One match to go!

Can statistics really predict the results of Premier League football matches? **David Spiegelhalter** and **Yin-Lam Ng** put their skills to the test, and their reputations on the line—and scored!

On May 23rd, 2009, the 20 teams in the English Premier League each had one match left to play. West Bromwich Albion (West Brom) were at the bottom of the league with 31 points and Manchester United (Man U) were at the top with 87. The bottom three teams would be relegated: West Brom were certain to be one of them, but there were four other teams trying to avoid the other two places. Man U were certain to end up the top team and so were not expected to play their strongest team in their away match against Hull City. Hull, though, were one of the teams up for relegation and so had everything to play for.

BBC Radio 4's *More or Less* is more or less the only series on British national broadcasting that deals with numbers in a serious way. Its producers had heard of the work that we had been doing on modelling European football results, and they asked us to produce predictions for these final ten matches using a statistical method that could be explained, to non-specialist listeners, on the radio. This was quite a tricky challenge. Prediction itself is easy. Predicting accurately is the hard bit. In this case we knew that our predictions would be announced on the radio before the matches and then afterwards compared with what really happened and how well other pundits did. Our reputations would be on the line; and so, to an extent, would be the reputation of statistics.

Complex statistical models are used extensively in the sports betting industry, but we wanted a fairly straightforward model that could be explained using familiar concepts derived from the raw data. We found that using some basic theory we could quite easily produce a reasonable probability for all the possible results of a game, although then we used a slightly more sophisticated analysis for our actual probabilistic predictions.

We can start by looking at the state of the league on May 22nd, 2009, with goals for and goals against (see Table 1).

The average number of goals scored by the teams in the season so far, and therefore also the average number of goals conceded, was 46. If we divide the number of goals scored by 46, we get a measure of "attack strength", so Arsenal's 1.39 shows they have scored 39% more goals than average. If we divide the number of goals conceded by 46 we get a measure of "defence weakness", so Stoke City's 1.11 shows they have let in 11% more goals than average.

We also need two other pieces of information: the average number of goals scored per match by a home team up till now and the average number scored by an away team. Home teams had scored, on average, 1.40 goals per match; for away teams the figure was 1.08. Home teams clearly do better.

Now suppose we want to predict the result of Hull *versus* Man U. We start by estimating how many goals Hull will score. They are playing at home, so if they were an average team in an average match we expect them to score 1.40. But this is not an average team; over the season they have scored only 85% of the average number of goals, and so their "attack strength" is 0.85. Multiplying up we get $1.40 \times 0.85 = 1.19$ as the number of goals we might expect them to score against an average team. But their opposition is not average either: Man U's defence weakness is 0.52, since they have conceded only 52% of the average. So we get a total of $1.40 \times 0.85 \times 0.52 = 0.62$ expected goals by Hull, which does not look too good for their supporters.

For Man U, their baseline is 1.08, the average number of goals scored by an away team. But by

The crucial match: Hull City *vs* Manchester United. Photo courtesy Hull City AFC 2009. Not to be reproduced without permission

Table 1. State of the Premier League before final matches played on May 24th, 2009

Team	*Points*	*Goals for*	*Attack strength*	*Goals against*	*Defence weakness*
Manchester United	87	67	1.46	24	0.52
Liverpool	83	74	1.61	26	0.57
Chelsea	80	65	1.41	22	0.48
Arsenal	69	64	1.39	36	0.78
Everton	60	53	1.15	37	0.80
Aston Villa	59	53	1.15	48	1.04
Fulham	53	39	0.85	32	0.70
Tottenham	51	44	0.96	42	0.91
West Ham	48	40	0.87	44	0.96
Manchester City	47	57	1.24	50	1.09
Stoke	45	37	0.80	51	1.11
Wigan	42	33	0.72	45	0.98
Bolton	41	41	0.89	52	1.13
Portsmouth	41	38	0.83	56	1.22
Blackburn	40	40	0.87	60	1.3
Sunderland	36	32	0.70	51	1.11
Hull	35	39	0.85	63	1.37
Newcastle	34	40	0.87	58	1.26
Middlesbrough	32	27	0.59	55	1.20
West Brom	31	36	0.78	67	1.46

'Attack strength' = 'goals for'/46, where 46 is the average number of goals scored by a team. Similarly 'Defence weakness' = 'goals against'/46. 370 games have been played with 919 goals: 518 scored by home teams (average 1.40 per game); 401 by away teams (average 1.08 per game)

the time we adjust this for Man U's attack strength and Hull's defence weakness, we get $1.08 \times 1.46 \times 1.37 = 2.16$ expected goals in the match against Hull.

But, just as nobody has 2.4 children, nobody scores 2.16 goals—this is only an expected value, the average if the match were played again and again, heaven forbid. But we can use the Poisson distribution to distribute 100% of probability across the possible number of goals, which gives the probability distributions shown in Table 2.

So, if the next match follows past performance, there is a 54% probability that Hull will not score at all, and 63% (100 – 25 – 12) probability Man U will get at least two goals, even though they are playing away.

To get the probability of an actual final score we might assume the goals scored by each team are independent, in the sense that if we knew how many Man U scored it would not give us any additional information about Hull's performance. This is a strong assumption and we will come back to it in a moment, but it means that to find, for example, the probability of a 0–2 result, which is the most likely outcome, we multiply 54% by 27% to get 15%—so even the most likely result is still not very likely.

This independent Poisson model was developed by Maher[1], who also investigated a model that allows for correlation between teams' results. Versions of such bivariate Poisson distributions have been used by, for example, Dixon and Coles[2] and Karlis and Ntzoufras[3]. Estimating probabilities allowing for correlations is more complicated and requires special software: we used the bivpois function in R provided by Karlis and Ntzoufras[4].

Table 2. Expected number of goals, and percentage chance of getting a particular score for the two teams, assuming a Poisson distribution

Team	*Expected goals*	*% chance of achieving the following scores:* 0	1	2	3	4	5
Hull City	0.62	54	33	10	2	0	0
Manchester United	2.16	12	25	27	19	10	5

We have fitted models to all major league results in Europe over the last 15 years, and the predictions here are based on the best model found, which had a single parameter that allowed for matches to have a small tendency to be either high or low scoring, which we might call a "pitch effect".

These statistical models are very simplistic in that they assume that past performance throughout the season predicts future results and they do not take into account recent factors. For example, Hull City were trying to avoid relegation, Man U were conserving their strength having already topped the league and so it could be argued that Hull City stood a much better chance of winning than the 9% we had given them—some people obviously thought so, as the odds offered by the bookies were more like 2 to 1 against, i.e. a 33% chance of Hull winning.

Table 3 shows the three most likely results for each match according to the statistical model—the actual results are shown in bold and in the final column.

Note that the highest probability is 20%, and for most matches there is only around 40% chance that any of these three "most-likely" specific scores occur. So it is rather misleading to treat the "most-likely" results as predictions—all this model does is to produce (what we hope are) reasonable probabilities. If we add up the probabilities for all the score combinations that lead to a win, draw or lose we get the probabilities shown in Table 4. Some of these become quite high, for example, the 72% probability of a home win in the Arsenal–Stoke match, but even these could not be considered as firm predictions. The Fulham–Everton match was very finely balanced with the three possible outcomes almost equally likely.

The "most likely" results were read out on the *More or Less* broadcast on May 22nd, 2009, and, somewhat to our consternation, were reported as definite predictions without any qualifying probabilities. They were also given on the BBC website (`http://news.bbc.co.uk/1/hi/programmes/more_or_less/8062277.stm`), this time with probabilities (although we mistakenly said the Fulham–Everton most-likely 0–0 prediction had probability 10% whereas we should have said 19%, and Liverpool–Tottenham's most-likely prediction was given probability 10% instead of 16%).

So what happened? The day of the matches was nerve-wracking, but when the results were announced we were very relieved to find that using the "best predictions", we got 9 out of 10 correct, in terms of win, draw or lose, plus two exact scores. This was par-

Table 3. The three most likely final scores for each match, with actual results shown in bold and in the final column

Home	*Away*	*Most likely result*	*Second most likely result*	*Third most likely result*	*Actual result*
Arsenal	Stoke	2–0 (14%)	1–0 (13%)	2–1 (9%)	**4–1**
Aston Villa	Newcastle	**1–0 (10%)**	2–0 (10%)	2–1 (10%)	**1–0**
Blackburn	West Brom	1–1 (10%)	2–0 (10%)	2–1 (10%)	**0–0**
Fulham	Everton	0–0 (19%)	1–0 (16%)	0–1 (14%)	**0–2**
Hull	Mancheser United	0–2 (14%)	**0–1 (14%)**	1–2 (9%)	**0–1**
Liverpool	Tottenham	1–0 (16%)	2–0 (15%)	3–0 (10%)	**3–1**
Manchester City	Bolton	2–1 (10%)	1–1 (10%)	**1–0 (10%)**	**1–0**
Sunderland	Chelsea	0–1 (20%)	0–2 (15%)	0–0 (13%)	**2–3**
West Ham	Middlesbrough	1–0 (19%)	0–0 (14%)	2–0 (13%)	**2–1**
Wigan	Portsmouth	**1–0 (16%)**	0–0 (14%)	1–1 (13%)	**1–0**

ticularly gratifying as Mark Lawrenson, the official BBC football expert, only got seven correct results and only one exact score.

This was a very good result for statistics! But perhaps a bit lucky—in particular it is very difficult to predict draws and it was rather fortunate that the "most-likely" 1–1 Blackburn–West Brom predicted score turned out in fact to be a 0–0 draw, since a draw was not the most likely outcome. One possible advantage of the statistical method is that it is not influenced by emotion—for example in the Hull–Man U match, Hull had been considered as having a reasonable chance of a win, but we firmly went for a Man U win and were proved correct—this cool-headed statistical approach is helped by the fact the neither of us support a team or even know much about football. In terms of the exact scores, the Sunderland–Chelsea 2–3 result was the most "surprising", having been given a probability of just over 1%.

Our results were only judged by the BBC in terms of the "most likely" prediction, but a more subtle analysis would evaluate the quality of the whole probability distributions provided for each match. For example, adding up the columns of Table 4 reveals that we expected 4.7 home wins, 2.4 draws and 2.9 away wins and in fact there were 6, 1 and 3 respectively, which is only one draw away from almost perfect calibration! We can also use a "Brier scoring rule", developed in the field of weather forecasting, to check how accurate our probability distributions were. This is a penalty measured by the squared distance between the probability vector and the outcome and is conveniently expressed as

$$\text{Brier} = 1 + p^2(\text{home}) + p^2(\text{draw}) + p^2(\text{away}) - 2p(\text{actual outcome})$$

These are shown in Table 4: a Brier = 0 corresponds to a perfect prediction and Brier = 2 to a useless prediction that put 100% probability on an outcome that did not occur.

The total Brier penalty was 3.5. If we had used a uniform (0.33, 0.33, 0.33) prediction for all matches the total Brier penalty would have been 6.7—about double the penalty acheived. A more sensible "default" prediction would have been (0.45, 0.26, 0.29) since these are the proportions of home wins, draws and away wins throughout the season: this would have given a Brier penalty of 5.9. So our model has allowed us to reduce our penalty by 40% compared with a "no-skill" prediction. We can also check whether our observed Brier penalty of 3.5 is around what we would have expected were our assessed probabilities the "true" chances of the outcomes: the Z-statistic for testing this null hypothesis is –1.56, suggesting that our Brier penalty was slightly less than we would have reasonably expected it to be, confirming our impression that we were lucky, although it could be interpreted as our being a little too cautious and that our probabilities could have been closer to 0 or 1.

These types of models have been refined over the years and are now used by bookies and sports betting companies, who employ experienced statisticians and make use of the latest computational methods: in particular it is natural to allow for a team's abilities to change over the season, and so "discount" historical evidence to allow recent performance to dominate. And, not surprisingly, they don't tell anyone exactly what they do! One thing you can bet on—simple models like those above will be unlikely to out-perform the odds being offered by bookies, so we said that people should not use them to spot good bets. We have heard that some people did make money from our predictions, and we have since been approached by people wanting to work with us on sports modelling, but we are not likely to take this up as a sideline—it could be much too engrossing.

Table 4. The assessed probabilities for the 10 matches, with the actual results in bold: the Brier score is an overall assessment of the accuracy of the probability distribution: high scores indicate poor predictions

Home	*Away*	*Percent probability of*			*Brier penalty*
		Home win	*Draw*	*Away win*	
Arsenal	Stoke	**72**	19	10	0.12
Aston Villa	Newcastle	**62**	21	17	0.22
Blackburn	West Brom	54	**23**	23	0.94
Fulham	Everton	35	35	**30**	0.74
Hull	Manchester United	9	19	**72**	0.12
Liverpool	Tottenham	**72**	20	9	0.13
Manchester City	Bolton	**59**	22	19	0.25
Sunderland	Chelsea	10	25	**65**	0.20
West Ham	Middlesbrough	**57**	28	15	0.29
Wigan	Portsmouth	**44**	32	25	0.48

References

1. Maher, M. J. (1982) Modelling association football scores. *Statistica Neerlandica*, **36**, 109–118.
2. Dixon, M. J. and Coles, S. G. (1997) Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics*, **46**, 265–280.
3. Karlis, D. and Ntzoufras, I. (2003) Analysis of sports data by using bivariate Poisson models. *The Statistician*, **52**, 381–393.
4. Karlis, D. and Ntzoufras, I. (2007) Package 'bivpois'. (Available from `http://www.jstatsoft.org/v14/i10/supp/1`.)

David Spiegelhalter is Winton Professor of the Public Understanding of Risk at the Statistical Laboratory, University of Cambridge. This work formed part of Yin-Lam Ng's project for an MPhil in Statistical Science, for which she was awarded a distinction. A version of this article has appeared in *Plus*, the on-line mathematical magazine.