Modelling the Behaviour and Performance of Australian Football Tipsters

Matthew Amor and William Griffiths

University of Melbourne

February 3, 2003

Abstract

The forecasting performance of newspaper tipsters who predict the outcomes of English soccer matches has recently been assessed by Forrest and Simmons (2000). In this paper we extend their work to forecasts of AFL matches by five newspaper tipsters in Melbourne, Australia. These tipsters are assessed against some simple performance criteria as well as against the forecasts from a logit model designed to predict match outcomes. We find that most tipsters satisfy simple performance criteria. However, they do not fully exploit publicly available information and only two appear to successfully use independent information relevant to match outcomes.

1. Introduction

Forecasts of uncertain future outcomes are abundant in today's society. Whether it is for the weather, in business, macroeconomics or sport, predictive advice is readily available as a valued source of guidance for future events. Professional forecasts can come in the form of independent and specialised advice offered to individuals, or, alternatively, as widely accessible advice released in the media. In Australia, forecasting the outcomes of Australian Football League (AFL) matches is a popular pastime. All branches of the media, radio, television and newspapers, have so-called expert tipsters who make weekly predictions. In the city of Melbourne, generally recognized as the home of AFL football, almost all workplace offices have their own 'footy-tipping' competitions, with a particular employee assigned the task of running the competition. Legalized betting on AFL matches is possible through TAB betting shops, bars and the internet. This paper is concerned with analysing the quality of forecasts made by newspaper tipsters for the outcomes of AFL matches. These experts offer predictions of match outcomes for the 22 weeks of an AFL season. By modelling their performance and behaviour, we assess the value of the information provided by these expert tipsters. If their predictions offer valuable information, bettors incorporating this information into their own predictions can enhance their probability of making profits from gambling or winning tipping competitions.

The various tests and criteria we use for assessing the tipsters' performance can be categorized according to the following hypotheses:

1. There is no relevant information captured in the tipsters' forecasts.
2. Public information does not contribute additional information relevant to match outcomes beyond the information contained in tipsters' forecasts.
3. The tipsters' forecasts do not incorporate private information beyond the information already contained in public information variables.

To test whether the tipsters' forecasts contain *any* relevant information we compare the success rates of the tipsters to the success rates that would be obtained if some simple naïve forecasting strategies were adopted. In addition, we use a simple logit model to relate match outcomes to tipsters' forecasts. Nonsignificance of the forecasts in such an equation suggests they do not contribute any information about match outcomes. The results from these assessments are reported in Section 3.

For testing the second and third hypotheses, public information variables need to be defined and related to both match outcomes and tipsters' forecasts. The public information variables are described in Section 4; a comparison is made between the coefficients of the estimated logit model where match outcome is the dependent variable with the coefficients of the logit models designed to explain the tipsters' forecasts.

In Section 5 the logit models estimated in Section 4 and variations of them are used to test the second and third hypotheses about the contributions of public and private information to the information content of tipsters' forecasts. Consider a logit model where match outcomes are dependent on tipsters' forecasts and the public information variables. If the public information variables do not contribute any more information on match outcomes over and above that already provided by the tipsters' forecasts, then the tipsters' forecasts must embody all relevant public information. Conversely, significance of the public information variables implies that they contain relevant information not included in the forecasts. Also, if the tipsters' forecasts have significant explanatory power even when the public information variables are included, we can conclude that the tipsters have relevant private information not contained in the public information variables. Nonsignificance of a tipster's forecast suggests that a tipster has no special private information. Another way to assess the ability of the tipsters to process public and private information is to compare predicted match outcomes from the logit model with the tipsters' predictions. Comparisons of this kind also appear in Section 5.

Our methods for assessing AFL tipsters are modifications of those of Forrest and Simmons (2000) who examined the performance of newspaper tipsters in English football. They compared the performance of these tipsters to some naïve forecasting strategies and assessed the public and private information content of the tipsters' forecasts. There is, in general, an impressive body of literature on modelling sports outcomes, and the use of such models for forecasting and the development of betting strategies. Some examples of sports and their related publications that have appeared in the literature are the U.S. National Football League (Craig and Hall 1994, Gray and Gray 1997, Glickman and Stern 1998, Vergin 2001), American college football (Lebovic and Sigelman 2001), English football (Lee 1997, Crowder et al 2002), major league baseball in the U.S. (Albert 1994, Gandar, Zuber and Lamb 2001, Schall and Smith 2000), tennis (Klaassen and Magnus 2001), cricket (de Silva et al 2001) and horse racing (Gandar, Zuber and Johnson 2001). Studies involving Australian football include Stefani and Clarke (1992), Clarke (1993) and Brailsford et al (1995). Other examples that consider modelling or betting strategies for a number of sports are Gill (2000), Knorr-Held (2000), Mallios (2000) and Haigh (2001).

2. Data

Tipsters' predictions of AFL matches are published once a week in daily metropolitan newspapers. The sample used for this study was obtained from daily Melbourne newspapers. Data pertaining to matches played in the premiership seasons of 1999 and 2000 was collected for rounds 5 to 22. The first four rounds from each season were omitted so that a sufficient number of rounds had been played to provide adequate information on short-term form. The remaining 18 rounds, for each of two seasons, with 8 games per round, lead to 288 match outcomes. Two drawn results were omitted, leaving a total of 286 observations. Each match is viewed as a match between a lower-ranked team and a higher-ranked team, where a team's ranking is based on its position on a ladder. Ladder position is defined in terms of a team's win-loss record, with total points for and against used to rank teams who have the same win-loss record. Match outcomes are recorded as a win for a higher-ranked team or a win for a lower-ranked team. Similarly, a tipster's forecast is recorded as a tip for a higher-ranked team or a tip for a lower-ranked team. The match result, *ex ante* ladder positions, prior results, the venue and tipsters' predicted match outcomes were recorded for all 286 observations.

A sample of five expert tipsters was chosen: Michael Sheehan, Gerard Healy and Mark Robertson from the Herald Sun newspaper, and Robert Walls and Dermott Brereton from The Age newspaper. Walls, Healy and Brereton all had outstanding football careers whereas Sheehan and Robertson have won respect in football circles for their analytical skills. The odd number of tipsters implies a consensus forecast, defined as the most popular tip, can always be obtained from the individual tipster predictions. On average, over the 1999 and 2000 seasons the five selected tipsters correctly forecast 65.7% of matches. The average for all tipsters from the Herald Sun and The Age newspapers was 65.1%. Hence, the five sample tipsters offer a consistent representation of the population of expert tipsters. In what follows the percentage of matches correctly forecast will be referred to as a tipster's success rate or strike-rate.

3. Do Tipsters' Forecasts Contain Relevant Information?

Assessing the performance of the tipsters requires the setting of some explicit criteria. Having established some benchmarks, what then constitutes a satisfactory forecasting success rate? What standards should the tipsters satisfy? We answer these questions in terms of a number of criteria and benchmarks. Our initial tests are weak ones that will focus on a tipster's ability to offer guidance to bettors with little knowledge of football. In a later section more stringent tests of a tipster's performance will then explore the use they make of publicly available information and the additional use of private information.

Given that someone knows nothing about football, a relevant criterion is whether forecast strike-rates outperform purely random forecasting techniques. Someone who simply 'guesses' each match outcome would expect to predict correctly 50% of the time. Over the data set, the percentages of matches forecast correctly by the five tipsters, and from the consensus forecast, are:

Michael Sheehan	66.46%
Gerard Healy	63.99%
Mark Robertson	66.78%
Dermott Brereton	66.43%

Robert Walls	64.69%
Consensus	68.2%

Evidently, all tipsters' strike-rates are superior to guesswork. Also, a strategy of choosing the consensus forecast has performed better than following a single tipster.

Another simple strategy is to predict the higher-ranked team to win every match. Over the data set, the higher-ranked team won 67.8% of matches. Employing such a method yields a strike-rate superior to the strike-rates of all individual tipsters, but not superior to the strike-rate of the consensus forecast. Thus, if choosing the higher-ranked team is used as a benchmark strategy for assessing performance, the tipsters fail individually, but, collectively, they provide useful information. There are other relevant benchmarks, however. The tipster or selection rule with the highest strike rate is not necessarily the one that is providing the most valuable information. If the major concern is winning the office footy-tipping competition, then maximizing strike rate is a useful forecasting strategy. However, for betting on match outcomes, the most valuable information is that which correctly predicts a win by a lower-ranked team. Wins by lower-ranked teams will generally provide larger payouts than wins by higher-ranked teams. A strategy of always betting on the higher-ranked team is likely to lose money, even with a success rate of 67.8%.

The experts on average predicted the higher-ranked team to win on 73.4% of occasions. This was well above the actual winning rate of the 67.8%, suggesting the experts selected the higher-ranked team too regularly. On the other hand, since the experts managed an average strike rate of 65.7% when selecting 73.4% higher-ranked winners, and selecting 100% higher-ranked winners increases the strike-rate only marginally to 67.8%, the experts must have had some success forecasting lower-ranked winners.

These considerations can be formalized into another test for assessing the value of information provided by tipsters' predictions. Recognizing the advantages of picking lower-ranked winners, this benchmark strategy is to select the winner at random, subject to the restriction that the proportions of predicted higher-ranked winners and predicted lower-ranked winners correspond to the actual proportions of higher and lower-ranked winners found in practice. If higher-ranked teams are selected 67.8% of the time, and these teams win 67.8% of the time, and lower ranked teams are selected 32.2% of the time and these teams win 32.2 % of the time, the expected success rate of this strategy is

$$(0.678)^2 + (0.322)^2 = 0.563 \quad \text{or} \quad 56.3\%$$

If the expert tipsters are constrained to select 73.4% higher-ranked winners and 26.6% lower-ranked winners, but they otherwise choose randomly (one higher-ranked winner is not preferred to any other), their expected success rate is

$$(0.678)\times(0.734) + (0.322)\times(0.266) = 0.583 \quad \text{or} \quad 58.3\%$$

Given their success rates are considerably higher than both 56.3% and 58.3%, we can conclude their tips provide additional information in the sense that their proportion of correctly selected lower-ranked winners is greater than the average. Tipsters' assessment of objective information is likely to be more useful early in the season, when team strengths are less clear cut, than later in the season. Within the data set, the higher-ranked teams win 64.1% of games in the first half of the season and 71.6% in the second half. As the season progresses, ladder positions better reflect a team's overall strength; each team has played a greater number of other teams and performance has been measured over a longer period of time.

The above criterion can be applied to each tipster individually. For example, Mike Sheehan predicts the higher-ranked team on 69.9% of occasions. If someone with no knowledge chose the higher ranked team this frequently, but otherwise chose randomly, their expected proportion of correct predictions would be

$$(0.699)\times(0.678) + (0.301)\times(0.322) = 0.571 \quad \text{or} \quad 57.1\%$$

Since Mike Sheehan's strike-rate is 66.5%, his football knowledge leads to a better success rate for picking lower-ranked winners than would be achieved from random selection. A similar result holds for the other tipsters. Their expected winning percentages, given their proportions of higher-ranked selections, and an otherwise random selection, are Gerard Healy 60.0%, Mark Robertson 58.9%, Robert Walls 58.1% and Dermott Brereton 57.7%.

Another way to assess whether tipsters' forecasts contain any useful information about match outcomes is to test whether they have any explanatory power in an equation used to explain match outcomes. A suitable equation for such a test is a logit model where match outcome is the dependent variable and a tipster's forecast is the explanatory variable. Match outcomes are assigned a value of 1 for wins by lower-ranked teams and 0 for wins by higher-ranked teams. Similarly, the explanatory variable is a binary variable equal to 1 when a tipster tips a lower-ranked team and 0 when a higher-ranked team is selected. Maximum likelihood estimates for the logit models for each of the tipsters and for the consensus forecast are reported in Table 1. The last row of this table contains the χ^2 values for the likelihood-ratio tests for testing the significance of the forecast variable.

Using both *t*-tests and likelihood ratio tests, the hypothesis that no relevant football knowledge is embodied in the forecasts is conclusively rejected at the 5% level in all cases except for Gerard Healy. Apart from Healy, the tipsters pass this weak requirement of possessing some AFL knowledge. It is interesting to note that the two analytically based journalists, Sheehan and Robertson, provide the predictions with the highest degree of significance (highest t and χ^2 values). Whether their approach is more methodical and/or technical is still to be determined. Potentially, applying higher proportions of casual, less technical knowledge might cause the three former players' tips to exhibit lower levels of explanatory power. Gerard Healy's poorer performance could be attributable to less of an ability to pick lower-ranked winners. Although his strike-rate was only 1.7% below the mean strike-rate of the five tipsters, he predicted the higher-ranked team 4.5% more often than the mean and 8% more often than Sheehan. As already discussed, forecasting the higher-ranked team regularly could mask a lack of knowledge of football. Intuitively, one would expect a tipster's forecasts to display more explanatory power if that tipster demonstrates an ability to successfully predict both higher and lower-ranked winners.

Overall, we conclude that all tipsters pass the simple benchmarks established to assess whether their forecasts contain relevant information about AFL. With the exception of Healy, this conclusion is supported by estimated logit models designed to test whether the tipsters' forecasts can help explain match outcomes.

4. Modelling with Public Information Variables

In this section we are concerned with estimating two types of logit models. In the first the probability of a win by a lower-ranked team is related to a number of 'public-information' variables that are used generally to assess team performance. This model can be viewed as one that can be used to predict match outcomes. In the second type of logit model the probability of a tipster tipping a lower-ranked team is related to the same variables. Models of this kind are estimated for each of the tipsters. They are designed to summarize the way in which tipsters process relevant information to come up with a forecast. The results for the different models and different tipsters are compared in this section. In the next section the models are used to test hypotheses about the way in which tipsters use public and private information.

Our use of logit models to capture the behaviour of the tipsters is not intended to suggest that the tipsters themselves have their own logit models that they use for making predictions. Milton Friedman's 'as if' theory (1966) implies a tipster's forecast is determined by passive adaptations to external circumstances that lead to the same forecasts that would have been achieved by complex modelling techniques. In context, Friedman's hypothesis stipulates that tipsters determine their forecasts 'as if' they know the complicated formulas that yield the optimal prediction of a match outcome by calculating formulas that link various team strength indicators to a team's actual performance. Confidence in this hypothesis is not founded on a tipster's use of the above process, but rather emanates from a belief that, unless they were in some way capable of reaching essentially the same result, the tipsters would not be football experts. Nevertheless, it is inevitable that speculative beliefs as well as objective judgement will play some role in the tipsters' forecasts. Estimating the logit models should shed some light on the relative contributions of these factors.

Selecting the most relevant public information variables to make informed forecasts of AFL matches requires considerable thought, questioning and research. Those we believe to be important are described below. Clearly, other researchers may settle on a different list. The estimated coefficients of these variables in the various logit models are reported in Table 2.

HRSTF and *LRSTF*: These variables represent the short-term form of the higher and lower-ranked teams, respectively. A team's short-term form is based on results from the previous four matches and is calculated as follows. A winner by more than 30 points receives 3 points, winners by less than 30 points receive 2 points, a loser by less than 30 points is allocated 1 point and a loser by more than 30 points receives no points. A draw receives 1.5 points. Given the previous four rounds are being considered, the short-term performance indicator takes a value between 0 and 12. From Table 2 we see that both these variables are significant in the equations for all tipsters. Recent team performance is a good indicator of perceived future performance. However, for the actual match outcome, only the short-term form of the higher-ranked team is significant. The coefficient of short-run form for the lower-ranked team is much smaller than it is in the tipsters' equations and it is not significant. It appears that the tipsters place undue weight on the recent form of the lower-ranked team.

HRR and *LRR*: These variables are used to capture the 'reputations' of the higher and lower-ranked teams. Reputation is defined as the weighted aggregate number of games won by the higher and lower-ranked teams over the past two seasons. One point is awarded for winning a regular season game, two points for a final and four points for a grand final. 'Reputation' is intended to measure historical information embedded implicitly in the tipsters' forecasts. The estimates in Table 2 suggest that reputation has no influence on match outcomes; the coefficients of *HRR* and *LRR* are small and insignificant. On the other hand, the estimates for two of the tipsters (Sheehan and Robertson) are significant for the reputation of the higher-ranked team. The reputation of the lower-ranked team has little impact on tipster forecasts, except perhaps for Walls, whose coefficient of *LRR* is significant at the 10% level. Intuitively, the coefficients on the reputation variables display relatively low levels of significance for two reasons. The high importance tipsters place on recent form dictates that performances dating back to earlier seasons will have a reduced impact. Secondly, each tipster's implicit approximation of the reputation variable will be rather indistinct. They may draw opinions on reputation differently, say by according higher weights to single events in a season rather than overall season performance, and they may assess reputation over different time periods. Also, historical head-to-head outcomes for two particular teams could be relevant. Thus, the potential to misspecify the reputation variable is high.

HRHF, HRAF, LRHF & LRAF: These four variables are designed to measure the 'home form' and the 'away form' of the higher and lower-ranked teams. All games played at neutral venues are assigned a value of 0. The remaining matches are played at stadiums where one team has an acknowledged home-ground advantage. Teams playing in matches where a home-ground advantage exists are assigned a numerical value based on their previous performances at 'similar' venues. For instance, a team playing away at an interstate venue is assigned an aggregated value from its past five games played at interstate venues. Using the same points system as the short-term form variable, teams can earn from 0 to 3 points from each game. With the exception of matches played or to be played on neutral venues, the variables, when aggregated, will have a numerical value ranging from 0 to 15. Home and away-form has generally been recognized as a team performance indicator because of the perceived advantage of playing at home. Crowd support, familiarity with physical characteristics of particular stadiums and travel fatigue have been given as reasons for a home-ground advantage (Courneya and Carron 1992). Stefani and Clarke (1992) highlight the difficulty in winning matches interstate. Between 1980 and 1986 home teams (with home-ground advantage) won 57.2% of matches. National expansion of the league in 1987 saw the introduction of Perth and Brisbane teams. The home teams' winning percentage increased to 62.4% between 1987-89. Courneya & Carron (1992) attribute this change to the effects of unfamiliar routines (staying in hotels) and hostile home fans. The results in Table 2 suggest the tipsters place substantial weight on the home-form of the higher-ranked team (*HRAF*). Three of its estimated coefficients are significant at the 5% level, indicating that the past five performances of the higher-ranked home team are exploited by three of the five tipsters. Two tipsters consider the lower-ranked team's away-form (*LRAF*) to be an important determinant of match outcomes. The coefficients on the away-form of the higher-ranked team and the home-form of the lower-ranked team are insignificant, except for Walls' coefficient on *LRHF*, which is significant at the 10% level. Surprisingly none of the home and away variables have a significant impact on the actual match outcome.

$HRL10+$, $HRL7-9$ and $HRL4-6$: Relative team strength is represented by the difference in the relative ladder positions of the higher and lower-ranked teams. Rather than use a single variable to measure this difference, four bands of differences were set up, namely $1-3$, $4-6$, $7-9$ and $10+$. The dummy variables $HRL10+$, $HRL7-9$ and $HRL4-6$ were assigned values of one for matches that fell within their respective bands and $1-3$ was chosen as the reference band. The effect of relative ladder positions on a match outcome or a perceived match outcome is unlikely to be a linear one. The dummy variables are used to try and capture any nonlinearities. One would expect the probability of a win by a lower-ranked team to be smallest when $HRL10+=1$, to be slightly higher when $HRL7-9=1$, and to be even higher for $HRL4-6=1$. For this outcome to hold, the coefficients of these variables should all be negative, and, using obvious notation, they should satisfy the following inequality

$$\beta_{10+} < \beta_{7-9} < \beta_{4-6}$$

From Table 2, we see that Sheehan's and Wall's estimates satisfy this expectation, and those of Brereton almost do so. Sheehan, Healy and Brereton all have some coefficients that are significant at a 10% level of significance. Thus, the results suggest some of the tipsters exploit ladder position in the framing of their forecasts, but some do not; the general impact of these variables is not a uniform clear cut one. Some wrong signs and the insignificance of some of the coefficients highlights the unpredictable nature of the AFL. This unpredictably is also borne out by the estimated model for actual match outcomes. In this case the coefficients of the dummy variables all have the wrong sign.

Our comparison of the logit model for the actual match outcomes with those for tipster forecasts highlights several differences in the determination of the two processes. Intuitively, random on-field events make the actual results less methodical than the tipsters' forecasts. However, the comparison serves the important purpose of demonstrating where tipsters over emphasise and neglect particular variables in predicting match outcomes. The estimated model for match outcomes showed recent form of the higher-ranked team to be highly significant. Accordingly, the tipsters also place substantial weight on the recent form of the higher-ranked team. The estimates associated with the remaining variables show that the tipsters do use a number of the public information variables to formulate their forecasts, but they do not do so in a manner that is consistent across all tipsters, or necessarily consistent with the impact of the public information variables on the actual match outcomes. The additional uncertainty associated with match outcomes relative to predicted outcomes is likely to be the major reason for the lack of significance of many of the coefficients in the match-outcome equation.

5. Assessing the Public and Private Information Content of Tipsters' Forecasts

In this Section we assess whether the tipsters' forecasts effectively embody public and private information. Three distinct assessments are made. Public information content is considered first by testing the explanatory power of public information variables in logit models for each tipster. Private information content is then assessed by testing the relevance of the tipsters' forecasts in the same logit models. The third assessment makes comparisons of the predictive ability of a logit model, and different decision rules, with the success rates of the tipsters.

Predictions of future events in all fields invariably draw on a large proportion of public information. Assessing its role in determining tipster forecasts is the first task in this section. If the tipsters are found to embrace public information efficiently, then bettors who are aware of team strength indicators but do not have the time to process the data could benefit from following tipsters' forecasts. To test the hypothesis that team strength indicators do not add predictive power above and beyond the tipsters' forecasts, logit models where match outcomes (the probability of a lower-ranked winner) are related to a tipster's forecasts, and the public information variables, are estimated. A likelihood ratio test is used to test the hypothesis that the coefficients of the 11 public information variables are all zero. Failure to reject this hypothesis suggests the information in the public information variables is already embodied in a tipster's forecast and hence considering only that tipster's prediction is an adequate strategy. The test was performed on six logit models, one for each of the tipsters and one for the consensus forecast. The 5% critical value for a χ^2 distribution with 11 degrees of freedom is 19.67. The test values obtained are:

Michael Sheehan	26.37
Gerard Healy	29.03
Mark Roberston	26.01
Dermott Brereton	26.97
Robert Walls	25.69
Consensus	25.96

The hypothesis that the public information variables have no additional information about match outcomes is decisively rejected for all tipsters and for the consensus forecast. This test outcome suggests the 11 team-strength variables contain information that accounts for the pattern of match outcomes and which is not contained in the tipsters' forecasts. A single 'tipster variable' is not sufficient to capture the information in all the public information variables.

Not all information embodied in a tipster's forecast is publicly available. For instance, tipsters could possess knowledge of an injury to a key player or obtain technical performance indicators (such as recent tackling or marking averages) showing one team having a distinct advantage in a vital facet of the game. Such events are regarded as 'broken leg cues' and provide opportunities for contextual information to be used (Webby and O'Connor 1996). We are interested in whether tipsters use such events, in addition to public information variables, to help frame their forecasts. Effective use of contextual information potentially gives a tipster's forecasts additional explanatory power above and beyond that from modelling of public information variables only. Despite being able to identify the presence of private information, one can only speculate as to the content and the processing efficiency of such information. To test whether each of the tipsters effectively uses private information in their forecasts, we return to the logit models where match outcomes are related to a tipster's forecasts, and the public information variables. If a tipster's forecast does not contain any information on match outcomes that is not already contained in the public information variables, then the tipster's forecast is expected to be insignificant in the estimated logit model. Conversely, if a tipster's forecast does add significant explanatory power to the estimated model, it suggests that tipster is effectively using private information. The tests are performed using likelihood ratio tests; the 5% critical value for a χ^2 distribution with 1 degree of freedom is 3.84. The test values obtained are:

Michael Sheehan	6.47
Gerard Healy	0.00
Mark Roberston	4.47
Dermott Brereton	0.68
Robert Walls	1.22
Consensus	3.57

For two of the five tipsters the null hypothesis of no private information is rejected at the 5% level; it appears only Sheehan and Robertson's forecasts increase the explanatory power above and beyond the ability of the public information variables to account for the pattern of match results. Checking the corresponding *t*-statistics is an alternative approach. Sheehan and Robertson's forecasts remain statistically significant at the 5% level in the presence of public information variables; the other tipsters' forecasts continue to be insignificant. The career journalists appear able to effectively employ judgemental analysis in their forecasts. In contrast, all three ex-footballers appear unable to effectively capture independent information in their predictions. In the model with the consensus forecast the χ^2 value of 3.57 is below the critical value of 3.84 at the 5% level, but above the 10% critical value of 2.71. The consensus forecast improves the explanatory power of public information variables somewhat, even though the forecasts of three of the five tipsters do not.

Another way to assess the public and private information content of the tipsters' forecasts is to compare their success rates with the performance of forecasts obtained from an estimated logit model. Comparisons of this kind are an extension of those pursued in Section 3. In Section 3 the tipsters' performance was assessed against what we called naïve forecasting strategies. In what follows, the assessment is relative to model-based forecasts. Consider the logit model that relates match outcomes to the public information variables. One decision rule for forecasting is to predict a lower-ranked winner when the probability $\hat{p}$ estimated from the logit-model is greater than 0.5, and to predict a higher-ranked winner when $\hat{p} < 0.5$. If the in-sample $\hat{p}$'s are computed, and this decision rule is used to forecast, the logit model accurately predicts 68.1% of matches, an outcome superior to that of all individual tipster forecasts and just below the consensus forecast strike-rate of 68.2%. However, choosing 0.5 as the forecast threshold for $\hat{p}$ assumes picking lower-ranked winners has the same value as picking higher-ranked winners. As explained in Section 3, correctly picking a lower-ranked winner is likely to be more valuable, both in terms of information content and a potential betting payout. To accommodate the more valuable lower-ranked forecasts, it is reasonable to assume that the tipsters implicitly use a forecast threshold below 0.5. The fitted logit model predicted a higher-ranked win in 88.3% of matches.

To capture implicit forecast thresholds ($\bar{p}$'s) for each of the tipsters, we used the logit models estimated for each tipster, where their tips were related to the public information variables. From each estimated model, the in-sample predictions ($\hat{p}$'s) were calculated and a $\bar{p}$ was set such that 72.4% of the forecasts were higher-ranked winners. That is, 72.4% of the $\hat{p}$'s were such that $\hat{p} < \bar{p}$. The percentage 72.4% was chosen because it is the median proportion of higher-ranked winners from the five tipsters. It is equivalent to 207 higher-ranked forecasts from the 286 observations. The in-sample $\hat{p}$'s from the logit model where the dependent variable is match outcome were compared to each of the tipsters' thresholds to obtain the logit-model based forecasts that would be obtained using each of the tipsters' thresholds. Comparing these forecasts with actual outcomes produced the following strike-rates.

Michael Sheehan	62.59%
Gerard Healy	65.38%
Mark Robertson	63.64%
Dermott Brereton	64.69%
Robert Walls	63.29%

Apart from the results for Gerard Healy, these logit-model generated strike-rates are inferior to the tipsters' actual strike-rates. They suggest that the majority of tipsters' predictions capture some contextual information beyond their implicit modelling of public information variables.

6. Concluding Remarks

Any conclusions one makes about the performance of the selected AFL tipsters depends on the criterion or benchmarks against which their performance is assessed. All tipsters outperform a random selection rule that assumes no knowledge of football. However, when considered individually, each tipster does not outperform the simple selection strategy of always picking the higher-ranked team. This result changes when the tipsters' forecasts are considered collectively, through a consensus forecast; the consensus forecast does outperform the higher-ranked-team selection rule. If the objective of collecting information is to win the office footy-tipping competition, where the prize is for selecting the greatest number of winners, then always choosing the tipster-consensus forecast is the best strategy. It can be argued that an ability to pick lower-ranked winners is more valuable than an ability to pick higher-ranked winners. If one views the tipsters as being constrained to select a particular percentage of lower-ranked teams to win, and then assesses their performance conditional on this constraint, all tipsters do better than someone who randomly selects winners under the same constraint. A logit-model based approach for testing whether tipsters' forecasts contribute information about match outcomes suggests their forecasts are indeed valuable, although we were unable to establish this fact for Gerard Healy. In terms of statistical significance, the information from the two journalist tipsters (Sheehan and Robertson) provides more guidance about outcomes than that from the former football players. The introduction of public information variables into logit models (designed to explain tipster's forecasts and match outcomes) highlights the relative importance the tipsters place on these variables when framing their forecasts. The short-term form of the higher and lower-ranked teams, home form of the higher-ranked team and relative ladder position all have some relevance. Tests using logit models that contain both tipster forecasts and public information variables suggest that the tipsters' forecasts do not embody all the information in the public information variables. On the other hand, they also suggest that the forecasts from the two journalist tipsters contain private information not contained in the public information variables. Comparisons based on selection strategies that use estimated logit models supported earlier conclusions. When the potential extra value from picking lower-ranked winners is ignored, model-based forecasts outperform individual tipsters, but not the tipsters' consensus forecast. Allowing for the value of picking lower-ranked winners suggests all tipsters are performing well, with the possible exception of Gerard Healy.

References

Albert J. 1994. Exploring baseball hitting data: What about those breakdown statistics? *Journal of the American Statistical Association* **89**: 1066-1074.

Brailsford TJ, Easton SA, Gray PK, Gray SF. 1995. The efficiency of Australian football betting markets. *Australian Journal of Management* **20**: 167-195.

Clarke SR. 1993. Computer forecasting of Australian Rules football for a daily newspaper. *Journal of the Operational Research Society* **44**: 753-759.

Courneya KS, Carron AV. 1992. The home advantage in sports competitions: A literature review. *Journal of Sports and Exercise Psychology* **14**: 13-27.

Craig LA, Hall AR. 1994. Trying out for the team: Do exhibitions matter? Evidence from the National Football League. *Journal of the American Statistical Association* **89**: 1091-1099.

Crowder M, Dixon M, Ledford A, Robinson M. 2002. Dynamic modelling and prediction of English Football League matches for betting. *Journal of the Royal Statistical Society Series D (The Statistician)* **51**: 157-168.

De Silva BM, Pond GR, Swartz TB. 2001. Estimation of the magnitude of victory in one-day cricket. *Australian and New Zealand Journal of Statistics* **43**: 259-268.

Forrest DK, Simmons R. 2000. Forecasting sport: the behaviour and performance of football tipsters. *International Journal of Forecasting* **16**, pp: 317-331.

Friedman M. 1966. *Essays in Positive Economics*, 5th edn, The University of Chicago Press: Chicago.

Gandar JM, Zuber RA, Johnson RS. 2001. Searching for the favourite-longshot bias down under: an examination of the New Zealand pari-mutuel betting market. *Applied Economics* **33**: 1621-1629.

Gandar JM, Zuber RA, Lamb RP. 2001. The home field advantage revisited: a search for the bias in other sports betting markets. *Journal of Economics and Business* **53**: 439-453.

Gill PS. 2000. Late game reversals in professional basketball, football and hockey. *American Statistician* **54**: 94-99.

Glickman ME, Stern HS. 1998. A state-space model for National Football League scores. *Journal of the American Statistical Association* **93**: 25-35.

Gray PK, Gray SF. 1997. Testing market efficiency: Evidence from the NFL sports betting market. *Journal of Finance* **52**: 1725-1737.

Haigh J. 2001. The Kelly criterion and bet comparisons in spread betting. *Journal of the Royal Statistical Society: Series D (The Statistician)* **49**: 531-539.

Klaassen FJGM, Magnus JR. 2001. Forecasting the winner of a tennis match. CentER Discussion Paper No.2001-38: Tilburg University.

Knorr-Held L. 2000. Dynamic rating of sports teams. *Journal of the Royal Statistical Society Series D (The Statistician)* **49**: 261-276.

Lebovic JH, Sigelman L. 2001. The forecasting and determinants of football rankings. *International Journal of Forecasting* **17**: 105-120.

Lee AJ. 1997. Modelling scores in the premier league; Is Manchester United really the best? *Chance* **10**: 15-19.

Mallios WS. 2000. *The Analysis of Sports Forecasting: Modeling Parallels Between Sports Gambling and Financial Markets*. Kluwer: Boston.

Schall T, Smith G. 2000. Do baseball players regress toward the mean? *American Statistician* **54**: 231-235.

Stefani RT, Clarke SR. 1992. Predictions and home advantage for Australian Rules football. *Journal of Applied Statistics* **9**: 251-261.

Vergin RC. 2001. Overreaction in the NFL point spread market. *Applied Financial Economics* **11**: 497:509.

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

Table 1. Logit Model Estimates: Match Outcomes Dependent on Tipsters' Forecasts

	Michael Sheehan	Gerard Healy	Mark Robertson	Dermott Brereton	Robert Walls	Consensus
Constant	-1.046^{**}	-0.845^{**}	-0.989^{**}	-0.930^{**}	-0.950^{**}	-0.968^{**}
	(0.161)	(0.146)	(0.154)	(0.155)	(0.155)	(0.151)
Forecast	0.906^{**}	0.426	0.878^{**}	0.606^{*}	0.692^{*}	0.850^{**}
	(0.270)	(0.296)	(0.282)	(0.273)	(0.276)	(0.286)
LR χ^2 value	11.26	2.04	9.63	4.85	6.23	8.77

Standard errors shown in parentheses; single and double asterisks indicate statistical significance at the 5% and 1% levels, respectively.

Table 2 – Estimated Logit Models with Public Information Variables

	Michael Sheehan	Gerard Healy	Mark Robertson	Dermott Brereton	Robert Walls	Match Result
Constant	0.529 (0.804)	-1.076 (0.820)	0.325 (0.823)	-0.563 (0.766)	-0.164 (0.801)	-0.659 (0.696)
HRSTF	-0.186^{***} (0.064)	-0.237^{***} (0.069)	-0.304^{***} (0.069)	-0.166^{**} (0.065)	-0.210^{***} (0.067)	-0.162^{***} (0.058)
LRSTF	0.318^{***} (0.074)	0.240^{***} (0.077)	0.330^{***} (0.077)	0.295^{***} (0.075)	0.389^{***} (0.080)	0.084 (0.065)
HRR	-0.050^{**} (0.021)	-0.011 (0.020)	-0.058^{***} (0.022)	-0.007 (0.018)	-0.008 (0.018)	-0.011 (0.016)
LRR	-0.007 (0.021)	0.019 (0.021)	-0.006 (0.021)	0.017 (0.021)	0.044^{*} (0.023)	0.028 (0.018)
HRHF	-0.150^{**} (0.074)	-0.134 (0.086)	-0.120 (0.076)	-0.263^{***} (0.101)	-0.187^{**} (0.087)	-0.040 (0.061)
HRAF	-0.173 (0.115)	0.061 (0.099)	0.051 (0.094)	-0.003 (0.094)	-0.077 (0.099)	-0.006 (0.081)
LRHF	0.149 (0.100)	0.024 (0.090)	0.040 (0.088)	0.116 (0.088)	0.163^{*} (0.092)	0.069 (0.079)
LRAF	0.178 (0.117)	0.078 (0.134)	0.236^{*} (0.123)	0.266^{*} (0.148)	0.168 (0.136)	-0.060 (0.108)
HRL 10+	-1.544^{*} (0.799)	-1.936^{*} (1.088)	0.636 (0.581)	-1.044 (0.654)	-0.743 (0.647)	0.152 (0.510)
HRL 7-9	-0.743^{*} (0.443)	0.130 (0.442)	0.414 (0.434)	-1.047^{**} (0.479)	-0.529 (0.449)	0.663^{*} (0.386)
HRL 4-6	-0.012 (0.347)	-0.010 (0.376)	0.089 (0.381)	-0.058 (0.356)	-0.318 (0.374)	0.376 (0.339)
Log Likelihood	-138.77	-126.32	-135.47	-135.297	-132.37	-164.061