

Ordinal versus nominal regression models and the problem of correctly predicting draws in soccer

Hvattum, L. M.

Faculty of Logistics, Molde University College, Molde, Norway

Abstract

Ordinal regression models are frequently used in academic literature to model outcomes of soccer matches, and seem to be preferred over nominal models. One reason is that, obviously, there is a natural hierarchy of outcomes, with victory being preferred to a draw and a draw being preferred to a loss. However, the often used ordinal models have an assumption of proportional odds: the influence of an independent variable on the log odds is the same for each outcome. This paper illustrates how ordinal regression models therefore fail to fully utilize independent variables that contain information about the likelihood of matches ending in a draw. However, in practice, this flaw does not seem to have a substantial effect on the predictive accuracy of an ordered logit regression model when compared to a multinomial logistic regression model.

KEYWORDS: ASSOCIATION FOOTBALL, FORECASTING, ORDERED REGRESSION

Introduction

Predicting the outcome of a soccer match not only amounts to determining the relative strength of the two teams involved. Since the match may end in a draw, factors such as the risk aversion of the teams involved may influence the probability of a decisive result. Many different methods have been proposed to predict the outcomes of soccer matches, but one of the most common methods is to model match outcomes directly using a regression model with an ordinal valued dependent variable. This paper points out how the assumption of proportional log odds in the common ordinal regression models prevent the exploitation of independent variables conveying information regarding the likelihood of draws.

Academic literature has only to a small extent considered the peculiarities surrounding the prediction of draws. Pope and Peel (1989) found that experts lacked the ability to forecast draws. Cain, Law, and Peel (2000) noted that bookmakers' odds for draws were approximately constant, and had no significant explanatory power when predicting the score of a match. They also concluded that draws did not offer profitable betting opportunities on the whole. Dixon and Pope (2004) noted that bookmakers' draw odds were almost constant, whereas a Poisson model provided greater dispersion. Van Calster, Smiths, and Van Huffel (2008) studied the relationship between selected performance indicators and the occurrence of scoreless draws, without focusing on the prediction of draws. They found that matches between average and bad-to-average teams more often resulted in scoreless draws, in particular if matches with these teams in general had few goals. Furthermore, they found indications that there is an increased rate of scoreless draws for matches with few spectators.

Ordered logit regression (OLR) and ordered probit regression (OPR) have been used in several publications to predict match results in soccer. Kuypers (2000) argued that ordinal regression makes sense, since match outcomes are naturally ordered, and therefore did not test any non-ordered regression models. Ordered regression models have later been used for prediction of match results by Koning (2000), Forrest and Simmons (2000), Dobson and Goddard (2001, 2003, 2008), Audas, Dobson, and Goddard (2002), Goddard and Asimakopulos (2004), Goddard (2005), Forrest, Goddard, and Simmons (2005), Graham and Stott (2008), Hvattum and Arntzen (2010), and Hvattum (2015). Some research has relied on the use of ordered regression models as a means to normalize bookmakers' odds, such as by Štrumbelj (2014, 2016).

The use of multinomial logit regression (MLR) models as an alternative to OLR or OPR models has been limited. Nyberg (2014) used an MLR model to evaluate market efficiency, by using as independent variables the averages of normalized implied probabilities from bookmakers' odds. Vlastakis, Dotsis, and Markellos (2009) similarly used an MLR model with bookmaker odds as independent variables, trying to obtain profitable forecasts when combined with other models. Searching in academic journals, no positive identification has been made for work where multinomial models have been used with independent variables not based on bookmaker odds, where the problems of predicting draws using ordered models have been discussed in detail, or where an empirical comparison has been made between multinomial and ordered models for the prediction of soccer matches.

This work intends to demonstrate that the common use of standard ordered regression is potentially inappropriate for modelling match result outcomes, and to show that MLR avoids the problem identified for OLR and OPR. While it is not surprising that the ordered regression models are unsuitable for predicting draws, from a theoretical perspective, this paper quantifies how big the potential gap can be. Furthermore, the paper evaluates both existing independent variables from the literature as well as new independent variables, in terms of being able to improve predictions of drawn soccer matches.

The remainder of this paper is structured as follows. In the next section, the specification of an OLR model and an MLR model for the prediction of soccer results are given. The following section describes the experimental setup used in this paper. The penultimate section then provides results from the experiments. This includes testing whether MLR is indeed superior to OLR in terms of being able to utilize information pertaining specifically to the probability of draws occurring. Furthermore, the tests examine a range of existing and new independent variables, trying to establish whether any of them contain information that is important when determining the probability of draws. Concluding remarks are presented in the last section.

Background

Two different types of regression models, OLR and MLR, are evaluated in this work. For both model types, the idea is to estimate a model based on historical data that can be used to predict the outcomes of future soccer matches. The outcome is encoded as a dependent variable (response variable), y , which is taken to be ordinal in OLR and nominal in MRL. Several measurements are available before a match is played. These are encoded as independent variables (predictor variables), x , which are typically real-valued or binary. The models then aim to describe the relationship between the independent variables and the possible values of the dependent variable, such that the probability of a given value for y is a function of x and a number of parameters that must be estimated.

The following description of the standard ordinal regression models used for the prediction of soccer match outcomes is based on (Dobson and Goddard, 2001) and (Greene, 2012). Assume that M historical matches have been recorded, and that the result of a match is denoted by y . The potential results, y , of a soccer match can be ordered from 1 to $K = 3$, with $y = 1$ representing a home win, $y = 2$ representing a draw, and $y = 3$ representing an away win.

Assume that there are V independent variables, and that a value x_i for each independent variable i has been calculated prior to each match. It is of interest to find the probability $\pi_j(x)$ of each potential result $j = 1, \dots, K$ as a function of the independent variables $x = (x_1, \dots, x_V)^T$. This is done by introducing one parameter β_i for each independent variable i , and parameters θ_j for $j = 1, \dots, K - 1$. To simplify the notation, write $\beta = (\beta_1, \dots, \beta_V)$, and let $\theta_0 = \infty$ and $\theta_K = -\infty$.

Furthermore, denote by F the cumulative probability distribution that describes the error term in an unobserved process that links the independent variables and the dependent variable. When F is chosen as the logistic distribution, the OLR model is formed, whereas choosing a standard normal distribution gives an OPR model. The logistic distribution, which will be used in this paper, is given by

$$F(z) = \frac{1}{1 + e^{-z}}$$

The conditional probabilities of each result can now be stated, for each potential result j , as

$$\pi_j(x) = F(-\theta_j - \beta x) - F(-\theta_{j-1} - \beta x)$$

which can be written explicitly for soccer matches as

$$\pi_1(x) = F(-\theta_1 - \beta x), \pi_2(x) = F(-\theta_2 - \beta x) - \pi_1(x), \pi_3(x) = 1 - \pi_1(x) - \pi_2(x).$$

Figure 1 illustrates the OLR model, showing how independent variables x are turned into

probabilities. While this class of models has been frequently used to model and predict outcomes of soccer matches, it is evident that independent variables that only influence the likelihood of draws, without influencing the relative proportion of home wins to away wins, are of limited utility: while the probability of draws is not constant for varying x , their distribution has to follow the assumption of proportional log odds. Figure 2 illustrates the shape of estimated probabilities using the difference in Elo ratings, $E_i - E_j$, as the only independent variable. The model was estimated based on results of 36,648 matches from the four divisions in the English league system between the 1997/1998 and the 2014/2015 season, using the method of maximum likelihood.

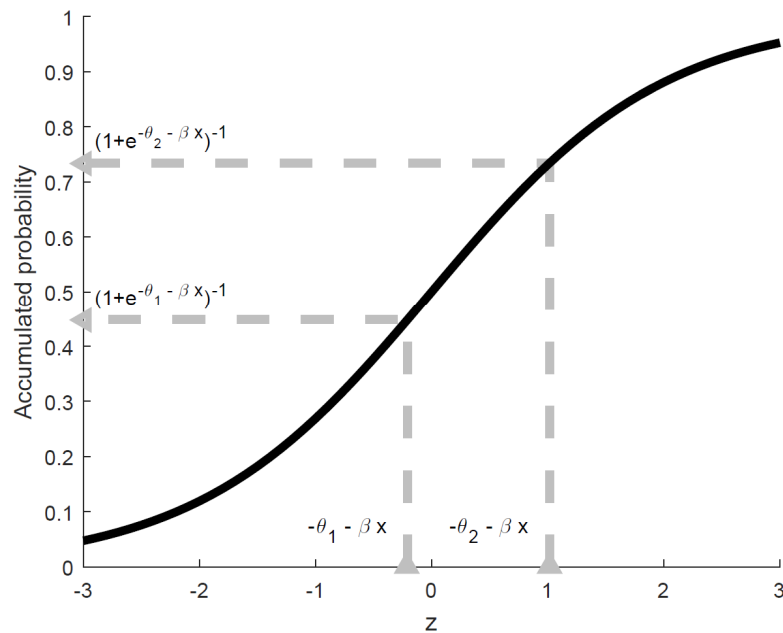


Figure 1. Ordered logit regression does not facilitate independent variables that only influence the relative likelihood of a draw result.

In this work, the OLR model is compared with the MLR model (Greene, 2012)[Chapter 18]. The MLR model has additional parameters that allows the probability of the middle result, the draw, to be influenced without necessarily changing the ratio of probabilities for the extreme results. For the ease of presentation, define $x_0 = 1$ for all matches. This is to avoid introducing a separate parameter θ for the constant term in the regression, by directly including the constant term in the linear combination of the independent variables. Hence, in this case we have $x = (x_0, x_1, \dots, x_V)^T$. Parameters α_{ji} are introduced for each independent variable i , including the one representing the constant term, for each possible result j , with $\alpha_j = (\alpha_{j0}, \alpha_{j1}, \dots, \alpha_{jV})$. Conditional probabilities of each result can now be written, for each potential result $j = 1, \dots, K$ as

$$\pi_j(x) = \frac{e^{\alpha_j x}}{\sum_{k=1, \dots, K} e^{\alpha_k x}}$$

where one potential result k is chosen, setting $\alpha_{ki} = 0$ for all independent variables i to ensure model identifiability. For soccer matches, taking $k = 3$ as the reference result, this can be written explicitly as

$$\pi_1(x) = \frac{e^{\alpha_1 x}}{1 + e^{\alpha_1 x} + e^{\alpha_2 x}}, \pi_2(x) = \frac{e^{\alpha_2 x}}{1 + e^{\alpha_1 x} + e^{\alpha_2 x}}, \pi_3(x) = \frac{1}{1 + e^{\alpha_1 x} + e^{\alpha_2 x}}.$$

The number of parameters in an OLR model is $V + K - 1$, and the number of parameters in an MLR model is $(V + 1)(K - 1)$. Hence, in the context of predicting soccer match outcomes, with $K = 3$, using MLR instead of OLR roughly amounts to doubling the number of parameters. In models with many independent variables, this increases the risk of overfitting, resulting in worse predictions for future matches. In addition, the assessment of a single independent variable in the multinomial regression model is less straightforward. However, this may be an acceptable sacrifice in the pursuit of better predictions of drawn matches. For both types of regression models, maximum likelihood estimation is used to determine the regression coefficients.

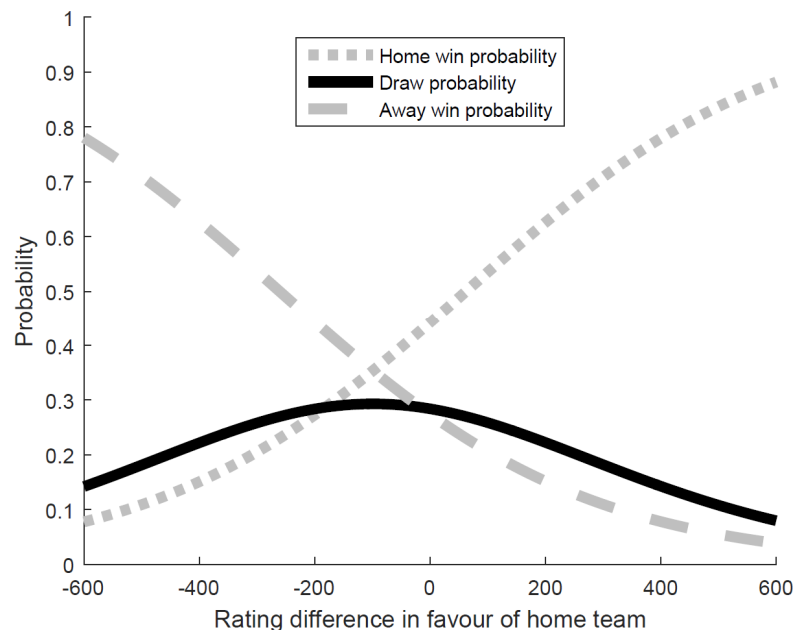


Figure 2. Probabilities of match outcomes as a function of Elo rating differences, estimated using ordered logit regression.

Experimental setup

The experiments presented in this paper can be divided in two parts. In the first part, the OLR and the MLR models are compared in terms of their ability to incorporate independent variables that contain information regarding the likelihood of a matches ending in a draw. This is performed by generating an artificial independent variable that contains a controlled amount of information regarding the likelihood of a draw. In the second part, a wide range of existing and novel independent variables are investigated, using both the OLR and the MLR models, trying to identify variables that can be used to improve the models' ability to predict drawn matches.

To evaluate predictions, two directions are pursued. First, the predictions are compared by calculating their quadratic loss (Witten and Frank, 2005): taking $\pi_j(x)$ as the estimated probability of outcome j for a match with given values for the independent variables x , and taking r_j as a binary indicator equal to 1 if the match ended with outcome j and 0 otherwise, the quadratic loss of a single match is $\sum_{j \in \{1,2,3\}} (\pi_j(x) - r_j)^2$. The quadratic loss is then averaged over all predicted matches, and the resulting averages from using different prediction methods can be compared.

Second, predictions are evaluated by considering the return on bets placed. A bookmaker provides decimal odds o_j for each outcome j , such that a successful bet gives a profit of $(o_j - 1)$ times the stake and an unsuccessful bet gives a loss equal to the stake. A prediction method is

assumed to place a unit bet at an odds of o_j for all outcomes where $o_j\pi_j > 1$. The average return on investment is then calculated over all predicted matches. Odds data are available from up to 13 different bookmakers, from which the best available odds (the highest value of o_j) is considered for each outcome of each match.

Data

The data used in the experiments are publicly available. Match results and odds data were downloaded from <http://www.football-data.co.uk>. The main experiments are based on data from English league matches. Matches from four divisions, currently named the Premier League, the Championship, League One, and League Two, are included, starting from the 1997/1998 season up to and including the 2014/2015 season. The total number of matches in this data set is 36,648. The two first seasons are used only for initial calibration required for the calculation of some of the independent variables studied, whereas seasons 1999/2000 through 2009/2010 are used solely as historical observations. The remaining five seasons are used for making forward predictions: the regression coefficients are estimated based on observations up to the start of a given season, and then the predictions for that season are evaluated.

In additional experiments, to verify the findings from the tests using English league matches, additional leagues are also considered. In particular, for the same seasons as for the English league, matches from the two highest divisions of the French, the Italian, and the Spanish league systems are considered. Considering all 18 seasons from 1997/1998 to 2014/2015, these additional data sets contain 13,392 matches, 14,236 matches, and 15,156 matches, respectively. The same methods to evaluate the regression models are used throughout.

Independent variables

Table 1 lists sets of independent variables that are known from existing literature and that have been used when predicting soccer match outcomes. Different independent variables are considered in the different parts of the experiments, but one variable is always included: the difference in Elo rating between the two teams, E_{ij} as included in the set V_1 . The Elo rating of a team is updated dynamically after each match. Let ξ_0^H and ξ_0^A be the rating of the home team and the away team before a match. Define the actual outcome of a match, from the perspective of the home team, as $\mu^H = 1$ for a home win, $\mu^H = 0.5$ for a draw, and $\mu^H = 0$ for an away win. According to the Elo-ratings before the match, an expected outcome for the home team is calculated as $\gamma^H = \left(1 + c^{(\xi_0^H - \xi_0^A)/d}\right)^{-1}$, and for the away team as $\gamma^A = 1 - \gamma^H$. After the match, new Elo-ratings are calculated as $\xi_1^H = \xi_0^H + k(\mu^H - \gamma^H)$ and $\xi_1^A = \xi_0^A + k(\gamma^H - \mu^H)$, so that the rating points gained by one team is equal to the points lost by the other team. The calculations involve three parameters, where $c = 10$, $d = 400$, and k is equal to $10(1 + |g|)$, where $|g|$ is the absolute value of the goal difference in the match.

The values of the parameters were determined in a computational study reported by Hvattum and Arntzen (2010). Initial Elo ratings are obtained in a bootstrapping process: all teams are first given the same ratings. Two seasons of data are then used to update the ratings. If the ratings obtained at the end of the two seasons are similar enough to the ratings taken at the start, the process is halted. Otherwise, the ratings obtained at the end of the two seasons are taken as new ratings at the start and the process is repeated.

For a more detailed description of the variables previously used in the literature, see (Goddard, 2005) and (Hvattum, 2015). Regarding the variables in sets V_8 and V_9 , the definition in (Goddard and Asimakopoulos, 2004) is followed: a match is significant if it is still possible, before the

match is played, for the team in question to either win the league, to be promoted to a higher division, or to be relegated to a lower division, when assuming that all other teams obtains on average one point from their remaining matches.

Table 2 lists sets of independent variables that have not been found in existing publications using regression to predict soccer matches. The first of these, in N_1 , is only used in the first part of the experiment: it is an *ex post* calculated value that is specifically designed to contain information only about whether or not a given match ended in a draw. The binary variable Z is only useful to illustrate the difference between OLR and MLR in terms of being able to incorporate information regarding draws, and cannot be used for making *ex ante* predictions. The level of information included in the variable is represented by a parameter ρ , such that when $\rho = 0.5$ the variable consists of random noise, being arbitrarily assigned either the value 0 or the value 1. When $\rho = 1$ (0), the variable takes the value 1 (0) if the match ended in a draw and takes the value 0 (1) otherwise. For intermediate values of ρ , the value of the variable is randomly chosen with a bias that depends on the value of ρ .

Variables in $N_2 - N_4$ are calculated on a similar basis as V_1 , and are included to check whether the probability of draws differs systematically for matches with large favorites or for matches with better teams (higher average ratings). The variable in N_5 supplements V_8 and V_9 , to check whether draws are more likely in unimportant matches. The calculation of the variable in N_6 is based on previous encounters between the two teams involved, and equals the weighted average number of goals scored in past matches between the two teams. Matches with team i as the home team are weighted by 0.8, and matches with team j as the home team are weighted by 0.2. If no previous matches between the teams are available, the variable equals the average number of goals for all the matches involving the two teams. This variable is included to see if there is a connection between draws and the amount of goals in past matches between two particular teams.

Set N_7 consists of four new variables, two for the home team and two for the away team, representing the average number of goals scored and conceded for both teams. This means that N_7 is somewhat similar to $V_2 - V_7$, but on a different time scale and without splitting into home and away performance. The variable in N_8 represents the average of the drawing rates for the two teams involved in a match, calculated based on matches in the current and the previous season. This variable is included to see if the historic drawing rates of teams influence the probability of draws in a given match. In an attempt to classify local derbies, N_9 is a binary variable equal to 1 if the distance between two teams' home grounds is less than 15 kilometres beeline. This may give additional information compared to just using the natural logarithm of the geographical distance, as in V_{10} , to check if the probability of draws is different in derbies from other matches.

The variable in N_{10} measures whether the match is in an early or a late phase of the season, coded as a binary indicator variable being 1 if the teams have played less than three matches so far this season or if the teams have less than three matches left to play this season. The variable is included to see if there are more draws in the beginning of seasons, when the league is not yet settled, or in the end of the season, when league positions are more or less decided. The variable in the set N_{11} is used to indicate whether a match is played in the weekend, including Friday, or on any day from Monday to Thursday. This is based on an indication that the level of home advantage may depend on the day of the week in which a match is played (Krumer and Lechner, 2016).

Table 1. Selection of independent variables used in extant literature, including (Goddard, 2005) and (Hvattum, 2015).

Set	Description
V_1	E_{ij} is the difference in Elo rating prior to the match between the home team i and the away team j .
V_2	F_{idys} equals f_{idys}/n_{iy} where f_{idys} is the total number of goals scored by team i in matches played 0–12 months ($y = 0$) or 12–24 months ($y = 1$) before the current match; within the current season ($s = 0$), the previous season ($s = 1$), or two season ago ($s = 2$); in a division $d \in \{-2, -1, 0, 1, 2\}$ steps from the team's current division; and n_{iy} is the number of matches considered when counting the number of goals.
V_3	A_{idys} equals a_{idys}/n_{iy} where a_{idys} is the total number of goals conceded by team i defined for (d, y, s) as above.
V_4	S_{im}^H is the number of goals scored in the m th most recent home match by team i .
V_5	S_{im}^A is the number of goals scored in the m th most recent away match by team i .
V_6	C_{im}^H is the number of goals conceded in the m th most recent home match by team i .
V_7	C_{im}^A is the number of goals conceded in the m th most recent away match by team i .
V_8	I_{ij}^H equals 1 if the match is important for championship, promotion, or relegation issues for home team i but not for away team j , and 0 otherwise.
V_9	I_{ij}^A equals 1 if the match is important for championship, promotion, or relegation issues for away team j but not for home team i , and 0 otherwise.
V_{10}	D_{ij} is the natural logarithm of the geographical distance between the home grounds of team i and team j .

Table 2. Selection of novel independent variables introduced in this paper.

Set	Description
N_1	Z is randomly chosen, <i>ex post</i> , to be 1 with probability of ρ if the match ended in a draw and with a probability of $(1 - \rho)$ if the match did not end in a draw, and 0 otherwise.
N_2	E_{ij}^2 is the square of the difference in Elo rating prior to the match between the home team i and the away team j .
N_3	E_{ij}^{AVG} is the average Elo rating of the home team i and the away team j prior to the match.
N_4	E_{ij}^{AVG2} is the square of the average Elo rating of the home team i and the away team j prior to the match.
N_5	I_{ij} equals 1 if the match is not important for either the home team i or the away team j , and 0 otherwise.
N_6	M_{ij} equals the weighted average number of goals scored in previous matches between teams i and j , with a higher weight for matches where i was the home team.
N_7	G_i^{HS} , G_i^{HC} , G_j^{AS} , and G_j^{AC} are the average number of goals scored or conceded for the home team and away team over the current and previous season.
N_8	W_{ij} equals the average number of draws in matches involving teams i or j over the current and previous season.
N_9	D_{ij}^B is a binary variable equal to 1 if the distance between the home grounds of teams i and j is less than 15 kilometres, and 0 otherwise.
N_{10}	R_{ij} is a binary variable equal to 1 if the match is played in an early or a late round, that is one of the first three or one of the last three matches for the teams in the current season, and 0 otherwise.
N_{11}	W_{ij} equals 1 if the match is played on a Friday, Saturday, or a Sunday, and 0 otherwise.

Results

This section presents the results from the two parts of the experiment.

Using ex post information to evaluate OLR and MLR

The first part is designed to highlight the practical difference between OLR and MLR when including independent variables that contain information about the likelihood of draws. To this end, the independent variables in V_1 and N_1 are combined, with the parameter ρ indicating the amount of information regarding draws that is contained in the variable of N_1 referred to as Z . The evaluation is performed using the five seasons from 2010/2011 to 2014/2015, which contain 10,180 matches. Some matches are not included in the evaluation, as the Elo ratings could not be calculated for all teams, leaving 9,730 matches for which the calculations are performed.

The expected return on random betting (selecting a match and an outcome using a uniform probability distribution) is 0.989 across the 9,730 matches included in the test. The basic regression models with only the Elo difference included as an independent variable performs worse than this, with 0.967 for OLR and 0.962 for MLR, suggesting that the odds are possibly biased against the type of information included in Elo ratings. As there are some matches with arbitrage opportunities, a betting rule based on the Shin normalization of the best available odds (Štrumbelj, 2016) can also be used, resulting in a return on investment of 0.980.

Figure 3 illustrates how the quadratic loss changes for predictions using OLR and MLR when the information content of Z changes with ρ . With $\rho = 0.5$, Z contains only noise, and OLR and MLR have similar performance, both slightly worse than the probabilities obtained by using the Shin normalization of the best available betting odds. However, as the information content increases, either by increasing or decreasing the value of ρ , MLR starts to perform noticeably better than OLR. The same trend can be observed in Figure 4, where the return on investment is shown for different values of ρ . It can be seen though, that even OLR obtains positive returns on betting when the information content in Z is large enough: this is presumably as the variable is also indirectly related to the probability of a home win.

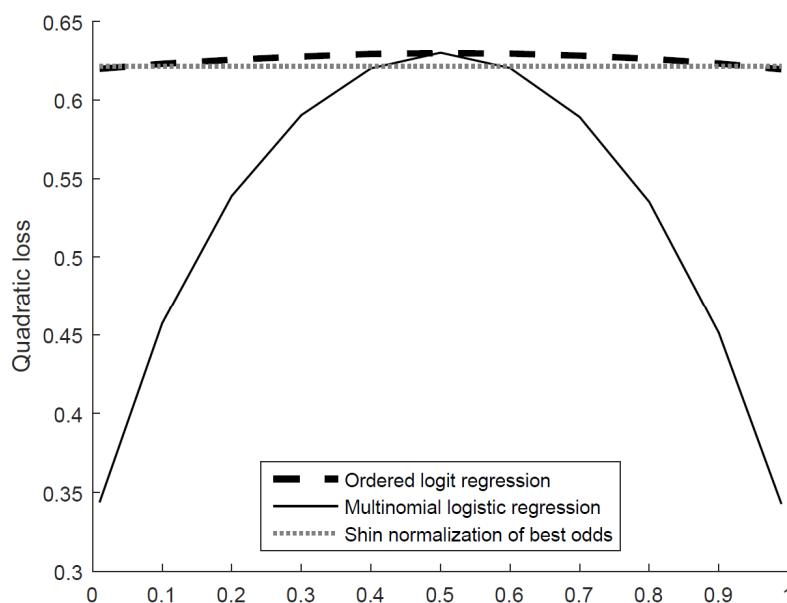


Figure 3. Quadratic loss when using OLR and MLR with independent variable Z calculated *ex post* with information about the occurrences of draws, together with E_{ij} .

Figure 5 further illustrates the situation when ρ is close to 0.5: For each regression model we consider whether the return on draw bets is higher than 1, and calculate the P-value from a one-sided one-sample t-test. Thus, the plotted P-values correspond to the probability that a higher return on bets, considering only draw bets, would be observed, given that the expected return on draw bets is 1. The figure illustrates that these P-values decrease much faster for MLR than for OLR when information regarding draws is available from the independent variables. Furthermore, the difference between MLR and OLR becomes evident even for smaller deviations of ρ from 0.5 when considering P-values from a one-sided two-sample unequal variance t-test. The P-values here signify the probability of observing a relatively higher return on draw bets for MLR given that the expected return on draw bets is equal for OLR and MLR.

The first part of the experiments has illustrated how MLR is better than OLR at exploiting information regarding the occurrence of draws. This difference can be seen by observing lower values for the quadratic loss function or by observing a higher return on investment when considering bets placed when model probabilities are higher than the inverse of the best odds available. In addition, calculating P-values as in Figure 5 may give an even clearer indication as to whether a given set of independent variables actually contain information regarding draws that is better exploited by MLR than by OLR. In the following, each of these measures will be used when evaluating the difference between MLR and OLR for both a set independent variables that have been previously presented and used in the literature and a set of new independent variables.

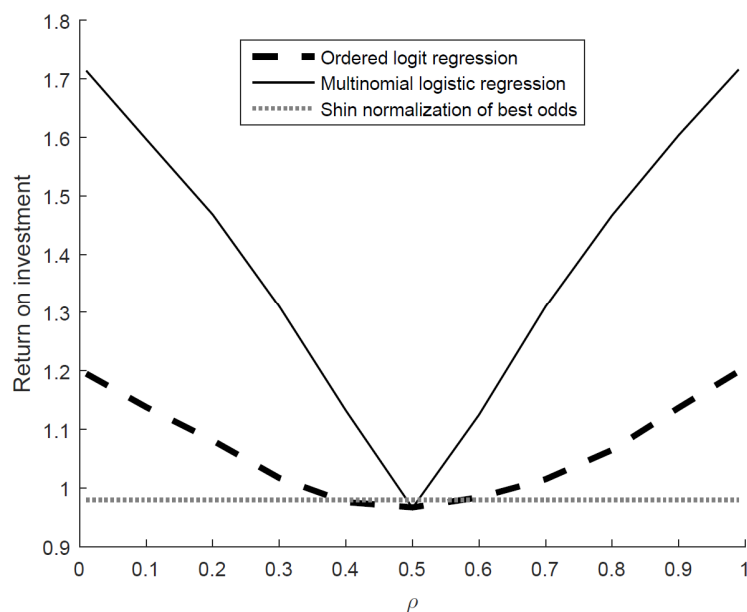


Figure 4. Return on investment when using OLR and MLR with independent variable Z calculated *ex post* with information about the occurrences of draws, together with E_{ij} .

Using *ex ante* information to evaluate OLR and MLR

Having illustrated the extent to which MLR is able to use information regarding the likelihood of draws better than OLR, the second part of the experiments examines the performance of MLR and OLR on a variety of independent variables as summarized in Table 1 and Table 2. Some of the matches from 2010/2011 to 2014/2015 cannot be considered, as not all of the independent variables in sets V_2-V_7 and N_1-N_{11} can be calculated, in particular for teams being newly promoted to League Two. This leaves a total of 9,308 matches that is the basis for the following comparisons.

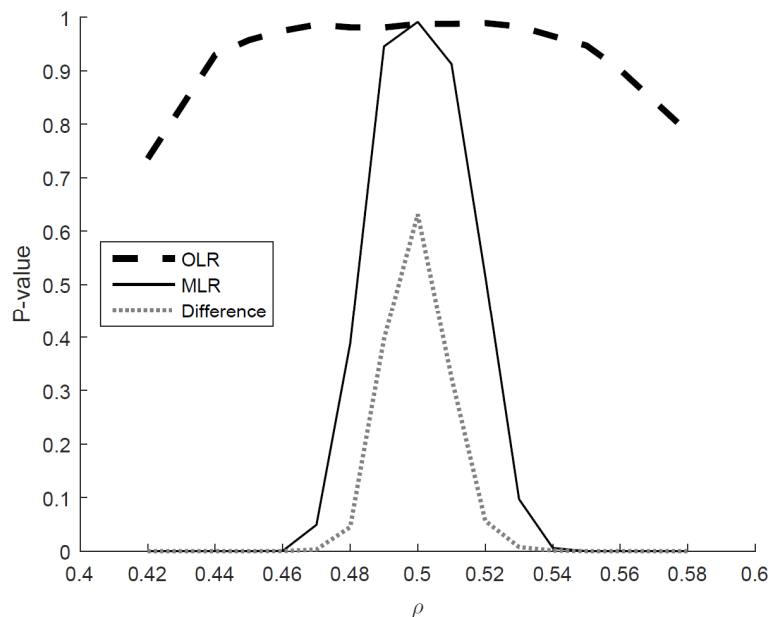


Figure 5. P-values for comparing the return on draw bets based on OLR and MLR with independent variable Z calculated *ex post* with information about the occurrences of draws, together with E_{ij} .

Table 3 summarizes the results, presenting the quadratic loss for predictions, the return on investments (ROI) on all bets, the ROI on draw bets, and the same P-values as calculated in Figure 5. Numbers in bold indicate that the results are better than when just using the Elo rating difference (V_1) as the sole independent variable. Using Shin normalization of odds yields predictions resulting in an average quadratic loss of 0.6205, which is far better than the loss of the regression models.

The first tests only consider independent variables already known from the literature. The independent variables in sets V_2 and V_3 are tested together, as is the case for variables in sets $V_4 - V_7$, as well as V_8 and V_9 . Most combinations of independent variables improve the performance when it comes to ROI, in particular when considering only draw bets, but in no case is MLR better than OLR with statistical significance. Nevertheless, except for the combination with V_8 and V_9 , the two-sample t-test comparing MLR and OLR have lower P-values than the base case of only using V_1 . This means that some of the other independent variables may contain some information that is relevant for predicting draws, but not enough to make a statistically significant difference. On the other hand, the combination with V_8 and V_9 is the only one where the quadratic loss of MLR is improved. From this it seems that the independent variables hitherto used in academic studies that have applied ordered regression on soccer match prediction do not provide much information that is relevant for the prediction of draws.

The next tests consider a collection of novel independent variables, as summarized in Table 2. The situation is similar as for the existing variables: for many of the new variables there is an improved ROI for draw bets and an ROI for bets in general, and also a relative improvement of the ROI for draw bets using the MLR models compared to the OLR models. However, if any of the variables contained information relevant for prediction of draws, it would be expected that the ROI on draw bets for MLR would be significantly better than for OLR. This is not the case, as for most of the combinations of independent variables, OLR has an equal or better return.

Finally, two new combinations of independent variables are considered, where existing and new variables are combined. Combination C_1 consists of the sets of variables that obtained improved P-values from the two-sample t-test comparing MLR and OLR. That is, C_1 comprises $V_1 - V_7$,

V_{10} , N_2 , and $N_5 - N_{11}$, for a total of 114 independent variables. The second combination C_2 extends C_1 by also including V_8 and V_9 which were seen to improve the quadratic loss of MLR in the tests reported in Table 3. Interestingly, this larger combination shows the best performance so far in terms of ROI on draw bets for the MLR model, whereas the OLR model performs in line with previous observations. Again, though, the difference between MLR and OLR is not statistically significant. However, the ROI of draw bets for the MLR model with C_2 is better than the ROI of draw bets from the corresponding model with only V_1 , with a P-value of 0.03 from a one-sided two-sample t-test.

In a set of additional tests, data sets from the French, Spanish, and Italian leagues are considered. The data sets span the same years as the English data set, but have fewer divisions included. The number of matches eligible for calculating predictions in the test seasons from 2010/2011 to 2014/2015 is 2,890 for the French league, 2,935 for the Spanish league, and 2,910 for the Italian league. Since only two divisions are available, some of the independent variables included in V_2 and V_3 must be omitted. This pertains to those combinations of F_{idys} and A_{idys} that involve divisions $d = 2$ or $d = -2$ steps from a team's current division.

With some slight differences, tests using V_1 , C_1 , and C_2 yields the same results as for the English league. For example, for the French league, all six models evaluated provide a positive return on draw bets. The best ROI on draw bets is for the MLR models including many independent variables. The P-values from comparisons of MLR and OLR are in line with the observations from the English league, and thus not statistically significant even though the direction of the difference in performance indicates that MLR is better than OLR.

Having examined four separate data sets, from four different league systems, an overall trend is quite clear: when including a large set of independent variables, several of which were designed to probe for predictive power with respect to forecasting draws, the performance of MLR seems better than OLR when looking at the return on draw bets. However, even though the trend is clear, a basic one-sided two-sample t-test is unable to determine a statistically significant difference in the performance. A final test was therefore conducted in which the data sets were joined together, considering at once 79,432 matches out of which 18,043 are used to calculate quadratic loss and return on bets. The results indicate that the return on bets is not in general improved by combining several different leagues. A possible partial explanation is that the base rates for draws are different for each league, and that there are no variables included to adjust for this. For example, 30.6 % of the matches ended in a draw in the French data set, whereas only 27.3 % of the matches ended in a draw in the English data set. To facilitate the fact that different leagues have different distributions of results, binary indicator variables are also added to the mix of independent variables. This seems to improve the return on bets in general, and the ROI on draw bets for MLR improves more than for OLR. The test again fails to indicate a statistically significant difference between OLR and MLR, suggesting that there is indeed very little information in the 119 different variables included that can help to improve draw predictions.

Table 3. Main results.

V	$ V $	Quadratic loss		ROI (#bets)		ROI draws (#bets)		P-values						
		OLR	MLR	OLR	MLR	OLR	MLR	OLR	MLR	Diff				
V_1	1	0.629	0.629	0.967	(12831)	0.962	(11966)	0.936	(4106)	0.910	(3337)	0.99	1.00	0.74
$V_1 - V_3$	49	0.628	0.629	0.985	(12849)	0.988	(12636)	0.959	(4201)	0.961	(3337)	0.95	0.93	0.48
$V_1, V_4 - V_7$	53	0.628	0.629	0.967	(12934)	0.969	(12770)	0.949	(4262)	0.966	(4004)	0.98	0.90	0.32
V_1, V_8, V_9	3	0.628	0.628	0.976	(12895)	0.968	(12176)	0.951	(4175)	0.925	(4054)	0.97	1.00	0.75
V_1, V_{10}	2	0.629	0.629	0.973	(12825)	0.968	(11996)	0.954	(4098)	0.938	(3541)	0.96	0.98	0.66
V_1, N_2	2	0.629	0.629	0.967	(12886)	0.959	(12145)	0.939	(4135)	0.925	(3342)	0.99	1.00	0.64
V_1, N_3	2	0.628	0.629	0.970	(12828)	0.957	(11936)	0.956	(4115)	0.925	(3504)	0.96	0.99	0.78
V_1, N_4	2	0.628	0.629	0.969	(12822)	0.957	(12073)	0.942	(4100)	0.908	(3317)	0.99	1.00	0.81
V_1, N_5	2	0.629	0.629	0.966	(12854)	0.970	(11877)	0.935	(4120)	0.938	(3425)	0.99	0.98	0.47
V_1, N_6	2	0.629	0.629	0.969	(12853)	0.967	(12324)	0.935	(4120)	0.928	(3228)	0.99	1.00	0.58
V_1, N_7	5	0.628	0.628	0.973	(12799)	0.972	(11859)	0.948	(4105)	0.942	(3797)	0.98	0.97	0.56
V_1, N_8	2	0.629	0.629	0.968	(12820)	0.963	(11899)	0.935	(4094)	0.914	(3225)	0.99	1.00	0.71
V_1, N_9	2	0.629	0.629	0.969	(12836)	0.970	(11817)	0.938	(4107)	0.934	(3277)	0.99	0.99	0.55
V_1, N_{10}	3	0.629	0.629	0.967	(12793)	0.970	(11912)	0.941	(4086)	0.943	(3173)	0.99	0.97	0.48
V_1, N_{11}	2	0.629	0.629	0.969	(12833)	0.968	(11915)	0.939	(4100)	0.923	(3256)	0.99	1.00	0.65
C_1	114	0.629	0.630	0.982	(13188)	0.984	(12923)	0.955	(4461)	0.981	(3277)	0.97	0.77	0.23
C_2	116	0.629	0.630	0.985	(13226)	0.987	(12931)	0.953	(4515)	0.985	(4198)	0.97	0.73	0.18

Concluding remarks

Ordered regression is commonly used in academic literature to determine probabilities for the outcomes of soccer matches. However, this use is only justifiable when the influence of an independent variable on the log odds is the same for each outcome. For the prediction of soccer matches, this implies that independent variables only containing information about the relative propensity of drawn results cannot be fully exploited. Hence, standard ordered regression models cannot utilize variables characterizing situations where the risk aversion of teams (the probability of a draw) vary but their relative strength (the ratio of probabilities for home wins and draw wins) do not.

This paper showed that multinomial logit regression (MLR) is a viable alternative to ordered logit regression (OLR) when independent variables related to the probability of draws are present. A significant effort was then expended to evaluate the ability of independent variables to improve predictions of draws in soccer, including both existing variables from the literature as well as a range of novel variables. The results were uniform: no observations made indicated a statistically significant difference in the quality of predictions made using MLR or OLR. Despite this, most of the tests did show that MLR led to better returns on draw bets than OLR when including many different independent variables.

It therefore seems that the use of ordered regression to estimate outcome probabilities in soccer matches is acceptable, despite having some theoretical shortcomings. With the independent variables typically used in studies to forecast soccer matches, there is no evidence that using nominal regression is significantly better in actual practice. The forecasts derived from the regression models considered in this paper are unlikely to directly benefit actors in the betting market, given that the return on investment from bets placed based on the derived probabilities is never above 1 with statistical significance. However, both the OLR models and the MLR models are ostensibly equally good choices for analysing a range of related situations, such as the competitive balance within a league (Koning, 2000), the impact of managerial changes (Audas et al., 2002), or the effect of playing on artificial turf (Hvattum, 2015).

Although this paper has focused on the much used ordered regression models and only considered a multinomial regression model as an alternative, many other methods have been examined in the literature. The ability of alternative methods, such as other non-proportional odds ordinal models, random forests, or Gaussian process ordinal regression, to predict draws in soccer remains unexplored. However, based on the results in this work, it does seem that as of now, we are simply not aware of any independent variables that are effective at discriminating matches that are likely to end in a draw.

Acknowledgements

The author is grateful for the insightful comments from two anonymous reviewers that helped to improve the initial version of the paper.

References

- Audas, R., Dobson, S., & Goddard, J. (2002). The impact of managerial change on team performance in professional sports. *Journal of Economics and Business*, 54, 633–650.
- Cain, M., Law, D., & Peel, D. (2000). The favourite–longshot bias and market efficiency in UK football betting. *Scottish Journal of Political Economy*, 47, 25–36.

- Dixon, M. & Pope, P. (2004). The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting*, 20, 697–711.
- Dobson, S. & Goddard, J. (2001). *The Economics of Football*. Cambridge University Press, Cambridge.
- Dobson, S. & Goddard, J. (2003). Persistence in sequences of football match results: A Monte Carlo analysis. *European Journal of Operational Research*, 148, 247–256.
- Dobson, S. & Goddard, J. (2008). Forecasting scores and results and testing the efficiency of the fixed-odds betting market in scottish league football. In: Albert, J. & Koning, R., eds., *Statistical Thinking in Sports*, Boca Raton, Florida, USA: Chapman & Hall, 91–110.
- Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: the case of English football. *International Journal of Forecasting*, 21, 551–564.
- Forrest, D. & Simmons, R. (2000). Forecasting sport: the behavior and performance of football tipsters. *International Journal of Forecasting*, 16, 317–331.
- Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21, 331–340.
- Goddard, J. & Asimakopoulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, 23, 51–66.
- Graham, I. & Stott, H. (2008). Predicting bookmaker odds and efficiency for UK football. *Applied Economics*, 40, 99–109.
- Greene, W. (2012). *Econometric Analysis*, Harlow, England: Pearson, 7th edition.
- Hvattum, L. (2015). Playing on artificial turf may be an advantage for Norwegian soccer teams. *Journal of Quantitative Analysis in Sports*, 11, 183–192.
- Hvattum, L. & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*, 26, 460–470.
- Koning, R. (2000). Balance in competition in Dutch soccer. *The Statistician*, 49, 419–431.
- Krumer, A. & Lechner, M. (2016). Midweek effect on performance: evidence from the German soccer Bundesliga. Economics Working Paper Series 1609, University of St. Gallen, School of Economics and Political Science.
- Kuypers, T. (2000). Information and efficiency: An empirical study of a fixed odds betting market. *Applied Economics*, 32, 1353–1363.
- Nyberg, H. (2014). A multinomial logit-based statistical test of association football betting market efficiency. Discussion paper 380, HECER, Helsinki.
- Pope, P. & Peel, D. (1989). Information, prices and efficiency in a fixed odds betting market. *Economica*, 56, 323–341.
- Van Calster, B., Smiths, T., & Van Huffel, S. (2008). The curse of scoreless draws in soccer: the relationship with a team's offensive, defensive, and overall performance. *Journal of Quantitative Analysis in Sports*, 4, Article 4.
- Vlastakis, N., Dotsis, G., & Markellos, R. (2009). How efficient is the european football betting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting*, 28, 426–444.
- Štrumbelj, E. (2014). On determining probability forecasts from betting odds. *International Journal of Forecasting*, 30, 934–943.
- Štrumbelj, E. (2016). A comment on the bias of probabilities derived from betting odds and their use in measuring outcome uncertainty. *Journal of Sports Economics*, 17, 12–26.
- Witten, I. & Frank, E. (2005). *Data mining: practical machine learning tools and techniques*, San Francisco, CA: Elsevier.