Evaluating the efficiency of the association football transfer market using regression based player ratings

Olav Drivenes Sæbø* and Lars Magnus Hvattum**

* Department of Industrial Economics and Technology Management, NTNU, Norway

** Faculty of Logistics, Molde University College, Norway, hvattum@himolde.no

Abstract

In recent times, the use of quantitative methods to improve decisions within sports has increased. In association football, large amounts of match data has become available. This work first shows how simple match data describing the players on pitch and the time for events such as goals and red cards, can be used to derive an objective player rating. The rating is based on solving a large linear regression model. The resulting player ratings are in turn used as input to a regression model for analyzing transfer fees. It is shown that the performance of players, as reflected in the player ratings, is an important predictor of transfer fees. At the same time, several other important factors that determine the size of transfer fees are identified.

1 Introduction

Association football is packed with tradition, but in the last few years computer science and big data has played an increasingly important role: players are tracked using cameras or sensors, detailed video analysis is employed, game theory is used to model in-game events, and large data bases of results and match data have become available. This opens up possibilities for exploiting data to gain further insight into the mechanics of the game [13], as well as to analyze information efficiency in related markets [7].

While there are several published methods for rating and ranking teams in association football [2, 8, 10], the rating of players has received much less attention. In this work an objective rating of players is proposed, based on a regression model capturing the performance of players relative to their team mates and the opposition. Similar models have previously been presented for basketball [14] and ice hockey [11, 12].

The last two decades have seen the revenues of leading European association football clubs rising steadily, with broadcasting windfalls in particular soaring [4]. The continual

This paper was presented at the NIK-2015 conference; see http://www.nik.no/.

rise of popular and financial interest in association football puts higher pressure on professional clubs to invest wisely in their core competence: football players. To help assess investments, it helps to know the market value of the players and to identify inefficiencies in the pricing of players. Several past studies have used ordinary least squares regression to model transfer fees based on a range of independent variables [5], some of which describe the talent of the players, while others describe external factors related to the buying or selling club. To describe the talent of players, many different independent variables have been suggested, such as goals scored, games played, international caps, player age, and position [5].

In this work the objective regression based player ratings are used as a variable in a standard transfer fee regression model, together with a range of other explanatory variables, for analyzing transfer fees paid to release players from their current contracts. This allows the identification of some potential inefficiencies in the transfer market. The analysis of transfer fees also confirms that the objective player ratings are able to describe a significant part of the observed transfer fees.

Section 2 describes the calculation of objective regression based player ratings. Models for describing transfer fees are proposed in Section 3. The data used in the calculations are outlined in Section 4, together with some implementation details. Results and discussions are presented in Section 5, before concluding remarks follow in Section 6.

2 Regression based player ratings

Plus-minus player statistics measure the number of goals scored minus the number of goals conceded when the given player is used by a team, and is an idea that has been used in several sports. An adjusted plus-minus statistic was first proposed in basketball, taking into account the strength of team mates and the opposing side [14]. To describe the idea in the context of association football, consider a set of past matches and divide each match into a set of segments, defined as periods of time where the set of players is constant. For each segment i, define

$$\alpha_{ij} = \begin{cases} 1 & \text{player } j \text{ plays for the home team in the segment,} \\ 0 & \text{player } j \text{ does not play in the segment,} \\ -1 & \text{player } j \text{ plays for the away team in the segment,} \end{cases} \tag{1}$$

and let β_i be the goals scored minus goals conceded for the home team within the segment. Each player j is assigned a rating x_j, that describes the players relative contribution towards the goals differential, given by

$$\alpha x = \beta. \tag{2}$$

Equation (2) is unlikely to have any solutions whenever ratings are calculated based on a large set of historical match data. Therefore, ratings are instead found by minimizing the model errors, given as the sum of squared differences between the actual goals differences, β, and the model predictions $\hat{\beta} = \alpha x$. That is, adjusted plus-minus ratings are given by an x that minimizes $(\beta - \alpha x)^T(\beta - \alpha x)$, corresponding to an interpretation as a linear regression model estimated using ordinary least squares, and resulting in $x = (\alpha^T\alpha)^{-1}\alpha^T\beta$.

The adjusted plus-minus ratings for players with little playing time recorded are prone to large errors [11, 14]. When certain players join each other in the team for most or all of their minutes, the system struggles to differentiate their contributions from one another. This collinearity can lead to unfortunate, high errors for both players if the system puts too much weight on the few segments in which the players are separated. In an effort to reduce the errors, a technique called ridge regression, or Tikhonov regularisation, has been proposed for plus-minus ratings [12]. Rather than using ordinary least squares regression, minimizing the sum of squared errors, ridge regression adds a punitive factor, $\lambda x^T x$, to the target function, thereby discouraging values that differ strongly from 0. The ratings x that minimize the new target are now given by

$$x = (\alpha^T\alpha + \lambda I)^{-1}\alpha^T\beta, \tag{3}$$

where I is the identity matrix and λ is a parameter that signifies the strictness of the regularization. Setting $\lambda = 0$ reduces Equation (3) to an ordinary least squares problem, while increasing the parameter λ means some information is sacrificed in an attempt to tackle noise in the data.

Some refinements to the details above have been made to better fit in the context of association football. First, the duration in minutes, D_i, of different segments may vary significantly. To handle this, ratings are interpreted as the marginal contribution of a player to the goal difference of the whole team per 90 minutes, and the goal differences, β_i, are scaled accordingly. Second, a significant advantage for home teams is recognised in all studies of match outcomes. This effect must be accounted for in the rating system as well, otherwise players playing a disproportionate amount of minutes at home will be over-appreciated. In this work, the home advantage is recognised by instantiating a ”twelfth man” dummy player – a contributor to results that will be included in every home team's starting lineup. Third, a football match is affected by the showing of red cards. A similar solution to the one imitating home advantage is proposed: four dismissal dummy variables are instantiated. Whenever a team is shown their first red card, the player in question is replaced by the ”first dismissal” dummy player. A second dismissal leads to the substitution of the offending player for a ”second dismissal” dummy, and so forth. However, for each dismissal that is ”cancelled out”, i.e. a team loses one of its ”surplus” players, the relevant dismissal dummy is dismissed. For example, Team A receiving two red card sees them represented by nine players, the home advantage dummy, and dismissal dummies number 1 and 2. If Team B later receives its first red card, Team A lose their dismissal dummy number 1. That means Team A then consists of nine players, the home advantage dummy, and dismissal dummy number 2, while Team B consists of ten players.

In standard plus-minus ratings, no regard is shown for the chronology of performances. By producing a set of ratings using a time-independent observation set, no allowance is made for dynamic performance standards, as all past observations of performances are weighted identically. This limitation is similar to that detected in [3], and so a solution similar to theirs is proposed: all past observations are down-weighted exponentially, depending on the age of the observations, t, and a discounting parameter, k. This leads to the formation of a slightly different looking appearance matrix, α:

$$\alpha_{ij} = \begin{cases} e^{-kt} & \text{player } j \text{ plays for the home team in the segment,} \\ 0 & \text{player } j \text{ does not play in the segment,} \\ -e^{-kt} & \text{player } j \text{ plays for the away team in the segment,} \end{cases} \tag{4}$$

This discounting of older observations means greater emphasis is placed on recent performances. Further, this allows the computation of more dynamic ratings that change more quickly, staying in tune with recent trends. The definition of the goal differences, β_i, must also change accordingly. Considering that the home team scores H_i goals and concedes A_i goals in segment i, and when scaling the result up to 90 minutes, the new parameter β can be described as

$$\beta_i = \frac{90(H_i - A_i)e^{-kt}}{D_i}, \tag{5}$$

where the age of the observation, t, is computed as the fraction of a year passed before the date on which players are evaluated. Setting $k = 0$ leaves $e^{-kt} = 1$, and observations of all ages are once more weighted equally.

3 Transfer fee models

Several authors have proposed ordinary least squares regression models for describing observed transfer fees in association football, as summarized in [5]. As in other studies we use the logarithm of the transfer fee as the dependent variable. We propose two different models, built around the regularized adjusted plus-minus player ratings described in Section 2. That is, the rating of the player involved in a transfer is calculated for a point in time just prior to the transfer, and this rating is included as the primary player performance measure in the transfer fee models. In the following we describe the additional independent variables included in the models.

A player's versatility can to a large extent be attributed to his ability to play in different areas of the pitch or utilise both feet to kick the ball. Most players are right footed, so binary variables signalling whether a player prefers using his left foot or has no preference for either are included in both models. A maximum of one can be affirmative for any player – if both equal zero, the player is right footed.

Positional binary variables are aggregated, dividing positions on the pitch into five rough areas, fulfilling the roles as Goalkeeper, Defender, Midfielder, Winger, or Forward. A versatile player can possibly fulfill any number of these roles, and so there is no limit to the number of variables taking the value of 1 in this case. Indeed, every Winger is also classed as a Forward. A further measure is aggregated to capture versatility on the pitch, namely the characteristic of an Inverse Wide Player – a player that is comfortable using his left foot on the right hand side of the pitch or vice versa. A tendency to play inverse wingers in attack and encourage them to move inward has been observed, while traditionally wide players would prefer to stay wide and cross the ball from the touchline. Height as a parameter is included in interaction with the ability to play in defence or attack, rather than independently. This allows testing for significance of height in the positions directly responsible for trying to defend or attack the goals.

Nationalities are grouped into the following categories with associated binary variables: African, UK or Eire, EU national, non-EU European, Asian, other English speaking and South American. A maximum of one affirmative variable is allowed per entry. A transfer of a Scottish player will therefore not influence the estimation of the EU geographical coefficient. This allows us to study trends in the transfer market for English clubs more closely, as well as the market for British or Irish players on the continent. Using such a variety of classifications makes it possible to examine the implications of work permit regulations and to test for discrimination based on cultural background.

Two binary dummy variables are included as a proxy for experience and to some extent performance, detailing whether the player to this date has represented his country internationally on senior level, youth level, or neither. Again, a maximum of one variable is affirmative at a time. A lack of available detailed international caps' history precludes the use of a factor that accurately specifies transfer-date international experience. To compensate, this retrospective measure is used in combination with a backwards scaling of current total number of senior international caps. The factor ScaledCaps assumes a debut at the age of 18 and linearly scales the total to the age at the time of transfer. If the player ratings accurately reflects the quality of players the regression coefficient of these three variables should be 0, unless there is a bias in the transfer market with respect to capped players. An estimated current total of days spent injured is also included retrospectively and scaled linearly to provide a measure of how injury prone a player had been up until the time of transfer.

Player age at the time of transfer is included through both first- and second-order terms. Players are seen to rise in value as they improve through training and experience over time, until their performance and expected contribution deteriorates with age. Players' month of birth is also included as an independent variable, to test for a relative age effect. The year of the transfer is also included, and in case there is a steady inflation or deflation of the transfer market over the time period examined, it is hoped that this can be corrected to some extent by weighting the year of transfer.

The first model thus consists of the 26 variables described above, and will be used on a data set that covers transfers to the biggest European leagues. For a smaller data set, consisting of transfers to English clubs, additional information on the remaining contract duration with the selling club has been included in a 27th independent variable.

4 Data and implementation

For calculating player ratings match data from five seasons, 2009-2010 to 2013-2014, were collected from 14 competitions: English Premier League, English Championship, English FA Cup and League Cup, German Bundesliga, Italian Serie A, French Ligue 1, Spanish Primera Division, Portuguese Primeira Liga, Dutch Eredivisie, Belgian Pro League, Norwegian Tippeliga, UEFA Champions League, and UEFA Europa League. This yields $20,217$ matches and a total of $120,834$ segments, and includes in total 15,884 different players. The required data from each match consists of the match date, the starting line-ups, substitutions made, goals scored, and red cards handed out.

Positional attributes and biographical data for players, along with historical transfer data, were sourced from `transfermarkt.de` [6]. The first transfer fee model is based on 1,457 transfers involving players that at the time played in either the Spanish, German, Italian, French, or Portuguese top division, or in one of England's two top divisions. For the second transfer fee model, requiring the remaining contract duration to be known, all players currently playing in the top two English divisions that were bought by one of the 44 English clubs after 2009, have had their corresponding time-of-transfer contract duration sourced from club and media archives. This was done by manually searching for published contract renewal information for each of the 444 transfers in question.

The reading, parsing, and transformation of input data is coded in C#. Equation (3) is handled by creating a component object model-reference to Matlab, which performs the main calculations according to the following script:

% Input: α, β, λ.

$d = \texttt{size}(\alpha); A = \alpha' * \alpha + \lambda * \texttt{speye}(d(2)); b = \alpha' * \beta'; x = A \backslash b;$

The backslash-operator used to calculate x will automatically employ ordinary least squares estimation since an exact solution cannot be found. Furthermore, as the input matrix α is provided as a sparse matrix, and as `speye` generates a sparse identity matrix, the matrix A is also sparse, making the calculations efficient. The data for the transfer fee models were handled in Excel, while the Matlab function `fitlm` was used for the regression analysis.

5 Results and discussion

This section contains results from the player ratings model and from the two transfer fee models.

Player ratings

The player ratings model has two parameters, λ and k, whose values must be determined before use. Two criteria for model performance were used. One was the mean squared error from fitting the sum of appearing players' ratings to observed segment results, calculated out-of-sample using 50-fold cross validation. As cross validation is based on observations of different ages, this was only applied for λ, with $k = 0$. The other criterion was to calculate the quadratic loss of the estimated number of points for each team after simulating a full season of the English Premier League as in [7]. The latter approach showed that values of λ and k should not be selected independently, and the final values used were $\lambda = 3000$ and $k = 0.2$.

Having finalised the specification of an adjusted plus-minus rating system for footballers, we can compute ratings for any given date. Every available observation dating from before that date will be included in the rating system. Using the data set specified in Section 4, the most up-to-date player ratings will be calculated for the date 1 July 2014. With a data set of 120,834 segments and 15,884 players, the rating calculations are performed using Matlab on a standard laptop computer within 4 seconds of computing time. To achieve such computing times it is essential to use sparse matrix operations throughout the calculations, as any single observation will only feature $11+1+11 = 23$ non-zero cell entries. Without sparse data structures, the computer runs out of memory for a data set of only 10,000 players, whereas a larger data set of 19,000 players still requires less than 9 seconds when using sparse matrices. The data set contains 7,240 unique active players that participated in at least one match during the last season included. The twenty most highly rated active players featured in the data set and evaluated by the model are listed in Table 1.

Considering that the model includes a total of 15,884 players, it seems to do a good job at identifying which ones are valuable to their team. Out of the twenty nominees for the 2014 FIFA Ballon d'Or, eight feature in the top-20 of the model, including all of the three finalists. Considering the FIFA Team of the Year 2014, seven out of eleven players feature in the top-20 of the model. One could argue that Cristiano Ronaldo, the 2014 Ballon d'Or winner, ought to be recognised as the greatest individual player alongside Lionel Messi, but the model suggests that his teammates Sergio Ramos and Karim Benzema have been just as important to Real Madrid's results when considering their performance

Table 1: The top-20 highest rated players on 1 July 2014

Player	Nationality	Team	Position	Year of birth	Rating
Lionel Messi	ARG	Barcelona	F	1987	0.1957
Sergio Ramos	ESP	Real Madrid	D	1986	0.1867
Karim Benzema	FRA	Real Madrid	F	1987	0.1855
Víctor Valdés	ESP	Barcelona	G	1982	0.1790
Cristiano Ronaldo	POR	Real Madrid	F	1985	0.1757
Pepe	POR	Real Madrid	DM	1983	0.1727
Manuel Neuer	GER	Bayern München	G	1986	0.1683
Helton	BRA	Porto	G	1978	0.1603
Thibaut Courtois	BEL	Atlético Madrid	G	1992	0.1586
Thomas Müller	GER	Bayern München	F	1989	0.1580
Sergio Busquets	ESP	Barcelona	DM	1988	0.1571
Andres Iniesta	ESP	Barcelona	MF	1984	0.1566
Maxi Pereira	URU	Benfica	DM	1984	0.1564
Pedro	ESP	Barcelona	F	1987	0.1545
Philipp Lahm	GER	Bayern München	DM	1983	0.1518
Joao Moutinho	POR	Monaco	M	1986	0.1516
Zlatan Ibrahimovic	SWE	PSG	F	1981	0.1500
Marin Demichelis	ARG	Manchester City	DM	1980	0.1490
Angel di María	ARG	Real Madrid	MF	1988	0.1477
Ezequiel Garay	ARG	Benfica	D	1986	0.1471

from 2009 to 2014. Players of all kinds of preferred positions are featured in the top-20. While forwards are always likely to lead shots statistics, and defenders blocks and interceptions equivalents, this particular rating method shows no discrimination towards tactical or positional concerns.

The rating model also estimates the effect of home advantages and red cards. Home advantage is estimated to 0.408 goals per 90 minutes, according to the regression coefficient of the ”twelfth man” dummy player included for the home team, which is in line with other studies [1]. Regarding dismissals, the first red card has a value of 1.55 goals per 90 minutes, whereas additional dismissals are attributed a much smaller effect, with 0.37 and 0.02 goals per 90 minutes, for the second and third dismissal, respectively. This may make sense, as being shown a first red card is often the time when tactics and preparations become distorted. Further reductions should have an added negative effect, but second and third dismissals are very likely to occur late on in games, with an increased likelihood that the result is more or less settled already.

Transfer fees

The results from the two transfer fee models are presented below. Table 2 shows the estimated regression coefficients for the model covering 1,457 transfers from a broad range of European top leagues, and Table 3 presents the same data for the model covering 444 transfers to the English top divisions.

From the first model we can see that the player performance metric, in this case an adjusted plus-minus rating as explained in Section 2, is quite naturally proved to be the single most important driver of player value. However, the other factors also combine to explain a lot of variance in fees.

The first model indicates that there is a significant increase in transfer fees for players that can use both feet, thus rewarding their flexibility. The same sign of the regression coefficient is observed in the second, smaller, model, but the coefficient is no longer significantly different from 0. Perhaps surprisingly, few of the positional variables have coefficients that are significantly different from 0. The results indicate that wingers are either undervalued or less flexible than other players, as their regression coefficient is significantly different from 0 and negative. There is also a weak indication that inverted wide players demand higher transfer fees as a result of their preferred playing position.

The height of a player does not seem to influence the transfer fees, although it is possible that the benefits of height are already present in the player performance ratings. The two selected models show this for defenders and forwards separately, but similar findings were noted in other model variants where height was included directly.

Nationalities of players influence transfer fees for footballers. Since the models consider European top leagues, where non-EU players face clubs' quotas and strict work permit criteria, it is reasonable that EU players are associated with higher transfer fees. Also South American players demand higher transfer fees than players with otherwise similar characteristics. South American national teams are traditionally strong contenders for the FIFA World Cup, and many players have sought moves to European leagues since the 1970s. If the performance metric is accurate, there should not be a rational bias in favour of South American players for European clubs, and so it might be appropriate to speculate in a possible overestimation of such players' talents, as mentioned in [9]. Furthermore, British or Irish players are, according to the models, subjected to a significant devaluation. A tendency for British clubs to develop their own British players or recruit them relatively cheaply early or late in their careers, and instead spend more money on foreign recruits, might to some extent explain the negative coefficient.

Experience is valued through the significant variables measuring international experience. Senior debuts are greatly rewarded, with incremental caps also driving value positively. Players that only have youth international experience are somewhat devalued in the European market models. A possible explanation is clubs are less willing to pay high fees for players with youth caps that never made the transition to senior national teams. There is no similar significant effect in the English market model.

Variables regarding player age behave as expected, with a positive linear effect and a negative quadratic effect, giving a transfer fee that is smaller for younger players with a lot of uncertainty regarding their eventual prospects, and smaller for older players when there is less potential and possibly a higher risk of injuries.

In the second model, for the smaller set of transfers to the English top divisions, contract duration at the time of transfer was included. As expected, this is a highly important driver of player value. Whenever a contract nears expiration, and ignoring change in other parameters, value intuitively has to drop as a potential free Bosman transfer becomes an option open to the player. Players might not always enjoy the situation, as they may not enjoy the same remuneration at a new club, but their current clubs certainly have very little to gain in ways of a transfer fee.

Table 2: Proposed explanatory variables for modelling transfer fee, with estimated coefficients and degrees of significance (*** 1 %, ** 5 %, * 10 %), based on 1457 observations. Model goodness of fit = $R^2 = 0.331$

Variable	Type	Coefficient	P-value
(Intercept)		(21.744)	(0.571)
Rating	Continuous	8.343	0.000***
Left Footed	Binary	0.048	0.495
Both Feet	Binary	0.217	0.034**
Goalkeeper	Binary	0.165	0.257
Defender	Binary	−1.503	0.337
Midfielder	Binary	0.082	0.259
Forward	Binary	−0.891	0.490
Winger	Binary	−0.220	0.007***
Inverted Wide Player	Binary	0.158	0.059*
Defender: Height	Integer	0.008	0.379
Forward: Height	Integer	0.006	0.385
African	Binary	0.340	0.274
UK or Eire	Binary	−0.770	0.000***
EU	Binary	0.919	0.003***
Other European	Binary	0.267	0.404
Asian	Binary	0.326	0.428
Eng. Speak	Binary	−0.117	0.764
South American	Binary	1.185	0.000***
Capped	Binary	0.528	0.000***
Youth Capped	Binary	−0.244	0.006***
Scaled Caps	Continuous	0.014	0.000***
Scaled Injuries	Continuous	0.000	0.044**
Age	Integer	0.239	0.038**
Age^2	Integer	−0.007	0.002***
Month of Birth	Integer	−0.012	0.125
Transfer Year	Integer	−0.005	0.800

The explanatory power of the two transfer fee models will depend on the particular data set used, and the model including remaining time of the current contract has a much larger R^2 of 0.584, compared to 0.331 for the larger European model. However, most of the improved explanatory power is due to the reduced data set: the second model without the years left of the contract but using the same 444 transfers still has an R^2 of 0.481.

Table 3: Proposed explanatory variables for modelling transfer fee (English market), with estimated coefficients and degrees of significance (*** 1 %, ** 5 %, * 10 %), based on 444 observations. Model goodness of fit = $R^2 = 0.584$

Variable	Type	Coefficient	P-value
(Intercept)		(−111.110)	(0.097)
Rating	Continuous	3.939	0.002***
Left Footed	Binary	0.019	0.861
Both Feet	Binary	0.148	0.337
Goalkeeper	Binary	0.067	0.760
Defender	Binary	0.486	0.836
Midfielder	Binary	0.096	0.416
Forward	Binary	−0.106	0.953
Winger	Binary	−0.317	0.015**
Inverted Wide Player	Binary	0.225	0.081*
Defender: Height	Integer	−0.002	0.896
Forward: Height	Integer	0.003	0.799
African	Binary	0.334	0.440
UK or Eire	Binary	−0.783	0.000***
EU	Binary	0.884	0.033**
Other European	Binary	0.137	0.765
Asian	Binary	0.128	0.897
Eng. Speak	Binary	0.170	0.719
South American	Binary	1.085	0.015**
Capped	Binary	1.038	0.000***
Youth Capped	Binary	0.245	0.090*
Scaled Caps	Continuous	0.018	0.000***
Scaled Injuries	Continuous	0.001	0.077*
Age	Integer	0.319	0.084*
Age^2	Integer	−0.007	0.039**
Month of Birth	Integer	−0.017	0.153
Transfer Year	Integer	0.060	0.072*
Years Left	Integer	0.466	0.000***

6 Concluding remarks

The contributions of this paper is two-fold. First, a regression-based player rating for association football has been defined and its ability to identify the top players in European leagues has been demonstrated. The ratings model can be solved for a relatively large data set in a few seconds on a standard laptop computer. Second, the player ratings and other explanatory variables have been tested in two regression models for transfer fees in association football. The models are able to explain some of the variance in transfer fees, while at the same time hinting at possible biases in the market: South American footballers seem to demand a transfer fee that is too high compared to their on-field performances, and players from UK and Eire seem to be undervalued by the market.

References

[1] S.R. Clarke and J.M. Norman. Home ground advantage of individual clubs in english soccer. *The Statistician*, 44:509–521, 1995.

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

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

[4] S. Dobson and J. Goddard. *The Economics of Football*. Cambridge, 2001.

[5] B. Frick. The football players' labor market. *Scottish Journal of Political Economy*, 54:422–446, 2007.

[6] S. Herm, H.-M. Callsen-Bracker, and H. Kreis. When the crowd evaluates soccer players' market values: accuracy and evaluation attributes of an online community. *Sport Management Review*, 17:484–492, 2014.

[7] L.M. Hvattum. Analyzing information efficiency in the betting market for association football league winners. *The Journal of Prediction Markets*, 7:55–70, 2013.

[8] L.M. Hvattum and H. Arntzen. Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*, 26:460–470, 2010.

[9] S. Kuper and S. Szymanski. *Soccernomics*. HarperSport, 2014.

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

[11] B. Macdonald. An improved adjusted plus-minus statistic for NHL players. *Proceedings of the MIT Sloan Sports Analytics Conference*, 2011.

[12] B. Macdonald. Adjusted plus-minus for NHL players using ridge regression with goals, shots, Fenwick, and Corsi. *Journal of Quantitative Analysis in Sports*, 8(3), 2012.

[13] J. Vecer. Crossing in soccer has a strong negative impact on scoring: Evidence from the English Premier League the German Bundesliga and the World Cup 2014. http://dx.doi.org/10.2139/ssrn.2225728, 2014.

[14] W. L. Winston. *Mathletics*. Princeton University Press, 2009.