International Journal of Forecasting 16 (2000) 317–331
www.elsevier.com/locate/ijforecast

Forecasting sport: the behaviour and performance of football tipsters

David Forrest*, Robert Simmons

Centre for Sports Economics, University of Salford, Salford M5 4WT, UK

Abstract

Professional advice is available in several forecasting contexts, such as share prices, sales and the weather. English newspaper tipsters offer professional advice on the outcomes of English and Scottish football (soccer) matches. Such advice could potentially inform selections of bettors in fixed odds or pools betting. This paper investigates the effectiveness of the guidance given by newspaper tipsters. Employing a sample of three tipsters and 1694 English league games, we find that tipster success rates are higher than would follow from random forecasting methods. We identify some differences between the processes by which actual results and tipster forecasts are determined. Likelihood-ratio tests imply that the tipsters fail adequately to utilise easily obtainable public information on teams' strength. Further tests show that only one of three tipsters appears to make successful use of other unspecified information relevant to game outcomes. A consensus forecast across the three tipsters appears to outperform any single tipster.

Keywords: Football; Newspaper tipsters; Logit

1. Introduction

Professional advice on the outcomes of economic, political and other events is readily available in many settings. For example, independent forecasters may predict the aggregate inflation rate, sales, asset prices, outcomes of political elections and the weather. The judgements of forecasters may represent specialised, marketed advice or may be available as free advice in the media. This paper focuses on the provision of free professional advice, published in newspapers, in order to guide the behaviour of bettors interested in the outcomes of football matches. Tipsters are a familiar adjunct to many betting markets. Some of those predicting National Lottery winning numbers claim psychic powers but one may presume that sports tipsters profess a more conventionally defined form of expertise. However, Makradakis, Wheelwright, and Hyndman (1998) summarised the literature on expert forecasts as follows: 'in nearly all cases where the data can be quantified, the predictions of the (statistical) models are superior to those of the expert' (p. 492). In this paper, we investigate the nature of tipster expertise and assess the effectiveness of the

*Corresponding author. Tel.: +44-161-295-3674; fax: +44-161-295-2130.

E-mail address: d.k.forrest@salford.ac.uk (D. Forrest).

0169-2070/00/$ – see front matter
PII: S0169-2070(00)00050-9

guidance offered against a statistical model of actual football results.

Newspaper football tipsters are essentially involved in subjective probabilistic judgemental forecasting (Goodwin & Wright, 1993; Webby & O'Connor, 1996). Judgemental forecasting usually refers to the use of additional elements of judgement over and above techniques of time-series forecasting. Football tipsters offer definite selections of match outcomes which bettors might use in their selections. The selections offered are not probability-weighted. A football match has a well-defined termination point (at least in the eyes of bookmakers) and a well-defined outcome, in the form of either home win, draw (tie) or away win. Unlike most applications of judgemental forecasting, football tipsters do not apply formal statistical techniques to inform their predictions, relying instead on intuition. We can infer the tipster's implicit model, however, by regression analysis.

The information available to forecasters includes 'public' information, such as the league placings and form of the two competing teams, and 'private' information which is somewhat less tangible. Tipsters can use combinations of public and private information to make their selections. Hence, tipsters employ contextual information (information other than time-series patterns and general data on team performance) to make their selections. In this respect, the 'broken-leg cue' is important (Webby & O'Connor, 1996). This refers to unusual pieces of information which enter into a tipster's judgement over and above normal information. Here, the broken-leg cue has a literal meaning when injuries to key players may influence a tipster's prediction of a result.

Academic studies of sports tipster performance are limited in number and normally linked to tests of betting market efficiency. For example, Figlewski (1979) found that the forecasts of a group of newspaper tipsters (handicappers in American parlance) had considerable power in explaining the results of horse races in New York if no other variables were included in his regression equation but added nothing to a model containing betting odds. This finding supported the notion of market efficiency but said little about the performance of the tipsters themselves. Thus the failure of the tipsters' predictions to raise the explanatory power of an equation including (only) betting odds may have resulted (to take two possibilities) either from the tipsters being so respected that the odds responded so as to fully discount their selections or from the tipsters and bettors independently reaching the same conclusions from their reading of form (in a pari-mutuel betting market, odds are determined solely by the behaviour of the bettors). In the first case, one would regard the tipsters as truly expert and serving the function of aiding market efficiency but in the second, they would be judged inexpert because they know no more than the bettors themselves. Clarification of the issue would require the employment of objective data on the ability and form of each runner to check whether tipsters fully exploited the information and whether they added anything to it.

Bird and McCrae (1987) and Pope and Peel (1989) considered tipsters in the context of bookmaker rather than pari-mutuel betting markets, looking at horse racing in Australia and soccer in the United Kingdom, respectively. Each paper reported a simulation of bets being placed in accordance with the recommendations of a panel of newspaper experts. Neither found evidence for abnormal bettor returns (though there was some indication that experts' predictions became worth following towards the end of the football season). Again, the implication is that experts' selections contain the same information as betting odds. However, it would be interesting to know what sort of information was being captured and whether it was being exploited fully.

Goodwin and Corral (1996) adopted a more comprehensive approach which included explicit consideration of the question of whether newspaper tipsters possessed information additional to that included in the form records. The application was to greyhound racing at the Woodlands dog track in Kansas City. Employing a likelihood-ratio test to distinguish between different logit models, they found that dummy variables indicating newspaper tips of the winner and runner-up in each race were statistically significant in the presence of a large number of variables representing information recorded in the greyhound formbook. Tipsters thus appeared to know something (perhaps intangible) not included in the readily obtainable public information. On the other hand, a similar test revealed that tipsters themselves did not fully exploit the public information.

The context of our paper is English professional soccer. Like Goodwin and Corral, we test whether newspaper forecasts contain information not in the form listings and whether the tipsters themselves fully exploit the listings. We go further by modelling the role of the various form data in determining the tipsters' forecasts and whether these tipsters give the appropriate weight to each type of form information. In addition, we consider alternative approaches to measuring the help tipsters give to their readers. Our contribution is novel in directly assessing what variables determine tipster predictions. Our methodology can be generally applied to situations where forecasters make judgements on outcomes which can be ordered.

2. Data

In 1925, *The People* became the first newspaper to include a column offering forecasts of soccer results (Sharpe, 1997). It took six decades for some of the more up-market newspapers to follow them but now all London-based national daily newspapers except the *Financial Times* carry a weekly guide for football bettors that includes an expert's forecast of the outcome of each match (home win, away win or draw). Tabloid and middle-market newspapers carry ancillary information on, for example, recent form and league standings and some newspapers offer 'expert analysis' in the form of supporting commentary on predicted match outcomes by their tipster. Resource constraints prevented our studying the performance of every newspaper tipster, so we selected three newspapers for analysis. They were chosen to be representative of the range of titles available: we required a broadsheet, a middle-market and a tabloid publication and thus adopted (in view of their availability in library archives to which we had access) *The Times*, the *Daily Mail* and *The Mirror*. Each of the three tipsters' columns produced forecasts of the outcome of every professional fixture rather than pick 'best bet' games.

We attempted to collect information pertaining to all matches played within the English professional league structure on Saturdays between December, 1996 and April, 1998 (The professional leagues are the FA Premiership and the three divisions of the Football League; Cup and Scottish League fixtures were excluded). However, we omitted August and September of 1997 from our data set because these were the first two months of a new season and there was therefore limited recent form information on the teams in each match. For each included fixture, we noted the forecasts of the three tipsters and collected the information, listed in bettors' guides in the *Daily Mail* and *The Mirror*, on the recent home and away performances of the home and away team, respectively. We also collected the summary statistics of each club's cumulative (current season) performance embodied in the weekly tables of league standings, also published on the newspapers' sports pages. A small number of games were omitted from our data set because we were unable to obtain forecasts from all three tipsters (e.g. there may have been a production problem or strike at one of the newspapers). We were left with a data set of 1694 football matches.

3. Criteria for assessing performance

An examination of the success of the tipsters in forecasting the 1694 matches must be based on some benchmark level of performance which they would need to have exceeded for their employment to be judged worthwhile. What should they be expected to achieve? The answer depends on what sorts of bettors they are intended to help. We propose three levels at which they might give assistance and each requires different statistical tests to be performed.

The weakest test of tipster efficiency would be to check whether they offered any guidance to a bettor who would enjoy a gamble but who knows nothing about soccer. Such bettors would hope that, at a minimum, tipsters would call more matches correctly than would be expected to follow from randomly picking home wins, draws and away wins. In Section 4, we ask what 'random' selection would mean in this context and test the extent to which tipsters' forecasts are more accurate than a random method that we employ.

A second and higher level of guidance that tipsters could offer is to bettors who are aware of the significance of all the items of data printed in the newspaper guides but who have no taste or time for processing them themselves: they look to the forecaster to provide a single statistic (namely home, draw or away selection) that captures and summarises all this public information. In Section 5, we employ regression analysis to model how the experts process the public information and propose and implement tests of whether the processing is performed efficiently.

A yet higher level of expectation of tipsters would be implied if one supposed that they should have access to relevant information that was either private or semi-public (in the sense of being too costly for ordinary bettors to purchase). An example of private information might be a training injury to a key player. An example of semi-public information might be that the attacking set-pieces of a high-flying team would be likely to be ineffective against an apparently weaker team because of the defensive style the latter employs (such a conclusion might require reviews of video evidence). If tipsters possess such information (and tipsters interviewed in Sharpe (1997, Ch. 4) claim to have access to news that the public does not hear), then in appropriate regression analysis tipsters' forecasts should be statistically significant in explaining the pattern of match outcomes even when all public listed information is included. In Section 6, we confront tipsters with this toughest test of their usefulness. Section 7 summarises the conclusion of the several approaches adopted in the paper.

4. Tipster picks versus random picks

In our data set, the proportions of matches called successfully by the three tipsters were:

Daily Mail	42.56%
The Mirror	41.09%
The Times	42.86%

Can these strike-rates be considered high or low? Would pure guesswork be likely to yield worse or better strike-rates? Since we are beginning with a weak test of tipster efficiency, we require an appropriately naïve strategy to provide a benchmark against which to make judgements.

One extremely simple strategy would be to call every match a home win, this being the most frequently occurring of the three outcomes. In our data set, 0.475 of matches played were won by the home team, 0.281 were draws, and 0.244 were away victories. Therefore, always forecasting a home win would have yielded a 47.5% strike-rate, comfortably superior to that of any of the three tipsters.

This line of thought is no doubt unfair to tipsters. Their editors would scarcely pay them for providing a 100% home win list of forecasts each week and it would be of little use to bettors who could hardly dare hope for a betting market so inefficient as to yield a positive return to betting on the same outcome every time. The tipsters therefore produce a mixture of all three possible forecasts in each week's list of games. Indeed, an important function of tipsters is to provide guidance to bettors in the choice of draws in the pari-mutuel football pools game, and so tipsters must feel obliged to report some predicted draws.

What might be proposed as an alternative naïve strategy is to choose the outcome of each match at random within the constraint that the proportions of home wins, draws and away wins should correspond to the proportions found in real life results. These proportions are relatively stable over time. For simplicity, we use the proportions from our data set. If h, a and d are these proportions, then the strike-rate from such a strategy would have been $h^2 + d^2 + a^2 =$ 36.4%. All the tipsters outperformed this naïve random strategy.

Of course, it would be easy for those with little knowledge of soccer to pass this test, simply by moving some way towards selecting all matches as home wins. In fact, all three tipsters over-predicted homes relative to the frequency of home wins in real results; respectively, they called home wins in 0.563, 0.539 and 0.533 of games.

Suppose a prospective tipster has no expertise or no wish to make the effort required to use his expertise. Might he obtain as good a strike-rate as one of our experts by random forecasting within the constraint that he adopts the same win–draw–away proportions as the expert himself? If so, the efforts of the real-life expert might be judged wasted.

If one had allocated the forecasts of tipsters randomly across the 1694 matches, the strike-rate achieved would have been $h_i h + d_i d + a_i a$, where h_i, d_i and a_i are the proportions adopted by tipster i and h, d and a are the relative frequencies in real match outcomes. Depending on whether tipster i was the *Mail, Mirror* or *Times* expert, realised strike-rates would have been 38.3, 37.9, and 37.5%, respectively. In each case, therefore, the newspaper tipster achieved a higher strike-rate than would have resulted from random picking within the constraint of predicting the same proportion of each outcome as the tipster himself. To this extent, the achieved strike-rates indicate that some knowledge of soccer is embodied in the published forecasts.

This finding may be checked by regression analysis with the advantage that the statistical significance of the superiority of a tipster's forecasts over random selection could be assessed easily from a likelihood-ratio test. In Table 1, we report our first regression results. Because the three possible outcomes of a match follow a natural ordering, the ordered logit model (first proposed by Zavoina & McElvey, 1975) is appropriate. The dependent variable is match outcome (home = 0, draw = 1, away = 2) and tipster forecasts are embodied in dummy variables corresponding to draw and away predictions. An alternative approach to modelling soccer results, applied by Maher (1982) and Dixon and Coles (1997), is to assume that goals scored by opposing teams are each determined independently by a Poisson process. Our approach takes a different tack and explicitly models what tipsters do. We use our model of tipster predictions as a basis for an assessment of their performance. Tipsters are unlikely to use the Poisson process in making their selections, given the formidable computational requirements involved and their preference for intuition over formal modelling.

Table 1
Ordered logit regression analysis[a]

	(1) Daily Mail		(2) The Mirror		(3) The Times	
	Coefficient	Impact	Coefficient	Impact	Coefficient	Impact
Constant	−0.063	+0.016	−0.071	+0.018	−0.083	+0.021
	(1.02)		(1.12)		(1.27)	
Draw tip	0.252*	−0.063	0.348*	−0.087	0.329*	−0.082
	(2.26)		(3.25)		(2.80)	
Away tip	0.628*	−0.157	0.510*	−0.128	0.460*	−0.115
	(5.11)		(4.12)	(1.27)	(4.23)	
Log-likelihood	−1773.18		−1776.00		−1776.80	
Restricted log-likelihood	−1786.75		−1786.75		−1786.75	

[a] Dependent variable: match result; independent variables: tipster forecasts in the relevant newspaper; 'Impact' shows the marginal effect on home win; absolute *t*-statistics shown in parentheses; * indicates significance at the 5% level.

In the regression estimates of actual results in Table 1, the coefficients on both tipster dummies are significant in the equations for all three newspapers. The positive signs indicate that in every case, if the tipster calls a draw or an away win, this is associated with a lower probability of a home win occurring on the field. The model implicitly places each match on a scale ranging from 'home win extremely likely' to 'away win extremely likely' and, when a tipster calls other than a home win, the match is pushed along the scale from the home win end towards draw and away win. The marginal effects (calculated as in Greene, 1997) are interpreted as follows. Taking *The Times* as an example, if *The Times* expert predicts a draw, this lowers the probability that there will be a home win by eight percentage points. If his forecast is for an away win, this lowers the probability of a home win by 11 percentage points. However, it is worth noting that, even in the presence of tipsters' forecasts, the most probable outcome in every single match (regardless of tipster forecasts) remains a home win, reflecting the large magnitude of home advantage in English soccer.

In the equation for each tipster, the statistical significance of the newspaper forecasts may be assessed by comparing the log-likelihood function value with that for a model in which the coefficients on the relevant variables are constrained to be zero. In a chi-squared likelihood ratio test, the test statistic is $2(L_F - L_R)$, where L_R is the maximised value of the log-likelihood function for the restricted model and L_F is that for the unrestricted model. The number of degrees of freedom is given by the number of restrictions imposed on the restricted model (here two). For our three equations, the test statistic values are 27.14, 21.50 and 19.88. The critical value is 5.99 at the 5% level. The hypothesis that there is no relevant information captured in the tipster's forecasts is therefore rejected decisively in all three cases. This confirms our conclusion that tipsters pass the weak requirement of knowing something. From the preceding analysis, it appears that they know enough for copying the forecasts of any one of them to yield more success than if one just guessed the outcomes of matches.

5. Tipsters and public information

We now inquire as to the source of tipsters' knowledge. The broadsheet newspapers such as *The Times* excepted, there is little variation across newspapers in the choice of facts and figures displayed in the weekly bettor guides. Presumably, this reflects some consensus about what matters in the determination of the likely outcome of a soccer game. In Table 2, we report the results of ordered logit estimation of the relationship between each tipster's forecasts and the information that newspapers often present alongside tipsters' columns. Recent form is represented by variables (last five home and last five away) measuring the home team's performance in its last five home games and the away team's performance in its last five away games, as reported in the *Daily Mail*. Our measures were constructed by awarding two points for each win and one point for each draw so that the variables take values in the range zero to ten. The cumulative season-to-date home

Table 2
Ordered logit regression analysis[a]

	(1) Daily Mail		(2) The Mirror		(3) The Times		(4) Result	
	Coefficient	Impact	Coefficient	Impact	Coefficient	Impact	Coefficient	Impact
Constant	0.431	−0.104	0.485*	−0.119	0.810*	−0.199	0.278	−0.067
	(1.37)		(2.22)		(3.89)		(1.37)	
Last 5	−0.193*	+0.091	−0.139*	+0.067	−0.134*	+0.065	0.011	−0.005
home	(5.97)		(4.40)		(4.21)		(0.38)	
Last 5	0.054	−0.025	−0.029	+0.014	−0.028	+0.014	−0.009	+0.004
away	(1.95)		(0.99)		(1.03)		(0.37)	
Home goals	−0.118	+0.025	−0.135	+0.029	0.075	+0.016	−0.152*	+0.033
ratio	(1.40)		(1.59)		(0.92)		(2.26)	
Away goals	0.912*	−0.099	1.000*	−0.110	0.301*	−0.033	0.337*	−0.038
ratio	(7.82)		(7.80)		(2.29)		(2.87)	
Home lead	−2.255*	+0.543	−3.236*	+0.795	−3.207*	+0.787	−0.287	+0.072
17+	(4.39)		(4.37)		(5.26)		(1.02)	
Home lead	−2.047*	+0.493	−3.315*	+0.814	−3.426*	+0.840	−0.928*	+0.232
13–16	(5.27)		(5.52)		(6.49)		(4.01)	
Home lead	−1.624*	+0.391	−1.843*	+0.453	−2.061*	+0.506	−0.454*	+0.113
9–12	(5.69)		(6.64)		(7.50)		(2.36)	
Home lead	−1.212*	+0.292	−1.242*	+0.305	−1.139*	+0.279	−0.260	+0.065
5–8	(5.88)		(6.27)		(6.20)		(1.57)	
Home lead	−0.498*	+0.120	−0.773*	+0.190	−0.500*	+0.123	−0.301	+0.075
1–4	(2.87)		(4.30)		(3.04)		(1.85)	
Away lead	0.487*	−0.117	0.581*	−0.143	0.311	−0.076	−0.137	+0.034
5–8	(2.99)		(3.49)		(1.93)		(0.86)	
Away lead	0.902*	−0.217	1.347*	−0.331	0.606*	−0.149	−0.139	+0.034
9–12	(4.78)		(7.15)		(3.39)		(0.77)	
Away lead	1.184*	−0.285	1.812*	−0.445	1.134*	−0.278	−0.137	+0.034
13–16	(5.41)		(7.86)		(5.04)		(0.65)	
Away	1.549*	−0.373	2.283*	−0.561	1.518*	−0.372	0.105	−0.026
lead 17+	(4.65)		(7.15)		(5.08)		(0.39)	
Log-likelihood	−1332.22		−1262.02		−1442.25		−1758.63	
Count-R^2	0.638		0.658		0.596		0.479	

[a] Dependent variable: (1) to (3) tipster forecasts; (4) match results; independent variables: public listed information; 'Impact' shows the marginal effects on home win (dummy variables) or the standard deviation marginal effect on home win (continuous variables); absolute *t*-statistics in parentheses; * indicates significance at the 5% level.

and away performances of the two teams are captured by 'home goals ratio' and 'away goals ratio', in each case calculated by dividing goals scored by goals conceded. The relative all-round strength of the clubs measured over the season to date is embodied in the difference in their league positions. To allow for non-linearity, relative league position is entered via banding with a series of dummies such as 'home lead 13–16' (the current league position of the home team is between 13 and 16 places higher than the visiting team) and 'away lead 5–8' (the away team currently leads its hosts by between 5 and 8 places in the standings); the reference band is 'away 1–4', chosen because this was, by a small margin, the most frequently occurring grouping in the data set.

In Table 2, we also show for comparison the result of ordered logit estimation of an equation relating actual outcomes to all these team strength indicators. As one might expect, given the scope for matches to be decided by random on-the-field events, this equation is less well determined than those which model tipster selections, but it nevertheless serves the function of allowing a check on whether the items of data either emphasised or neglected by tipsters in fact matter in accounting for the actual pattern of game outcomes.

In an ordered logit model, interpretation of the size of coefficients (as opposed to their statistical significance) requires computation of marginal effects. In Table 2, in columns headed 'impact', we show for the binary variables the marginal effect on home win of the variable taking on the value of one rather than zero. For example, we estimate that if the home team has a lead of 1–4 places in the league standings, the probability that the *Daily Mail* columnist will call a home win is raised by 12.0 percentage points relative to the situation represented by the reference (or excluded) category of the away side enjoying a lead of 1–4 places. Of course, the model also permits calculation of marginal effects on draws and away wins but, for conciseness, we omit these from the table.

The 'impact' shown in Table 2 in respect of the set of continuous variables (last five home, last five away, home goals ratio, away goals ratio) was obtained by taking the calculated marginal effect on home win and multiplying it by the standard deviation of the variable.

This 'standard deviation marginal effect' shows how the probability of a home win is affected when the value of the relevant variable is raised by one standard deviation from its mean. For example, in the estimates of actual results, the impact shown for 'away goals ratio' is -0.038. This implies that if the visiting team has an away goal ratio one standard deviation better than the mean away goal ratio (measured across all observations), then the probability of the home team winning is thereby lowered by 3.8 percentage points.

It is clear from Table 2 that tipsters give considerable implicit weight to our proposed indicators derived from data listed in newspapers. The coefficients on most of the variables are highly significant. However, they are not uniformly so and the following principal points emerge.

First, all three tipsters accord significance to the recent home performances of the home team while appearing not to take seriously the recent road record of the visiting club. Possibly they have in mind that home advantage stems in part from the home team's familiarity with the physical characteristics of its stadium and from the degree of crowd support from which it benefits (Courneya & Carron, 1992). The extent to which the team can exploit these advantages may be thought to be embodied in the recent results obtained at that stadium. The away team's recent away record will, on the other hand, relate to games at a variety of stadia where the physical dimensions, surface characteristics and crowd behaviour may be different in every case from the conditions to be encountered in the next fixture. For this or for whatever reason, recent away form appears not to be taken as relevant to the outcome of the next away trip.

Second, and by contrast, the season-long goal ratio variable is shown in Table 2 to have a significant effect on newspaper forecasts only in the case of the away club. The information selected as mattering as a summary of the away team's past performances has at least the advantage that it will be unaffected by whether the recent away games have been played on grounds with particularly idiosyncratic characteristics.

Inspection of the regression results for these variables in respect of real match outcomes indicates that recent form is in fact of little relevance whether it refers to the home or away team. The coefficients on home goals ratio and away goals ratio are, however, both highly significant and almost symmetric in their impact on the predicted outcome of a match. These variables measure cumulative form and are more reliable predictors of actual results given the difficulty in separating signal and noise when only a small number of matches are considered. The tipsters perhaps suggest an overly deterministic process where match results are seen to some extent as extrapolations of recent form. Note that these contrasts between the determination of tipster and actual results are unlikely to be explained by collinearity between the recent form and goals ratio variables which, from an inspection of the correlation matrix, appeared not to be unduly high (0.52 (last five home results and home goals ratio) and 0.25 (last five away results and away goals ratio)).

Another possible source of home advantage noted by Courneya and Carron (1992) is the travel fatigue and disruption to routine experienced by the away team. Clarke and Norman (1995) related the degree of home advantage to the distance between the two teams' grounds. We experimented with the inclusion of distance in all the equations estimated in Table 2 but, while correctly (negatively) signed in the results equation, it was not statistically significant; nor did tipsters appear to give this information any weight. Distance was therefore omitted from the final estimated equations.

The third feature of interest in Table 2 is the large weight tipsters accord to relative league positions. The relevant coefficients are all highly significant in each of the three tipster equations. Furthermore, inspection of the marginal effects shows it to be almost uniformly true that each extra step on the spectrum from large away lead to large home lead raises sharply the probability of a home win being tipped. This faith in the guidance offered by the standings is not quite supported by the equation for actual match outcomes where substantial home leads certainly improve the chances of a home win but where the size of an away lead is not a useful predictor: the coefficients on the away lead bandings are all insignificant, showing no impact on the result relative to the base situation of a small away lead (away lead 1–4 being the reference banding in the specified equation). Relative to the statistical process determining actual results, it appears that the tipsters give undue weight to both small home team leads and to away leads. Indeed, the tipsters appear to make their predictions as if matches can be ranked according to a scale between large away lead (away win) and large home lead (home win). Actual results contain far more noise than this procedure would warrant; only large home leads, of 9–16 places, significantly affect the probability of a home win. A similar comparison appears in a study of the verdicts of the Pools Panel when matches are postponed (Forrest & Simmons, 2000).

There are then differences between the way tipsters model the relationship between outcomes and listed information and the way in which the information appears to matter in practice. This could of course raise a suspicion that tipsters might not be processing the information accurately. That in itself, one should note, would not necessarily bring about unsatisfactory overall performance by tipsters in that they may compensate by wise use of private information. For now, we shall analyse further the use they made of the public information.

Ordered logit estimation yields, for each observation, a fitted value of a continuous latent variable that is posited to underlie the discrete categories of the dependent variable that we actually observe. Predicted probabilities of each category occurring may then be obtained corresponding to each observation. In our case, we had, for each fixture, estimated probabilities for home win, draw and away win. In the calculation of count-R^2, the 'predicted outcome' is the category with the highest probability and this predicted outcome is then compared with observed outcome to obtain the proportion of observations 'explained'. Our regression equation on actual results in this sense 'explains' 47.9% of the observations. However, the figure cannot be taken too seriously. Though it is higher than the tipster strike-rates reported in Section 4 above, it is achieved only by 'predicting' that 96.5% of matches will be won by the home team. This reflects the substantial magnitude of home advantage in professional soccer since, notwithstanding the identification of several variables that might significantly pull the balance of advantage towards the away team, the single most probable outcome in almost every match remained a home win. As noted above, calling home win in almost every match is not an option for tipsters, so they cannot be expected to achieve strike-rates comparable with the proportion of observations 'explained' by the results regression equation.

As a first means of assessing the extent to which tipsters take adequate account of the patterns identified by regression analysis of actual outcomes, we adopted the following procedure, described for the *Daily Mail* but applied in turn to *The Mirror* and *The Times* as well. Of the 1694 matches in the data set, the *Daily Mail* called 954 in favour of the home team; 417 were forecast to be draws; the away side was tipped in the remaining 323 fixtures. We sorted the 1694 observations according to how probable a home win was according to the estimated real-results regression equation; probabilities of a home win ranged from 0.789 to 0.117. Taking the *Daily Mail* tipster predictions as constraints, we allocated as 'home predictions' the 954 matches with the highest probability of a home win; we called draws for the next 417 fixtures ordered in descending order of probability of home win; and the residual 323 contests were termed 'away predictions'. Our allocations were then compared with on-the-field results. 44.0% of match outcomes were 'explained' by this procedure of using the regression equation but imposing the same home–draw–away distribution as in the *Daily Mail* data set. The corresponding figures for *The Mirror* and *The Times* were 43.4 and 43.6%, respectively. All these figures were higher than the tipster strike-rates recorded earlier, strengthening a suspicion that the newspaper experts were not modelling as successfully as they might the role of the data embodied in public information. Thus, even when imposing the constraint that the set of results had to have the same distributions of outcomes as the tipster's forecasts, the statistical model was still more successful than the expert.

Of course, it is possible that the supposed superiority of the regression-based 'forecasts' over the subjective forecasts is attributable to their having the advantage of being made ex post rather than ex ante. The findings therefore needed to be checked against out-of-sample forecasting performance. Accordingly, we collected a fresh sample which included all matches played in the English leagues on the 12

Saturdays from October 10 to December 26, 1998. However, missing copies in our newspaper archive (microfilm records were not yet available) limited our comparisons to 139 matches in the case of *The Mirror* and 262 in the case of the *Daily Mail* (as against 367 for *The Times*).

We followed our previous procedure, employing the regression equation estimated from our main sample (reported in Table 2 above) to generate sets of forecasts made within the constraint that our home–draw–away proportions had to match those found in the columns of the relevant newspaper tipster over the second sample period. *The Times* and *The Mirror* achieved slightly higher strike-rates than the corresponding regression-based forecasts. However, the average tipster strike-rate across the three newspapers (weighted by sample size) was again less than the regression strike-rate: 41.8% compared with 43.4%.

Given this evidence, we returned to our original sample to conduct a more formal test of whether tipsters made full use of public information. We re-estimated the equation relating actual results to public information but with the addition of dummy variables representing the picking out of a game as 'draw' or 'away win' by the particular tipster. The results for each of the tipsters are exhibited as Table 3.

We again report our findings using the *Daily Mail* as reference point. In Table 1, we reported a regression of match results on *Daily Mail* tips. In Table 3, we regress match results on *Daily Mail* tips and on 13 variables representing public listed information. The model in the first regression is therefore nested within that used in the subsequent regression. Hence, nested hypothesis testing is appropriate. The log-likelihood value for the first equation (where the coefficients on the 13 public information variables are constrained to be zero) is compared with the log-likelihood value for the second (unrestricted) equation.

The quantity $2(L_F - L_R)$ is distributed as chi-squared with k degrees of freedom, where L_F and L_R are the log-likelihood values for the full and restricted models, respectively and k is the number of coefficients constrained in the restricted version. Here $2(L_F - L_R) = 35.64$. The critical value of chi-squared with 13 degrees of freedom (5% level) is 22.36. The restrictions are therefore decisively rejected: the 13 extra variables contribute information relevant to match outcomes that is not included in the *Daily Mail* tips. It appears that the *Daily Mail* tipster

Table 3
Ordered logit regression analysis[a]

	(1) Daily Mail		(2) The Mirror		(3) The Times	
	Coefficient	Impact	Coefficient	Impact	Coefficient	Impact
Constant	0.198	−0.049	0.197	−0.049	0.162	−0.040
	(0.96)		(0.96)		(0.78)	
Draw tip	0.046	−0.112	0.126	−0.031	0.143	−0.036
	(0.38)		(1.06)		(1.14)	
Away tip	0.142*	−0.088	0.175	−0.044	0.181	−0.045
	(2.50)		(1.14)		(1.43)	
Log-likelihood	−1755.36		−1957.78		−1757.45	

[a] Dependent variable: match results; independent variables: 13 public listed information variables and tipster forecasts; 'Impact' shows the marginal effect on home win; absolute t-statistics shown in parentheses; * indicates significance at the 5% level.

does not fully exploit all the information available in his own newspaper.

Results for *The Mirror* and *The Times* are similar. Test statistics in these cases were (respectively) 36.44 and 38.70. The formal tests therefore provide further support for our suspicion that there is some deficiency in the way in which each of the three tipsters employs the public listed information. This deficiency refers specifically to the role of variables which are widely publicised in newspapers and which add significantly to the explanatory power of our regression results in Table 2.

6. Tipsters and private information

Of course, any deficiencies in tipsters' processing of public information could be compensated if they were to employ insights attributable to their command of private information or semi-public information (information in principle in the public domain but not cheaply accessible or interpretable to non-specialists). In this section, we implement further nested hypothesis tests to illuminate this possibility. We are unable to ascertain what semi-public information the tipster may possess nor the effectiveness with which this information is processed. We can, though, assess whether extra insights offered, implicitly or explicitly, by the tipsters add explanatory power to the role of variables representing public information in our regression equation for actual results.

Consider the 'results' equation in Table 2 where match result is regressed on public information. This is a restricted version of the equation in Table 3 where match result is regressed on public information and on two dummy variables reflecting the *Daily Mail's* tips. Once again, we can test the restrictions. In this case, the test statistic is 6.54 against a critical value (2 degrees of freedom) of 5.99. The restrictions are therefore rejected and the *Daily Mail* expert appears able to add to the ability of the public information to account for the pattern of match results. However, the values of the test statistic in similar procedures including *Mirror* and *Times* forecasts were 1.70 and 2.36, respectively, and one could not therefore reject the hypothesis that the conclusions in these newspapers provided no information supplementary to that captured in the public information variables (The regression of match results on tipster forecasts yielded significant coefficients (Table 1) but for *The Mirror* and *The Times* the coefficients became insignificant in the presence of public information variables (Table 3)). Inspection of *t*-statistics implies therefore that *Mirror* and *Times* tip dummies may contain no independent information. However, jointly assessing the significance of 'tip draw' and 'tip away' in the nested hypothesis test reported here is necessary because we know that public information plays a substantial role in determining tips (Table 2) and collinearity between tip dummies and public information variables could therefore account for the rejection of the former in *t*-tests where both sets of variables are present).

It is possible that a consensus of tipsters' forecasts might perform better than any one individual tipster. Some studies suggest otherwise. Clarke (1993) set up a computer program to predict Australian Rules scores and found that this yielded superior forecasts to individual and consensus tips, although some of the 'tipsters' appeared to be more concerned with newspaper publicity by means of 'controversial' selections than with more objective assessment. Clemen and Winkler (1985) examined the independence of a group of North American sports forecasters and found a correlation between forecasts of 0.8 which was equivalent to only 1.25 independent forecasts. In our context of soccer predictions, we might similarly anticipate a problem of correlated forecasts and consequent lack of independence between tips-

Table 4
Ordered logit regression analysis[a]

	(1)		(2)	
	Coefficient	Impact	Coefficient	Impact
Constant	0.002	0	0.094	−0.023
	(0.02)		(0.44)	
Home tip	−0.109	0.027	0.130	−0.032
	(0.84)		(0.89)	
Draw tip	0.283	−0.071	0.270	−0.067
	(1.86)		(1.76)	
Away tip	0.600*	−0.150	0.521*	−0.130
	(3.91)		(3.15)	
Log-likelihood	−1769.6		−1753.7	

[a] Dependent variable: match results; independent variables: (1) consensus tipster forecast; (2) consensus tipster forecasts and three public information variables as shown in Table 2; 'Impact' shows the marginal effect on home win; absolute *t*-statistics shown in parentheses; * indicates significance at 1% level.

ter selections, given the findings in Table 2. On the other hand, if each of the three tipsters captures unique insights then the consensus should outperform any individual forecast.

To test for the effectiveness of a consensus forecast, we identified cases where two or more of the three tipsters agreed on the outcome of a match. Thus, if any two tipsters call a draw we code the dummy variable, *draw tip* as 1, and similarly for *home tip* and *away tip*. The excluded category here is no consensus, i.e. three different selections by the tipsters. The regression results using this consensus forecast are shown in Table 4. As before, we test for whether the consensus fully exploits all the public information present in our performance variables. The likelihood ratio chi-squared test statistic is 32.62. Set against a critical value of 22.36 for 13 degrees of freedom, this suggests that the coefficients on the added public information variables are jointly significantly different from zero. Hence, the public information variables add extra explanatory power to the consensus forecast in the determination of actual results.

Does a consensus forecast yield explanatory power over and above a regression of actual results on public information indicators only? Here, the likelihood-ratio test for addition of the three extra tipster dummy variables gives a chi-squared test statistic of 10.58. With a critical value for three degrees of freedom of 7.82, we cannot reject the hypothesis that the coefficients of the extra tipster variables are jointly significantly different from zero. Hence, we conclude that the consensus forecast adds something to our public information variables even though two of the tipsters considered on an individual basis do not.

7. Summary and conclusions

Our first statistical tests were concerned with whether the three newspaper tipsters sampled were able to predict more soccer games successfully than would someone employing a random process based on a very limited knowledge of soccer. All three tipsters passed these tests and therefore possessed some appropriate knowledge of soccer.

However, when we modelled the process by which tipsters took into account the information on teams' strength featured on soccer betting pages in two of the newspapers, we noted some differences from the process by which the results themselves appeared to be determined. From the results of likelihood-ratio tests conducted after estimating equations regressing match results on individual tipster forecasts (with and without the items of public listed information), these differences seem to be sufficient for all three tipsters to be judged inadequate in their processing of public listed information.

Finally, we implemented likelihood-ratio tests designed to find whether the tipsters possessed enough independent information for their tips to remain statistically significant predictors of match results even in the presence of publicly listed data. Only one of the tipsters appeared to exploit such private or semi-public information. Hence, there is no overwhelming evidence here that the predictions of match results from our regression models are inferior to those made by the experts. This conclusion is consistent with the received wisdom on the effectiveness of expertise in forecasting, as summarised in Makradakis et al. (1998). As a caveat, we found that taking into account the presence of any consensus among tipsters could add to the effectiveness of regression forecasts based on public information. We conclude that while individual tipsters' guidance is better than no guidance, the expertise they can claim to offer is limited.

Acknowledgements

The bulk of the research assistance was carried out thoroughly by Joanne Lord whose role is duly acknowledged. We are also grateful to Pam Carroll and Daniel Eisen for additional help. The comments of two anonymous referees and an editor facilitated improvements in the paper. The usual caveat applies.

References

Bird, R., & McCrae, M. (1987). Tests of the efficiency of racetrack betting using bookmaker odds. *Management Science 33*, 1552–1562.

Clarke, S. R. (1993). Computer forecasting of Australian rules football for a daily newspaper. *Journal of the Operational Research Society 44*, 753–759.

Clarke, S. R., & Norman, J. M. (1995). Home advantage of individual clubs in English soccer. *Journal of the Royal Statistical Society Series D (The Statistician) 44*, 509–521.

Clemen, R. T., & Winkler, R. L. (1985). Limits for the precision and value of information from dependent sources. *Operations Research 33*, 427–442.

Courneya, K. S., & Carron, A. V. (1992). The home advantage in sports competitions: A literature review. *Journal of Sport and Exercise Psychology 14*, 13–27.

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

Figlewski, S. (1979). Subjective information and market efficiency in a betting market. *Journal of Political Economy 87*, 75–87.

Forrest, D. K., & Simmons, R. (2000). Making up the results: The work of the Football Pools Panel 1963–97. *Journal of the Royal Statistical Society Series D (The Statistician) 49*, 253–260.

Goodwin, B. K., & Corral, L. R. (1996). Bettor handicapping and market efficiency in greyhound pari-mutuel gambling. *Applied Economics 28*, 1181–1190.

Goodwin, P., & Wright, G. (1993). Improving judgemental time series forecasting: A review of the guidance provided by research. *International Journal of Forecasting 9*, 147–161.

Greene, W. H. (1997). Econometric analysis, 3rd ed, Macmillan, New York.

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

Makradakis, S., Wheelwright, S., & Hyndman, R. (1998). Forecasting methods and applications, 3rd ed, Wiley, New York.

Pope, P. F., & Peel, D. A. (1989). Information, prices and efficiency in a fixed-odds betting market. *Economica 56*, 323–341.

Sharpe, G. (1997). Gambling on goals: a century of football betting, Mainstream, Edinburgh.

Webby, R., & O'Connor, M. (1996). Judgemental and statistical time series forecasting: A review of the literature. *International Journal of Forecasting 12*, 91–118.

Zavoina, R., & McElvey, W. (1975). A statistical model for the analysis of ordinal level variables. *Journal of Mathematical Sociology 2*(Summer), 103–120.

Biographies: David FORREST is Lecturer in Economics, University of Salford. He has research interests in the economics of sport, the arts and gambling and in valuation issues in cost–benefit analysis. Recent outlets include *Oxford Economic Papers*, *Journal of the Royal Statistical Society*, *Journal of Transport Economics and Policy* and *Journal of Gambling Studies*.

Robert SIMMONS is Lecturer in Economics and Director of the Centre for Sports Economics, at the University of Salford. His research interests are in labour economics, economics of professional team sports and economics of gambling. He has published in *Economica*, *Bulletin of Economic Research*, *Applied Economics* and several other economics journals. He is a member of the Board of Editors of the *Journal of Sports Economics*.