

EQUILIBRIUM TEAM SELECTION IN FOOTBALL UNDER WIN OR PROFIT MAXIMISATION

KJETIL K. HAUGEN

Abstract. This article applies mathematical methods to investigating (extremely) simplified versions of the team selection problem for football managers. The team selection problem – obviously a game problem – is analysed game theoretically, but some funny results of a combinatorial nature are also added. In general, the findings indicate that the problem of selecting a football team is a very demanding task. Still, a simple comparison between win and profit maximising game models indicate surprising complexity in game prediction.

1. INTRODUCTION

Team selection in team sports¹ is a fundamental problem in sports coaching. It holds similar importance alongside different team sports, but has probably gained most attention within European football (soccer).

A modern football manager's main area of responsibility is related to managing talent (buying and selling players), and picking the team before each match. This last activity is the focus of this paper.

Obviously, picking a team for a given match is related to a great deal of local characteristics, such as players' form, their spirit, their injury situation, their performance in previous matches and so on. In addition, the team selected for a given match should be linked to the given opponent. Certain players may be in bad form, but may have personal abilities and qualities which make them 'more valuable' against certain teams. As such, there is both a competitive and a 'local' dimension involved in team selection. As indicated by the title, it is the competitive dimension of team selection which is the main focus of this article.

In subsequent sections, we will, after a brief literature survey, discuss combinatorial as well as competitive properties of team selection (Sections 3 and 4) respectively.

MSC (2010): primary 91; secondary 97K20, 91A10.

Keywords: combinatorics, game theory, football, team selection.

Associate professor Halvard Arntzen is gratefully acknowledged for reading through the manuscript and adding helpful comments.

¹Actually, it is also important in individual sports, but perhaps for slightly different reasons than in team sports. An athlete devoting years and years to prepare for an Olympic competition will naturally be interested in actually meeting up on the starting line. At the same time, the team manager is interested in sending his most competitive team (of individuals) to maximise the probability of getting gold medals.

2. A BRIEF LITERATURE SURVEY

Sport management literature dealing with team selection is quite comprehensive. Unfortunately, most of the focus is on youth football, and whether all youngsters should be guaranteed playing time, or related to talent development. See for instance [1–3, 11, 12] for some examples.

Our focus, which is on elite sport and equilibrium team selection, is quite sparsely treated. Some noteworthy exceptions do however exist. Both books [15] and [7] discuss the topic, but more in the form of questions than answers and analysis.

In addition to some analysis and answers, subsequent sections may also implicitly answer the question of why these topics have received meagre analytic treatment: Team selection equilibria are really hard to find.

3. COMBINATORICS

Some relevant sets and cardinalities need to be defined:

\mathcal{K} : Set of Keepers in squad.

\mathcal{D} : Set of Defenders in squad.

\mathcal{M} : Set of Midfielders in squad.

\mathcal{A} : Set of Attackers in squad.

m : Number of players in the the team at startup (=11).

N : = $|\mathcal{K} \cup \mathcal{D} \cup \mathcal{M} \cup \mathcal{A}|$, Number of players in squad.

To illustrate some points, we need some notion on a representative squad size. According to [16], some observations from various European teams (2016/2017 season) may help: Hull City (UK), Barcelona (Spain), Bayern (Germany) and Paris St. Germain (France) have squads of 34, 32, 30 and 29 respectively. Hence, a choice of $N = 30$ does not seem too far from reality².

Now, assume³ that we investigate the number of possible 11-men first teams we can organise given $N = 30$. This is easily calculated by the binomial coefficient

$$C_{FULL} = \binom{30}{11} = 54627300.$$

The reality is of course not so complex (54.6 million is a fairly big number), not all players can play in all positions. The keeper position is for instance a position which will not fit all players. The same holds more or less for other player categories. Football squads are hence partitionable, typically within the groups $\{\mathcal{K}, \mathcal{D}, \mathcal{M}, \mathcal{A}\}$.

Now, assuming 3 potential keepers in a squad, as well as an equal distribution of the remaining 27 players into the groups above, the total number of classical 4-4-2 system set-ups can be found as

$$C^{4-4-2} = \binom{3}{1} \cdot \binom{9}{4} \cdot \binom{9}{4} \cdot \binom{9}{2} = 1714608. \quad (3.1)$$

²This number is only used for illustrative purposes.

³Fairly unrealistic of course.

Even if 1.7 million is significantly less than 54.6, it is still a substantial amount of first teams to choose from.

It turns out to be interesting to calculate a few other traditional set-ups; the 4-3-3 and 4-5-1 systems

$$C^{4-3-3} = \binom{3}{1} \cdot \binom{9}{4} \cdot \binom{9}{3} \cdot \binom{9}{3} = 2667168, \quad (3.2)$$

$$C^{4-5-1} = \binom{3}{1} \cdot \binom{9}{4} \cdot \binom{9}{5} \cdot \binom{9}{1} = 428652. \quad (3.3)$$

Comparing the results of (3.1), (3.2) and (3.3), we observe significant differences. For instance, the 4-3-3 system returns 6.2 times ($\frac{2667168}{428652} \approx 6.2$) more potential first teams compared to the 4-5-1 system.

This brings us to Egil ‘‘Drillo’’ Olsen [18], the most successful Norwegian national team manager in history. Olsen was known for 2 things, apart from his immense success on the pitch, his probability calculations and his immense love of the 4-5-1 system. Our simple calculations above may shed more light on this history, as the 4-5-1 system has nice probabilistic consequences compared to traditional alternatives. If you pick a team at random given a 4-5-1 system, you would get a probability more than 6 times higher to pick the correct team than a 4-3-3 alternative. Maybe this was the reason for Olsen’s affection for the 4-5-1 system?

A slight generalisation is helpful. Let us assume we keep the equal distribution between the sets $\{\mathcal{D}, \mathcal{M}, \mathcal{A}\}$. Let us furthermore assume l potential keepers and n , ($n = \frac{N-l}{3}$) potential Defenders, Mid-fielders and Attackers. Then, equations (3.2) and (3.3) generalise to

$$C_{n,l}^{4-3-3} = \binom{l}{1} \cdot \binom{n}{4} \cdot \binom{n}{3} \cdot \binom{n}{3},$$

$$C_{n,l}^{4-5-1} = \binom{l}{1} \cdot \binom{n}{4} \cdot \binom{n}{5} \cdot \binom{n}{1}.$$

The ratio, here defined as f_n , becomes

$$f_n = \frac{\binom{l}{1} \cdot \binom{n}{4} \cdot \binom{n}{3} \cdot \binom{n}{3}}{\binom{l}{1} \cdot \binom{n}{4} \cdot \binom{n}{5} \cdot \binom{n}{1}} = \frac{10}{3} \cdot \frac{(n-2)(n-1)}{(n-4)(n-3)}.$$

Furthermore, f_n is decreasing as

$$f_{n+1} < f_n \Leftrightarrow \frac{n}{n-2} < \frac{n-2}{n-4} \Leftrightarrow n(n-4) < (n-2)^2 \Leftrightarrow 0 < 4.$$

A plot of f_n^4 is shown in Figure 1.

Many practitioners would argue that a 30-man squad will never contain 30 players that can play. Due to injuries, illness or other circumstances, perhaps no more than 70% of a squad is available at any point in time. Hence, the problem exemplified above may be even simpler - combinatorially.

If we make the simplifying assumption that keepers always are fresh and able, while the other groups are victims of illness or injuries, the squad is reduced to 21 players ($0.7 \cdot 30$). Then the 9’s in (3.1), (3.2) and (3.3) are changed to 6’s and

⁴ f_n is only defined for $n \geq 5$.

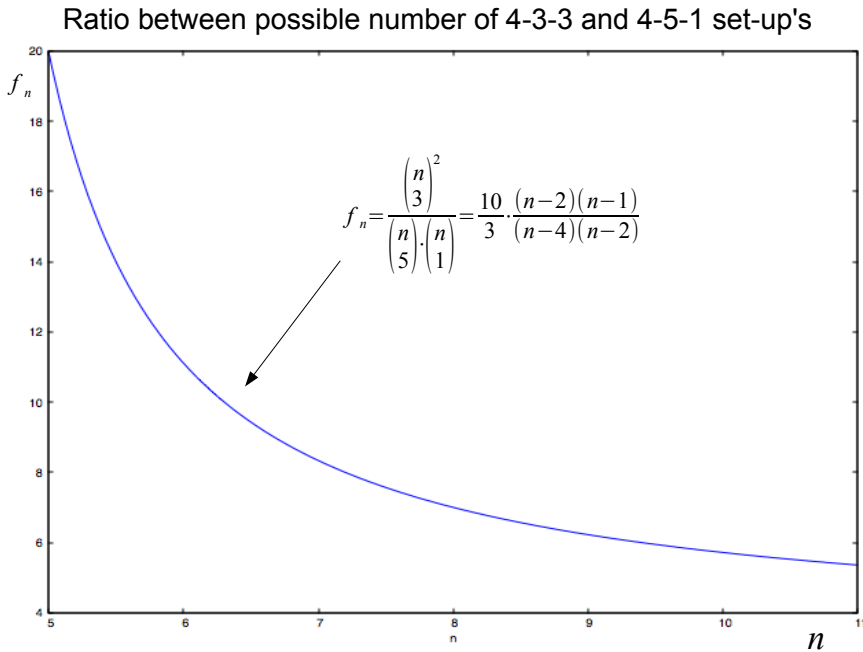


Figure 1. A plot of f_n .

the 4-5-1 system produces only 18.000 possible combinations. Still, Egil Olsen's argument holds even better than before, as f_6 increases to a more than 10 times higher success probability⁵ with the 4-5-1 than the 4-3-3 system.

4. GAME MODELS

In the introduction (Section 1), we discussed briefly that a team selection process should include judgements on what players the opposing team choose for an upcoming match. Hence, team selection in football fits a game framework. The nice thing about this gaming situation, is that it is reasonable to assume that both teams decide on their first team selections independently of each other, and that this decision (logically) is final⁶. That is, there are two players (teams or managers) deciding without observing each other's decisions, and these decisions are not open for change during the match. Hence, it seems reasonable to address the situation by a 2-player game of imperfect, but complete information.

Less nice are the results of Section 3, where the number of possible decisions is large. Players with thousands or millions of strategic choices are in general hard to

⁵See Figure 1.

⁶The question of choice of substitutes is left out for simplistic reasons, but it is straightforward to realise that inclusion of substitutes only will be a matter of changing the $m = 11$ to a bigger number. Hence, the methodology which is introduced in this section can still be applied.

handle in gaming situations. As a consequence, we will simplify the team selection game-definition by making the following assumptions:

- i) Two managers on behalf of two football teams, here named T_1 and T_2 , pick players for their teams before an upcoming match.
- ii) The qualities of the two teams are assumed different, and (without loss of generality) T_1 is assumed 'better' than T_2 . Our meaning of 'better' in this context is that, within a large set of potential matches between T_1 and T_2 , the share of wins of T_1 is larger than that of T_2 . This complies with the probabilistic definition of game outcome; say $p_{12} > p_{21}$, where p_{ij} is defined as the probability that T_i beats T_j . Implicitly, p_{ij} depends on team selection decisions for both teams. This definition follows the modelling framework in [7], [4], [5], [6] and [8].
- iii) The quality difference between T_1 and T_2 is explained by individual player quality differences. For T_1 , groups \mathcal{A}, \mathcal{M} contain better (and more expensive) players than equivalent groups \mathcal{A}, \mathcal{M} for T_2 . However, the defenders (\mathcal{D}) in T_2 are better than those in T_1 . We further assume that quality differences between groups of players are equivalent. That is, if T_1 , at a certain point in the team selection process, picks a member in \mathcal{A} or \mathcal{M} and T_2 picks a player in \mathcal{D} , both teams relative playing strength is unaffected.
- iv) The probability of a draw is assumed non-zero, and denoted by p_D . It is assumed that this draw-probability is unaffected by the team selection process⁷.
- v) We restrict the team selection problem to only include picking the last player on the first team. That is, we enter the team selection process at the very last stage for both managers, when 10 of the players (including both keepers) are already selected. Formally, this defines p_{ij} as the probability that T_i beats T_j if they only play with 10 players each. Finally, we assume that it is only player quality that affects win probability, not group membership. Hence, if the first 10 players chosen lack a forward, one should still pick a midfielder or defender if such have higher quality than the forwards available.
- vi) Both managers (teams) are win maximisers. That is, they want to achieve the highest probability possible of winning the match.
- vii) Both managers have access to the above information and choose to make their team selection without informing their opponent about their decisions until immediately before the match is played.

Given assumptions i) to vii), the resulting game, with best replies – red ellipses for T_1 and red rectangles for T_2 , can be formulated as in Figure 2.

Before we finish analysing the game, a few explanations of the pay-off values in Figure 2 may be helpful. If we examine the top left pay-offs (the strategy combination $\{\mathcal{A}, \mathcal{A}\}$), the pay-offs are $p_{12} + \epsilon$ and $p_{21} - \epsilon$ for T_1 and T_2 respectively. These values are found by utilising the above assumptions. As p_D is assumed

⁷This assumption is highly unrealistic, but is made in order to keep our arguments at a reasonably readable level. Relaxation of this assumption is possible, without interfering with our main results. The next assumption v) even makes this assumption iv) less of a problem.

		T_2		
		\mathcal{A}	\mathcal{M}	\mathcal{D}
T_1	\mathcal{A}	$p_{21}-\epsilon$ $p_{12}+\epsilon$	$p_{21}-\epsilon$ $p_{12}+\epsilon$	p_{21} p_{12}
	\mathcal{M}	$p_{21}-\epsilon$ $p_{12}+\epsilon$	$p_{21}-\epsilon$ $p_{12}+\epsilon$	p_{21} p_{12}
	\mathcal{D}	p_{21} p_{12}	p_{21} p_{12}	$p_{21}+\epsilon$ $p_{12}-\epsilon$

Figure 2. The normal form of the team selection game.

unchanged⁸, any increase in p_{12} must be followed by a similar decrease in p_{21} ⁹. Hence, as T_1 chooses a skilled player while T_2 chooses a less skilled player (see assumption iii), T_1 's playing strength increases, here by a convenient parameter ϵ , while T_2 must get a similar reduction. If we move to the top right corner, no change is observed in win probabilities, as both teams here choose their best player options¹⁰. All other pay-offs in Figure 2 are constructed by a similar logic.

Now, the final analysis of the game is straightforward. It contains 3 types of Nash equilibria, two in pure strategies; $\{\mathcal{A}, \mathcal{D}\}$ and $\{\mathcal{M}, \mathcal{D}\}$ as well as an infinite amount of mixed strategy equilibria where T_1 mixes with any probability between \mathcal{A} and \mathcal{M} , while T_2 chooses (the pure) \mathcal{D} strategy. Or, simpler formulated; both teams adds the best player remaining when the final player (the 11th) is selected.

This is definitely no remarkable result. On the contrary, this corresponds with any lay-man's logic and, hence, game theory brings no surprise here.

However, if we change the objective of the manager (team) slightly, it is straightforward to show that alternative equilibria exist, even as probable game predictions. Above, we assigned win maximising objectives to the two players T_1 and T_2 . Let us alternatively assume profit maximising behaviour for the managers (teams).

These two alternative preference choices have been a heavily debated subject in sports economic theory. Many authors argue that such differing preferences could explain the very different economic development which has been observed between

⁸See assumption iv).

⁹ $p_{12} + p_{21} + p_D = 1$.

¹⁰See assumption v).

European and US sport – see, e.g. [13,14,17]. The argument is related to an obvious point; if one is merely interested in winning matches, deficits may occur easier than if profits are the objective. Others [10] argue that such observed differences can be explained very well without introducing different preferences between European and US sports administrators. It is argued that it is questionable whether such differences should be present in the first place.

We will not engage further in this debate here, but instead analyse our game given such changed preferences. In order to move from win to profit maximisation, we must introduce some more model content. To make things as simple as possible, let us assume that generated revenue from playing a match is in some way increasing when the number of matches won increases. Again, to simplify, we assume that both teams generate the same income by winning a match¹¹. It is perhaps reasonable also to assume that increasing the winning percentage from 10 to 20 has more revenue generating potential than increasing it from 80 to 90¹². These assumptions lead to the introduction of a revenue function, say $R(\text{win probability})$ where $R' > 0$ and $R'' < 0$. Obviously, we also need to introduce some costs. To make it as easy as possible, and in accordance with previous definitions; say c_H, c_L where the high costs (c_H) are related to selecting the best players (i.e. players from groups \mathcal{A}, \mathcal{M} for T_1 and \mathcal{D} for T_2), while the low cost alternative involves choosing players from the remaining 'unskilled' player groups. One could of course ask what the meaning of these costs should be. Is a football player's income directly related to playing a match? Actually, today, this is the case in many football teams. Contracts defining the playing time as a part of both the buying price and the salary become more and more common. Hence, it does indeed make sense to assume that selecting good and expensive players for a given match actually may lead to higher costs than selecting less skilled players.

Now, given the above assumptions, it is straightforward to realise that the win maximising game in Figure 2 can easily be transformed to a profit maximising version by simply substituting the original pay-offs with $R(\text{original pay-off}) - \text{relevant cost}$. Where the relevant cost for T_1 is c_H if a player in the groups \mathcal{A} or \mathcal{M} is chosen and c_L if a player is chosen from \mathcal{D} , likewise and opposite for T_2 . This game is shown in Figure 3.

if the following two inequalities

$$R(p_{12}) - c_L > R(p_{12} + \epsilon) - c_H \Rightarrow c_H - c_L > R(p_{12} + \epsilon) - R(p_{12}) \quad (4.1)$$

and

$$R(p_{21}) - c_L > R(p_{21} + \epsilon) - c_H \Rightarrow c_H - c_L > R(p_{21} + \epsilon) - R(p_{21}), \quad (4.2)$$

are satisfied, the consequences for the best reply functions are as shown in Figure 4.

¹¹It is fairly obvious that a team-specific dimension is present in real life. Manchester United will certainly generate far more revenue from winning the Champions League final than Leicester City. But, we simplify and deliberately overlook this option here.

¹²Some may even argue that the R -function should have a zero-derivative at some point. After all, if a team becomes too good, the number of interesting opponents will decrease, and so will the number of relevant and interesting matches – See, e.g., [9]

		T_2		
		\mathcal{A}	\mathcal{M}	\mathcal{D}
T_1	\mathcal{A}	$R(p_{21}-\epsilon)-c_L$ $R(p_{12}+\epsilon)-c_H$	$R(p_{21}-\epsilon)-c_L$ $R(p_{12}+\epsilon)-c_H$	$R(p_{21})-c_H$ $R(p_{12})-c_H$
	\mathcal{M}	$R(p_{21}-\epsilon)-c_L$ $R(p_{12}+\epsilon)-c_H$	$R(p_{21}-\epsilon)-c_L$ $R(p_{12}+\epsilon)-c_H$	$R(p_{21})-c_H$ $R(p_{12})-c_H$
	\mathcal{D}	$R(p_{21})-c_L$ $R(p_{12})-c_L$	$R(p_{21})-c_L$ $R(p_{12})-c_L$	$R(p_{21}+\epsilon)-c_H$ $R(p_{12}-\epsilon)-c_L$

Figure 3. The normal form of the profit maximising team selection game.

		T_2		
		\mathcal{A}	\mathcal{M}	\mathcal{D}
T_1	\mathcal{A}	$R(p_{21}-\epsilon)-c_L$ $R(p_{12}+\epsilon)-c_H$	$R(p_{21}-\epsilon)-c_L$ $R(p_{12}+\epsilon)-c_H$	$R(p_{21})-c_H$ $R(p_{12})-c_H$
	\mathcal{M}	$R(p_{21}-\epsilon)-c_L$ $R(p_{12}+\epsilon)-c_H$	$R(p_{21}-\epsilon)-c_L$ $R(p_{12}+\epsilon)-c_H$	$R(p_{21})-c_H$ $R(p_{12})-c_H$
	\mathcal{D}	$R(p_{21})-c_L$ $R(p_{12})-c_L$	$R(p_{21})-c_L$ $R(p_{12})-c_L$	$R(p_{21}+\epsilon)-c_H$ $R(p_{12}-\epsilon)-c_L$

Figure 4. Best replies as a result of inequalities (4.1) and (4.2).

Actually, our assumptions of $R'' < 0$ and $p_{12} > p_{21}$ simplify as this assumption leads to

$$R(p_{21} + \epsilon) - R(p_{21}) > R(p_{12} + \epsilon) - R(p_{12})$$

Then, inequalities (4.1) and (4.2) can be represented by (4.2) alone. That is, if $c_H - c_L > R(p_{21} + \epsilon) - R(p_{21})$, the two pure Nash equilibria $\{\mathcal{D}, \mathcal{A}\}$ and $\{\mathcal{D}, \mathcal{M}\}$ with accompanying infinite mixed strategies (with any probability) for T_2 ¹³ are secured.

This can be interpreted as the existence of Nash equilibria of the “opposite type” compared to those identified in the win maximisation case. That is, we have identified the existence of Nash equilibria of the type ‘both teams select their worst players’.

Surely, the game in Figure 4 can have more Nash equilibria, but here, the only point was to demonstrate the existence of the type identified above.

The above results indicate that a simple change of player objectives has the power of turning a teams selection game up-side-down. The simple intuitive solution of the win maximising game, where both teams pick their best player(s) may suddenly change into a situation where the opposite equilibrium is realised. Again, of course, this is perhaps not counter intuitive, but the complexity difference (and potential added extra equilibria) are interesting.

5. CONCLUSIONS

The main, and perhaps most important result this paper demonstrates is the complexity of selecting players for a football team. The simple fact that most normal squad sizes induce very high possible line-ups indicate that the real team selection game is tough to solve.

Actually, the real team selection game involves 11 players to pick, not a single one as we have restricted our analyses to. Still, our extremely simplified examples show that even then, it is not obvious how these games turn out. Many practitioners would perhaps feel that picking a team cannot be so difficult. This article underlines the opposite, it is really hard, at least if one wants to use mathematics (and some economics) as decision support. Perhaps some of us may gain more respect for the jobs football managers do after reading this article.

The fact that the 4-5-1 system had such nice combinatorial characteristics, came as surprise to the authors. Maybe the “godfather” of Norwegian football, Egil “Drillo” Olsen, still has some tricks to pass on to the managers of tomorrow.

REFERENCES

- [1] J. Ashworth and B. Heyndels, *Selection bias and peer effects in team sports*, J. Sports Econ. **8** (2007), 355–377.
- [2] J. Baker, J. Cote and B. Abernethy, *Sport-specific practice and the development of expert decision-making in team ball sports*, J. Appl. Sport Psychol. **15** (2003), 12–25.
- [3] S. Gil, F. Ruiz, A. Irazusta, J. Gil and J. Irazusta, *Selection of young soccer players in terms of anthropometric and physiological factors*, J. Sports Med. Phys. Fitness **47** (2007), 25–32.

¹³Here, T_2 mixes between \mathcal{A} and \mathcal{M} , while T_1 chooses \mathcal{D} .

- [4] K. K. Haugen, *An improved award system for soccer: A (game-theoretic) comment*, *Chance* **20** (2007), 22–24.
- [5] K. K. Haugen, *Point score systems and competitive balance in professional soccer*, *J. Sports Econ.* **9** (2008), 191–210.
- [6] K. K. Haugen, *The “Norwegian soccer wonder” – a game theoretic approach*, *Scandinavian Sport Studies Forum* **1** (2010), 1–26.
- [7] K. K. Haugen, *Always Change a Winning Team: A Text on Soccer and Game Theory*, Tapir Academic Press, NO-7005 Trondheim, Norway, 2012.
- [8] K. K. Haugen, *Point score systems and football coaching secrecy*, *Math. Appl.* **5** (2016), 11–20.
- [9] K. K. Haugen, *Uncertainty of outcome and varying fan preferences – a game theoretic approach*, *Math. Appl.* **5** (2016), 1–10.
- [10] K. K. Haugen and H. A. Solberg, *The financial crisis in European football – a game theoretic approach*, *Eur. Sport Manag. Q.* **10** (2010), 553–567.
- [11] D. G. Hoare and C. R. Warr, *Talent identification and women’s soccer: An Australian experience*, *J. Sports Sci.* **18** (2000), 751–758.
- [12] J. Keogh, *The use of physical fitness scores and anthropometric data to predict selection in an elite under 18 Australian rules football team*, *J. Sci. Med. Sport* **2** (1999), 125–133.
- [13] S. Kesenne, *League management in professional team sports with win maximizing clubs*, *European Journal for Sports Management* **2** (1996), 14–22.
- [14] S. Kesenne, *The objective function of a team*, in: W. Andreff and S. Szymanski (eds.), *Handbook of the Economics of Sport*, Edward Elgar Publishing, Cheltenham, UK, 2006, 601–609.
- [15] I. Palacios-Huerta, *Beautiful Game Theory: How Soccer Can Help Economics*, Princeton University Press, Princeton, New Jersey 08540, USA, 2014.
- [16] Soccerbase, Teams, <http://www.soccerbase.com/teams/>, 2017, Accessed: 2017–06–02.
- [17] M. Terrien, N. Scelles, S. Morrow, L. Maltese and C. Durand, *The win/profit maximization debate: Strategic adoption as the answer?*, *Sport, Business and Management: An international Journal* **7** (2017), 121–140.
- [18] Wikipedia, Egil Olsen, https://en.wikipedia.org/wiki/Egil_Olsen, 2017, Accessed: 2017–06–06.

Kjetil K. Haugen, Faculty of Logistics, Molde University College, Specialised University in Logistics, Molde, Norway
e-mail: kjetil.haugen@himolde.no