

WORKSHOP ABSTRACTS

AUEB SAW 2022

6th Aueb Sports Analytics Workshop





Sponsors





• 26-27 MAY 2022 • HYBRID CONFERENCE

Featuring A Short Course on Basketball Data Science

Day 1 - Thursday 26 May 2022 (Virtual)

Remembering Stefan Kesenne (1950–2021)

Thomas Peeters, *Erasmus School of Economics, Belgium*

Stefan Kesenne has passed away on June 21, 2021. Dr Kesenne was an emeritus professor at the University of Antwerp and the University of Leuven. He will be remembered as an inspiring, honest, and courageous man with strong ethical convictions. His groundbreaking research and unwavering support for the academic community will leave a permanent mark on the sports economics research landscape

The Impact of Absent Coworkers on Productivity in Teams

Thomas Peeters, *Erasmus School of Economics, Belgium*

We study the impact of the temporary absence and replacement of a coworker on the productivity of the remaining workers in a production team. Our analysis distinguishes between the absence of coworkers who perform the same task in the team (substitutes) and those who perform a complementary task. The absence of both types of coworkers causes remaining workers to produce less output per minute worked, but for substitute workers they (partially) compensate by increasing their working time. As a result, the absence of complementary workers has a more severe impact on the total output of remaining workers. The productivity loss for both types of absences is strongest for the least able remaining workers, who also compensate more in terms of working time. All in all, absenteeism decreases total team production through a combination of the productivity loss of remaining workers and the lower ability of replacement workers compared to the absent worker.

Assessing competitive balance in the English First Division/PremierLeague for over forty seasons using a stochastic block model

Georgios Nalbantis, *Faculty of Economics and Social Sciences, Institute of Sports Science, University of Tübingen*

By linking the literature strands on star employees and specialization, we analyze the impact of specialist and generalist stars on organizational performance. Since the degree of specialization and diversification can change over time, we propose a dynamic framework for our analysis, allowing specialist and generalist stars to switch their roles. In a first step, we develop a game-theoretical model from which we derive our main hypotheses. In a second step, we empirically test these hypotheses using data from professional basketball. In line with the theoretical predictions of our model, we find that teams with generalist stars outperform teams with specialist stars. Moreover, while a generalist star switching role to a specialist always worsens team performance, a specialist star in basketball switching role to a generalist may improve team performance.

Joint work with Christian Manger, Tim Pawlowski and Philip Yang

Revisiting financial fair play and competitive balance.

Elisavet Manoli, *University of Loughborough, UK*

Against the backdrop of the financially unstable and volatile environment threatening the long-term viability of European club football, UEFA introduced the financial fair play regulations in 2011. FFP was met with severe criticism regarding its potential adverse effects on competitive balance across European football, while simultaneously being welcomed for the exact opposite reason, the potential increase of competitive balance through reducing the importance of financial power. The lack of clarity and cohesion on this issue has since drawn academic interest on the matter, in order to better clarify what the effects of FFP on competitive balance have been. The studies examining this issue have predominantly focused on the five big football leagues in Europe, offering little clarity due to their mixed results. At the same time, to this day, the potential effects of FFP on competitive balance in smaller European football leagues remain a rather uncharted territory.

In this study, the attention is drawn on the Greek Super League (the men's football top division of the country) in order for its competitive balance before and after the introduction of FFP to be examined. The Herfindahl Index of Competitive Balance as the primary method is used, and is supported by standard deviation of points analysis, which together provide an indication of the level of competitive balance for each of the 15 seasons examined. This examination allows for the trends in competitive balance to be identified, with emphasis drawn on the seasons after the introduction of the regulations.

The results provide an indication that FFP regulations might have had an adverse effect on competitive balance in the Greek Super League, since a notable decline in competitive balance is noted. This decline is not unprecedented in the league under examination, but it nevertheless points towards a potentially adverse effect of the implementation of FFP. While FFP were not introduced with the primary aim of improving competitive balance within European football, the decline of competitive balance noted can be considered an indication that they might not have worked as a measure to improve European club football.

With practical examples highlighting that FFP can be overcome and research indicating that adverse effects might have been brought upon, the recent news suggesting that FFP are to be scrapped might indeed be of no surprise.

Assessing competitive balance in the English First Division/PremierLeague for over forty seasons using a stochastic block model

Nial Friel, *School of Mathematics and Statistics, University College Dublin, Ireland*

Competitive balance is a desirable feature in any professional sports league and encapsulates the notion that there is unpredictability in the outcome of games as opposed to an imbalanced league in which the outcome of some games are more predictable than others. In this paper, we develop a model-based clustering approach to provide an assessment of the balance between teams in a league. We propose a novel Bayesian model to represent the results of a football season as a dense network with

nodes identified by teams and categorical edges representing the outcome of each game. The resulting stochastic block model facilitates the probabilistic clustering of teams to assess whether there are competitive imbalances in a league. A key question then is to assess the uncertainty around the number of clusters or blocks and consequently estimation of the partition or allocation of teams to blocks. We apply our model to each season in the English first division/premier league from 1978/79 to 2019/20. A key finding of this analysis's evidence which suggests a structural change from a reasonably balanced league to a two-tier league which occurred around the early 2000's.

The hot hand in the wild

Konstantinos Pelechrinis, *University of Pittsburgh, USA*

Streaks of success have always fascinated people and a lot of research has been conducted to identify whether the “hot hand” effect is real. While sports have provided an appropriate platform for studying this phenomenon, the majority of existing literature examines scenarios in a vacuum with results that might or might not be applicable in the wild. In this study, we build on the existing literature and develop an appropriate framework to quantify the extent to which success can come in streaks—beyond the stroke of chance—in a natural environment. Considering in-game basketball game situations, our analysis provides statistical evidence that individual players do indeed exhibit the hot hand in varying degrees, that is, individual players can consistently get in a streak of successful shots beyond random chance. However, as a whole, the average player exhibits shooting regression, that is, after consecutive makes he tends to perform below expectations. Even though our results are based on a sports setting, we believe that our study provides a path towards thinking of the hot hand beyond a laboratory-like, controlled environment. This is crucial if we want to use similar results to enhance our decision making and better understand short- and long-term outcomes of repeated decisions.

Orthogonal schedules in Round Robin tournaments

Frits Spieksma, *Eindhoven University of Technology, Netherlands*

A basic and popular format for a competition is the well-known Single Round Robin (SRR) format. Given a set of n teams (n even), an SRR format prescribes that each team plays against each other team once in a set of $n-1$ rounds, such that each team plays once in each round. Moreover, in each match there is one team that plays home (H), and one team that plays away (A). In this presentation we focus on a question that is relevant for a set of strategies that are known as first-break-then-schedule (fbts). These are strategies that follow a 2-step approach: first decide upon the home/away designation of each team in each round (thereby specifying a home-away pattern (HAP) for each team), then allocate all the matches in a way that is compatible with these HAPs. A key question is to what extent specifying the HAP-set in Step 1 impacts the set of possible schedules in Step 2. Or in other words, what is the diversity of schedules compatible with a given HAP-set?

One measure that has been proposed to assess the flexibility of a HAP-set is called the width. Informally, the width of a HAP-set equals the number of schedules such each pair of these schedules has no match in the same round. Clearly, the larger the width, the more flexible the HAP-set. We give upper and lower bounds for the width. And we prove that the upper bound on the width can be achieved for a particular HAP-set when the number of teams is a power of 2.

This is joint work with Roel Lambers

The use of technology in football clubs' operations

Michael Anagnostou, *University Campus of Football Business, UK*

The role and importance of technology in football increases exponentially. New on and off-the pitch applications are activated. They range from Goal Line Technology, VAR, talent identification and tracking performance systems to improved stadia experience, accessibility, safety and sports retailing and merchandising. The football clubs organize special labs and hubs for data collection. Specialist sports technology companies test new models of artificial intelligence for decision making. We have entered a new era in the use of technology for professional football, but at the same time new dilemmas arise. Who really decides? Humans or machines? The modern sports managers and executives must be aware of the new technologies. They also must be able to distinguish the potential and limits of technology applications in their field of action.

Generating data cultures in sports organizations on and off the field of play

Hisham Shehabi, *N3XT Sports, USA*.

Data is the new oil, and organizations across all industries that invest in their data capabilities and infrastructure are developing a competitive advantage. In the sports industry, whilst data has always been important it is gaining significant traction on and off the field of play to generate new operating, business & commercial models. To properly leverage data, technology and processes are key. However, without the right data culture from the top to the bottom of the organization, it cannot be a sustained competitive advantage. This talk will bring insights from across the Olympic Movement, football and other professional sports as to how data culture is being generated (and hindered) within organizations.

Training in lockdown. The impact of stringency measures on the Olympic games of Tokyo 2020

Vincenzo Alfano, *University of Messina, Italy*

The XXXII edition of the summer Olympic games was postponed due to the COVID emergency. Did the pandemic also affect the results? We aim to understand the impact of stringency measures over the previous year on the performances of athletes in the

Olympic games. It is the appointment of life for many world-level athletes. They often train and structure their agenda around this date, to arrive at peak physical condition at the crucial moment. Stringency measure may have affected the results of this competition, by influencing the daily routine of top-tier athletes. By making harder to effectively train, to access sportive infrastructure, to meet teammates, and more in general to live the life of an athlete, countries that imposed tougher NPIs to fight COVID-19 may have hindered the chances of their athletes to win an Olympic medal. Our quantitative analysis shows an effect of stringency measures on Olympic medals. It is especially relevant in team sports. This is consistent with the idea that stricter NPIs made harder for teams to train together, and to reach the necessary chemistry and harmony to succeed in such a competitive event. Also, women were more penalized by higher stringency than men. This is true both looking at the total medals won, and by restricting the analysis at team events.

Day 2 - Friday 27 May 2022 (Hybrid)

Leveraging the power of Data to create Sustainable Sports

Giannos Konstantopoulos, *The Sports Footprint, Greece*

Sustainability has become a buzzword in the sports industry as sport organizations are realizing the benefits of more sustainable operational models and they are following their fans' demand for Climate Action. However, the "greening" process is long and requires transformation in all key operations of a sport organization. Following the principle "What gets measured gets managed", sports environmental data are necessary for managers leading this transformation.

In this presentation, we will discuss how to acquire and use data in the areas of:

- Sustainable transportation
- Waste and water management
- Energy

Also, we will analyse how to use those data for marketing purposes and sponsorship deals. Lastly, we will attempt to prove, with the use of data, that sustainable sport organizations have a competitive advantage against their competition.

The Higher-Order PLS-SEM Confirmatory Approach for Composite Indicators applied on Football

Mattia Cefis, *University of Brescia, Italy*

Supporting the strategic decisions of a football team's management is becoming crucial. We create some new composite indicators to measure the performance, applying Confirmatory Composite Analysis (CCA) to a Third-Order Partial Least Squares

Structural Equation Model (PLS-SEM). To do this, data provided by Electronic Arts(EA) Sports experts and available on the Kaggle data science platform has been used; in particular, the dataset was composed of 29 Key Performance Indices (KPIs) defined by EA Sports experts, concerning the top 5 European leagues. After a preliminary players' Observed Heterogeneity Analysis (season, role and league effects), a PLS-SEM for each player's role was developed, relying on the most recent season, 2021/2022. In order to confirm each model, a CCA was applied. The results underline how some sub-areas of performance have different significance weights depending on the player's role; as concurrent and predictive analysis, our third-order player indicator (PI) overall was compared with the existing EA overall (as benchmark) and with some performance quality proxies, such as the player's market value and wage, showing interesting and consistent relations.

Trends in competitive balance and the long-term success of European football

Vasilios Manasis, *University of Athens & AUEB, Greece*

Competitive balance refers to the homogeneity of the teams in terms of their athletic ability. The importance of the notion is unquestionable through time since it is closely associated with the uncertainty of outcome in games and championship prizes. Despite its thriving success the last few decades, the long-term success of European Football is threatened by the deteriorating levels of competitive balance both in domestic and European championships. Since none of the actions taken so far seem to be efficient to revert this trend, more generous measures, mainly borrowed from US professional sports, should be applied in favor of competitive balance.

Enhancing the fight against doping: using copulas for multivariate testing analysis on Athlete Biological Passport profiles

Brunero Liseo, *Sapienza Università di Roma, Italy*

Doping control, or testing, is one of the essential components adopted by anti-doping organizations to protect clean sport competitions. Alongside the evaluation of athletes' samples for prohibited substances or methods, the Athlete Biological Passport (ABP) has been established, for some specific disciplines, as a complementary pillar in the detection of doping, since its introduction in 2009. The fundamental principle of the ABP is to monitor over time athletes' individual profiles – with respect to certain doping biomarkers such as testosterone – that may reveal anti-doping rule violations. Significant variations from an athlete's established levels can be further assessed for possible manipulation.

Currently, the practical implementation of the ABP framework is based on a Bayesian approach (Sottas et al., 2007) called ADAPTIVE. Specifically, given a biomarker, the predictive posterior distribution of a future sample is used to determine individual limits discriminating between normal and anomalous values. These individual limits are then

continuously updated as additional samples are taken, and the observed values are compared against them to identify potential anomalies.

However, the ADAPTIVE approach is implemented on longitudinal profiles of single markers (e.g., testosterone), of single combinations of two markers (e.g., testosterone over epitestosterone; the so-called T/E ratio), or, alternatively, of a few biomarkers following a univariate approach. Notably, a recent concept study (Alladio et al., 2016) evaluated the simultaneous combination of 13 steroid markers using principal component analysis, suggesting that a multivariate strategy may assist and enhance doping detection.

Inspired by the use of copulas for modeling multivariate datasets, in this work we extend the established ADAPTIVE method to multivariate testing of longitudinal ABP profiles. We express the multivariate joint distribution of a set of biomarkers through a separate modeling of the marginal distributions and of their dependence structure. Focusing on a parametric setting, we evaluate the performance of such a multivariate approach in a number of simulation studies, varying according to the copula family and the number of biomarkers.

Flexible marked spatio-temporal point processes with applications to event sequences from association football

(Petros Dellaportas, *AUEB & UCD, Greece & Ireland*)

We develop a new family of marked point processes by focusing the characteristic properties of marked Hawkes processes exclusively to the space of marks, providing the freedom to specify a different model for the occurrence times. This is possible through a decomposition of the joint distribution of marks and times that allows to separately specify the conditional distribution of marks given the filtration of the process and the current time. We develop a Bayesian framework for the inference and prediction from this family of marked point processes that can naturally accommodate process and point-specific covariate information to drive cross-excitations, offering wide flexibility and applicability in the modelling of real-world processes. The framework is used here for the modelling of in-game event sequences from association football, resulting not only in inferences about previously unquantified characteristics of the game dynamics and extraction of event-specific team abilities, but also in predictions for the occurrence of events of interest, such as goals, corners or fouls, in a specified interval of time.

Effects of gender, age and match status on the tactical development of shooting opportunities during the U17, U20 and Senior FIFA World Cup. A multilevel analysis.

Vasilis Armatas, *School of Physical Education, Kapodistrian University of Athens Athens, Greece & Olympiakos FC Academy*

The aim of the present study was to explore the combined effects of gender, age and match status on the tactical development of shooting opportunities (SO) during the U17, U20 and Senior Soccer FIFA World Cup tournaments. The sample included 1992 team

possessions that led to shooting opportunities in 96 matches (48=men; 48=women) from the U17, U20 and Senior World Cup tournaments. Multilevel logistic regression models revealed that men's teams had lower odds of progressing by fast attacks (OR=0.519; 95% CI: 0.382-706; $p<0.001$) and higher odds of finishing by means of volleys (OR=2.842; 95% CI: 1.929-4.188, $p<0.001$) and headers (OR=1.911; 95% CI: 1.264-2.890; $p<0.01$) than women's teams, in comparison to combinative attacks and finishing on the ground, respectively. Men's teams also registered higher odds of completing more passes and implementing a higher passing tempo than women's teams. As far as the effect of the age, senior and U20 teams presented higher odds of progressing by counterattack than U17 teams. When drawing or losing, teams had higher odds of counterattacking, in comparison to progressing by combinative attack. This study revealed interactive tactical effects of gender, age and match status on the development of SO during the diverse World Cup tournaments.

A Performance Evaluation of the EURO 2020 Soccer Players based on Fantasy Sports Points

Vasilios Palaskas, *Fantasy Sports Interactive Ltd (Ltd) & AUEB Analytics Research Team, Athens University of Economics and Business, Athens*

In Fantasy Sports, users select athletes from real-life sports events to create their own Fantasy Teams, which compete between them based on the actual performance of their athletes on the field. The performance of the athletes on the pitch is represented by an index (Performance Points) based on several actions performed during the match.

In this analysis we follow a research path which aims at the performance evaluation of the soccer players who participated in EURO 2020, using the performance points they received during the matches they played in, based on the FSI evaluation scoring system as means to approach their overall performance. Essentially, the performance points that athletes receive in their matches constitute a general index of their performance on the field, since this is a weighted index rating each skill (action) according to its importance and the position the athlete plays.

Finally, we suggest our EURO 2020 best XI according to the footballers' Fantasy Points Performance and compare our results with the ones suggested by UEFA.

Joint work with V. Mexias and I. Ntzoufras.

footBayes: an R package for fitting Bayesian and frequentist football models

Leonardo Egidì, *University of Trieste, Italy*

Modeling football outcomes became incredibly popular over the last years. However, an encompassing computational tool able to fit in one step many alternative football models is missing yet. With the footBayes package we want to fill the gap and to give the possibility to fit, interpret and graphically explore the following goal-based Bayesian football models using the underlying Stan (Stan Development Team (2020))

environment: Double Poisson (Baio and Blangiardo (2010), Groll, Schaubberger, and Tutz (2015), Egidi, Pauli, and Torelli (2018)), Bivariate Poisson (Karlis and Ntzoufras, 2003), Skellam for goals' differences: (Karlis and Ntzoufras, 2009), Student-t for goals' differences: (Gelman, 2014), Precisely, we'll learn how to: fit static maximum likelihood and Bayesian models; fit dynamic Bayesian models; change the prior distributions and perform some sensitivity tests; interpret parameters' estimates; retrieve predictive intervals for team-specific football abilities; check the models; through graphical posterior predictive checks; obtain out-of-sample predictions; reconstruct the final rank's league; compare models. The package is also available at the following link: <https://github.com/LeoEgidi/footBayes>.

Bayesian Skill Importance for Volleyball Set Determination

Ioannis Ntzoufras, *AUEB Sports Analytics Group, Department of Statistics, AUEB, Athens, Greece*

The aim of this work is to evaluate the contribution of each Volleyball skill in the final match outcome. Here, we focus on modeling the difference of winning sets for the Greek A1 men's League of the regular season 2016-2017. We use two alternative models: a) an ordinal multinomial logistic regression model and b) the zero deflated and truncated Skellam (ZDTS) model introduced in a previous work of ours (Ntzoufras et. al, 2021, IMA J. Manag. Maths). In these models we further consider standard volleyball skills as explanatory features. These skills capture the main characteristics of the game such as the serve, the block, etc. We use Bayesian variable selection algorithms in order to evaluate which skills contribute to the final score and estimate their corresponding posterior effect size.

Joint work with V.Palaskas, L. Egidi and S. Drikos

Post-tournament analysis. Methods to investigate teams performance. The Volleyball example.

Sotitios Drikos. *University of Athens & AUEB Sports Analytics Group, AUEB, Athens, Greece.*

The results of a post-tournament analysis lag in time and the coaches and practitioners always look more forward than backward. Nevertheless, post-tournament analysis is an essential tool to develop the game's philosophy further and to generate solutions to specific problems. The use of accumulative data from match analyses especially in elite-level tournaments can help coaches and managers to specify Key Performance Indicators (KPI) and their correlation with the teams' success. This presentation consists of a bouquet of already published studies using data from the male EuroVolley tournament in three ways in order to define the KPIs of the tournament. Initially, to investigate which KPIs are differentiated between winners and losers of balanced sets. In continue, which KPIs are differentiated between winners and losers of typical sets taking into account the quality of the rival and finally which KPIs distinguished teams' performance levels.

Basketball & Analytics: How analytics shaped the Sport

Christos Marmarinos. *AUEB Sports Analytics Group, AUEB & Kosovo National Basketball Team*

In this presentation we will examine the connection of Data Analysis and the game of Basketball. The movement of analytics has had a drastic effect on the way the game is played, perhaps more than anything in the last years. However, this relationship of opposite disciples has not always been without problems. We will discuss the nature of these controversies. In addition, we will provide ideas that would help bridge the gap between coaching and analytics, for the benefit of the players and the Sport

Evaluating the performance of NBA players: A survival analysis study

Ambra Macis, *University of Brescia, Italy*

Data analysis for sport has developed consistently over the last years. The aim of this work is to analyze the performance of NBA players during a given period of time. The interest was focused on studying, for each player, the time taken to exceed a given point threshold during the post All-Star season segment, on the basis of a baseline set of covariates. More in detail, the achievement of the threshold (dichotomous event: achieved or not) and the minutes played for reaching it (time-to-event) were extracted from play-by-play data relative to the post All-Star segment of the 2020-2021 NBA season. Baseline covariates included the NBA statistics of each player related to the pre All-Star season segment and other information regarding the player, e.g. if he had been selected for playing the All-Star game. Analyses were performed through statistical methods for survival analysis, with particular attention to variable selection. The obtained results allowed us to identify the main players' features leading to a higher probability of reaching the point threshold.

Modelling the outcome of basketball game: some new results.

Dimitrios Karlis, *AUEB Sports Analytics Group, Department of Statistics, AUEB, Athens, Greece*

Basketball is a sport of large exposure in many countries and especially in USA. The blooming of sports analytics has increased a lot the demand for models about basketball covering a wide range of aspects of the game. In this talk we focus on problems related to the prediction of the score of the match and hence its outcome. This is of considerable interest for fans, trainers, teams as companies, betting purposes and more. Existing models are based mainly on a bivariate Gaussian model for the joint distribution of the scores. We make use of historical data from NBA, that lead to alternative, more advanced models, for this purpose. We use historical data to build more advanced models, while at the same time we compare the new models with the typical ones based on bivariate Gaussian distribution.

SAW 2022

Organized by



Ioannis Ntzoufras

Dimitris Karlis

Sotiris Drikos

Administrative Assistant

Argiro Damoulaki

Panayotis Zoris