Modelling football match results and the efficiency of fixed-odds betting

John Goddard*

University of Wales Swansea

Ioannis Asimakopoulos

University of Wales Bangor

* **Corresponding author**

Address: Department of Economics
University of Wales Swansea
Singleton Park
Swansea SA2 8PP
UK

Tel: +44 (0)1792 295168
Fax: +44 (0)1792 295872
E-mail: j.a.goddard@swansea.ac.uk

Key words: Sports betting, efficiency, ordered probit

Abstract

An ordered probit regression model estimated using 15 years' data is used to model English league football match results. As well as past match results data, the significance of the match for end-of-season league outcomes; the involvement of the teams in cup competition; the geographical distance between the two teams' home towns; and the average attendances of the two teams all contribute to the model's performance. The model is used to test the weak-form efficiency of prices in the fixed-odds betting market, and betting strategies with a positive expected return are identified.

Modelling football match results and the efficiency of fixed-odds betting

1. Introduction

The predictability of match results is the main concern of research on the efficiency of sports betting markets. The recent applied statistics literature has focussed primarily on modelling goal scoring (Dixon and Coles, 1997; Rue and Salvesen, 2000; Crowder et al., 2002). Recently some econometricians have suggested modelling match results directly (rather than indirectly through scores) using discrete choice regression models (Forrest and Simmons, 2000a,b; Koning, 2000; Kuypers, 2000; Dobson and Goddard, 2001). A focus on match results rather than scores can be justified partly on grounds of simplicity: fewer parameters are required; estimation procedures are simpler; and the resulting models lend themselves to the inclusion of a variety explanatory variables. This paper presents a regression-based model to explain and predict match results that is more extensive and comprehensive than any previously developed, as regards both the range of explanatory variables that are incorporated, and the extent of the data set used to estimate the model.

Research into the efficiency of prices set by bookmakers in betting markets has provided a small but increasing contribution to the financial economics literature on market efficiency. Much of this literature focuses on racetrack betting, but betting on team sports match results has also attracted attention. Early researchers sought evidence of inefficiencies in the form of systematic biases in bookmakers' prices, such as home-away team or favourite-longshot biases. More recently forecasting models have been used to establish whether historical information available in previous match results can be extrapolated to formulate profitable betting strategies. This paper uses the model outlined above to test the weak-form efficiency of the prices quoted by high street bookmakers, and to identify potentially profitable betting strategies.

The paper is structured as follows. Section 2 reviews the literature on modelling and forecasting football match results. Section 3 describes the specification and estimation of an ordered probit regression model to explain and predict match results. Section 4 reviews the literature on betting market efficiency. Section 5 investigates the efficiency of the prices quoted by high street bookmakers over four English football seasons from 1998-9 to 2001-2, using both regression-based tests and direct economic tests of the profitability of betting strategies based on the model's evaluation of expected returns. Section 6 concludes.

2. Modelling football match results

A limited but increasing number of academic researchers have attempted to model match results data for football. Early contributions by Moroney (1956) and Reep et al. (1971) use the poisson and negative binomial distributions to model at an aggregate level the distributions of the numbers of goals scored per game. This aggregate approach precludes the generation of specific forecasts for individual matches based on information about the respective strengths of the two teams. By comparing final league placings with experts' pre-season forecasts, however, Hill (1974) demonstrates that individual match results do have a predictable element, and are not determined solely by chance.

Maher (1982) develops a model in which the home and away team scores follow independent poisson distributions, with means reflecting the attacking and defensive capabilities of the two teams. A full set of attacking parameters and a set of defensive parameters for each team are estimated *ex post*, but the model does not predict scores or results *ex ante*. A tendency to underestimate the proportion of draws is attributed to interdependence between the home and away scores, and corrected using the bivariate poisson distribution to model scores.

Dixon and Coles (1997) develop a forecasting model capable of generating *ex ante* probabilities for scores and match outcomes. The home and away team scores follow independent poisson distributions, but for low-scoring matches an *ad hoc* adjustment allows for interdependence. Using a similar framework, Rue and Salvesen (2000) assume that the time-varying attacking and defensive parameters of all teams vary randomly over time. The prior estimates of these parameters are updated as new match results information is obtained. Markov chain Monte Carlo iterative simulation techniques are used for inference. Crowder et al. (2002) develop a procedure for updating the team strength parameters that is computationally less demanding.

Researchers who have examined the impact of specific factors on match results include Barnett and Hilditch (1993), who investigate whether artificial playing surfaces, used by several clubs during the 1980s and early 1990s, conferred an additional home-team advantage. Ridder et al. (1994) show that player dismissals have a negative effect on the match result for the teams concerned. Clarke and Norman (1995) use a range of non-parametric techniques to identify the effect of home advantage on match results. Dixon and Robinson (1998) investigate variations in the scoring rates of the home and away teams during the course of a match. The scoring rates at any time depend partly upon the number of minutes elapsed, but also upon which (if either) team is leading at the time.

Recently, several researchers have used discrete choice regression models to model match results directly, rather than indirectly through scores. Apart from its computational simplicity, a major advantage of this approach is its avoidance of the thorny problem of interdependence between the home and away team scores.[1] Forrest and Simmons (2000a,b) investigate the predictive quality of newspaper tipsters' match results forecasts, and the performance of the pools panel in providing hypothetical results for matches that were postponed. Koning (2000) estimates a model to describe a set of match results *ex post*, as part of a broader analysis of changes in competitive balance in Dutch

[1] If scores are the focus, the difference between a 0-0 draw and a 1-0 home win is the same as the difference between a 1-0 and a 2-0 home win, or between a 2-0 and a 3-0 home win. If results are the focus, there is a large difference between a 0-0 draw and a 1-0 win, but no difference between home wins of different magnitudes. What is crucial is which (if either) team won; the precise numbers of goals scored and conceded by either team are incidental.

football. Kuypers (2000) uses a variety of explanatory variables drawn from current-season match results to estimate an *ex ante* forecasting model.[2]

3. An ordered probit regression model for match results

In Section 3, ordered probit regression is used to model and predict football match results.[3] The result of the match between teams i and j, denoted $y_{i,j}$, depends on the unobserved variable $y^*_{i,j}$ and a normal independent and identically distributed (NIID) disturbance term, $\varepsilon_{i,j}$, as follows:

Home win:	$y_{i,j} = 1$	if	$\mu_2 < y^*_{i,j} + \varepsilon_{i,j}$	
Draw:	$y_{i,j} = 0.5$	if	$\mu_1 < y^*_{i,j} + \varepsilon_{i,j} < \mu_2$	
Away win:	$y_{i,j} = 0$	if	$y^*_{i,j} + \varepsilon_{i,j} < \mu_1$	(1)

$y^*_{i,j}$ depends on the following systematic influences on the result of the match between teams i and j:

$P^d_{i,y,s}$ = Home team i's average win ratio (1=win, 0.5=draw, 0=loss) from matches played 0-12 months (y=0) or 12-24 months (y=1) before current match; within the current season (s=0) or previous season (s=1) or two seasons ago (s=2); in the team's current division (d=0) or one (d=±1) or two (d=±2) divisions above or below the current division.

$R^H_{i,m}$ = Result (1=win, 0.5=draw, 0=loss) of m'th most recent home match played by home team i (m=1…M).

$R^A_{i,n}$ = Result of n'th most recent away match played by home team i (n=1…N).

[2] These include the average points per game and the cumulative points attained by the home and away teams in the current season; the league positions and goal differences of the two teams; and the points and goal differences obtained by the two teams from the last three matches.

[3] Dobson and Goddard (2001) describe an early prototype of the present model.

$SIGH_{i,j}$ = 1 if match has championship, promotion or relegation significance for home team i but not for away team j; 0 otherwise.

$SIGA_{i,j}$ = 1 if match has significance for away team j but not for home team i; 0 otherwise.

CUP_i = 1 if home team i is eliminated from the FA Cup; 0 otherwise.[4]

$DIST_{i,j}$ = natural logarithm of the geographical distance between the grounds of teams i and j.

$ATTPOS_{i,k}$ = residual for team i from a cross-sectional regression of the log of average home attendance on final league position (defined on a scale of 92 for the PL winner to 1 for the bottom team in FLD3) k seasons before the present season, for k=1,2.

$LAST1_{i,j}$ = result (1 for a home win, 0.5 for a draw and 0 for an away win) of the corresponding fixture between teams i and j if that fixture took place in the preceding season. $LAST1_{i,j}$=0 if the fixture did not take place in the previous season.

$LAST0_{i,j}$=1 if the fixture did not take place in the previous season, and 0 if it did.

$P^{d}_{j,y,s}$, $R^{H}_{j,n}$, $R^{A}_{j,m}$, CUP_j, $ATTPOS_{j,k}$ = as above, for away team j.

Match results data for the Premier League (PL) and three divisions of the Football League (FLD1 to FLD3) were obtained from various editions of *Rothmans Football Yearbook*. Forecasts for the 2001-2 season were obtained using a version of the ordered probit model estimated over the previous 15 seasons from 1986-7 to 2000-1 inclusive, and reported in Table 1. Forecasts for the 1998-9, 1999-2000 and 2000-1 seasons were obtained using versions of the same model, each estimated using data for the previous 15 seasons. These estimations are not reported. The contribution to the model of each set of explanatory variables is now considered.

[4] The FA Cup is a sudden-death knock-out tournament involving both league and non-league teams. Up to and including the 1999 season, teams from FLD2 and FLD3 entered the cup in November, and teams from the PL and FLD1 in January. The final is played at the end of the league season, usually in early- or mid-May.

Team quality indicators

The win ratio variables $P^{d}_{i,y,s}$ (for team i, and their counterparts for team j) are the main team quality indicators. The higher the value of $y^{*}_{i,j}$ the higher is the probability of a home win, so positive home team and negative away team coefficients are expected. It is assumed that team i's underlying quality is captured by its win ratio over the previous 12 months, $P^{0}_{i,0,0} + \sum_{d=-1}^{+1} P^{d}_{i,0,1}$, and its win ratio between 12 and 24 months ago, $\sum_{d=-1}^{+1} P^{d}_{i,1,1} + \sum_{d=-2}^{+2} P^{d}_{i,1,2}$. The model allows the individual components of these sums to make different contributions to the team quality measure. For example, if current-season results are a better indicator of current team quality than previous-season results in the same division within the same 12-month period, the coefficient on $P^{0}_{i,0,0}$ should exceed the coefficient on $P^{0}_{i,0,1}$. If previous-season results from a higher division indicate higher quality than those from a lower division, the coefficient on $P^{+1}_{i,1,1}$ should exceed the coefficient on $P^{0}_{i,1,1}$; and so on. In general, the estimates of the coefficients on $\{P^{d}_{i,y,s},\ P^{d}_{j,y,s}\}$ reported in panel 1 of Table 1 are well defined, and conform accurately to prior expectations.

Experimentation indicated that the coefficients on $\{P^{d}_{i,y,s},\ P^{d}_{j,y,s}\}$ were strongly significant for y=0,1; but not for y=2 (defined in the same way as described above for matches that took place 24-36 months prior to the match in question). $\chi^{2}(20)$ in panel 5 of Table 1 is the omitted variables version of the Lagrange Multiplier (LM) test described by Weiss (1997) for the joint significance of the 20 additional coefficients on $\{P^{d}_{i,2,s},\ P^{d}_{j,2,s}\}$ for d= –2 , ... ,+2, s=2,3 (using the scoring form of the information matrix). The test is not significant.

Recent performance indicators

The recent match results variables $R^{H}_{i,m}$ and $R^{A}_{i,n}$ (for team i, and their counterparts for team j) allow for the inclusion of each team's few most recent home and away results in the calculation of $y^{*}_{i,j}$. Although $R^{H}_{i,m}$ and $R^{A}_{i,n}$ also contribute towards the values of $P^{d}_{i,y,s}$ and are to some extent correlated with these variables, the possibility of short-term persistence in match results suggests $R^{H}_{i,m}$ and $R^{A}_{i,n}$ may have particular importance (over and above the information they convey about team quality) in helping predict the result of the current match. Generally the estimated coefficients on $\{R^{H}_{i,m}, R^{A}_{i,n}, R^{H}_{j,n}, R^{A}_{j,m}\}$ reported in panel 2 of Table 1 are more erratic than those on $\{P^{d}_{i,y,s}, P^{d}_{j,y,s}\}$, in terms of both their numerical magnitudes and the statistical significance of individual coefficients.

Experimentation indicated that the home team's recent home results are more useful as predictors than its recent away results; and similarly the away team's recent away results are more useful than its recent home results. Statistically significant estimated coefficients are obtained for some (but not all) values of m≤9, and for some n≤4. M=9 and N=4 are the chosen lag lengths. $\chi^2(10)$ in panel 5 of Table 1 is the omitted variables LM test statistic (Weiss, 1997) for the joint significance of the 10 additional coefficients on $\{R^{H}_{i,m}, R^{A}_{i,n}, R^{H}_{j,n}, R^{A}_{j,m}\}$ for m=10,11,12 and n=5,6. The test is not significant.[5]

Other explanatory variables, cut-off parameters and diagnostic tests

The identification of matches with significance for championship, promotion and relegation issues has been an important issue in the literature on the estimation of the demand for match attendance (Jennett 1985, Peel and Thomas, 1988). It is also relevant in the present context, since match outcomes are likely to be affected by incentives: if a match is significant for one team and insignificant for the other, the incentive difference is likely to influence the result. The algorithm used here to assess whether or not a match is significant is crude but simple. A match is significant if it is still possible

[5] Other similar tests were carried out using various permutations of additional variables for m>9 and n>4, with the same result.

(before the match is played) for the team in question to win the championship or be promoted or relegated, assuming that all other teams currently in contention for the same outcome take one point on average from each of their remaining fixtures.[6] Estimated coefficients on $SIGH_{i,j}$ and $SIGA_{i,j}$ that are positive and negative (respectively) are consistent with incentive effects of the kind described above. Both coefficients reported in Table 1 are signed accordingly and significant at the 1% level.

Early elimination from the FA Cup may have implications for a team's results in subsequent league matches, although the direction of the effect is ambiguous. On the one hand, a team eliminated from the cup may be able to concentrate its efforts on the league, suggesting an improvement in league results. On the other hand, elimination from the cup may entail loss of confidence (while cup progress fosters team spirit), suggesting a deterioration in league results. In Table 1 the estimated coefficient on CUP_i is negative, the coefficient on CUP_j is positive, and both are significant at the 1% level. This suggests that the second of the two effects described above dominates. Contrary to football folklore, elimination from the cup appears to have a harmful effect on the team's subsequent league results.

Clarke and Norman (1995) demonstrate that geographical distance is a significant influence on match results. This finding is confirmed by a positive and highly significant coefficient on $DIST_{i,j}$ in Table 1. The greater intensity of competition in local derbies may partially offset home advantage in such matches, or the psychological or practical difficulties of long distance travel for both teams and spectators may increase home advantage in matches between teams from opposite ends of the country.

$ATTPOS_{i,k}$ and $ATTPOS_{j,k}$ are positive for teams that tend to attract higher-than-average attendances after controlling for league position (and negative in the opposite case). These variables allow for a 'big team' effect on match results: regardless of the values of other controls, 'big' teams are more likely (and 'small' teams less likely) to win. The advantages of a large support might be psychological

[6] Several alternative definitions of significance were considered, but the chosen algorithm produced values of $SIGH_{i,j}$ and $SIGA_{i,j}$ with the most explanatory power in the ordered probit model. The algorithm succeeds in identifying those matches, mostly played during the last few weeks of the season, in which there is a major difference between the incentives facing the teams.

(direct influence of the crowd on the match result) or material (teams with a large revenue base have more resources to spend on players). Either way Table 1 suggests the 'big team' effect is in fact a highly significant determinant of match results. To reduce the effect of temporary variation in the attendance-performance relationship, the values of these variables for two previous seasons are included in the model. Since the values over successive seasons tend to be highly correlated, $\Delta ATTPOS_{i,1} = ATTPOS_{i,1} - ATTPOS_{i,2}$ (and its counterpart for team j) is used in place of $ATTPOS_{i,1}$.

Finally, as well as their forecasts many newspaper (and other) tipsters report the results of previous meetings between the same two teams in recent seasons. Previous results may be relevant if there are influences on matches involving particular teams which persist from year to year ('jinxes' or 'horses for courses' effects); or if a defeat inspires a team to raise its efforts in an effort to exact 'revenge' next time. $LAST1_{i,j}$ is the result of the corresponding fixture if that fixture took place in the preceding season. A second dummy, $LAST0_{i,j}$=1 if the fixture did not take place and 0 if it did, is also required as a control, to distinguish between these two cases. The negative reported coefficient on $LAST1_{i,j}$=0 in Table 1 suggests a tendency for the fortunes of the two teams to be reversed if they meet in consecutive seasons, perhaps due to a 'revenge' effect as decsribed above. However, the coefficient is only significant at the 10% level, so the effect appears to be relatively weak.[7]

Panel 3 of Table 1 also reports the estimated cut-off parameters in (1), $\hat{\mu}_1$ and $\hat{\mu}_2$; Glewwe's (1997) LM test of the normality of $\varepsilon_{i,j}$ in (1); Weiss's (1997) LM test (using the scoring form of the information matrix) of the null hypothesis that $\varepsilon_{i,j}$ in (1) are homoscedastic;[8] and the omitted variables

---

[7] The result from two seasons before was insignificant and is not included in Table 1. The inclusion of the result of the first fixture in the current season among the predictors of the result of the return fixture between the same teams also failed to produce a significant coefficient.

[8] In the homoscedasticity test, $\varepsilon_{i,j} \sim N(0, \exp(2\sigma z_{i,j}))$ under the alternative hypothesis of heteroscedasticity of known form. $\varepsilon_{i,j}$ are homoscedastic under H_0:σ=0. A convenient choice for $z_{i,j}$ is the uncertainty of match outcome measure $z_{i,j} = \hat{\rho}_{i,j}(1-\hat{\rho}_{i,j})$ where $\hat{\rho}_{i,j} = 1 - 0.5[\Phi(\hat{\mu}_1 - \hat{y}^*_{i,j}) - \Phi(\hat{\mu}_0 - \hat{y}^*_{i,j})]$ is the 'expected result' on a scale of 0 ('certain' away win) to 1 ('certain' home win). The alternative hypothesis allows the variance of the unsystematic component in the match result to vary directly (σ>0) or inversely (σ<0) with uncertainty of outcome.

LM tests described above. The normal and homoscedastic null hypotheses are both accepted at any reasonable significance level, suggesting that the ordered probit model provides a suitable representation of the match results data.

4. The efficiency of the fixed-odds betting market

Section 4 reviews previous literature on the efficiency of team sports betting markets. Pankoff (1968) developed the first regression-based test of efficiency in the National Football League (NFL) betting market, by regressing match outcomes (measured by the score differential) on bookmakers' spreads. Intercept and slope coefficients insignificantly different from zero and one respectively suggest that the spread was an unbiased predictor of the match outcome (see below). Gandar et al. (1988), however, point out the relatively low power of regression-based efficiency tests, and propose a series of economic tests involving direct evaluation of the returns that would have been earned by implementing technical trading rules (betting on the basis of the past performance of the teams); and behavioural rules (betting to exploit certain hypothesised behavioural patterns of the public).[9] In an NFL data set some behavioural rules are found to be profitable, but technical rules are not.

Golec and Tamarkin (1991) test the efficiency of the spreads posted on the outcomes of NFL and college football matches. For NFL betting they find evidence of inefficiencies favouring bets on home wins and bets on underdogs. No evidence of bias is found in the college football betting spreads. Dare and MacDonald (1996) generalise the empirical methodology for regression-based tests. Several earlier tests were based on specifications that imposed implicit restrictions on the general model. Tests that search only for evidence of home-away team or longshot biases, without recognising the interdependence between these team characteristics, are liable to produce misleading results.[10]

---

[9] A typical behavioural rule is to back underdogs against favourites in cases where the favourite covered the spread by a large margin the previous week.

[10] In other recent US studies, Badarinathi and Kochmann (1996) show that a strategy of betting regularly on underdogs in American football was systematically profitable. Gandar et al. (1998) find that the bookmakers'

The prices for bets on the results of English football matches are fixed by the bookmakers several days before the match, and are not adjusted as bets are placed even if new information is received. Bookmakers' prices are in the form: a-to-b home win; c-to-d draw; and e-to-f away win. If b is staked on a home win, the overall payoffs to the bettor are +a (the bookmaker pays the winnings and returns the stake) if the bet wins, and –b (the bookmaker keeps the stake) if the bet loses. These quoted prices can be converted to the home win, draw and away win 'probabilities': $\theta^{H}_{i,j}$= b/(a+b); $\theta^{D}_{i,j}$= d/(c+d); $\theta^{A}_{i,j}$= f/(e+f). The sum of these expressions invariably exceeds one, however, because the prices contain a margin to cover the bookmaker's costs and profits. Implicit home win, draw and away win probabilities which sum to one are $\phi^{H}_{i,j} = \theta^{H}_{i,j}/(\theta^{H}_{i,j}+\theta^{D}_{i,j}+\theta^{A}_{i,j})$, and likewise for $\phi^{D}_{i,j}$ and $\phi^{A}_{i,j}$. The bookmaker's margin is $\lambda_{i,j}=\theta^{H}_{i,j}+\theta^{D}_{i,j}+\theta^{A}_{i,j} - 1$.

Pope and Peel (1989) investigate the efficiency of the prices set by four national high street bookmakers for fixed-odds betting on English football. A simple (though as before not very powerful) test of the weak-form efficiency hypothesis is based on regressions of match outcomes against implicit bookmaker's probabilities. Consider the linear probability model:

$$r_{i,j} = \alpha_r + \beta_r \phi^{r}_{i,j} + u_{i,j} \qquad (2)$$

where $r_{i,j}$=1 if the result of the match between teams i and j results is r, for r=H (home win), D (draw) or A (away win) and 0 otherwise. In (2) a necessary weak-form efficiency condition is $\{\alpha_r = 0, \beta_r = 1\}$. Since ordinary least squares (OLS) estimation of the linear probability model produces a heteroscedastic error structure, Pope and Peel use weighted least squares (WLS) estimation, using

closing prices provided a closer approximation to basketball match outcomes than their opening prices. Price movements before close of trade suggest that informed traders were active and influential. Gandar et al. (2001) are sceptical over the existence of systematic biases favouring bets on home teams in baseball and basketball.

$\hat{r}_{i,j}(1-\hat{r}_{i,j})$ as weights, where $\hat{r}_{i,j}$ are the fitted values of the dependent variable obtained from (preliminary) estimation of the model using OLS.[11] There is some evidence of departures from $\{\alpha_r=0, \beta_r=1\}$. Recently Cain et al. (2000) report evidence of longshot bias in the fixed-odds betting market for match results and scores in English football. Comparisons of estimated fair prices with actual prices for specific scores suggest some evidence of biases.

5. Efficiency of prices in the fixed-odds betting market: empirical results

In Section 5, tests are presented for weak-form efficiency in the prices quoted by high street bookmakers for fixed-odds betting on match results during four English league football seasons, from 1998-9 to 2001-2 inclusive. If the ordered probit model produces information about the match outcome probabilities that is not already reflected in the bookmakers' prices, then the latter fail to satisfy standard weak-form efficiency criteria: that all historical information relevant to the assessment of the match outcome probabilities should be reflected in the quoted price.

The fixed-odds betting data set comprises the prices quoted by five bookmakers (denoted B1 to B5) for all league matches played during the four seasons.[12] From a total of 8,144 league matches played, the model is capable of generating predictions for 7,782. All matches involving teams that entered the league within the previous two calendar years are discarded, because the model requires match results for a full two-year period prior to the match in question. In 2001-2, one FLD3 fixture was rescheduled with insufficient notice for the bookmakers to quote prices, and is also discarded. The final sample comprises 7,781 matches. Estimated *ex ante* probabilities for the 1,945 matches available for the 2001-2 season are obtained by substituting the full set of covariate values for the home and away teams for each match into the ordered probit model estimated using data for the 15 seasons 1986-7 to

[11] Alternatively, the model can be estimated as a logit regression, in which case different numerical estimates of the coefficients ρ_1 and ρ_2 are expected.

[12] The data on the bookmakers' prices were obtained from www.mabels-tables.co.uk

2000-1 (inclusive) as reported in Table 1, generating a fitted value for the match between home team i and away team j, denoted $\hat{y}^*_{i,j}$. The estimated home win, draw and away win probabilities are: $p^H_{i,j}$ =1–$\Phi(\hat{\mu}_2 - \hat{y}^*_{i,j})$, $p^D_{i,j}$ =$\Phi(\hat{\mu}_2 - \hat{y}^*_{i,j})$–$\Phi(\hat{\mu}_1 - \hat{y}^*_{i,j})$ and $p^A_{i,j}$ =$\Phi(\hat{\mu}_1 - \hat{y}^*_{i,j})$, where Φ is the standard normal distribution function. Probabilities for matches in the other three seasons are obtained in the same way, in each case using the ordered probit model estimated over the previous 15 seasons.

Table 2 summarises features of the the implicit bookmaker probabilities, the model's estimated match result probabilities, and the match outcomes. In Table 2 the implicit bookmaker probabilities are obtained by applying the procedure described in Section 4 to the arithmetic mean of the prices quoted by the five bookmakers for each outcome.[13] When the bookmaker and the model's probabilities are disaggregated by month and by division, there is little systematic variation in the mean probabilities for each outcome, but the cross-sectional standard deviations of both the bookmaker and the model's probabilities increase systematically during the course of the season. It is possible to make a more specific assessment of the probabilities for any individual match (based mainly on the current season's results) towards the end of the season than at the start (when results from previous seasons provide the only guidance). The cross-sectional standard deviations are highest in the PL and lowest in FLD3. This apparently reflects a lesser degree of competitive balance within the PL than is the case within each division of the Football League.

Overall the model appears effective in replicating the main features of the bookmaker probabilities. As an overall indicator of forecasting accuracy, Rue and Salvesen (2000) suggest the pseudolikelihood measure, calculated as the geometric mean of the probabilities for the observed

[13] Several important changes affecting the UK fixed-odds betting market took place during the observation period. The UK government abolished betting duty previously levied at the rate of 10 pence per £1 bet, for bets placed via the Internet in 2000, and for bets placed in the high street in 2001. At the start of the period, bookmakers required bettors to place combination bets on at least five matches (for bets on home wins) or three matches (for draws or away wins). This restriction was progressively relaxed, and individual bets on any match are currently accepted. Apparently in response to these changes, the bookmakers have increased their margins by between 1% and 2% (see Table 4 below).

results. This can be calculated using both the bookmaker and the model's probabilities. In all cases the difference in forecasting performance appears very small.

Regression-based weak-form efficiency tests

Panel 1 of Table 3 reports WLS estimates of α_r and β_r in (2) for r=H,D,A. Estimations are carried out over all four seasons and for each season individually. In the estimations over all seasons, the acceptance or rejection of H_0:α_r=0; H_0:β_r=1; and H_0:{α_r, β_r}={0,1} is borderline in most cases. In the estimations for the individual seasons, H_0:β_r=1 is rejected at the 5% level for r=D in one of the four seasons; but elsewhere H_0:α_r=0, H_0:β_r=1, and H_0:{α_r, β_r}={0,1} are always accepted. Overall there seems to be little or no evidence of systematic departure from these necessary weak-form efficiency conditions.

The availability of the evaluated probabilities obtained from the model permits an extension of these conventional regression-based weak-form efficiency tests. If the model produces no additional relevant information (beyond what is already contained in the bookmaker probabilities) a term in $(p^r_{i,j} - \phi^r_{i,j})$ should be insignificant when added to the regressions described above. The additional weak-form efficiency conditions are γ_r=0 and {α_r, β_r, γ_r}={0,1,0} for r=H,D,A in the linear probability model:

$$r_{i,j} = \alpha_r + \beta_r \phi^r_{i,j} + \gamma_r(p^r_{i,j} - \phi^r_{i,j}) + u_{i,j} \qquad (3)$$

Panel 2 of Table 3 reports the WLS estimation results of (3). In the estimations over all seasons, H_0:γ_r=0 and H_0:{α_r, β_r, γ_r}={0,1,0} are rejected at the 5% or 1% levels for r=H,D and A. In the estimations for individual seasons the results are similar, although slightly less consistent due to the smaller number of observations used in each regression. Overall this appears to constitute reasonably strong evidence that the model does contain additional information that is not impounded into the bookmaker odds, and that the latter are therefore weak-form inefficient.

Economic weak-form efficiency tests

It is also possible to test for weak-form efficiency directly, by calculating *ex post* the returns that could have been generated by following various betting strategies. For simplicity, the economic weak-form efficiency tests reported below are based on an assumption that individual bets could have been placed on all matches throughout the sample period. Table 4 analyses the returns available from all available bets, while Table 5 analyses the returns from bets placed only on the match outcome (home win, draw or away win) with the highest expected return according to the probabilities generated from the ordered probit model.

Panel 1 of Table 4 shows the average returns attainable by placing £1 bets on each possible outcome of every match, for each of the five bookmakers. The negative returns of between about 10% and 12% reflect the size of the bookmakers' margins, but provide a benchmark against which to assess the performance of selective betting strategies. B3's margin appears to be slightly but consistently higher than those of the other four bookmakers. Panel 2 of Table 4 shows the average returns attainable by placing £1 bets on each possible outcome of every match with the bookmaker offering the most favourable price for that outcome. A comparison between panels 2 and 1 provides an indication of the potential for profitable exploitation of anomalies between the prices offered by different bookmakers. Average losses are reduced by about 5% in 1998-9, and by about 4% in 2001-2. This suggests some tendency for convergence between the prices offered by the five bookmakers over the course of the sample period.

Panel 3 of Table 4 shows how the returns from 23,343 available bets (three bets on each of 7,781 matches with the bookmaker offering the most favourable price for each outcome) vary if the bets are ranked in descending order of expected return according to the ordered probit model within each season, and grouped into nine bands (from the top 5% to the bottom 30%). Despite the presence of some random variation in the average returns between adjacent groups, the actual returns are correlated with the rankings by expected return according to the model.[14] A strategy of selecting only those bets appearing in the top 15% of the expected returns distribution would have produced a (relatively generous) positive return of at least 4% in all four seasons. A (perverse) strategy of selecting only those bets appearing in the bottom 30% of the expected returns distribution would have produced a loss of between 14% and 17% in all four seasons, significantly worse than would be expected from random betting after allowing for the effect of the bookmakers' margins.

In Table 5 it is assumed that bets are placed only on the match outcome (home win, draw or away win) with the highest available return according to the probabilities obtained from the model. Panel 1 of Table 5 shows the average returns attainable if this strategy is used with each of the five bookmakers individually, and panel 2 shows the average returns if the same strategy is used with the most favourable price available for each outcome. A comparison of panels 1 and 2 of Table 5 with their counterparts in Table 4 suggests that this strategy improves the expected return by around 7% or 8%. All of the returns shown in panel 2 of Table 5 are non-negative.

Panels 3, 4 and 5 of Table 5 show how the returns from the 7,781 available bets vary if the bets are disaggregated by calendar month (panel 3); by division (panel 4); and according to whether either team had played its previous league match within five days of the current match (panel 5). For both 1998-9 and 1999-2000, according to panel 3 the model performs badly during the first part of the season, but well during the second part (from January onwards). For 2000-1 and 2001-2 this pattern is reversed, with positive returns reported for most months up to December, but predominantly negative returns from January onwards. It is not clear whether this variation is purely random, or whether it might be due to adjustments in the bookmakers' own price-setting rules. In particular, it seems plausible that the large positive returns the analysis suggests were available towards the end of the 1998-9 and 1999-2000 seasons might have prompted some reappraisal on the part of the bookmakers, leading to an improvement from their perspective in 2000-1 and 2001-2.

[14] For the four seasons, the correlation coefficients are 0.060, 0.055, 0.042 and 0.046, respectively. All are significantly different from zero at the 1% level.

Panel 4 of Table 5 suggests there is little or no systematic variation in average returns by division. Panel 5 tests the importance of one potential source of inefficiency, which arises from the fact that the bookmakers' prices are usually compiled and published at least five days before the match in question is played. This means that if either team's previous match took place within this five-day period, information about the result will not be impounded into the bookmakers' prices for the match in question. In contrast, the model takes account of the results of all previous matches, whenever they were played. The relative performance of model (in comparison with the bookmakers) should therefore be higher for matches in which either team had played its previous match within the last five days than elsewhere. Panel 5 provides some evidence that this was the case, although the difference between the average returns for the two cases seems to have narrowed (or disappeared altogether) over time. For the first three seasons the average return differential was around 6%, 1% and 5%, respectively. For the 2001-2 season the pattern was reversed, with a differential of –3%. As before, it is unclear whether the apparent trend is systematic, random or partly both. However, a tendency for inefficiencies to diminish over time would be consistent with several other of the findings reported in Section 5.

6. Conclusion

This paper has reported the estimation of an ordered probit regression model, which has been used to predict English league football results. This paper is the first to quantify the predictive quality not only of past match results data, but also of a wide range of other explanatory variables. The significance of each match for championship, promotion or relegation issues; the involvement of the teams in cup competition; the geographical distance between the teams' home towns; and a 'big team' effect are all found to contribute to the model's performance. The ordered probit model is considerably easier to implement than several of the team scores forecasting models that have been developed recently in the applied statistics literature, but appears capable of achieving comparable forecasting performance.

The model has been used to test the weak-form efficiency of the prices quoted by high street bookmakers for fixed-odds betting on match results during four recent football seasons. Regression-based tests indicate that the model contains information about match outcomes that is not impounded into the bookmakers' prices. The latter are therefore weak-form inefficient. A strategy of selecting bets ranked in the top 15% by expected return according to the model's probabilities would have generated a positive return of at least 4% in each of the four seasons. A strategy of betting on the match outcome for which the model's *ex ante* expected return is the highest would have generated positive or zero gross returns in each of the four seasons. There is some evidence, however, that inefficiencies in the bookmakers' prices have diminished over time.

References

Badarinathi R. and Kochman L. 1996. Football betting and the efficient market hypothesis, *American Economist* **40**: 52-55.

Barnett V. and Hilditch S. 1993. The effect of an artificial pitch surface on home team perfromance in football (soccer). *Journal of the Royal Statistical Society Series A* **156**: 39-50.

Cain M., Law D. and Peel D. 2000. The favourite-longshot bias and market efficiency in UK football betting. *Scottish Journal of Political Economy* **47**: 25-36.

Clarke S.R. and Norman J.M. 1995. Home ground advantage of individual clubs in English soccer. *The Statistician* **44**: 509-521.

Crowder M., Dixon M., Ledford A. and Robinson M. 2002. Dynamic modelling and prediction of English Football League matches for betting. *The Statistician* **51**: 157-168.

Dare W.H. and MacDonald S. 1996. A generalized model for testing the home and favorite team advantage in point spread markets, *Journal of Financial Economics* **40**: 295-318.

Dixon M.J. and Coles S.C. 1997. Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* **46**: 265-280.

Dixon M.J. and Robinson M.E. 1998. A birth process model for association football matches. *The Statistician* **47**: 523-538.

Dobson S. and Goddard J. 2001. *The economics of football*. Cambridge: Cambridge University Press.

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

Forrest D. and Simmons R. 2000b. Making up the results: the work of the football pools panel, 1963-1997. *The Statistician* **49**: 253-260.

Gandar J.M., Dare W.H., Brown C.R. and Zuber R.A. 1998. Informed traders and price variations in the betting market for professional basketball games, *Journal of Finance* **53**: 385-401.

Gander J.M., Zuber R.A. and Lamb R.P. 2001. The home field advantage revisited: a search for the bias in other sports betting markets. *Journal of Economics and Business* **53**: 439-453.

Gandar J.M., Zuber R.A., O'Brien T. and Russo B. 1988. Testing rationality in the point spread betting market, *Journal of Finance* **43**: 995-1008.

Glewwe P. 1997. A test of the normality assumption in the ordered probit model. *Econometric Reviews* **16**: 1-19.

Golec J. and Tamarkin M. 1991. The degree of inefficiency in the football betting market: statistical tests. *Journal of Financial Economics* **30**: 311-323.

Hill I.D. 1974. Association football and statistical inference, *Applied Statistics* **23**: 203-208.

Jennett N. 1984. Attendances, uncertainty of outcome and policy in Scottish League Football, *Scottish Journal of Political Economy* **31**: 176-198.

Koning R.H. 2000. Balance in competition in Dutch soccer. *The Statistician* **49**: 419-431.

Kuypers T. 2000. Information and efficiency: an empiurical study of a fixed odds betting market. *Applied Economics* **32**: 1353-1363.

Maher M.J. 1982. Modelling association football scores. *Statistica Neerlandica* **36**: 109-118.

Moroney M.J. 1956. *Facts from figures,* 3rd edn. Penguin: London.

Pankoff L.D. 1968. Market efficiency and football betting, *Journal of Business* **41**: 203-214

Peel D.A and Thomas D. 1988. Outcome uncertainty and the demand for football, *Scottish Journal of Political Economy* **35**: 242-249.

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

Reep C., Pollard R. and Benjamin B. 1971. Skill and chance in ball games. *Journal of the Royal Statistical Society Series A* **131**: 581-585.

Ridder G., Cramer J.S. and Hopstaken P. 1994. Estimating the effect of a red card in soccer. *Journal of the American Statistical Association* **89**: 1124-1127.

Rue H. and Salvesen O. 2000. Prediction and retrospective analysis of soccer matches in a league. *The Statistician* **49**: 399-418.

Weiss A.A. 1997. Specification tests in ordered logit and probit models. *Econometric Reviews* **16**: 361-391.

Table 2 Bookmaker's implicit probabilities, and forecast probabilities: descriptive statistics

	Bookmaker				Model				Actual		
	$\phi^{H}_{i,j}$	$\phi^{D}_{i,j}$	$\phi^{A}_{i,j}$	PsL	$p^{H}_{i,j}$	$p^{D}_{i,j}$	$p^{A}_{i,j}$	PsL	H(%)	D(%)	A(%)
1998-9	0.452 *0.103*	0.268 *0.014*	0.280 *0.095*	0.353	0.476 *0.128*	0.267 *0.031*	0.257 *0.105*	0.354	0.451	0.280	0.269
1999-2000	0.455 *0.118*	0.264 *0.018*	0.281 *0.106*	0.359	0.470 *0.136*	0.267 *0.032*	0.263 *0.113*	0.359	0.456	0.274	0.270
2000-1	0.455 *0.111*	0.264 *0.017*	0.281 *0.099*	0.359	0.468 *0.130*	0.270 *0.030*	0.262 *0.108*	0.357	0.465	0.272	0.263
2001-2	0.453 *0.113*	0.266 *0.018*	0.281 *0.102*	0.360	0.465 *0.134*	0.271 *0.031*	0.264 *0.111*	0.360	0.464	0.266	0.270
Prem	0.453 *0.135*	0.264 *0.023*	0.284 *0.122*	0.368	0.475 *0.152*	0.263 *0.038*	0.262 *0.126*	0.368	0.464	0.269	0.267
Div 1	0.453 *0.109*	0.266 *0.016*	0.281 *0.099*	0.358	0.468 *0.141*	0.267 *0.033*	0.265 *0.118*	0.357	0.461	0.276	0.264
Div 2	0.454 *0.104*	0.266 *0.015*	0.280 *0.094*	0.354	0.469 *0.127*	0.270 *0.029*	0.260 *0.104*	0.354	0.444	0.275	0.281
Div 3	0.455 *0.101*	0.266 *0.015*	0.279 *0.091*	0.354	0.468 *0.108*	0.274 *0.024*	0.258 *0.088*	0.353	0.280	0.272	0.259
Aug-Oct	0.451 *0.098*	0.269 *0.013*	0.280 *0.089*	0.355	0.465 *0.111*	0.274 *0.024*	0.261 *0.091*	0.354	0.456	0.275	0.268
Nov-Dec	0.455 *0.108*	0.267 *0.016*	0.278 *0.097*	0.362	0.470 *0.130*	0.269 *0.030*	0.258 *0.106*	0.362	0.484	0.263	0.253
Jan-Feb	0.455 *0.114*	0.266 *0.016*	0.279 *0.103*	0.357	0.469 *0.136*	0.269 *0.031*	0.265 *0.114*	0.358	0.451	0.282	0.266
Mar-May	0.455 *0.126*	0.260 *0.021*	0.285 *0.113*	0.359	0.475 *0.152*	0.263 *0.037*	0.261 *0.125*	0.357	0.449	0.273	0.279
All	0.454 *0.111*	0.265 *0.017*	0.281 *0.101*	0.358	0.470 *0.132*	0.269 *0.031*	0.261 *0.109*	0.357	0.459	0.273	0.268

Notes: Data for $\phi^{r}_{i,j}$ and $p^{r}_{i,j}$ (r=H, D, A) are cross sectional means, with standard deviations in italics.

PsL is Rue and Salvesen's (2000) pseudolikelihood measure of forecasting accuracy: the geometric mean of the bookmaker's or model's probabilities for the actual results.

H(%), D(%) and A(%) are the actual proportions of home wins, draws and away wins.

Table 3 Weak-form efficiency: regression-based tests

	All	1998-9	1999-2000	2000-1	2001-2
Observations	7781	1944	1946	1946	1945
1. TESTS BASED ON $r_{i,j} = \alpha_r + \beta_r \phi^r_{i,j} + u_{i,j}$ for r = H,D,A					
Home wins					
Constant	-0.051** (0.021)	-0.048 (0.046)	-0.060 (0.040)	-0.070 (0.043)	-0.023 (0.042)
$\phi^H_{i,j}$	1.123*** (0.046)	1.105 (0.101)	1.136 (0.085)	1.174 (0.092)	1.073 (0.091)
F_1	3.93**	0.55	1.28	2.12	0.73
Draws					
Constant	-0.150** (0.068)	-0.296 (0.151)	-0.136 (0.124)	-0.215 (0.137)	-0.028 (0.136)
$\phi^D_{i,j}$	1.595** (0.258)	2.147** (0.567)	1.554 (0.472)	1.848 (0.523)	1.107 (0.511)
F_1	3.57**	2.51*	1.13	1.55	0.02
Away wins					
Constant	-0.014 (0.013)	0.020 (0.028)	-0.016 (0.024)	-0.009 (0.026)	-0.042 (0.024)
$\phi^A_{i,j}$	1.005 (0.046)	0.890 (0.101)	1.014 (0.088)	0.971 (0.094)	1.114 (0.089)
F_1	3.58**	0.95	0.85	1.65	1.92
2. TESTS BASED ON $r_{i,j} = \alpha_r + \beta_r \phi^r_{i,j} + \gamma_r(p^r_{i,j} - \phi^r_{i,j}) + u_{i,j}$ for r = H,D,A					
Home wins					
Constant	-0.054** (0.021)	-0.046 (0.046)	-0.069* (0.039)	-0.083* (0.042)	-0.015 (0.042)
$\phi^H_{i,j}$	1.115** (0.046)	1.065 (0.101)	1.142* (0.084)	1.195 (0.091)	1.042 (0.091)
$p^H_{i,j} - \phi^H_{i,j}$	0.432*** (0.078)	0.628*** (0.163)	0.395*** (0.153)	0.324** (0.141)	0.502*** (0.170)
F_2	13.20***	5.49***	3.24**	3.35**	3.37**
Draws					
Constant	-0.107 (0.072)	-0.201 (0.166)	-0.101 (0.131)	-0.182 (0.143)	-0.028 (0.143)
$\phi^D_{i,j}$	1.427 (0.272)	1.797 (0.621)	1.419 (0.498)	1.701 (0.548)	1.106 (0.542)
$p^D_{i,j} - \phi^D_{i,j}$	0.540** (0.227)	0.696 (0.432)	0.372 (0.454)	0.957** (0.447)	-0.000 (0.515)
F_2	4.31***	2.48*	0.94	3.07**	0.01
Away wins					
Constant	-0.015 (0.013)	0.023 (0.028)	-0.018 (0.023)	-0.015 (0.026)	-0.041* (0.024)
$\phi^A_{i,j}$	1.041 (0.046)	0.923 (0.101)	1.061 (0.088)	1.008 (0.096)	1.134 (0.088)
$p^A_{i,j} - \phi^A_{i,j}$	0.422*** (0.084)	0.537*** (0.180)	0.533*** (0.165)	0.264* (0.153)	0.400** (0.183)
F_2	11.33***	3.72**	4.24***	2.18*	3.16**

Notes: Standard errors of estimated coefficients are shown in parentheses.

t-tests on individual coefficients are for $H_0:\alpha_r=0$; $H_0:\beta_r=1$; and $H_0:\gamma_r=0$.
F_1 is an F-test for $H_0:\{\alpha_r, \beta_r\}=\{0,1\}$ and F_2 is an F-test for $H_0:\{\alpha_r, \beta_r, \gamma_r\}=\{0,1,0\}$.

*** = significant at 1% level (2-tail test); ** = significant at 5% level; * = significant at 10% level.

Table 4 Economic tests for weak form efficiency: comparison of all possible bets

	1998-9	1999-2000	2000-1	2001-2
1. Average returns from each bookmaker				
B1	-0.100	-0.108	-0.114	-0.116
B2	-0.100	-0.110	-0.114	-0.119
B3	-0.102	-0.112	-0.116	-0.121
B4	-0.093	-0.108	-0.110	-0.115
B5	-0.099	-0.108	-0.110	-0.114
2. Average returns from selecting best available price for each bet				
	-0.053	-0.064	-0.071	-0.076
3. Average returns from selecting best available price for each bet, and ranking all bets in descending order of expected return (according to the model)				
Top 5%	0.116	0.008	-0.008	0.160
5-10%	0.072	0.148	0.007	-0.094
10-15%	-0.111	-0.067	0.061	-0.020
15-20%	0.089	0.065	0.024	-0.076
20-25%	0.031	0.083	-0.032	-0.024
25-30%	-0.142	0.072	-0.014	-0.058
30-50%	-0.005	-0.045	-0.053	-0.040
50-70%	-0.023	-0.121	-0.104	-0.057
70-100%	-0.167	-0.146	-0.139	-0.170

Note: Data are average returns per £1 bet.

Table 5 Economic tests for weak form efficiency: comparison of bets placed only on match outcome with highest expected return (according to the model)

	1998-9	1999-2000	2000-1	2001-2
1. Average returns from each bookmaker				
B1	-0.016	0.017	-0.057	-0.014
B2	-0.057	-0.017	-0.045	-0.035
B3	-0.002	-0.053	-0.062	-0.080
B4	-0.048	-0.020	-0.027	-0.035
B5	0.003	0.011	-0.068	-0.051
2. Average returns from selecting best available price for each bet				
	0.016	0.026	0.000	0.008
3. Average returns from selecting best available price for each bet, by calendar month				
Aug	0.029	-0.031	0.095	0.098
Sept	-0.060	0.123	0.012	0.123
Oct	0.053	0.077	-0.112	-0.009
Nov	-0.124	-0.114	0.245	0.075
Dec	-0.068	-0.134	0.030	0.018
Jan	0.121	0.104	-0.041	-0.076
Feb	0.004	0.003	0.110	-0.128
Mar	0.113	-0.003	-0.102	-0.052
Apr/May	0.060	0.150	-0.089	-0.029
4. Average returns from selecting best available price for each bet, by division				
Prem	0.004	0.053	0.029	-0.080
Div One	-0.013	0.031	-0.052	0.055
Div Two	0.002	-0.037	-0.026	0.050
Div Three	0.072	0.072	0.068	-0.056
5. Average returns from selecting best available price for each bet, matches split according to whether either team had played its last match within five days of current match				
Yes	0.049	0.033	0.026	-0.013
No	-0.008	0.021	-0.028	0.018

Note: Data are average returns per £1 bet.

Table 1 Ordered probit estimation results

1. WIN RATIOS OVER PREVIOUS 24 MONTHS ($P^{d}_{i,y,s}$, $P^{d}_{j,y,s}$)

	Home team (i)				Away team (j)			
	0-12 months (y=0)		12-24 months (y=1)		0-12 months (y=0)		12-24 months (y=1)	
Matches played:	Current season (s=0)	Last season (s=1)	Last season (s=1)	Two seasons ago (s=2)	Current season (s=0)	Last season (s=1)	Last season (s=1)	Two seasons ago (s=2)
Two divisions higher (d=2)				-0.027 (0.554)				-0.268 (0.565)
One division higher (d=1)		1.883^{***} (0.250)	0.856^{***} (0.211)	0.558^{***} (0.203)		-1.541^{***} (0.248)	-0.747^{**} (0.210)	-0.452^{**} (0.202)
Current division (d=0)	1.726^{***} (0.151)	1.214^{***} (0.137)	0.716^{***} (0.130)	0.481^{***} (0.130)	-1.233^{***} (0.148)	-0.927^{***} (0.136)	-0.551^{***} (0.130)	-0.299^{**} (0.128)
One division lower (d=-1)		0.878^{***} (0.122)	0.506^{***} (0.117)	0.446^{***} (0.108)		-0.588^{***} (0.123)	-0.320^{***} (0.118)	-0.115 (0.108)
Two divisions lower (d=-2)				-0.058 (0.196)				-0.422^{**} (0.201)

2. MOST RECENT MATCH RESULTS ($R^{H}_{i,m}$, $R^{A}_{i,n}$, $R^{H}_{j,n}$, $R^{A}_{j,m}$)

Number of matches ago (m,n)	1	2	3	4	5	6	7	8	9
Home team (i) Home matches (HH)	0.012 (0.008)	0.008 (0.008)	0.037^{***} (0.008)	-0.004 (0.008)	0.005 (0.008)	-0.009 (0.008)	0.006 (0.008)	-0.005 (0.008)	0.010 (0.008)
Away matches (HA)	0.008 (0.008)	0.021^{**} (0.008)	0.020^{**} (0.008)	-0.006 (0.008)					
Away team (j) Home matches (AH)	-0.020^{**} (0.008)	-0.018^{**} (0.008)	-0.014 (0.008)	-0.010 (0.008)					
Away matches (AA)	-0.017^{**} (0.008)	-0.013 (0.008)	-0.021^{**} (0.008)	-0.021^{**} (0.008)	-0.014 (0.008)	-0.000 (0.008)	-0.008 (0.008)	-0.012 (0.008)	-0.026^{***} (0.008)

3. OTHER EXPLANATORY VARIABLES, CUT-OFF PARAMETERS AND DIAGNOSTIC TESTS

	$SIGH_{i,j}$	$SIGA_{i,j}$	$CUP_{i,j}$	$CUP_{i,j}$	$DIST_{i,j}$	$\Delta ATTPOS_{i,1}$	$ATTPOS_{i,2}$	$\Delta ATTPOS_{j,1}$	$ATTPOS_{j,2}$
	0.164^{***} (0.031)	-0.056^{*} (0.032)	-0.116^{***} (0.024)	0.069^{***} (0.025)	0.056^{***} (0.008)	0.194^{***} (0.036)	0.141^{***} (0.022)	-0.188^{***} (0.036)	0.164^{***} (0.022)
	$LAST0_{i,j}$	$LAST1_{i,j}$	μ_1	μ_2	Normality	Hetero.	Omitted lagged win ratios & lagged results		
	-0.043^{**} (0.019)	-0.037^{*} (0.021)	-0.234 (0.090)	0.531 (0.090)	$\chi^2(2)=2.17$	$\chi^2(1)=2.58$	W.R., 24-36 mths (y=2): $\chi^2(24)=9.11$ Results, m=10-12(HH,AA); n=5-6(HA,AH) matches ago: $\chi^2(10)=12.80$		

Notes: Estimation period is seasons 1986-7 to 2000-1 (inclusive). Number of observations = 29,594.
Standard errors of estimated coefficients are shown in parentheses.
*** = significant at 1% level (one-tail test); ** = significant at 5% level; * = significant at 10% level.