of schedules should also follow a number of specific criteria to maintain fairness and attractiveness, such as team travelling, home/away game spacing, division series spacing and team byes, as well as criteria and rules external to NFL, such as stadium availability. A hierarchical Tabu Search approach is proposed to generate high quality heuristic solutions. It employs cyclical-exchange neighbourhood structures that compete to generate neighbouring solutions by sequentially swapping horizontally or

vertically two or more games, between two (or more) teams or two (or more) rounds, respectively. Computational experiments demonstrate the efficiency, scalability and robustness of the proposed heuristic approach.

Statistical modeling of availability to train in elite sports.

Ioannis Kosmidis, University College London

Recent technological advances have allowed the frequent and simultaneous measurement at small cost of multiple physiological and psychological aspects of individuals and of their fitness and everyday activity. Such data are increasingly being used, for example by elite sport teams to quantify and monitor athlete fatigue and identify the variables that are important for explaining it, with the aim of preventing non-functional overreaching. This talk will briefly discuss the challenges involved in recording and preparing fatigue-related data for statistical analysis, and present results from ongoing research on modeling availability-to-train in terms of other fatigue-related variables. Special care is being taken to use models that can account for both the heterogeneity between athletes and the natural time dependence between measurements.

Modeling football outcomes: a review.

Dimitris Karlis, Professor, Athens University Economics and Business

Modelling football outcomes have received considerable attention the last years both for betting purposes but also for better understanding the dynamics of the game. In this talk some of the most common models will be presented, focusing on most recent developments, discussing different approaches and also demonstrating the advantages and disadvantages of the models.

Wednesday 8 November 2017

Elite Players' Perceptions of Football (soccer) Playing Surfaces, their Impact on the Game and their Relationship with Injuries.

Alun Owen, Professor, University of Coventry

We consider the potential explanatory factors that may be associated with different attitudes amongst the global population of elite football (soccer) players to the use of different surfaces for football. A questionnaire was used to capture elite football players' perceptions of playing surfaces from over 1,000 players from 44 different countries. A mixed effects ordinal logistic regression model was used to explore potential explanatory factors of players' perceptions of the manner in which surfaces such as grass (Natural turf), artificial surfaces (Artificial Turf) and gravel like surfaces (hard surfaces often resulting from backed mud for example) impact on football. The work presented is based on the work in Owen et al. (2016). This same study is currently being used to assess the perceptions players have of the link between different surfaces and injuries in football and reports on the disconnect between these perceptions and the football injuries research literature.

Optimization in sports league scheduling: Experiences from the Belgian ProLeague soccer.

Dries Goosens, Professor, University of Ghent

Every sports competition needs a schedule of play, stating who will play whom, when, and where. A good schedule is important given its impact on the competition's fairness and outcome, public attendance, commercial interests, and cost of policing. In this presentation, we discuss our experiences with scheduling the Belgian Pro League soccer competition, for which we develop the official schedule since 2006. We present methods that have proven their value in real-life sports scheduling, and discuss how we continuously improved our scheduling approach, in order to accommodate changes of the competition format and requirements of the stakeholders. Equally important, we show how we improved the transparency of the scheduling process. We discuss fairness issues, as well as a discrete choice experiment we carried out to estimate the schedule's impact on stadium attendance and TV viewership.

Consumer-based Brand Equity of Celebrity Footballers: Antecedents & Consequences.

Eugenia Tzoumaka, PhD, The American College of Greece

Celebrities have been recently included in the list of brands, "one of several operationalizations of the broader concept of brand" (Thomson, 2006, p. 104). In the Forbes Celebrity 100 List, five out of 30 celebrity athletes are footballers with earnings varying from \$32 to \$93 million (Forbes, 2016). Given that the active players around the globe count to 250 million (Weil, 2016), apparently only 0.000002% of the King's ministers attain the status of global celebrity brands. Academic literature and industry proprietary indices have long limited sports celebrities' brand valuation to their capacity as endorsers; while a measurement approach on their consumer-based brand equity is missing. The present study addresses this gap by proposing and testing the first model of celebrity footballer consumer-based brand equity (CBEE). A quasi-experimental online survey approach, namely a one-group post-test protocol, was utilized for data collection from football fans in Greece. 1,459 participants were randomly assigned to different versions of a questionnaire that included eight international celebrity footballers. Using the Structural Equation Modeling framework for data analysis, the relationship between the antecedents, and consequences of the celebrity footballer CBEE were computed. The results of the latent variable structural model demonstrate the large effect of social identification that mediates brand associations' effects on the CBEE consequences.

References

Forbes. (2016). The World's Highest-Paid Celebrities. In D. Pomerantz (Ed.), *Forbes*.

Thomson, M. (2006). Human Brands: Investigating Antecedents to Consumers' Strong Attachments to Celebrities. *Journal of Marketing*, 70(3), 104-119. doi: 10.2307/30162104

Weil, E. (2016). Football *Encyclopædia Britannica*. Retrieved from <https://www.britannica.com/sports/football-soccer>

How customer analytics can boost sales in sports industry? A basketball team case study.

Sotirios Drinias, Accenture

Sports teams aim to increase revenues which are mainly coming from sponsors, TV channels, fans and retail. In order to optimize fans engagement to campaigns, sports teams want to build

a proper model to manage customers (fans) relationship, given the fact that fans are traceable through their data (posts, transactions, loyalty cards etc.). Sotiris describes the strategy and the process while presenting a case study from a basketball team which significantly increased sales.

Statathlon: Setting up your successful Sports Analytics startup.

Dimitrios Sakellaris (Statathlon)

Statathlon is a website focused on the scientific validation of all sports. Applying statistical analysis, data visualization and deep research we quantify and illustrate every sector of sports industry. Statathlon is the first Greek startup initiative in its field and it is consisted of three main categories: Data Analysis: Using statistical methodology (Descriptive Statistics, Bayesian statistics, Hypothesis testing, etc.) and data visualizations we produce innovative sports analysis. Bio-athletics: applying researches to prove in which way can the brain improve the athlete's skills. Business: applying meticulous researches on financial, marketing, technological, and other industry verticals researches to analyze sports from a business spectrum. There are four ways for this project to be viable as a company: Sponsored content (clubs, leagues, universities, etc.), Premium Content, Monetization (e-shop) and Adds. On the contrary, except the classic fixed costs, there is the data acquisition cost. There are plenty open datasets in the web, yet the most qualitative are not free. Until this day, the team processes the data that are free. Statathlon has also a unique type of operating process. Every week the team meets and discusses the upcoming topics. After a brief presentation by the researcher, the other team members are proposing a different scientific approach about the same case study. By following this step, everyone in the team learns new techniques at the same time and the team quality grows efficiently. Next step of the team is to reach some agreements on sponsored content, attract more specialists to join the team and seek funding for its activities. At the same time, we maintain this fast-growing learning process and continue to provide unique scientific outcomes on the website. Statathlon is unique because of its idea, its operation procedures and its final goal: to emphasize the importance of Data Analysis in contemporary sports industry.

Shooting efficacy: is the most important factor in top level men's handball?

Panagiotis Meletakos, Ph.D., Member of AUEB Sports Analytics Group

Handball is a very popular team sport in Europe. However it has not been as excessively analyzed as other popular sports. Such analysis should include the effect of different game aspects on the teams' success, especially at the highest level. We analyzed seven world championships from 2005 to 2017 with the participation of 24 teams in each championship. The dependent variable was the ranking of the teams and the independent variables were the shooting efficacies, the goalkeeper efficacies and the technical parameters. Significance level was set at 0.05. The results have shown that practically all efficacy and technical parameters have a significant correlation with ranking. Stepwise linear regression analysis showed that six of these parameters can significantly predict the ranking of a team in the world championships. Logistic regression analysis showed that three of them (9m shooting efficacy, wing shooting efficacy and blocks) produce a high predictive value whether a team can achieve one of the first eight places. Among these top eight teams, 2-minutes suspensions together with 9m shooting efficacy play an important role for the team's final ranking. Evidently more rigorous analysis is warranted. The present findings could provide valuable insight for coaches concerning their team's tactical choices and training methods.

Data benchmarking as a tool for effective coaching in Volleyball.

Sotiris Drikos, Ph.D., Member of AUEB Sports Analytics Group

A benchmark is a reference point for comparisons established from the value of a performance indicator. Data benchmarking concerns quantitative comparisons of performance outcomes across different sections and/or different time periods (Bartlett, Gratton, & Rolf, 2006). In coaching the use of data benchmarking can identify top performers and what makes them succeed, set the right performance standards, select proper players and manage athletes and teams more effectively. In Indoor Volleyball a team's performance in skills is directly connected to the game result (Marcelino, Mesquita, & Sampaio, 2009) and with the sports success in a championship (Drikos, Kountouris, Laios, & Laios, 2009). In order to create benchmarks for volleyball, we used performance data from each team of Greek Volley League from 2005-06 until 2015-16 (N=

131). We considered the effectiveness of fourteen (14) parameters from 5 skills: Serve, Attack after passing, attack after defense, Block and Pass. The conclusions showed that high-level teams of the Greek championship, which finished in positions 1-4, had as benchmarks: for every five serve errors there could be two aces, for every four attack after passing kills there could be one error or one blocked, for every three attack after defense there could be one error or one attack blocked, and for every six passes on the target there could be one error or overpass.

References

Bartlett, R., Gratton, C., & Rolf, C. G. (2006). *Encyclopedia of international sports studies* (Vols. 1 A-E). New York: Routledge.

Drikos, S., Kountouris, P., Laios, A., & Laios, I. (2009). Correlates of team performance in Volleyball. *International Journal of Performance Analysis of Sports*, 9(2), pp. 149-156.

Marcelino, R., Mesquita, I., & Sampaio, J. (2009). Home advantage and set outcome in high-level volleyball. *Journal of Sports Sciences*(26), pp. S66-S67.

On the stochastic nature of indices of competitive balance.

Ioannis Ntzoufras, Professor, Athens University Economics and Business

Competitive balance is becoming an important issue in the sport research agenda since it is related to several aspects of the sports, including problems related to financing and economic development and success of the leagues. While the literature contains a lot of attempts to define, estimate and work with certain indices, little is known on their stochastic behavior. Calculations of such indices include random errors due the random nature of sports and thus it is important to examine the behavior of certain indices of competitive balance with respect randomness. We adopt a model based Monte Carlo approach to examine the impact of randomness for such indices and thus to correct for the part that observed values can be attributed to randomness or they reflect something more systematic. Examples with European Leagues are provided. A discussion related to the interpretation of such indices in practice is also given.

SAW 2017 - Organized by



I. Ntzoufras

D. Karlis

S. Drikos

Administrative Assistants

K. Matzorakis

S. Mutsana