Rating Systems for Fixed Odds Football Match Prediction

What is a Rating System?

A rating system provides a quantitative measure of the superiority of one football team over their opposition in a match. Such superiority is determined by analysing and comparing one or more aspects of past performance for each of the sides. Rating systems differ in the way side superiority is calculated, but basically each method calculates a points difference for a forthcoming football match, by subtracting a points rating for the away side from a points rating for the home side. The home and away team points ratings are determined through a quantitative analysis of past performance involving different aspects of a team's strength. The simplest of these uses either league points, league positions or goals conceded and scored, whilst more complex ratings might be based on elaborate match statistics including shots on goal, corners, and perhaps even possession if such data are available.

For many simple rating systems, no account is made of the quality of the opposition. A recent form goal-difference rating system, for example, simply looks at the number of goals scored and conceded by the two teams for a specified number of matches preceding the contest under examination. One goal is always worth one goal, whether it is scored away to Manchester United or at home to Birmingham City. Power Ratings or Rateform, described by Bill Hunter in his book *Football Fortunes*, overcome this problem by proportioning the worth of each goal scored (or conceded) to the strength of the opposition against whom it was scored (or conceded to).

Once a set of ratings for a match has been calculated, the next step is to estimate from them the chances of each result occurring, from which betting predictions can be made. A match rating must in some way be translated into a probability distribution for the three possible results in a football match: the home win, the draw and the away win. To achieve this, one must analyse some historical data, of the sort available from Football-Data. This process is described for a simple goals superiority recent form rating system.

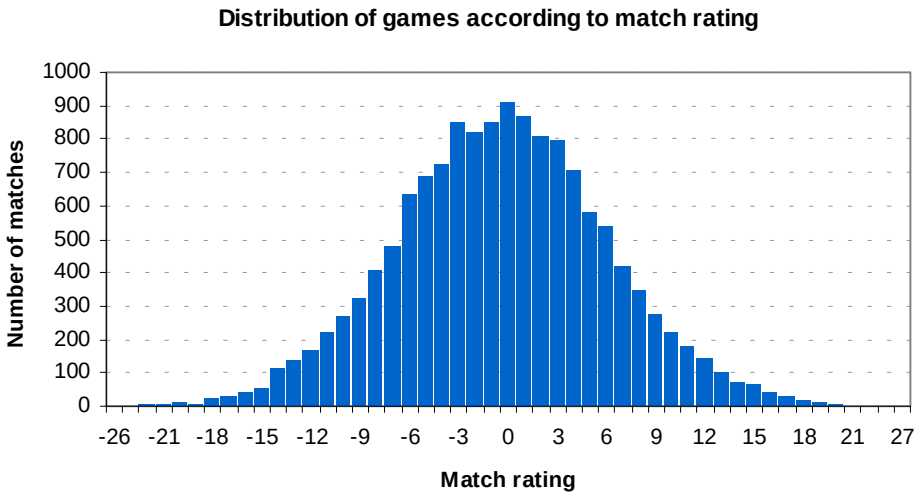
A Goals Superiority Rating System

Goal difference provides one measure of the dominance of one football side over another in a match. The assumption for a goals superiority rating system, then, is that teams who score more goals and concede fewer over the course of a number of matches are more likely to win their next game. That is to say that their match form, in terms of scoring and conceding goals, is potentially better than those teams with a lower rating by this reckoning. Typically, recent form means the last 4, 5 or 6 matches. Here, the last 6 matches are chosen.

To see how a goal superiority match rating is calculated, consider the following example for a game played between Tottenham and Leeds at White Hart Lane. In their last 6 games, Tottenham have scored 6 goals and conceded 9. Meanwhile, Leeds have scored 8 times and conceded 11 goals. Tottenham's goal superiority rating for the last 6 games is -3; for Leeds it is also -3. The match rating is simply given by the home side's rating minus the away side's rating, and for this match is therefore 0. With enough historical data we can translate this match rating into a home win-draw-away win probability distribution and subsequently the fair odds for one or more of

the three possible results. To begin this task we must look to see how frequently a match with a goals superiority rating of 0 finishes with a home win, a draw and an away win.

Using results data for the English Premiership and Divisions 1, 2, and 3 for seasons 1993/94 to 2000/01, goal supremacy ratings for the last 6 matches played by every team have been calculated. Of the 16,272 matches played during these 8 years, 14,002 of them were eligible for a rating calculation, with the matches played in the earlier weeks of each season obviously unsuitable for recent form analysis. The number of matches with each rating is shown in the figure below. Ratings vary from between -26 (a very weak home side and strong away side) to +27 (a very strong home side and weak away side). Obviously, there are far fewer games with large positive or negative ratings. Of these games, 46.2% finished with a home win, 28.1% with a draw, and 25.7% with an away win.



The number and percentage of home wins, draws and away wins for each match rating is tabulated on the next page. With a match rating of 0, we might estimate from this data that Tottenham to have roughly a 46% chance of winning the game, since of the 908 matches played between 1993 and 2001 that had the same rating, nearly 46% of them finished with a home win. The corresponding expectancy for a Leeds win is 26%, whilst that for the draw is 28%. This is virtually the same as the long-term average results distribution of 46.2%, 28.1%, and 25.7% for home wins, draws and away wins respectively, for the whole English Football League, unsurprising since a match rating of zero is roughly half way between the highest and lowest match ratings.

Since we have calculated a results probability distribution for every match rating, we could use the tabulated data on the next page to predict the most likely result for any match, provided at least 6 games have been played in the season to describe the recent form. According to the data, a match rating of +12, for example, would have a 63% chance of ending with a home win, whilst another game rated at -12 would give the away team a 39% expectancy of victory. In general, the higher the match rating the greater the probability of a home win. Conversely, the lower the rating, the greater the chance for an away win. It is initially not obvious how the match rating influences the likelihood of a drawn game.

Goal supremacy match ratings and historical result percentages

Match rating	*Number of home wins*	*Number of draws*	*Number of away wins*	*% of home wins*	*% of home draws*	*% of away wins*
-26	0	1	1	0.0%	50.0%	50.0%
-23	0	0	2	0.0%	0.0%	100.0%
-22	0	0	3	0.0%	0.0%	100.0%
-21	0	2	4	0.0%	33.3%	66.7%
-20	2	2	7	18.2%	18.2%	63.6%
-19	1	1	3	20.0%	20.0%	60.0%
-18	5	7	9	23.8%	33.3%	42.9%
-17	7	9	12	25.0%	32.1%	42.9%
-16	6	14	21	14.6%	34.1%	51.2%
-15	25	12	19	44.6%	21.4%	33.9%
-14	32	30	51	28.3%	26.5%	45.1%
-13	43	38	58	30.9%	27.3%	41.7%
-12	51	50	64	30.9%	30.3%	38.8%
-11	75	54	91	34.1%	24.5%	41.4%
-10	84	94	91	31.2%	34.9%	33.8%
-9	123	91	112	37.7%	27.9%	34.4%
-8	171	113	124	41.9%	27.7%	30.4%
-7	190	121	170	39.5%	25.2%	35.3%
-6	242	202	191	38.1%	31.8%	30.1%
-5	279	212	197	40.6%	30.8%	28.6%
-4	293	219	215	40.3%	30.1%	29.6%
-3	374	246	229	44.1%	29.0%	27.0%
-2	372	233	214	45.4%	28.4%	26.1%
-1	375	251	222	44.2%	29.6%	26.2%
0	414	259	235	45.6%	28.5%	25.9%
1	412	243	212	47.5%	28.0%	24.5%
2	401	220	189	49.5%	27.2%	23.3%
3	395	224	175	49.7%	28.2%	22.0%
4	391	177	137	55.5%	25.1%	19.4%
5	297	180	102	51.3%	31.1%	17.6%
6	260	146	131	48.4%	27.2%	24.4%
7	236	98	83	56.6%	23.5%	19.9%
8	197	94	56	56.8%	27.1%	16.1%
9	158	86	32	57.2%	31.2%	11.6%
10	125	57	42	55.8%	25.4%	18.8%
11	113	34	33	62.8%	18.9%	18.3%
12	90	30	22	63.4%	21.1%	15.5%
13	61	23	17	60.4%	22.8%	16.8%
14	48	15	11	64.9%	20.3%	14.9%
15	38	21	8	56.7%	31.3%	11.9%
16	30	9	2	73.2%	22.0%	4.9%
17	20	8	2	66.7%	26.7%	6.7%
18	15	1	1	88.2%	5.9%	5.9%
19	8	4	1	61.5%	30.8%	7.7%
20	5	1	0	83.3%	16.7%	0.0%
21	1	0	0	100.0%	0.0%	0.0%
22	1	0	1	50.0%	0.0%	50.0%
23	1	0	0	100.0%	0.0%	0.0%
27	1	0	0	100.0%	0.0%	0.0%
Total	**6468**	**3932**	**3602**	**46.2%**	**28.1%**	**25.7%**

But what about a rating of +15? Surely one might expect the chance of a home win to be greater than that for a match rating of +12, and yet only 57% of games rated +15 finished with such a result. Similarly, only 34% of games rated -15 finished with an away success. Of course, these discrepancies arise because the relationship between the match rating and the probability of a result is inherently “noisy” and imperfect. Such discrepancies become more apparent for the extreme ratings, for which, owing to the limited amount of match data, one or two results have a much greater influence on the results probability distribution. To accommodate this variance, we need to standardise our forecasting model. By doing so, we can make a practical attempt at defining the fair odds for a football game.

Defining the Fair Odds

The first task is to consider each result independently, and identify the "best-fit" relationship with the match ratings. The easiest way to determine this best-fit relationship is to draw the match ratings and result probabilities as three scatter plots, one for each result, as shown on the next page. This can be done simply using Microsoft Excel. For each scatter plot, the best fit line (with its equation) has been superimposed on the data points, representing what would statistically be considered to be the best relationship between match rating and result probability.

The equation for each best-fit line can also be easily calculated with Microsoft Excel, and these are shown for each figure. For each equation, y, the probability of a particular result occurring, is some function of x, the match rating. For home wins, the corresponding figure on the next page shows clearly that the best relationship is a straight line one. The value of R^2 shown for each of the three equations is simply a statistical measure of how closely the real data match the best-fit lines. A perfect relationship, denoted by $R^2 = 1$, would mean that the best-fit line and equation describe perfectly the real data. Consequently, a fairly good relationship exists between the match rating and home win probability, where as much as 86% of the variation in the real data is explained by its best-fit equation. For away wins and particularly draws the relationship is weaker.

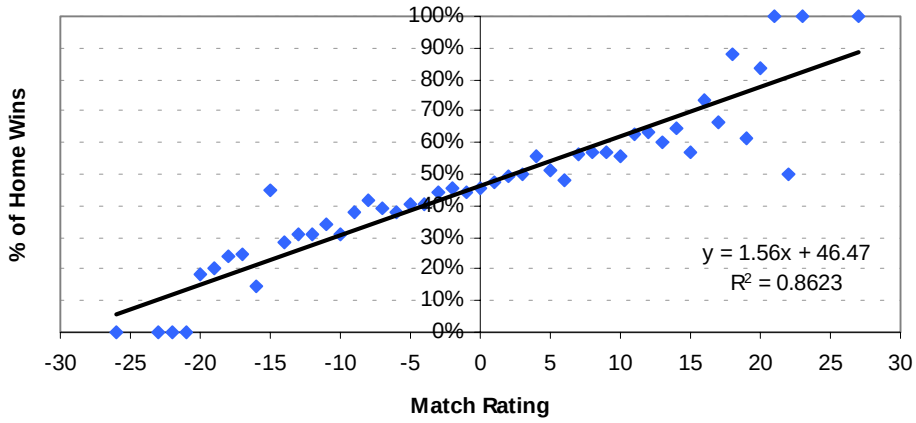
With each equation we can easily determine the expected probability of a home win, draw and away win for any match where we have calculated the goal supremacy rating. For the Tottenham-Leeds game, where the match rating was 0, the probability of a home win, for example, can be determined using:

$$y = 1.56x + 46.47$$

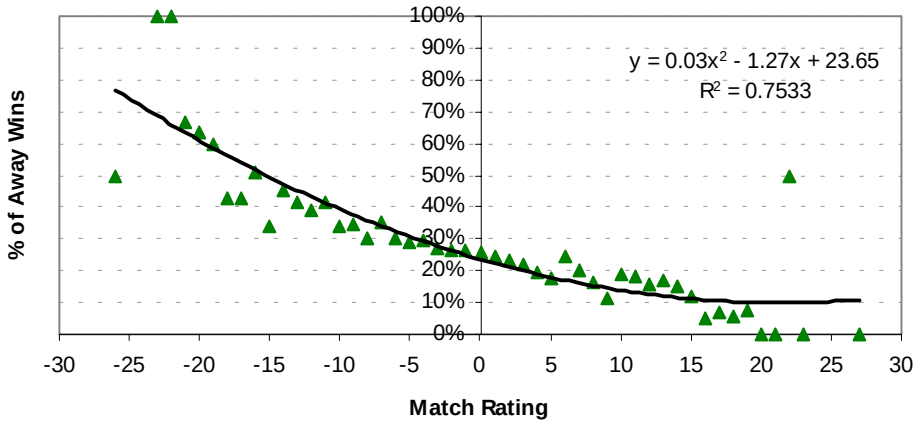
where y = the probability of a home win and x = the match rating.

For a match rating of 0, the probability of a home win will be 46.47%. This is very close to the percentage of historical games with a rating of 0 finishing with a home win. Such a close match arises because the best-fit relationship provides a particularly good fit for the real data ($R^2 = 0.8623$). The beauty of this equation is that we no longer need to remember all the information contained in the table on the previous page to calculate the home win probability for other games with different ratings. A game with a match rating of +10, for example, has a 62.1% likelihood of ending with a home win; one rated at -7, a 35.6% chance. To calculate these percentages, all we need to do is input the match rating into the equation above.

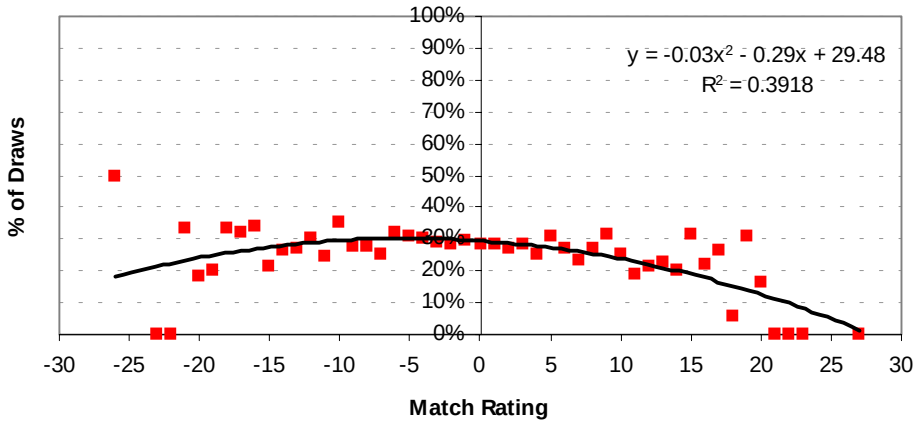
Distribution of home wins by match rating



Distribution of away wins by match rating



Distribution of draws by match rating



With an estimation of the probability or expectancy for a home win we can easily define the fair odds for a home win, using:

100 divided by probability of home win

Consequently, the fair odds according to this forecasting model for a Tottenham win are 100/46.47 or 2.15. We can define fair odds for the away win and draw in the same way, using the equations for away wins and for draws (shown with their corresponding figures on the previous page) to first calculate the expectancy of the away win and the draw. Their accuracy, of course, will be lower than for home wins, since the best-fit equations describe the real data less precisely. Thus, the fair odds for the Tottenham-Leeds game are 2.15, 3.39 and 4.23 for the home win, draw and away win respectively. Fair odds for some other match ratings are show tabulated below.

Result expectancies and their fair odds for a goal supremacy rating system

Match rating	*Home win expectancy*	*Draw expectancy*	*Away win expectancy*	*Total*	*Fair home odds*	*Fair draw odds*	*Fair away odds*
-16	22%	26%	52%	100%	4.65	3.78	1.94
-14	25%	28%	47%	100%	4.06	3.62	2.11
-12	28%	29%	43%	100%	3.60	3.49	2.31
-10	31%	29%	39%	100%	3.24	3.40	2.54
-8	34%	30%	36%	100%	2.94	3.35	2.80
-6	37%	30%	32%	100%	2.69	3.32	3.09
-4	40%	30%	29%	100%	2.49	3.32	3.42
-2	43%	30%	26%	100%	2.31	3.34	3.80
0	46%	29%	24%	100%	2.15	3.39	4.23
2	50%	29%	21%	100%	2.02	3.47	4.71
4	53%	28%	19%	100%	1.90	3.59	5.25
6	56%	27%	17%	100%	1.79	3.75	5.84
8	59%	25%	15%	100%	1.70	3.96	6.49
10	62%	24%	14%	100%	1.61	4.24	7.17
12	65%	22%	13%	100%	1.53	4.61	7.86
14	68%	20%	12%	100%	1.46	5.12	8.51
16	71%	17%	11%	100%	1.40	5.83	9.08

Identifying Value Bets

Once we have identified fair odds for a match, the final step is the easiest. By comparing our fair odds to the unfair odds of the bookmaker (with his overround built in to make a profit) we can determine whether he have identified a value bet. Where the bookmaker's odds are superior to our fair odds, we have potentially gained an edge, provided that our ratings system provides an accurate forecast, or more appropriately an accurate expectancy of the result. In such a case the bookmaker has under-estimated the probability of the result and is offering, by our forecast's assessment, higher than the mathematically fair odds. Over the long-term, provided the ratings analysis is accurate, one should make a profit. The distinction between result forecast and result expectancy is an important one, and is central to the principles of value betting. It is not enough simply to predict the most likely result. To potentially make a profit over a series of bets, we must always compare our fair odds to those of the bookmaker's, and think of the outcome of a match in terms of a probability distribution rather than as simply win or lose.

For the game between Tottenham and Leeds, the bookmaker Interwetten offer 2.2 for Tottenham to win. Since these are higher than our calculated fair odds of 2.15, this was a value bet. The game finished 2-0 to Tottenham and the bet was won. One winning bet, however, provides no measure of a profitable system. To ensure that we have one, any system should be tested over a much longer series of matches, for example over the course of a full season. With access to historical results and betting odds from Football-Data it is possible to test this retrospectively.

Testing a Match Prediction System

The results of an imaginary betting record using the goal supremacy rating system to identify value home wins for English league games played during the 2001/02 season are tabulated below. A total of 1746 games were given a match rating, with a team's first 6 games obviously ineligible for rating. Where a value bet was found, a unit stake was placed (1 point). Since a greater degree of uncertainty exists for the prediction of matches with very low or very high ratings (because far fewer matches with such ratings went into defining the best-fit forecasting equation), one might chose to avoid betting on such matches altogether. Conversely, forecasts for ratings closer to the central value of 0 are more reliably defined by the best-fit equation (see the top figure on page 5). Profit analyses contrasting the betting application of four different match rating ranges are therefore described.

Profit analysis for the goal supremacy match ratings forecast model, home wins only

	All ratings	Ratings -10 to +10	Ratings -5 to +5	Ratings -2 to +2	Blind betting on all home wins
Number of bets	526	459	330	171	1746
Profit from best odds	**+11.07**	**+16.32**	**+17.65**	**+17.32**	**-65.20**
Yield from best odds	**+2.10%**	**+3.56%**	**+5.35%**	**+10.13%**	**-3.73%**

By taking the best available odds (from 6 bookmakers), the goals supremacy rating system would have returned a profit over turnover (or yield) during the 2001/02 English football season of 2.1% for all match ratings. This figure increases as the range of match ratings is narrowed towards the central rating of 0. For matches rated between -2 and +2, over £1.10 would have been returned for every £1 stake. This compares to a -3.73% return betting on every home win blind.

A similar profits analysis can be performed for away wins and draws. Greater uncertainty in the forecasting model, as mentioned previously however, will make predictions less reliable and profits harder to achieve.

This guide to match prediction comes from reworked material from Joe Buchdahl's forthcoming book, *Fixed Odds Sports Betting: The Essential Guide*, to be published by High Stakes Ltd., and due for publication during Autumn/Winter 2003.