Plus-Minus Player Ratings for Soccer

Tarak Kharrat[1], Javier López Peña[2], and Ian G. McHale[1]

[1]Centre for Sports Business, University of Liverpool Management School, U.K.
[2]Department of Mathematics, University College London, U.K.

August 13, 2019

Abstract

The paper presents plus-minus ratings for use in association football (soccer). We first describe the general plus-minus methodology as used in basketball and ice-hockey and then adapt it for use in soccer. The usual goal-differential plus-minus is considered before two new variations are proposed. For the first variation, we present a methodology to calculate an *expected goals* plus-minus rating. The second variation makes use of in-play probabilities of match outcome to evaluate an *expected points* plus-minus rating. We use the ratings to examine who are the best players in European football, and demonstrate how the players' ratings evolve over time. Finally, we shed light on the debate regarding which is the strongest league. The model suggests the English Premier League is the strongest, with the German Bundesliga a close runner-up.

Keywords: OR in sports, soccer, ridge regression, point-process, sparse matrix, plus-minus.

1 Introduction

In sport, there is great interest in evaluating and measuring the performance of players. In team sports, owners, managers and coaches want to identify which players are key to their team's success, so that recruitment and retention of players can be properly informed. Unlike other industries, there is much external interest in the performance of a sports team's *employees* from, for example, fans and the media wanting to know which players to support, and which to berate. As such, one of the main tasks of sports analytics is to evaluate the performance of players and understand their contribution to the team's results.

In this paper we present two modifications of the well-known *plus-minus* (PM) ratings model previously used to identify key players in basketball (see, for example, Sill (2010)) and ice-hockey (see, for example, Macdonald (2012a)). The PM ratings system is simple and intuitive, and provides an answer to the question: 'how does a team perform with a player, compared to without the player?'. The modifications we propose are specific to soccer - a game in which it is notoriously difficult to rate players objectively.

For individual sports like tennis and chess, rating players is perhaps simpler than for team sports. Paired comparisons models are well-established and several variations exist. McHale and Morton (2011) provide a ratings system for tennis for example. A perhaps more complex task is to estimate time-varying ratings for individuals which update following new information (the latest results). Elo ratings have been used for over half a century for rating chess players. Similarly, the *Glicko* rating system (Glickman, 2012) provides a more theoretically justified model for estimating time-varying ratings of individuals.

More recently, attention has moved to using machine learning techniques to estimate player ratings. The *TrueSkill* rating (Herbrich et al., 2007) developed at Microsoft is a generalisation of the Elo ratings and is used for dynamically rating video game players.

Rating players in sports teams is more problematic. Players often have different responsibilities with some concentrated on offence (i.e. aiding scoring), whilst others are specialised in defence (i.e. helping to prevent scores for the opposition). A commonly used approach is to assign a value to each of a set of actions considered to be 'of interest' and to reward the player taking them with the associated value.

This method was used for example in the EA SPORTS Player Performance Indicator (McHale et al., 2012) and was used until the end of the 2017/2018 season by the English Premier League as its official player ratings system. Due to its additivity, the previous approach provides simple, user-friendly player ratings and rankings. However, a cost of the simplicity is the lack of context and a deeper understanding of the situations in which actions were committed. Further, the data requirement is not trivial.

Models have been used to rate players for specific tasks. For example, Sáez Castillo et al. (2013) and McHale and Szczepański (2014) present methods to identify the scoring ability of footballers whereas López Peña and Touchette (2012), López Peña and Sánchez Navarro (2015), Brooks et al. (2016) and Szczepański and McHale (2016) deal with passing. But identifying the overall contribution of a player to a team's success (or lack of it) has proven difficult in soccer. However, the concept of the PM ratings provides hope.

The concept of the PM rating is fundamentally different to the rating mechanisms discussed above. It directly measures the contribution a player has on a team's success as measured by (the differential) of a target metric (goals for example). It does not make use of event data, and is not concerned with the number of actions a player might have achieved. All that matters is *"whilst the player was on the pitch, what happened to the target metric?"*. The adjusted PM rating, first described in Rosenbaum (2004), answers this question whilst taking account of the strength of the other players on the pitch.

Plus-minus ratings have been used extensively in ice-hockey (Schuckers et al., 2011; Macdonald, 2012a; Spagnola, 2013; Gramacy et al., 2013) and basketball (Sill, 2010; Fearnhead et al., 2011). Indeed, PM ratings are now part of the statistics reported by the media (ESPN for example[1]) and professional leagues (since 1968 the NHL has kept track of each player's PM rating[2]) in US team sports. Plus-minus ratings have yet to be widely adopted in soccer. To the best of our knowledge, the first academic study presenting a method for calculating plus-minus ratings in soccer is Saebo and Hvattum (2015). More recently, Schultze and Wellbrock (2018) presented a plus-minus ratings system for soccer in which betting odds were used to proxy the strength of the two competing teams, and 'reward' players more or less ratings points based on the relative strengths of the two teams. The concept of plus-minus ratings being used in soccer is discussed in some specialised forums[3] but very little work has been done on the topic, despite it being such a widely used, and accepted metric in US sports.

In this paper we propose to fill the gap and adapt the plus-minus rating for use in soccer. We present the model currently used in basketball, and reported by ESPN before suggesting modifications to adapt the methodology for use in soccer. We then propose two extensions of the methodology using new target metrics measuring team success: first, we present an expected goals (xG) plus-minus rating (xGPM); and second, we present an expected points (xP) plus-minus rating (xPPM). For the xGPM ratings we use a model to calculate the probability of a shot resulting in a goal. For the xPPM ratings we use an 'in-play' model to estimate the probability of each match outcome (win, draw, loss) given the current game state at any moment during the game. These new targets are more suited to calculating plus-minus ratings in soccer than the goal differential, which produced a very sparse target, and will mean we can better discriminate between players. Both models are presented below.

The remainder of the paper is structured as follows: first, we describe the data used for this research (Section 2). In Section 3 we present the naive plus-minus rating and the *regularized adjusted* plus-minus rating currently used in basketball. In Section 4 we describe two new variations of the regularized adjusted plus-minus ratings: an expected goals plus-minus rating (xGPM), and an expected points plus-minus rating (xPPM). In Section 5 we use the ratings to look for the top players across European soccer, and see how their ratings evolve over time, before using the model to examine the relative strengths of European leagues, and investigate how the ratings might be used to inform recruitment decisions of football clubs. We conclude with some closing remarks in Section 6.

2 Data

We collected data from 11 European leagues over the last 8 seasons as detailed in Table 1. For every game in our data set, we collect the match date, the starting line-ups, timings of any goals, and timings and player names of any substitutions and red cards.

[1] http://www.espn.com/nba/statistics/rpm
[2] http://www.nhl.com/stats/player
[3] http://www.soccermetrics.net/player-performance/adjusted-plus-minus-deep-analysis

Table 1: Description of data used by league and season.

League	Seasons	Games
England Premier League	2009/10–2016/17	3,040
Germany Bundesliga I	2009/10–2016/17	2,448
Spain La Liga	2009/10–2016/17	3,039
Italy Serie A	2009/10–2016/17	3,037
Germany Bundesliga II	2015/16–2016/17	612
England Championship	2013/14–2016/17	2,227
Netherlands Eredivisie	2013/14–2016/17	1,242
Turkey Super Lig	2014/15–2016/17	918
Portugal Liga NOS	2016/17	306
France Ligue 1	2009/10–2016/17	3,039
Russia Premier League	2013/14–2016/17	960
Total		**20,868**

For the expected goals model developed in Section 4.1, additional information regarding shots is needed. Specifically, the shot time, the shooter (x, y) coordinates and the type of shot (penalty, free-kick, header or open play), are extracted from Opta F24 feed.

3 The Plus-Minus Rating

The plus-minus statistic, in some form or another, has been in use since the 1950s in ice-hockey but is most seen nowadays applied to basketball. Indeed, the complexity of the game of basketball has led to several developments of the original concept. In this section we will first describe the naive plus-minus statistic, before presenting modifications that have been introduced in the literature. In what follows we will define everything in terms of soccer.

3.1 The Naive Plus-Minus Statistic

In its simplest form, a player's plus-minus statistic can be used to answer the question: "what happens when the player is on the pitch, compared to when he is off it?". Historically, goals (or points in basketball) have been the preferred way to measure "what happened" and the raw plus-minus score calculates the player's contribution to the goal difference of his team (per ninety minutes) whilst he is on the pitch. For example, consider a player who makes two match appearances. In the first match, he plays the first 60 minutes during which the team concedes one goal and fails to score itself. The match finishes in a 1-0 loss. In the second match, the player comes to the field with 30 minutes remaining and his team is enjoying being 3-0 ahead. During the 30 minutes of play he is on the pitch, the score moves to 5-0. The player's plus-minus rating is then $((-1/60) + (2/30)) \times 90 = +4.5$. In other words, when the player was on the pitch the team scored 4.5 goals per 90 minutes more than the opposition.

The net raw plus-minus statistic (the naive plus-minus rating) can be used to measure the importance of a player to his team. This is simply the raw plus-minus statistic when the player is on the pitch minus the raw plus-minus statistic when the player is not on the pitch. In our example, the raw plus-minus statistic without the player is $((0/30) + (3/60)) \times 90 = 4.5$, so that the naive plus-minus statistic is 0. It appears then that the team is equally successful with and without the player.

This is, of course, a very simplistic picture and several pieces of information are not taken into account. For example, the effects of strengths of the other players on the pitch, or of the game situation (such as a reduction in the number of players on a team following a red card), or of any home advantage have not been accounted for. Further, if one was to use this naive plus-minus rating to compare players from different teams, the results would be almost meaningless. Consider two different players: one playing for the league champions and the other playing for the leagues worst team. Suppose both players had naive plus-minus ratings of 0. Who is likely the better player? Most sports fans would say that the player achieving a naive plus-minus of 0 whilst playing for the league's worst team probably deserves more credit in this example.

To account for these factors, the *regularized adjusted* plus-minus statistic was introduced, and is described next.

3.2 Regularized Adjusted Plus-Minus

The adjusted plus-minus player metric was first described in Rosenbaum (2004) who presented the plus-minus statistic as a regression problem. Doing so means 'adjustments' can be made to the naive plus-minus statistic to account for the strengths of the other players on a team, and of the opposition players. The set up is again simple. Define a segment of play to be one where the same set of players (usually two sides of 11 players) are on the pitch. A new segment is defined every time a new set of players are on the pitch. This may occur when a substitution is made, or when a sending off occurs, or for a different match. Each segment $t = 1, \cdots, T$ is an observation. The dependent variable is the goal difference $\mathbf{y} = (y_1, ..., y_T)$ per 90 minutes during segment t. Let there be N players in total (in the whole league), then the independent variables form a $T \times N$ design matrix X of dummy variables defined as

$$x_{tj} = \begin{cases} 1 \text{ if player } j \text{ plays for the home team in the segment} \\ -1 \text{ if player } j \text{ plays for the away team in the segment} \\ 0 \text{ if he doesn't play in the segment} \end{cases}$$

where each player in the league is identified by a unique numeric index j. The adjusted plus-minus statistics for the N players are given by the estimated α in the linear model $\mathbf{y} = \mathbf{X}\alpha$, where α is an $N \times 1$ vector of parameters measuring the contribution of each player to the response variable (in this case, goal difference).

In basketball, the number of segments within a game is much higher than the number of players used in the game, and the matrix $\mathbf{X}^\top \mathbf{X}$ is 'well-behaved' so that α can be estimated. In soccer however, the number of substitutions is limited to three per team and the number of segments is much smaller than the number of players on the pitch. Further, over the course of a match and season, the same groupings of players will play together for a large percentage of the total minutes the team plays. For example, a partnership between two centre backs is commonplace in soccer meaning they are on the pitch together for nearly every minute of play during an entire season. The result of all of this is that although the matrix $\mathbf{X}^\top \mathbf{X}$ is well-behaved for basketball, it is not so for soccer, and is singular, or near-singular, so that attempts to estimate α using ordinary least squares for example will fail.

Ice-hockey suffers from these same problems and Macdonald (2012a) presented a solution using ridge regularisation (also known as Tikhonov regularisation) instead of ordinary least squares to estimate the coefficients. The resulting methodology is known as the *regularized adjusted plus-minus statistic*. Ridge regularisation is known to work well in the presence of colinearity and attempts to solve the problem by making a trade-off between minimising the estimation error (suppressing noise) and minimising the magnitude of the estimate (risking loss of information). In other words, instead of minimising the objective function in the usual squared errors problem:

$$\min ||\boldsymbol{\alpha}\boldsymbol{X} - \boldsymbol{y}||_2^2$$
$$\boldsymbol{\alpha} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y},$$

an alternative objective function, given by:

$$\min ||\boldsymbol{\alpha}\boldsymbol{X} - \boldsymbol{y}||_2^2 + \lambda ||\boldsymbol{\alpha}||_2^2$$
$$\boldsymbol{\alpha} = (\boldsymbol{X}^\top \boldsymbol{X} + \lambda^2 \boldsymbol{I})^{-1} \boldsymbol{X}^\top \boldsymbol{y},$$

is used. The penalty term, λ, penalizes large values of the parameters of interest. The advantage of the ridge regression compared to other regularisation techniques such as the lasso for example is that it shrinks the coefficients of correlated predictors towards each other. In the extreme case of k identical predictors, the ridge regularisation will give each of them identical coefficients with a magnitude equal to $1/k$ of the magnitude that any single one would get if it were the only one of the perfectly correlated variables used as a covariate. This is very desirable in the case of estimating adjusted plus-minus ratings. Consider the situation in which two players always play together (a pair of centre backs for example).

Although the players may, in actual fact, have different abilities, and contribute a different amount to their team's performances, it is intuitively sound for the model to estimate their contributions to the team as equal and thus award them identical ratings. Given the data that the model 'sees', there is no reason to credit one player more than the other. The only way to identify a difference in the contributions of the two players to team performances would be to use more information such as event data, or player tracking data. The simplicity of the data requirements for calculating regularized adjusted plus-minus ratings comes at the cost of not being able to differentiate between the contributions of such players. Herein, we calculate only regularized adjusted plus-minus ratings but for brevity refer to plus-minus ratings.

4 New Plus-Minus Ratings for Soccer

As a consequence of ice-hockey being a low scoring game, the latest developments in the plus-minus metric have looked at using alternative dependent variables to measure the team's success. The dependent variable is often called the 'target' as it is in some sense what the players should be targeting to improve during the match. In ice-hockey, Macdonald (2012b) and Schuckers and Curro (2013) use expected goals rather than actual goals as the target variable, whilst Macdonald (2012a) presents plus-minus models for shots. Instead of modelling a target such as goals, or shots, Thomas et al. (2013) makes the hazard function for each team scoring in an ice-hockey match a function of the players on the ice at any moment in time. In this section we present two new versions of the plus-minus metric for use in soccer: (a) a plus-minus metric with difference in expected goals between the two teams as the target variable, and (b) a plus-minus metric with change in expected points as the target. These new targets solve a major problem with the goal-based plus-minus rating: the sparsity of the response. In our data, 72% of the segments of play (periods of play in which the set of players on the pitch remains constant) have a goal difference of 0. The two new targets we propose do not have this problem.

4.1 Expected Goals Plus-Minus

In recent years the concept of expected goals in soccer and ice-hockey has become popular in the media (see, for example, Green (2012)). In the academic literature there has been limited interest with the only exceptions being, to the best of our knowledge, Lucey et al. (2014) and Eggels et al. (2016).

The idea behind the notion of expected goals (xG) is simple: for each shot on goal that a team has, the expected number of goals is the probability of the shot resulting in a goal. The probability of the shot being successful depends on several factors: the location of the shot (proximity to the goal), the player, the position of the defenders, the weather conditions, and the fatigue of the player, for example. The reason xG has become a popular concept in soccer is that it has been shown to be more informative than actual goals when judging how well a team has played[4]. Since goals are a rare event, they do not always reflect properly a team's performance on the pitch. An alternative is to use shots, which are an order of magnitude more common, instead of goals, but this has the problem of considering all shots with equal standing, regardless of how good a chance they have of being successful. An expected goals model deals with this issue by assigning to each shot a measure of its quality, computed as the probability the shot had of resulting in a goal.

In order to create our expected goals model, we compare the out-of-sample performance of several probabilisitic classifiers trained on a large amount of shots. Earlier work on expected goals models has focussed on finding models that predict expected goals that are as close as possible to the actual number of goals scored which, in our opinion, defeats the purpose of calculating expected goals. Since we are interested in predicting an accurate *probability* that a given shot will result in a goal, we use the Brier Score as the target for model training, hyper-parameter tuning, and cross-validation. Note that the definition of the Brier score adopted here follows the original formulation given by Brier (1950) and is defined as $BS = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{R}\left(p_{ji} - o_{ji}\right)^2$ in which p_{ji} is the probability that was forecast for outcome i, o_{ji} is the dummy variable equal to one if outcome i is observed and $R = 2$ corresponds to the two possible outcomes of a shot: a goal, or no goal. Lucey et al. (2014), present an expected goals model for football but use the *mean absolute error* as their target metric.

[4]See, for example, http://11tegen11.net/2016/02/21/predicting-league-football-using-xg-and-more/

Shots in football come from many different situations. We have separated our shots into four different categories: *penalty*, *freekick*, *header*, and *open play*. The latter category contains all shots taken with the foot that did not result directly from a set piece. Since the nature of each of these four types of shots is different, we designed our expected goals model by fitting four *specialist models*: one to each shot category. This means the feature selection process can be refined for each type of shot, and any redundant information is removed from the model (for instance, there is no point in using shot location when designing a model for penalties).

Our dataset contains almost 400,000 shot events. Of those shots, roughly 41,000 resulted in a goal (a conversion rate of 10.4%). The breakdown of shots by type is provided in Table 2. The baseline Brier Score is determined by calculating the Brier Score for a model in which the estimated probability is equal to the empirical frequency of scoring a goal for that particular shot type.

Table 2: Shot types and baseline Brier Scores. The baseline Brier Score is equal to the calculated Brier Score for a model in which the predicted probability is equal to the empirical frequency for that type of shot.

	Shots	Goals	Baseline Brier Score
Free Kick	19,723	1,229	0.058
Header	65,369	7,320	0.099
Open Play	304,398	29,168	0.087
Penalty	4,420	3,339	0.185
Total	393,910	41,056	0.093

We consider the following features in order to train our models, all of them normalized so that they lie in the range from 0 to 1. The dataset is split into training and validation sets using an outer-loop cross validation with three folds. The final scores are the averages of the scores in the three folds.

- *Horizontal pitch coordinate*: x, 1 corresponds to the goal line on the attacking side.
- *Adjusted vertical coordinate*: y_{adj}, 0 corresponds to a central position, 1 corresponds to the edge of the pitch (on either side).
- *Goal view angle*: the angle between the coordinates of the shot origin and the two goalposts.
- *Inverse of the distance to goal*: measured to the centre of the goal, 1 corresponds to a shot originating from the centre of the goal, whilst 0 corresponds to a shot originating from the furthest position on the pitch.
- *Time of play*: 0 corresponding to the kickoff and 1 corresponding to the last minute of the match (including any injury time played, such that 1 can correspond to more than 90 minutes).
- *Goal value*: a measure of how the winning probability would be affected if a goal was scored, given as empirical frequencies, based on goal difference and game time remaining.

Although some of the features we consider are correlated (namely the pitch coordinates, the inverse distance to goal, and the goal view angle) the relationships are nonlinear, and therefore some families of classifiers benefit from the additional information.

We test four main families of machine learning models: logistic regression; Random Forests; Gradient Boosting; and Neural Networks (Multi-Layer Perceptron). In order to fine tune the models' hyper-parameters, an inner-loop 10 fold cross validation is performed on the training set; the resulting model is then evaluated on the validation set in order to get the out-of-sample score. Results are summarized in Table 3.

As is often found with machine learning algorithms, no one type of model always performs 'best', albeit random forest and gradient boosting models seem to present an overall performance superior to logistic regression or neural networks. As a point of comparison, the mean absolute value from our combined best models is very similar to the best model in Lucey et al. (2014). However, the model used by Lucey et al. (2014) included information on the position of the defending players on the opposition side.

Table 3: Summary of model fits (Brier Scores) for each shot type. The best performing model is highlighted in bold.

	Penalty	Freekick	Header	Open Play
Baseline	0.1848	0.0584	0.0994	0.0866
Logistic Regression	0.1848	0.0580	0.0904	0.0743
Random Forest	0.1850	0.05749	0.0875	**0.0738**
Gradient Boosting	0.1848	**0.05748**	**0.0872**	0.0739
Neural Network	0.1848	0.0584	0.0922	0.0743

It is worth noting some characteristics of the models for each shot type. Penalties require consideration separately to other shots. All penalties are taken from the same spot so shot location variables cannot be included in the model. Further, only a few models manage to outperform the baseline score, and the improvement is so small that it is probably not significant. The bottom-line here seems to be that the outcome of penalties are random conditional on the set of factors included in our dataset. Other factors, such as player-dependent variables, are unaccounted for in our data and upon inclusion in the model, might improve the predictive power of the model. Therefore penalties should all be awarded the same value for expected goals.

For the freekick model, most of the predictive power comes from the location based features. Similarly to penalties, the scores for all the models trained are very close to the baseline. Gradient boosting outperforms random forests by a tiny, non-significant, amount.

The outcome of headed shots is heavily influenced by shot location, with the goal view angle being the dominant variable in the model.

By far the largest subset of shots in our dataset is open play shots. All the features seem to add value to the models, with the exception of game time. The dominant features are inverse goal distance and goal view angle.

The resulting *net* expected goals for each segment of play (in which the same set of players is on the pitch) is used as the dependent variable (or target) in our expected goals plus-minus (xGPM) player rating.

4.2 Expected Points Plus-Minus

The ultimate objective of a soccer match is to win. Team managers and fans want to know which players perform well when the match is tight and the scoreline is close. Using the goals-based (regularized adjusted) plus-minus metric, or the xG plus-minus metric presented above, does not account for the match situation. As such, we propose a new plus-minus metric based on expected points. In soccer leagues around the world, a team is awarded 3 points for a win, 1 point for a draw and 0 points for a loss. The expected points for the home team in minute t of a match is then

$$xP_t^H = 3 \times \mathsf{P}_t^{HW} + 1 \times \mathsf{P}_t^D,$$

where P_t^{HW} is the probability of the home team winning the match evaluated at time t, taking into account the current scoreline and the number of players on each team. P_t^D is the probability of the team drawing the match evaluated at time t.

In calculating our new expected points plus-minus statistic, we compare the expected points at the start of a segment of play with those at the end of a segment of play. For example, suppose that the first change in team lineups in a particular match happened in minute 60 (through substitution(s) or a red card dismissal(s)). The change in expected points for the home team is $\Delta xP^H = xP_{60}^H - xP_0^H$, whilst the change in expected points for the away team is $\Delta xP^A = xP_{60}^A - xP_0^A$. The target variable we propose is then the change in expected points for the home team minus the change in expected points for the away team, $y_{[0,60]} = \Delta xP^H - \Delta xP^A$.

In order to calculate expected points variables we need an 'in-play' model to estimate the probabilities of the home team winning, a draw and the away team winning at any moment of the match.

The model used here is a simplification of the random point process model described in Volf (2009). This process is fully characterised by the scoring intensity functions (also known as hazards) of the home and away teams $\lambda_H(t)$, and $\lambda_A(t), t \geq 0$ which are non-negative, bounded, measurable functions of t. The intensity is allowed to depend on some covariates $Z(t)$. $Z(t)$ is in turn an observed random process that can depend on time. A common framework to model the effect of covariates on the intensity function is to use a proportional hazard model, first described in Cox (1962).

Here, the hazards of each team scoring depend on two categorical covariates describing the game context at time t. They are defined by:

- $z_{GD}(t) = -3 \leq, -2, \dots, 2, \geq 3$ defines the goal differential with respect to the home team. We found that a truncation at 3 goals difference works well in practice.
- $z_{MP}(t) = -3 \leq, -2, \dots, 2, \geq 3$ defines the man power advantage with respect to the home team.

The model basically assumes that each team scores goals at a rate that depends on the time of the match, the number of red cards received by the two teams, and home advantage. The simplification we adopt over Volf (2009) is to not take account of the strengths of the two teams playing in any particular match. As such, we are effectively using 'average' probabilities over all games. The justification for this is that the identities of the players is already the focus of the estimation exercise and taking them into account at this stage would, in some sense be double counting, and result in 'punishing' players on good teams with high probabilities of winning matches.

The initial (average) probabilities of a home win, a draw and an away win at $t = 0$ can be computed from the empirical frequency. Using the last eight years of results from the English Premier League, these probabilities are 0.46, 0.26 and 0.28 respectively. The corresponding expected points at $t = 0$ are then 1.63 for the home team and 1.11 for the away team. We computed similar quantities for every league we have in our data.

Returning to our example, we can calculate the target variable as

$$\begin{aligned} y_{[0,60]} &= (xP^H_{60} - xP^H_0) - (xP^A_{60} - xP^A_0) && (1) \\ &= (3 \times \mathsf{P}^{HW}_{60} + \mathsf{P}^D_{60} - 3 \times \mathsf{P}^{HW}_0 - \mathsf{P}^D_0) - (3 \times \mathsf{P}^{AW}_{60} + \mathsf{P}^D_{60} - 3 \times \mathsf{P}^{AW}_0 - \mathsf{P}^D_0) && (2) \end{aligned}$$

The model computes these probabilities and the corresponding target $y_{[0,60]}$ for this game segment can be computed. This model is fitted as explained in Volf (2009, Section 4), and estimated probabilities are obtained by simulation using the procedure detailed in Volf (2009, Section 5).

This new target directly rewards players for contributing to the final result. Previous plus-minus ratings, including the expected goals plus-minus rating described above credits players for creating chances and scoring goals irrespective of the influence of them on the final result.

4.3 Minor Modifications to Plus-Minus Ratings for Soccer

Adjusting for Man Power

The effect of receiving a red card has been studied in soccer (see, for example, Ridder et al. (1994) and Liu et al. (2016)) and has been found to be beneficial for the opposing team in terms of scoring rate. Further, the advantage is larger in the case of the home team benefiting from having more players on the pitch.

In ice-hockey, the effect of player expulsion in plus-minus ratings has been modelled using a situation specific coefficient for each player: a coefficient for even-strength situations and another one during shorthanded situations (Macdonald, 2011). This solution has the effect of doubling the number of estimated coefficients and is not suitable for large numbers of players, and given the extremely low frequency of red cards in soccer, is unnecessary.

The solution we use here is different and follows that of Saebo and Hvattum (2015). The effect of receiving a red card is accounted for using a dummy variable set equal to 1 when team is shown its first red card. In effect, the player in question is replaced by the 'first dismissal' dummy player. In the rare cases of a second red card, the offending player is replaced by a 'second dismissal' dummy variable, and so forth. In games when the opposing team also receives a red card, the highest dismissal dummy is reset to 0 as the team's man power advantage is cancelled out. We use three *dismissal* dummy variables to cover the maximum number of dismissals occurring in the data.

Home Advantage

Home advantage in soccer was first discussed in the academic literature by Pollard (1986) and many researchers have since measured its magnitude (Clarke and Norman, 1995) and tried to explain its variation over time (Pollard and Pollard, 2005) and space (Pollard, 2006).

When computing the plus-minus statistic for basketball Winston (2012) accounted for home advantage by adjusting the points differential (the dependent variable in the regression model) by the average number of points by which the home team defeats the away team (3.2 per 48 minutes). Rather than adjusting the dependent variable, the solution we propose here is to add an intercept term to the regression problem which represents the average home advantage over all teams in the competition. This is more in line with what has been done previously in the soccer literature (see, for example, Maher (1982), Dixon and Coles (1997), Koopman and Lit (2015), Boshnakov et al. (2017)).

Chronology of Performances

It is widely accepted in sports that recent performances are more informative when predicting future performances. Dixon and Coles (1997) for example found that including a time-weighting in their likelihood function such that more recent matches had a larger impact on the estimated team strength parameters improved out-of-sample forecast accuracy. Therefore, in order to increase the predictive power of our rating, we apply a weighting scheme to the different observations (segments) when fitting the ridge regression. The weights are computed as follows:

$$w_i = \exp\Big(\zeta(\text{date}_i - \text{ratingDate})/3.5\Big)$$

with ζ being the time-weighting parameter ($\zeta = 0$ corresponds to the non-weighted regression), date_i the date of the ith observation (segment) and ratingDate is the date when the rating is computed. Following standard practise in soccer modelling (Dixon and Coles, 1997; Boshnakov et al., 2017), time distances are scaled in half week units, hence the denominator 3.5 used for the weighting.

League Competitiveness

Since we have data covering several leagues across Europe, we must control for any differences in strengths of the leagues themselves. For example, some leagues may have stronger players on average than other leagues. Two players of equal ability will perform differently if one is in a strong league whilst the other players in a weak league. The Union of European Football Associations (UEFA) itself acknowledges the inequity of ability across leagues and publishes a ranking by country and awards slots in European competitions accordingly. The consequence on our ratings of this is that a bias could be introduced so that players in weak leagues have inflated ratings. This problem appears when data from various competitions are used to fit the ridge regression.

We correct for this bias using a method similar to that of Saebo and Hvattum (2017) and use the players traveling between leagues to compare the strengths of each league. To do so, we introduce one coefficient x_l per league in the data. Assume we have L leagues and let m_{il} be the number of home team players minus the number of away team players, considering only players who are considered at time of match i to be adapted to competition $l, l = 1, \ldots, L$. Unlike Saebo and Hvattum (2017), we consider a player to be adapted to a competition if he plays at least 6 games in the current season in that competition or if he played more games in this competition than in any other over the previous 18 months to the game date. Hence, x_l is the weight of m_{il} in the ridge regression and represents the adjustment we need to apply to a player joining a new league.

The final ridge regression will need to estimate $N+1+3+L$ parameters (N players, a home advantage parameter, three dismissal parameters, and L league parameters). The model's design matrix is very sparse with a limited number of non-zeros entries per row. The model also has two hyper-parameters (the ridge penalty λ and the time weighting ζ) which need to be fine-tuned using cross-validation.

5 Computation Results and Discussion

5.1 Computation

The game segmentation algorithm (Section 3.2) as well as the minor adjustments described in Section 4.3 are applied to the data described in Table 1 using R Core Team (2016) and the result is stored in sparse matrix object implemented in the contributed package `Matrix` (Bates and Maechler, 2017). The computation resulted in $129,988$ segments and $N = 10,983$ players' ratings to be estimated. The ridge regression was performed using the contributed package `glmnet` (Friedman et al., 2010) and a multi-response Gaussian model using a "group" penalty on the coefficients for each variable (also known as multi-task learning).

5.2 Results

5.2.1 Model Tuning

As mentioned in Section 4, the model has two hyper-parameters namely the ridge penalty λ and the time weighting ζ which need to be fine-tuned. The strategy adopted here is to use the new PM player ratings in an ordered probit regression model to predict the match outcomes (home win/draw/ away win) and use the value of the hyper-parameters that minimised the out-sample Brier score. Referring to the definition of the Brier Score given earlier, we now use $R = 3$ to account for the three possible outcomes: home win, draw, away win. A 10-fold cross-validation was used to split the data into training and testing sets and the process was repeated three times. The covariates used are the average PM ratings derived in Section 4 for the starting 11 players using data from the two years prior to the game date[5]. The best model achieved an average Brier score of 0.292 (sd = 0.003) which is similar to the accuracy achieved by the betting market for the same set of games[6] with $\lambda = \mathbf{0.042}$ and $\zeta = \mathbf{0.002}$.

Expected goals models from Section 4.1 were fitted using 10-fold cross-validation, with hyper-parameter tuning in the inner loop. Logistic regression and random forest models used the implementation in `scikit- learn` (Pedregosa et al., 2011). Gradient boosing models were fitted and tuned using `xgboost` (Chen and Guestrin, 2016). Neural network models were fitted and tuned using `Keras` (Chollet et al., 2015).

5.2.2 Fitting Results

Before we investigate the actual players ratings, we first look at the significance of the other adjustments we introduced in Section 4.3, namely *man-power* and *home advantage*. The ridge regression was fitted using the last two seasons and the results are summarised in Table 4. The first red card has a large

Table 4: Impact of red cards on the three plus-minus ratings.

Parameter	PM	xGPM	xPPM
Red Card 1	-1.25	-1.18	-0.12
Red Card 2	-0.16	-0.15	-0.01
Red Card 3	-0.012	-0.005	-0.001
Home Advantage	0.006	0.005	0.0004

negative effect on all three ratings, whereas additional dismissals contribute a much smaller effect. One explanation is that a first red card is very likely to be followed by a considerable change in team tactics, and may happen early enough in a match to leave the opposing team with enough time to take advantage of the extra man-power. Further reductions will have an added negative effect, but will not be associated with a further change in tactics, and are very likely to occur late on in a game, when there is less time to change the match result.

[5] Different length windows were tried and two years was found to perform best in terms of Brier score.

[6] The adjusted probabilities deduced from bet365 pre-match betting odds achieved a Brier score of 0.295, after removing the bookmaker vigorish, for the same set of games.

The estimated home advantage effect is surprisingly very small for the goal and expected goal based PM rating and almost zero for the xPPM one suggesting that players do not perform, on average, differently playing home or away. It is worth noting here that finding a home advantage of zero for the xPPM rating is expected as we have already accounted for it when setting the initial expected points as explained in Section 4.2.

5.2.3 Identifying Europe's Best Players

One 'test' of our plus-minus ratings is to simply look at which players are highly rated. In this section we present the best three players for each calendar year since 2011, and for 2017-18, the best five players for each playing position.

The Ballon d'Or[7] is the most prestigious individual distinction in soccer and is awarded to the player deemed to have performed the best over the previous calendar year, based on voting by expert journalists. Our plus-minus ratings provide us with an alternative way to make a top-player classification for every calendar year. As a proof of concept, we have computed the average of the three variations of the plus-minus rating, each of them previously normalized to the $[0, 1]$ range using min-max normalisation, and filtered out players who didn't play at least 900 minutes (the equivalent of 10 full games). The results are summarized in Table 5. Despite the fact that the Ballon d'Or award was awarded to either Lionel Messi or Cristiano Ronaldo in every one of the years between 2011 and 2017, our scores suggest that perhaps some other players might have deserved the recognition.

Year	Ranking	Player	Score
2011	1	Cristiano Ronaldo	0.7674
	2	Lionel Messi	0.7486
	3	Pedro	0.7335
2012	1	Eden Hazard	0.7635
	2	Lionel Messi	0.7610
	3	Mirko Vucinic	0.7535
2013	1	Cristiano Ronaldo	0.7576
	2	Robert Lewandowski	0.7465
	3	Mats Hummels	0.7397
2014	1	Jérôme Boateng	0.7774
	2	Manuel Neuer	0.7751
	3	Lionel Messi	0.7663
2015	1	David Alaba	0.8130
	2	Robert Lewandowski	0.7855
	3	Claudio Bravo	0.7465
2016	1	Luis Suárez	0.7446
	2	N'Golo Kanté	0.7373
	3	Nick Viergever	0.7282
2017	1	Marc-André ter Stegen	0.7787
	2	Luis Suárez	0.7689
	3	Mohamed Salah	0.7590

Table 5: Top three players for each calendar year according to the average of the three plus-minus ratings calculated using matches played during the 12 months from January to December of that year.

The top five players by position during the 2017-18 season are shown in Table 6. In general, for the outfield positions in Table 6, the ratings seem reasonable. Each of the players featured are known to be very good. Perhaps most intriguing is that one of the hottest properties in world football, as at the time of writing, Matthijs de Ligt, features in the top five defenders in the 2017-18 season. Following Ajax's

[7]https://en.wikipedia.org/wiki/Ballon_d%27Or

run to the semi-finals of the UEFA Champions League (2018-19), every big club in Europe is trying to recruit him. It is promising that the plus-minus ratings recognised his performances a whole year earlier. For the goalkeepers however, we would encourage more caution to be taken when looking at the ratings. Goalkeepers typically play a very high proportion of minutes for the team. Further, if they do not play, they are typically replaced by a single reserve goalkeeper. As such, the plus-minus rating for a goalkeeper becomes a comparison of the team's performances when that player plays versus when the reserve goalkeeper plays. If the reserve goalkeeper is poor, or the team's results are relatively poor when he is on the pitch, then this will inflate the plus-minus rating of the first choice goalkeeper. Effectively, in these circumstances, the plus-minus rating becomes a measure of the comparative importance of the two players to their team. Nevertheless, the ratings are informative, and as of the 2017-18 season, strong candidates for Europe's top players are Lionel Messi, Kevin De Bruyne and Jordi Alba.

Position	Ranking	Player	Team	Score
Goalkeeper	1	André Onana	Ajax	0.6999
	2	Marc-André ter Stegen	Barcelona	0.6995
	3	Sven Ulreich	Bayern Munich	0.6934
	4	Alphonse Areola	PSG	0.6871
	5	Ederson	Manchester City	0.6777
Defender	1	Jordi Alba	Barcelona	0.7539
	2	Matthijs de Ligt	Ajax	0.7149
	3	Gerard Piqué	Barcelona	0.7051
	4	Thiago Silva	PSG	0.6734
	5	Nicolás Otamendi	Manchester City	0.6717
Midfielder	1	Kevin De Bruyne	Manchester City	0.7248
	2	Hakim Ziyech	Ajax	0.7119
	3	Fernandinho	Manchester City	0.6937
	4	Ivan Rakitic	Barcelona	0.6835
	5	Sergio Busquets	Barcelona	0.6767
Forward	1	Lionel Messi	Barcelona	0.7346
	2	Luis Suárez	Barcelona	0.7048
	3	Thomas Muller	Bayern Munich	0.6823
	4	Edison Cavani	PSG	0.6807
	5	José Callejón	Napoli	0.6780

Table 6: Top five players for each position for the seasons 2017-18 according to the average of the three plus-minus ratings.

Lastly in this section, we look at how the performance of selected players has evolved over time. In Figure 1 we plot the relative contributions of three of football's biggest stars to their respective teams: Cristiano Ronaldo, Lionel Messi, and Neymar. The plus-minus ratings have been calculated using a rolling 12-month window of matches played up to, but not including, the date shown on the x-axis.

Over the last three seasons, the ratings of all three players have been remarkably close, though in recent months Messi is rated as contributing more to his team than both Neymar and Ronaldo.

5.2.4 Using plus-minus ratings to inform recruitment decisions

Nowadays, the transfer market on footballers involves mind-boggling sums of money. In January 2018, Liverpool player Philippe Coutinho was sold to Barcelona for a reported £146m. At the time this was somewhat controversial as many pundits and fans believed Coutinho to be Liverpool's best player, and an essential element in the team being successful. Figure 2 plots the relative contributions of Liverpool's four main forwards, and the players around which the debate centred: Philippe Coutinho, Mo Salah, Sadio Mane, and Roberto Firmino.

Contrary to the belief at the time of the transfer, Figure 2 provides clear evidence that of the four players, Coutinho was the least valuable to Liverpool as both his expected goals plus-minus rating, and his expected points plus-minus rating is the lowest of the four players. Hence, it would appear that, based on footballing reasons only, it was not a poor decision by the Liverpool management to sell their 'best' player. Coutinho's relatively unsuccessful season at Barcelona in 2018-19, and the fact that Liverpool won the Europe's most prestigious trophy, the UEFA Champions League, in the same season, is perhaps further evidence of Liverpool's decision being a good one.

Using the plus-minus ratings to inform recruitment decisions (both incoming and outgoing) is an area of research we hope to do in future work. For now, we present the results of a simple experiment. When recruiting players, it is important to try to predict the future performance levels of players, not to describe past performances. As a preliminary investigation into whether this is possible using our plus-minus ratings, for players playing in consecutive seasons, we looked at the correlations of the ratings from one season to the next. This correlation was 0.35. The correlation from one season to two seasons ahead was 0.30. This is promising as the correlation reveals that performances one season, or even two seasons, ahead can be predicted by the plus-minus ratings. Of course, in practice, one would wish to use additional information to predict future performances such as the age of the player and the position the player plays in.

5.2.5 Comparing League Strengths

In this section we examine the results of adjusting for league strength in the PM ratings (Section 4.3). An interpretation of the estimated coefficients on each league is that they serve to shift a player's estimated plus-minus rating up (or down) when a player moves from a league that is of higher (lower) average strength to a league that is of lower (higher) average strength. As such, the parameter estimates can be used to compare the average strength of players in a league and how a particular player might perform if he is transferred to another league. Such insights are of great importance to clubs considering recruiting a player from another league and will help them understand the likely performance of a new player once playing in the new league. Table 7 shows the results where the parameters are again normalized to the $[0, 1]$ range (using min-max) for each of our PM ratings in. The normalized values show how far each league lies between the weakest and the strongest leagues. For example, for the xGPM rating, the Italian Serie A league lies 64% of the way between the worst league (the Portuguese Liga NOS) and the best league (Spain's La Liga). The final column of the table shows the mean league strength across the three plus-minus ratings. There is no theoretical reason to take a simple average, other than to identify which league is, on average, the strongest.

	Competition Name	PM	xGPM	xPPM	mean PM
1	England Premier League	1.00	0.67	0.97	0.88
2	Germany Bundesliga	0.92	0.32	1.00	0.75
3	Spain La Liga	0.43	1.00	0.49	0.64
4	Italy Serie A	0.61	0.64	0.66	0.64
5	Russia Premier League	0.49	0.52	0.61	0.54
6	Germany Bundesliga II	0.55	0.18	0.86	0.53
7	England Championship	0.63	0.27	0.53	0.48
8	Portugal Liga NOS	0.69	0.00	0.49	0.39
9	France Ligue 1	0.25	0.45	0.18	0.29
10	Turkey Super Lig	0.12	0.10	0.38	0.20
11	Netherlands Eredivisie	0.00	0.32	0.00	0.11

Table 7: League Ranking according to the PM rankings.

The English Premier League dominates the ranking with high scores in goals and points based PM ratings. The second strongest league appears to be the German Bundesliga. The Spanish league scores the highest in terms of expected goals PM but is slightly behind in terms of goals and expected points which may suggest that players trained in this league have a worse conversion ratio (converting opportunities to goals). We note that these league strength ratings are estimates of *average* league strength, and do not reveal the strength of the very best teams in the league. As such, although the Spanish giants Barcelona and Real Madrid would likely have very high strengths, these are tempered by the weaker strengths of other teams in the Spanish league. Surprisingly, the second divisions in Germany and England seem to perform better than the top divisions in France and Portugal. One possible explanation for this result is that teams get promoted from the second tier divisions in Germany and England and perform better in the top leagues than the players moving from Ligue 1 into these leagues. This may be a result of the players from the second tiers divisions being more familiar with the environment as they have not had to move countries to move leagues. Netherlands seem to be the 'weakest' league among the set of leagues we analysed.

In this analysis we have assumed the league strengths are static over time. Future work could investigate whether league strengths have changed over time.

6 Conclusions

The paper presents plus-minus ratings adapted to soccer. We have proposed two new versions of the plus-minus model designed to react to particular aspects of the game. Our first new plus-minus rating identifies players who change the net expected goals of a team. We have called this the expected goals plus-minus rating, xGPM. The second new plus-minus rating we propose is designed to identify players who change the results of teams by affecting the expected points of a team. We call this rating the expected points plus-minus rating, xPPM.

We have used the new ratings to look at the best players in Europe and to compare three of football's biggest stars: Ronaldo, Messi and Neymar. At the time of writing, Lionel Messi has the highest average plus-minus rating in Europe. We also used the model to examine the transfer of Philippe Coutinho from Liverpool to Barcelona. According to the plus minus ratings, Coutinho was not Liverpool's most important player, and the transfer made more sense than many fans and pundits assumed at the time. Lastly, we used the model to explore the relative strengths of the top leagues across Europe. It appears the English Premier League is slightly stronger than the German Bundesliga, followed by Spain's La Liga. Somewhat surprisingly, France Ligue 1 is rated as weaker than the second tier divisions in both England and Germany.

Plus-minus ratings in football are not likely to produce meaning ratings if calculated over a single game, or even a small number of matches since team line-ups do not change as frequently as in basketball for example. However, when estimated over a longer period of time, the ratings are meaningful. Indeed, we have begun to work with a Premier League club to use these ratings to help identify talent across Europe and aid its player acquisition department. We believe this is an indication that the ratings are informative and useful: that industry experts find the ratings useful is a promising sign. However, in certain circumstances the ratings should be treated with caution. There are four circumstances that we can think of:

- *a player plays all, or almost all, minutes*: in this circumstance, the player's plus-minus rating will converge on the plus-minus rating of his team;
- *a player plays a very small number of minutes*: in this circumstance, as a consequence of the ridge regression, the player's plus-minus rating will lie close to the global average plus-minus rating for all players;
- *a group of players always play together*: in this circumstance, the group of players will have identical plus-minus ratings, and the ratings cannot be used to differentiate between these players;
- *a player is always replaced by the same player*: in this circumstance, the plus-minus rating is informative as a comparison of the importance of these two players to their team, but it may produce an inflated rating for one of the players if his replacement is particularly weak. This can happen in the case of goalkeepers for example.

Future work may look at using these ratings as part of a forecasting model for match results. Alternatively, to aid those who make decisions regarding team lineups, one could investigate how pairings of players perform together. For example, a coach may be interested in knowing which central defensive pairing is the most effective.

Another area for development of plus-minus ratings in sport is to use alternative target metrics to measure team performance. For example, instead of using the change in expected points as the target as done in our xPPM rating here, one might look at using targets set further in the future such as the change in the probability of winning a league, or avoiding relegation for example.

For now, we hope that the objectivity of these new ratings and the seemingly 'expected' results may mean that plus-minus ratings are used more readily in the soccer industry - both by clubs, fans and the media.

Acknowledgements

We would like to thank Rick Parry for pointing us in the direction of plus-minus ratings. Further, we appreciate the comments of the three anonymous reviewers whose suggestions have certainly improved our work.

References

Bates, D. and Maechler, M. (2017). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-8.

Boshnakov, G., Kharrat, T., and McHale, I. G. (2017). A bivariate Weibull count model for forecasting association football scores. *International Journal of Forecasting*, 33(2):458–466.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3.

Brooks, J., Kerr, M., and Guttag, J. (2016). Developing a data-driven player ranking in soccer using predictive model weights. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 49–55. ACM.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754.

Chollet, F. et al. (2015). Keras. https://github.com/fchollet/keras.

Clarke, S. R. and Norman, J. M. (1995). Home ground advantage of individual clubs in English soccer. *The Statistician*, pages 509–521.

Cox, D. R. (1962). *Renewal theory*, volume 4. Methuen London.

Dixon, M. J. and 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.

Eggels, H., van Elk, R., and Pechenizkiy, M. (2016). Expected goals in soccer: Explaining match results using predictive analytics. In *The Machine Learning and Data Mining for Sports Analytics workshop*, page 16. Technische Universiteit Eindhoven.

Fearnhead, P., Taylor, B. M., et al. (2011). On estimating the ability of NBA players. *Journal of quantitative analysis in sports*, 7(3):11.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.

Glickman, M. E. (2012). Example of the Glicko-2 system. *Boston University*.

Gramacy, R. B., Jensen, S. T., and Taddy, M. (2013). Estimating player contribution in hockey with regularized logistic regression. *Journal of Quantitative Analysis in Sports*, 9(1):97–111.

Green, S. (2012). Expected goals in context. *optapro-blog*.

Herbrich, R., Minka, T., and Graepel, T. (2007). Trueskill$^{\text{TM}}$: a bayesian skill rating system. In Schölkopf, P. B., Platt, J. C., and Hoffman, T., editors, *Advances in neural information processing systems*, pages 569–576. MIT Press.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95.

Koopman, S. J. and Lit, R. (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1):167–186.

Liu, H., Hopkins, W. G., and Gómez, M.-Á. (2016). Modelling relationships between match events and match outcome in elite football. *European journal of sport science*, 16(5):516–525.

López Peña, J. and Touchette, H. (2012). A network theory analysis of football strategies. In Clanet, C., editor, *Sports Physics: Proceedings 2012 Euromech Physics of Sports Conference*, pages 517–528. École Polytechnique Paris.

López Peña, J. and Sánchez Navarro, R. (2015). Who can replace Xavi? a passing motif analysis of football players. *arXiv preprint arXiv:1506.07768*.

Lucey, P., Bialkowski, A., Monfort, M., Carr, P., and Matthews, I. (2014). Quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. In *Proc. 8th Annual MIT Sloan Sports Analytics Conference*, pages 1–9.

Macdonald, B. (2011). An improved adjusted plus-minus statistic for NHL players. In *Proceedings of the MIT Sloan Sports Analytics Conference, URL http://www. sloansportsconference. com*.

Macdonald, B. (2012a). Adjusted Plus-Minus for NHL players using ridge regression with goals, shots, Fenwick, and Corsi. *Journal of Quantitative Analysis in Sports*, 8(3):1–24.

Macdonald, B. (2012b). An expected goals model for evaluating NHL teams and players. In *Proceedings of the 2012 MIT Sloan Sports Analytics Conference, http://www. sloansportsconference. com*.

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118.

McHale, I. and Morton, A. (2011). A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting*, 27(2):619–630.

McHale, I. G., Scarf, P. A., and Folker, D. E. (2012). On the development of a soccer player performance rating system for the English Premier League. *Interfaces*, 42(4):339–351.

McHale, I. G. and Szczepański, Ł. (2014). A mixed effects model for identifying goal scoring ability of footballers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 177(2):397–417.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. *Journal of Sports Sciences*, 4(3):237–248.

Pollard, R. (2006). Worldwide regional variations in home advantage in association football. *Journal of sports sciences*, 24(3):231–240.

Pollard, R. and Pollard, G. (2005). Long-term trends in home advantage in professional team sports in north america and england (1876–2003). *Journal of Sports Sciences*, 23(4):337–350.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ridder, G., Cramer, J. S., and Hopstaken, P. (1994). Down to ten: Estimating the effect of a red card in soccer. *Journal of the American Statistical Association*, 89(427):1124–1127.

Rosenbaum, D. T. (2004). Measuring how NBA players help their teams win. *82Games.com (http://www.82games.com/comm30.htm)*, pages 4–30.

Saebo, O. D. and Hvattum, L. (2015). Evaluating the efficiency of the association football transfer market using regression based player ratings. *NIK Proceedings.*

Saebo, O. D. and Hvattum, L. (2017). Modelling the financial contribution of soccer players to their clubs. *Working paper.*

Sáez Castillo, A., Rodríguez Avi, J., and Pérez Sánchez, J. M. (2013). Expected number of goals depending on intrinsic and extrinsic factors of a football player. an application to professional Spanish football league. *European Journal of Sport Science*, 13(2):127–138.

Schuckers, M. and Curro, J. (2013). Total hockey rating (thor): A comprehensive statistical rating of national hockey league forwards and defensemen based upon all on-ice events. In *7th annual MIT sloan sports analytics conference.*

Schuckers, M. E., Lock, D. F., Wells, C., Knickerbocker, C., and Lock, R. H. (2011). National hockey league skater ratings based upon all on-ice events: An adjusted minus/plus probability (ampp) approach. *Unpublished manuscript*, 59.

Schultze, S. R. and Wellbrock, C.-M. (2018). A weighted plus/minus metric for individual soccer player performance. *Journal of Sports Analytics*, 4(2):121–131.

Sill, J. (2010). Improved NBA adjusted+/-using regularization and out-of-sample testing. In *Proceedings of the 2010 MIT Sloan Sports Analytics Conference.*

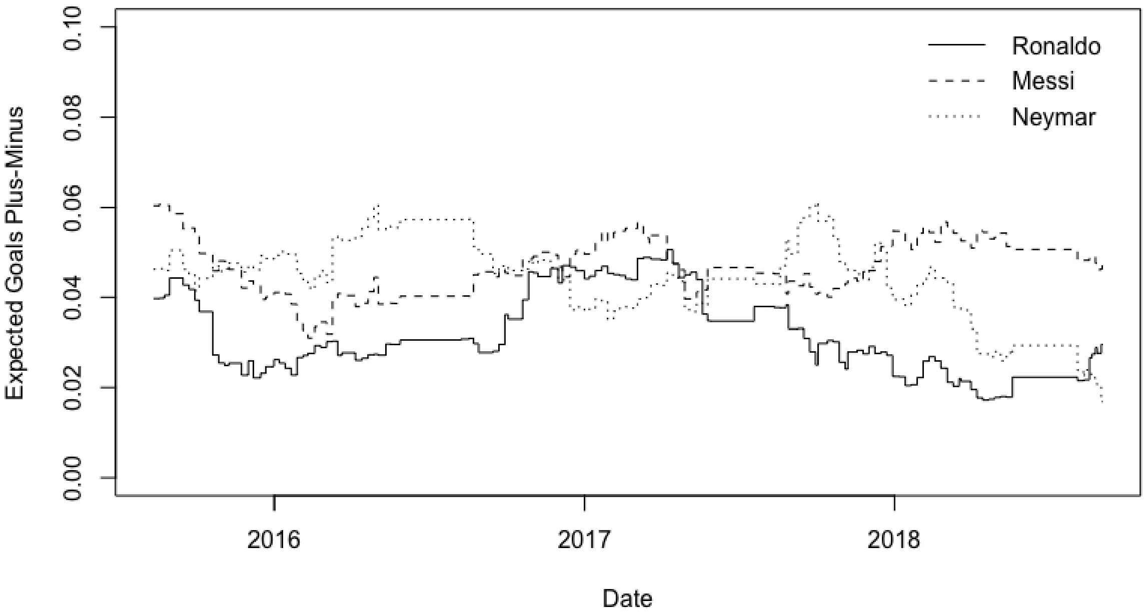
Spagnola, N. (2013). The complete plus-minus: A case study of the Columbus Blue Jackets. *Scholar Commons.*

Szczepański, Ł. and McHale, I. (2016). Beyond completion rate: evaluating the passing ability of footballers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(2):513–533.

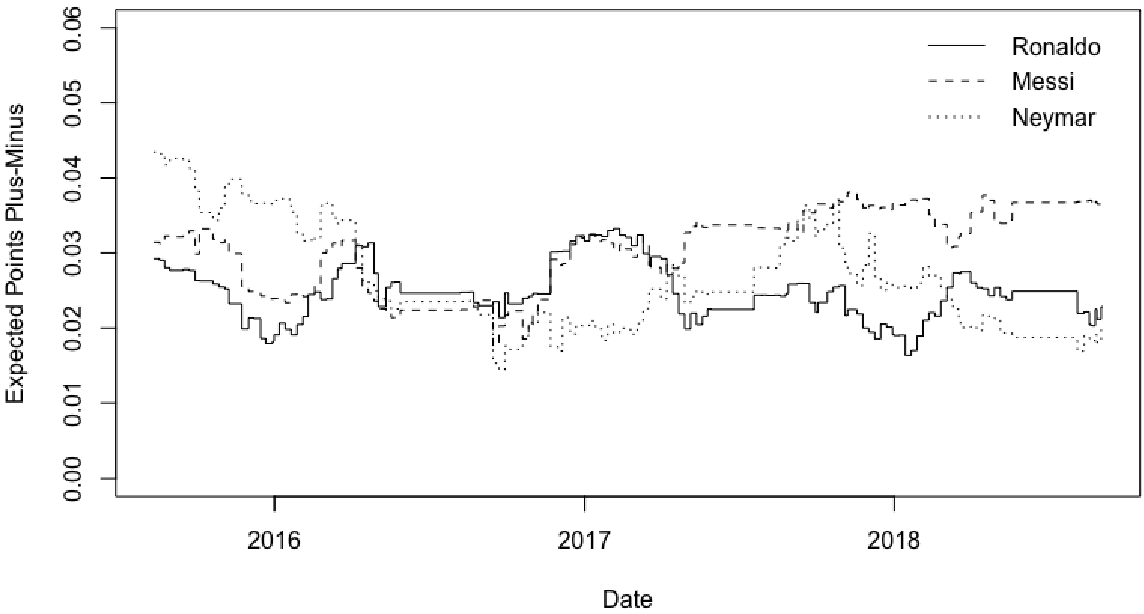
Thomas, A., Ventura, S. L., Jensen, S. T., and Ma, S. (2013). Competing process hazard function models for player ratings in ice hockey. *The Annals of Applied Statistics*, pages 1497–1524.

Volf, P. (2009). A random point process model for the score in sport matches. *IMA Journal of Management Mathematics*, 20(2):121–131.

Winston, W. L. (2012). *Mathletics: How gamblers, managers, and sports enthusiasts use mathematics in baseball, basketball, and football.* Princeton University Press.

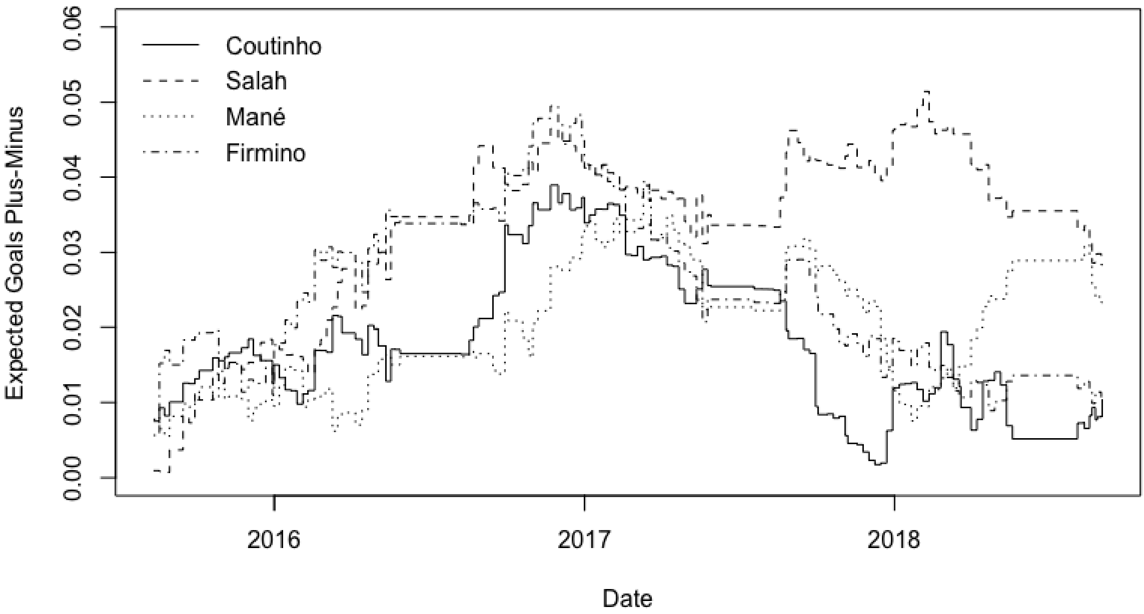


(a) Expected goals plus-minus, xGPM.

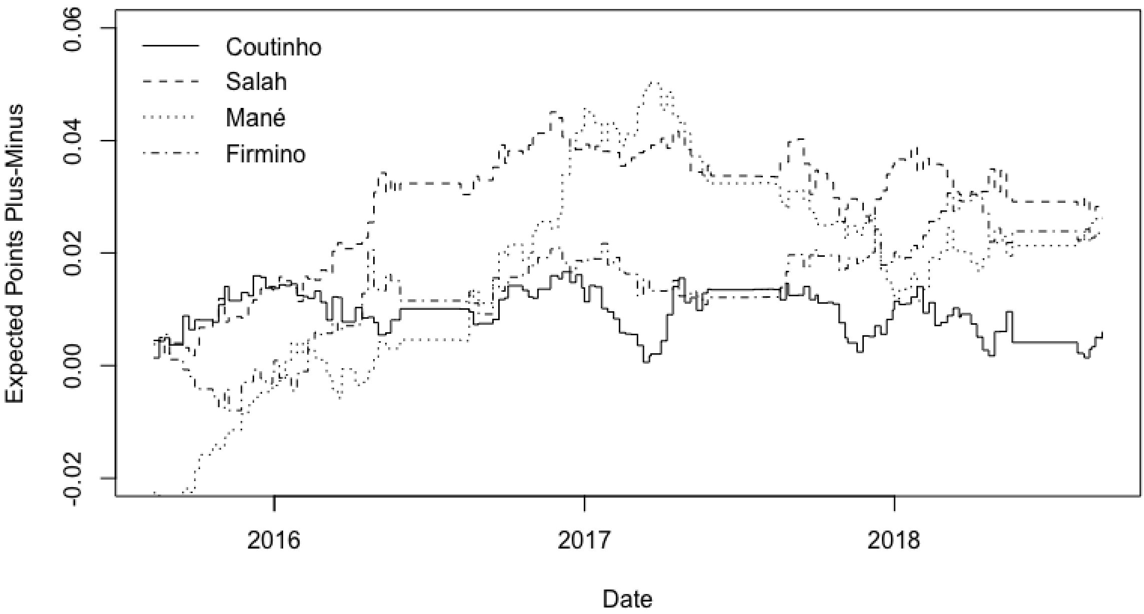


(b) Expected points plus-minus, xPPM.

Figure 1: Evolution of expected goals plus-minus, and expected points plus-minus ratings for football's "big three": Ronaldo, Messi and Neymar. The ratings are calculated using rolling 12-month windows of matches up to the date shown on the x-axis.



(a) Expected goals plus-minus, xGPM.



(b) Expected points plus-minus, xPPM.

Figure 2: Evolution of expected goals plus-minus, and expected points plus-minus ratings for Coutinho, Salah, Mane and Firmino. The ratings are calculated using rolling 12-month windows of matches up to the date shown on the x-axis.