The 12th man?: refereeing bias in English and German soccer

Babatunde Buraimo,

University of Central Lancashire, Preston, UK

David Forrest

University of Salford, UK

and Robert Simmons

Lancaster University, UK

Summary. The paper investigates potential bias in awards of player disciplinary sanctions, in the form of cautions (yellow cards) and dismissals (red cards) by referees in the English Premier League and the German *Bundesliga*. Previous studies of behaviour of soccer referees have not adequately incorporated within-game information. Descriptive statistics from our samples clearly show that home teams receive fewer yellow and red cards than away teams. These differences may be wrongly interpreted as evidence of bias where the modeller has failed to include within-game events such as goals scored and recent cards issued. What appears as referee favouritism may actually be excessive and illegal aggressive behaviour by players in teams that are behind in score. We deal with these issues by using a minute-by-minute bivariate probit analysis of yellow and red cards issued in games over six seasons in the two leagues. The significance of a variable to denote the difference in score at the time of sanction suggests that foul play that is induced by a losing position is an important influence on the award of yellow and red cards. Controlling for various pre-game and within-game variables, we find evidence that is indicative of home team favouritism induced by crowd pressure: in Germany home teams with running tracks in their stadia attract more yellow and red cards than teams playing in stadia with less distance between the crowd and the pitch. Separating the competing teams in matches by favourite and underdog status, as perceived by the betting market, yields further evidence, this time for both leagues, that the source of home teams receiving fewer cards is not just that they are disproportionately often the favoured team and disproportionately ahead in score. Thus there is evidence that is consistent with pure referee bias in relative treatments of home and away teams.

Keywords: Bivariate probit; Favouritism; Red cards; Referees; Soccer; Yellow cards

'It was like playing against 12 men'—Sir Alex Ferguson on the performance of referee Herbert Fandel after Manchester United's 2–1 defeat, away to Roma in a Champions' League match, April 2007.

1. Introduction

In professional soccer, referees are appointed to regulate matches under the Laws of Association Football, which are determined by the governing body of world soccer, the Fédération Internationale de Football Association. In applying these laws referees have sanctions in the form of cautions (henceforth called 'yellow cards') and expulsions of players from the field

Address for correspondence: Robert Simmons, Department of Economics, Management School, Lancaster University, Lancaster, LA1 4YX, UK.
E-mail: r.simmons@lancaster.ac.uk

('red cards'). Although yellow cards are issued for less heinous offences, such as dissent, deliberate handball, persistent fouling, obstruction and shirt pulling, this sanction offers an important disincentive to persist in illegal behaviour as a second caution to the same player is accompanied by dismissal ('second yellow card'). A red card results from serious misconduct such as hitting a player or a dangerous tackle or the so-called 'professional foul' where a player deliberately prevents a clear goal scoring opportunity for an opponent by unfair means. Red cards are relatively infrequent.

Fans, players and head coaches worldwide often complain about both inconsistent application of rules by referees and alleged bias against their team. Critical refereeing decisions can be pivotal for a team's prospects of winning championships, qualifying for lucrative European competition or avoiding relegation. As revenue streams, especially sales of broadcast rights, have grown in European football, so criticism of referees' behaviour has intensified. This has begun to be reflected in academic research and various references have investigated particular sources of bias, inconsistency and favouritism offered by referees in various European leagues, including England, Germany and Spain. These papers focus on bias in terms of the differential treatment by referees of home and away teams. Of course, other species of bias could be investigated: for example, Price and Wolfers (2007) considered racial bias by umpires in American basketball.

One strand in the football literature examines the amounts of time that are added on by referees at the end of each half of a game. Referees will stop their clocks immediately if they perceive a player to be sufficiently injured to warrant treatment on the field. They also receive official guidance on amounts of time to add for substitutions and are instructed to resist attempts at time wasting, with cautions if necessary. But the referee takes sole responsibility for timekeeping and has some discretion over amounts of time played. An influential paper by Garicano *et al.* (2005) presented evidence from the Spanish top division that referees awarded less added time after 90 minutes in games where the home team was ahead and more added time when the home team was behind. These results were obtained after controlling for numbers of substitutions, cautions and injuries that would tend to interrupt a match.

Subsequently, other researchers offered broad support for home team favouritism in terms of added time. Lucey and Power (2004) could replicate the results of Garicano *et al.* (2005) for Italian and US Major League soccer. Scoppa (2008) found that referees in Italy's top division, *Serie A*, added significantly greater injury time if home teams were losing. Dohmen (2008) and Sutter and Kocher (2004) reported that, for the German *Bundesliga*, referees added more time in games where the home team was behind. Rickman and Witt (2008) discovered that home team favouritism in the English top division, the Premier League, as revealed by discretionary time added by referees, was present at the beginning of their sample period but appeared to be eradicated when officials were hired on professional employment contracts, with an annual salary rather than a fee per match. When remuneration incentives changed, referees appeared to become more conscious of career implications, in terms of renewal of employment, of making biased judgements regarding time added at the end of games.

Dohmen (2008) added weight to the case that referees are biased by also investigating 'disputable' and 'incorrect' decisions as determined after matches by an independent panel of consultants appointed by the German Football Federation. It appeared that home teams were more likely to benefit from decisions where a goal was disputable or incorrectly allowed or a penalty could or could not have been awarded. Intriguingly, though, bias was less evident in grounds where a running track separated the crowd from the pitch. Referees also seemed, according to Dohmen's statistical analysis, not to offer superfluous extra time on grounds with running tracks and this suggests that the extent and effectiveness of pressure on referees depends on

the design of the stadium. The supposed mechanism is that referees respond, in presumably a subconscious way, to the preferences of the crowd when the physical proximity of the crowd is close.

Dawson *et al.* (2007) took forward the analysis of possible home bias by referees in a study based on seven seasons of data from the English Premier League. They distinguished between home and away team cautions and dismissals in a bivariate negative binomial model. They amalgamated yellow and red cards into a discrete disciplinary points measure. Although the results from the yellow cards regression are stated to be similar to those from this points measure, they did not offer an explicit separate analysis of yellow and red card offences. Their findings are indicative of bias in favour of home teams in that fewer cards are given to home teams and more to away teams, after controlling for relative team strengths and the importance of fixtures for league outcomes. Supporting evidence of referee bias in favour of home teams in the English Premier League was offered by Boyko *et al.* (2007). They examined yellow card awards and penalty decisions in 5244 games and found that there was inconsistent and favourable treatment of home teams across the 50 referees who were considered. However, a subsequent replication study by Johnston (2008) that used more recent data from the Premier League failed to confirm this result.

The prior literature on disciplinary sanctions that was reviewed above uses the match as the unit of observation and this raises issues about controlling for within-game effects. At the match level, for example, it might be thought that part of the reason for more cards on average for away teams is that they are more often than not the underdog, given home advantage, and will therefore typically play more defensively, tackle more often and thereby be penalized more frequently. Dawson *et al.* (2007) found residual bias against away teams even accounting for the match level factors of which team is the underdog and the extent to which it is the underdog. These characteristics were determined from pre-match probabilities of a win for the respective teams as generated from a match result forecasting model. In the majority of matches in any league, the home team will be the favourite. However, the style of play of each team will vary not only by the pre-match prospects based on home advantage and team strength but also by events unfolding as the game progresses. A team that is behind in score may become more desperate and commit more offences. Controlling more precisely for the effect of playing style on award of sanctions therefore requires the use of information on within-game score. Only then can the effect of playing style on sanctions be separated from refereeing bias. Otherwise what is attributed as referee bias may simply result from excessive effort by the offending teams. In particular, away teams may receive more yellow cards in the aggregate just because, on average, they spend more of the game trailing their opponents in score and therefore resort to more foul play. Taking into account events within the game requires a finer level of analysis than the match, which was the unit of observation in Dawson *et al.* (2007). Of course, the team attribute of being 'underdog' does not vary through the game. But the manifestation of that attribute, in terms of player misbehaviour and hence referee sanctions, will vary through games according to the state of the match score and it is this within-game variation in sanctions that we analyse in this paper.

Our contribution to the study of referee bias in football therefore departs from the existing literature by acknowledging the potential importance of within-game dynamics. We switch from match to minute of game as the unit of observation. Rather than model the number of cautions and dismissals as count variables (as in Dawson *et al.* (2007)), we model the *probability of a caution or dismissal within a specific minute of a match*. This permits us to introduce a full set of relevant within-game covariates such as the number of yellow and red cards that are issued up to the minute and the difference in score at the start of the minute. Specifically,

we set up a bivariate probit model (see Greene (2003), pages 710–719, for a formal exposition) in which the likelihood that a card is issued is determined for home and away teams jointly. Thus the two events of home team sanction and away team sanction are modelled simultaneously, and conditionally on sets of explanatory variables they are considered jointly dependent.

A further innovation in the paper is to bring together results from two leagues over the same time period. The literature to date reads as a set of cases, with one league analysed at a time. Here, we can present evidence of refereeing bias, using an analysis of yellow and red cards from two major European football leagues, not just one. The application to more than one league facilitates a search for corroborating evidence. Here, we undertake parallel analyses of the two countries, since differences in playing styles and referees' behaviour are likely to be systematic across the two leagues, rendering pooling of games invalid. Our sample period for each league covers the 2000–2001 to 2005–2006 seasons, which are a period of time when there were no radical changes in the rules of football.

2. Data and model

We propose two pairs of dependent variables. The first pair comprises binary indicators of the award of a specific type of card to home and away teams in a given minute of a match. For example, we model the joint responses of a home team given a yellow card and an away team given a yellow card. We then repeat the procedure for a home team given a red card and an away team given a red card, again modelled jointly. Unlike the Premier League, the *Bundesliga* features some stadia where a running track separates the crowd from the pitch and we wish to investigate whether this structural feature of football grounds has any effect on referees' behaviour in the award of sanctions. For this reason, models for the *Bundesliga* incorporate this feature but not the models for the English Premier League.

Following Dohmen (2008), we hypothesize that crowd pressure on referees will decrease when a track is present. We conjecture that referees will feel less intimidated in their decisions to award yellow and red cards to home teams when a track separates referees from the crowd. We construct a dummy variable, *track*, to capture the presence of a running track in a stadium. Then, the hypothesis to be tested is that home teams have a higher probability of receipt of a card and/or away teams have a lower probability of receipt of a card when a track is present.

The top division of the *Bundesliga* has 18 teams and our data set spans six seasons, giving a total of 108 team–season observations. Data on tracks in stadia, which were kindly provided by Joachim Prinz of the University of Paderborn, show that the number of team–season observations with tracks is 39 out of 108. The clubs that played in a stadium with a track varied over time. Three teams actually changed their ground structure during our sample period. Both Schalke, in 2001, and Bayern Munich, in 2005, moved to a new stadium without a running track. Hannover, in 2003, renovated its old stadium, removing the existing running track.

In a paper on referee bias in Italian football, Pettersson-Lidbom and Priks (2007) argued that referees were more likely to award yellow and red cards to home team players where the authorities had ordered games to be played in empty stadia following previous crowd trouble. However, it should be noted that, of their sample of 842 games across the top two divisions in Italian football, only 24 were played in empty stadia. Further, Pettersson-Lidbom and Priks did not control for within-game influences on the award of cards. Most importantly for our analysis, the empty Italian stadia in their study tended to be those belonging to clubs with a record of crowd trouble at home games and the 'empty stadium' effect on the award of cards may simply be a team-specific effect. By contrast, in our case of the *Bundesliga*, we have a substantial number

of 'with-track' game observations. Crucially, with some of the *Bundesliga* teams switching stadium design in the sample period, our track variable is not confounded with home-team-specific effects on treatment by referees and enables their separate identification.

Another dimension of stadium architecture is the size of the pitch, which is not homogeneous. The Laws of Association Football merely specify the minimum and maximum length and width of approved pitches. However, the variation in pitch dimensions across the *Bundesliga* is small, with the range of both length and width being just 5 m. Further, the presence of a running track does not appear to be correlated with the size of the pitch. Our conclusion is that any effects that we attribute to the presence of a track are unlikely merely to be reflecting factors that are associated with the size of the pitch as for example in any tendency in football for conflict to be greater where teams confront each other in a confined space.

As indicated previously our initial focus will be on the joint modelling of home and away team responses of yellow and red cards. Later in the paper, we shall also offer empirical results from estimation where the home and away classification is replaced by a different configuration: that of favourite team and underdog team. There the pairs of dependent variables will be the probabilities of cards of a particular colour being awarded to the *favourite* and *underdog* teams, as assessed by the bookmaker betting market for match results.

This alternative classification of team responses, between favourite and underdog, allows us to address the issue of home team bias directly by modelling the probability that a team will be awarded a particular card in the next minute of a match with a specific dummy variable included to reflect whether a team is home or away. To estimate the two equations, including the dummy of whether or not the underdog team is at home, which we call the 'home underdog' or equivalently whether the favourite team is the away team, requires that the favourite teams have a mix of home and away teams among them. We emphasize that, although the majority of favourite teams in our two samples are home teams, there are still some games in each league where the home team is the underdog, and we can capture this feature by using a dummy variable. Accordingly we adopt an alternative perspective on each match from that taken earlier in the home–away analysis. A convenient dichotomy is between teams that were favourites or underdogs in the betting market. A small number of matches had equal win probabilities for the two teams so there was no favourite and these games were excluded from the sample.

As noted above, the use of minute of game as the unit of observation allows us to control for within-game influences on the award of cards. We have data on times of yellow and red cards and goals scored for the English Premier League and *Bundesliga 1*, obtained from `www.11v11.co.uk` and `http://www.bundesliga.de`. In general, the two sites offer consistency in timing of cards and goals. The *Bundesliga* site has a peculiar feature in that it records minute 0 as the beginning and minute 89 as the end of a match. We added 1 minute on for each card and goal taken from the German site to provide consistency with the English data. There are occasions where more than one card is issued to the same team in a particular minute and, when that happens, we simply record '1'; there is therefore an extent to which this recording of cards slightly understates the total.

Our sources permit us to separate the data into yellow card, second yellow card and red card categories. By yellow card, we mean the first caution that is awarded in a game to a given player. Our variable 'yellow card' does not include instances of a second yellow card that is shown to the same player since this has different implications: automatically the player is shown a red card as well and therefore is dismissed from the field. Below, our category 'all red cards' includes both these 'second yellows' and the 'straight reds' that are handed out for offences that are sufficiently serious to merit immediate dismissal. Clearly, yellow cards occur much more frequently than red cards.

Table 1 shows the incidence of yellow and red cards per match in the top division of the *Bundesliga* and in the English Premier League, separated into 15-minute segments. It is clear that, generally, the frequency of cards is greater with the passage of time in a match. Also, the frequency of yellow cards is higher for the *Bundesliga* than for the English Premier League, possibly reflecting a more severe interpretation of the Laws of the Game in Germany.

Table 1 also reveals the rarity of dismissals, relative to cautions. This is itself a consequence of incentives. When a player receives a red card, his team plays with 10 men and evidence shows that teams with 11 men have a greater chance of winning against 10 men compared with 11, although this depends on the timing of dismissal and hence how much time is left for the team with full strength to exploit its advantage (Ridder *et al.*, 1994; Torgler, 2004; Caliendo and Radic, 2006). If the depleted team loses the game then the dismissed player may receive blame from fans and coaches for the defeat. For instance, in press conferences head coaches sometimes speak of an 'unnecessary dismissal' when a player performs a reckless act that induces a red card. Unless the dismissal is found to be unfair on appeal, the player will also serve a suspension with a minimum of three games for a straight red card and may receive a fine if the offence was very serious. The suspension has career implications for the player in that a replacement may claim and retain the player's place, even when the suspended player becomes available again. As a result, the kind of severe offences that are found in amateur football, such as fighting between players, are far less prevalent in the professional game.

We distinguish between control variables for within-game and pre-game influences. In the former category, we include *minute* and *minute squared* as covariates since it appears that, the longer the match continues, the more likely it is that a card will be issued. Also, we should note that neither Web site records time added at the end of each half and so the 45th and 90th minutes will typically last longer than others because they include 'injury time'. We account for this feature of the data by using dummy variables, *45th minute* and *90th minute*. The dynamics of previous yellow cards are included by separating numbers of cards that are issued to home and away teams in the preceding 3 minutes, *home yellow last 3 minutes* and *away yellow last 3 minutes*, from numbers of cards issued earlier in the game than the preceding 3 minutes, *home yellow prior* and *away yellow prior*.

Table 1.   Referees' awards of cards in the *Bundesliga* and Premier League in all games

Minutes	Home yellow	Away yellow	Home 2nd yellow	Away 2nd yellow	Home red	Away red
Germany (1836 games)						
1–15	239	267	0	0	1	3
16–30	437	615	4	7	7	7
31–45	673	758	8	23	14	22
46–60	521	631	7	17	10	24
61–75	622	766	6	31	13	27
75–90	767	942	25	35	25	41
England (2280 games)						
1–15	174	258	0	1	3	9
16–30	344	494	3	3	16	9
31–45	587	771	6	15	12	14
46–60	474	631	6	11	17	16
61–75	573	768	14	29	26	34
75–90	746	984	31	49	31	35

The choice of 3 minutes as the timeframe within which to treat earlier yellow card incidents as recent was determined from preliminary investigation. In the 1- or 2-minute period following a caution or dismissal, there is little opportunity for either team to react in terms of illegal behaviour. The referee may have to calm players down following protests about decisions and usually a free kick or penalty will be awarded, all of which takes up time. Our results are robust to alternative choices of 4, 5 and 6 minutes and regressions with these longer intervals yielded slightly lower pseudo-R^2-values. On this basis we selected 3 minutes as our threshold for recent events.

Numbers of straight red cards issued in the game before the subject minute are captured by *home red* and *away red*. Numbers of second yellow cards issued previously in the match are similarly captured by *home 2nd yellow* and *away 2nd yellow*. The effects of these within-game dynamic effects cannot be signed *a priori*. On the one hand, an extra card to a team may reduce the probability that a further card will be issued to the same team: a deterrence effect. On the other hand, an extra card to a team may be part of an escalation in illegal conflict between teams. Also, observers sometimes claim that referees have a tendency to 'even out' decisions so that a caution that is given to one team is followed by another to its opponent, but we cannot distinguish this from conflict escalation that is generated by the players themselves. Our set of variables covering previous yellow and red cards serves to control for dynamics of conflict during a game.

Illegal activity may also increase as teams fall behind and we register the *goal difference*, which is defined as home team current score minus away team current score as the match status at any point in time.

Teams that are behind in score, which are more often away than home teams, are hypothesized to generate extra effort in an attempt to negate the deficit and some of this extra effort will spill over into illegal activity, which is punishable by cautions or dismissals. This illegal effort was characterized by Garicano and Palacios-Huerta (2006) as 'sabotage in tournaments', where players attempt to reduce the effectiveness of opponents by unfair means. The propensity to undertake sabotage activity will be enhanced for a team that is behind in score.

The effect of goal difference on the probability of receiving a card of a particular colour may therefore depend, first, on which team is ahead in score and, second, on the difference in score at the time. For example, one extra goal from a 3–0 scoreline may have a different influence on the probability of receiving a card for the home team compared with one from a scoreline of 0–0. In the former case, the effort of the home team may be less intense as the game is virtually decided and a lower probability of home card may be a consequence. The effort of the away team could also of course become less intense as its deficit becomes large. To capture this possible non-linearity of effects, we introduce the quadratic term, *goal difference squared*.

Some football matches are notable for the intense rivalry that they generate among supporters and players; the result of tradition and independent of current team league standings and prospects. These matches are generally played between two local teams that are a short distance apart. Witt (2005) used distance between stadia of competing teams to capture the influence of local rivalry. Here we nominate a particular set of rivalrous games denoted by *derby* and predict that these will generate a higher probability of a caution or dismissal for each team. In Germany, there is typically much greater average distance between teams than in England and so far fewer local derbies are to be found in the *Bundesliga*, which has just six team pairs, following the list that was provided by Benz *et al.* (2009).

The extent of pressure that is exerted by fans may be positively related to the size of crowd and we proxy crowd intensity by the logarithm of match attendance (*log attendance*). Dawson

et al. (2007) used this measure as a covariate in their analysis of disciplinary points.

We also include a measure of *ex ante* relative team strength. Dawson *et al.* (2007) used an elaborate statistical forecasting model to generate win probabilities for the respective teams in a match. An alternative is to derive a relative team strength measure from betting odds. This has the advantage that odds will incorporate not only information from previous matches, as in the statistical model, but also fresh news such as that pertaining to absences of players from injury or suspension. Of course, reliance on bookmakers' odds to capture relative team strengths, as modified by home advantage, depends on the betting market being efficient. In the sample period that was employed by Dawson *et al.* (2007), there is evidence that it was not fully so (Forrest and Simmons, 2002) and a statistical model may indeed have been a more appropriate basis for deriving win probabilities.

However, since the abolition of betting tax in 2001, and as the growth of Internet competition has put pressure on bookmakers' margins, there is evidence (Forrest *et al.*, 2005) that the betting market has moved strongly towards displaying efficiency, i.e. towards odds capturing accurately all the factors that are relevant to the outcome of a match. Accordingly we choose to exploit odds data, from Ladbrokes, the largest UK bookmaker, and include in our model *difference in bookmaker probability*, which is defined as home win probability minus away win probability and, to capture non-linearity, its square. This variable is a proxy for *ex ante* relative team strengths. The larger the value of this variable, the stronger the relative strength of the home side and, we predict, the less or more likely that the home or away team respectively is to be awarded cards.

Using the bivariate probit link function we then jointly model the probabilities that the home team will receive a yellow card and the away team a yellow card in a minute period. The probit model incorporates for both probabilities the covariates which have been fully defined above: minute, minute squared, 45th minute, 90th minute, home yellow last 3 minutes, away yellow last 3 minutes, home yellow prior, away yellow prior, home 2nd yellow, away 2nd yellow, home red, away red, goal difference, goal difference squared, track, log attendance, derby, difference in bookmaker probability and difference in bookmaker probability squared.

Additionally, to account for grouping of observations into referees, covariates for fixed effects of referees were included in all the model specifications. Similarly included were sets of dummy variables for effects of the specific home team, the specific away team and the season of observations. A separate bivariate probit analysis was undertaken for home and away straight red cards by using the same set of covariates. As before, these models include home team, away team, referee and season effects.

Subsequently in the paper, models for both yellow and straight red cards and an additional model for the category of red cards induced by a second caution are fitted, but now the two joint probabilities in each model are for the favourite team and the underdog team. The control dummy variables for team, season and referee fixed effects are again included in these models. The covariates that are used in this set of models are now minute, minute squared, 45th minute, 90th minute, favourite yellow last 3 minutes, underdog yellow last 3 minutes, favourite yellow prior, underdog yellow prior, favourite 2nd yellow, underdog 2nd yellow, favourite red, underdog red, goal difference, goal difference squared, home underdog match, derby and difference in bookmaker probability. Here, of course, the goal and bookmaker probability variables are defined in terms of favourite team value minus underdog team value.

Our data are multilevel and are clustered by match. This induces correlation between observations within matches. The model estimation which we discuss in the next section assumes independence of error terms across matches but incorporates an adjustment for interdependence of error terms within matches to produce robust standard error estimates.

3. Empirical results

We begin with a descriptive analysis of referees' propensities to award yellow and red cards in the *Bundesliga* and the Premier League. The matches in our samples were officiated by 35 referees in the *Bundesliga* and 36 referees in the Premier League. Some referees were in charge of small numbers of games: just one in some instances. The average number of games officiated is 52 in the *Bundesliga* and 63 in the Premier League. The standard deviations are 36 and 47 respectively. In each league, referees are obliged to retire at the age of 48 years, so our sample includes some referees who had just begun careers at the top level and others who retired or were demoted during the sample period. An analysis of duration of career and referees' performance in the *Bundesliga* was presented by Frick *et al.* (2008).

Focusing on referees who officiated at least 25 games, we consider the mean number of each type of card for home and away teams. There are two immediate results from *t*-tests, which were conducted with samples of unequal variance. First, referees in each league tended to award fewer cards per game to home teams than to away teams, though this was not necessarily due to referee bias as away teams are more often involved in defensive play as they struggle to overcome home advantage. The only exception to this generalization was the case of straight red cards in England where the null hypothesis of equality of cards per game for home and away teams is not rejected ($p = 0.31$). Second, English referees tended to award fewer yellow cards per game than their German colleagues, irrespectively of whether the team was home or away. For both home and away teams these differences between England and Germany are statistically highly significant. However, for straight red cards and also, for home team only, second yellow cards, English referees were no less severe than German referees. Differences in incidence of cards across leagues could reflect different degrees of aggression; the industry stereotype is that the Premier League is faster paced, more physical and more attack oriented than the *Bundesliga*. But differences in interpretation of the Laws of the Game may also play a part. Offences meriting straight red cards involve more clear-cut decisions by referees compared with yellow cards, where there is greater scope for discretion.

Table 2 shows estimates of the parameters of bivariate probit models for the probabilities of yellow cards given to home and away teams, separately for each league. Table 3 has the corresponding results for the red card responses. The parameter estimates for the covariates in all the tables are their coefficients in a linear function, which in a probit model predicts a monotonic transformation $F^{-1}(p)$ of a probability p where F^{-1} is the inverse of the cumulative distribution function of the standard normal distribution. Estimation is maximum likelihood implemented in Stata 10 by the *biprobit* command with the option *cluster* to adjust standard error estimates to account for within-match dependences of observations (Stata, 2007). The tables also report an estimate of ρ, which represents the residual correlation between the joint responses in the model, after fitting the covariates. If ρ is 0, then separate probit models for the response would be appropriate rather than their being jointly dependent. A test for the null hypothesis of zero ρ is provided by comparing the log-likelihood for the bivariate probit models with sum of log-likelihoods of the separate univariate probit models. In all our bivariate probit results in Tables 2–6, this null hypothesis is always rejected at the 1% level. In some cases the estimate of ρ is also quite large. This justifies our use of bivariate probit rather than separate probit models for home–away and favourite–underdog equations.

Our control variables show statistically significant, plausible and important effects from within-game dynamics in both leagues. The explanatory power of the models drops markedly when within-game control variables are removed. For example, in Table 2, the pseudo-R^2 drops from 0.026 to 0.005 when within-game controls are removed from the *Bundesliga* yellow card model.

Table 2. Bivariate probit regression results for yellow card†

Covariate coefficient	Results for Bundesliga model		Results for Premier League model	
	Home team	*Away team*	*Home team*	*Away team*
Within game				
Minute	0.015	0.015	0.015	0.015
	(12.02)	(13.00)	(11.38)	(13.07)
Minute squared	−0.000098	−0.000096	−0.000099	−0.000101
	(7.51)	(7.86)	(7.41)	(8.32)
45th minute	0.346	0.273	0.437	0.534
	(6.55)	(5.23)	(8.99)	(12.63)
90th minute	0.526	0.550	0.638	0.735
	(10.34)	(11.61)	(13.81)	(17.86)
Home yellow last 3 minutes	−0.249	0.120	−0.032	0.195
	(6.90)	(4.33)	(0.92)	(7.04)
Away yellow last 3 minutes	0.102	−0.174	0.156	−0.104
	(3.77)	(6.17)	(5.44)	(3.64)
Home yellow prior	−0.088	0.016	−0.043	0.050
	(8.42)	(1.78)	(4.36)	(5.48)
Away yellow prior	0.024	−0.084	0.046	−0.066
	(2.64)	(9.53)	(5.19)	(7.13)
Home 2nd yellow	−0.114	0.132	0.023	−0.004
	(1.16)	(2.06)	(0.37)	(0.05)
Away 2nd yellow	0.057	−0.042	0.017	−0.032
	(1.09)	(0.84)	(0.31)	(0.53)
Home red	0.039	0.105	0.010	0.078
	(0.71)	(1.99)	(0.19)	(1.89)
Away red	0.043	−0.022	−0.023	−0.035
	(0.88)	(0.48)	(0.44)	(0.66)
Goal difference	−0.025	0.019	−0.028	0.030
	(3.49)	(2.96)	(3.82)	(4.43)
Goal difference squared	−0.015	−0.015	−0.011	−0.020
	(4.79)	(5.14)	(3.19)	(6.22)
Pre-game				
Track	0.102	0.038		
	(2.40)	(0.90)		
Log attendance	0.040	−0.010	−0.014	−0.025
	(1.03)	(0.27)	(0.17)	(0.37)
Derby	−0.002	0.057	0.074	0.111
	(0.03)	(1.47)	(3.00)	(4.88)
Difference in bookmaker probability	0.047	0.256	−0.014	0.093
	(0.76)	(4.19)	(0.27)	(1.92)
Difference in bookmaker probability squared	−0.423	−0.366	−0.334	−0.213
	(3.15)	(3.18)	(3.52)	(2.57)
Constant	−2.906	−2.210	−2.186	−2.144
	(6.50)	(5.23)	(2.63)	(3.00)
Pseudo-R^2	0.026	0.022	0.037	0.036
ρ		0.188		0.128
		(11.73)		(6.88)
Log-likelihood		−32916		−32844
Number of observations (minutes)		159210		204480

†Dependent responses are teams receiving a card in a given minute. Estimates are for coefficients of covariates in the probit model linear predictor of $F^{-1}(p)$ where p is the probability of receiving the card. In parentheses are absolute values of *t*-statistics, computed by using robust standard errors adjusting for clustering of observations within matches. Models are estimated jointly for home and away teams in each league but separately for the *Bundesliga* and Premier League. Models also include sets of dummy variables for the fixed effects of referee, home teams, away teams and year. Results for these are not reported.

Table 3. Bivariate probit regression results for straight red card†

Covariate coefficient	Results for Bundesliga model		Results for Premier League model	
	Home team	*Away team*	*Home team*	*Away team*
Within game				
Minute	0.013	0.018	0.010	0.007
	(2.25)	(3.09)	(2.25)	(1.30)
Minute squared	−0.000073	−0.000136	−0.000062	−0.000011
	(1.11)	(2.41)	(1.31)	(0.22)
45th minute	0.113	0.243	0.189	0.170
	(0.40)	(1.41)	(0.89)	(0.84)
90th minute	0.730	0.679	0.331	0.538
	(4.50)	(5.49)	(2.04)	(4.42)
Home yellow last 3 minutes	0.276	0.154	0.115	−0.050
	(2.55)	(1.62)	(1.07)	(0.39)
Away yellow last 3 minutes	−0.205	0.117	0.011	0.250
	(1.42)	(1.32)	(0.10)	(2.44)
Home yellow prior	0.013	0.067	−0.020	−0.031
	(0.36)	(2.21)	(0.55)	(0.88)
Away yellow prior	0.010	0.017	0.068	0.053
	(0.25)	(0.60)	(2.20)	(1.59)
Home 2nd yellow	−4.797	0.348	−4.539	0.190
	(19.55)	(2.06)	(27.33)	(0.95)
Away 2nd yellow	−0.046	0.175	0.102	−4.835
	(0.24)	(1.25)	(0.54)	(32.48)
Home red	−0.177	0.226	−0.156	0.316
	(1.21)	(1.46)	(1.35)	(2.55)
Away red	0.232	−0.293	0.314	−0.168
	(1.35)	(1.94)	(2.44)	(1.31)
Goal difference	−0.125	0.094	−0.111	0.082
	(3.41)	(3.32)	(3.98)	(1.99)
Goal difference squared	−0.057	−0.021	−0.004	−0.046
	(2.39)	(1.68)	(0.38)	(2.38)
Pre-game				
Track	0.267	0.163		
	(1.72)	(1.11)		
Log attendance	−0.282	−0.419	0.381	0.290
	(1.76)	(2.82)	(1.18)	(0.98)
Derby	−0.082	0.078	0.073	0.021
	(0.43)	(0.51)	(0.73)	(0.24)
Difference in bookmaker probability	−0.541	−0.247	0.449	0.107
	(2.24)	(1.25)	(1.97)	(0.56)
Difference in bookmaker probability squared	−0.336	−0.156	−1.142	−0.188
	(0.63)	(0.32)	(2.58)	(0.55)
Constant	−1.199	0.331	−8.262	−6.597
	(0.66)	(0.20)	(2.42)	(2.15)
Pseudo-R^2	0.117	0.093	0.091	0.071
ρ		0.760		0.589
		(8.32)		(6.15)
Log-likelihood		−1349		−1665
Number of observations (minutes)		159210		204480

†The footnote for Table 2 applies here also.

An extra yellow card that is received by the home team previously in the match, whether in the immediately preceding 3 minutes or before that, is associated with a reduced probability of receiving a home team yellow card in the current minute. Similarly, the probability of away team yellow cards is negatively related to prior incidence of away team yellow cards. These results are consistent with the intended deterrent effect of cautions.

By contrast, an extra yellow card received previously in the match by either a home or an away team, whether in the immediately preceding 3 minutes or before that, is associated with an *increased* probability of receiving a yellow card for its opponents in the current minute. In Table 2, the number of previous second yellow cards in a match is only significant with positive coefficient for the probability that a *Bundesliga* away team receives a yellow card. However, Table 3 shows that the number of previous second yellow cards in a match has particular relevance for the likelihood of another red card.

An extra goal scored by the home team, with away score constant, is shown to lead to a reduced probability that a home team will receive a yellow card in each league. The negative coefficient on goal difference squared shows that the effect of increasing goal difference increasingly lowers the probability that a home team will receive a yellow card. Thus an extra goal for the home team results in a lower probability that a home team will receive a yellow card at both 1–0 and 3–0 but there is a higher absolute marginal effect in the latter case.

An extra home team goal also leads to an increased probability that an away team will receive a yellow card if the away team is behind in score, this time with an inverted U-shaped relationship between yellow card probability and goal difference. Turning points in both the *Bundesliga* and the Premiership are between zero and a one-goal deficit for the away team. Hence, a change in goal difference from 0–2 to 0–1 is associated with an increased probability of receiving a yellow card for the away team. But a change in goal difference from 1–0 to 2–0 is associated with a lower probability of receiving a yellow card for the away team. These results therefore confirm that controlling for goal supremacy during the match is an important feature of our model and that use of a non-linear functional form yields interesting and plausible results. They suggest, for example, that much of the aggression that is seen in top level football is controlled: as the home club's lead becomes decisive, both teams become less likely to attract sanctions, presumably because the expected pay-off to committing offences is now low.

Turning to pre-game covariates, we note that derby matches in England generate an increased probability of cautions for each team, but this effect is absent in Germany where there are far fewer fixtures of intense local rivalry. Attendance is not a significant predictor of the likelihood of caution in either league. We also tested for significance of crowd density measured as attendance divided by ground capacity and, further, for attendance interacted with track in the *Bundesliga* model. Neither of these additional attendance terms was significant and they are excluded from the model. Another factor that might exert an influence is the composition of the crowd between home and away supporters. Unfortunately, data on the composition of crowds are not available in either England or Germany and so we could not test for crowd composition effects on the likelihood of sanctions. However, in the case of England, crowd composition does not seem likely to vary very much. A high proportion of games are played in stadia where all tickets are routinely sold with only a small and fixed percentage of them made available to the visiting club.

An increase in difference in the bookmaker's win probability for the home team relative to the away team is found to be associated with a reduced likelihood that the home team will receive a yellow card. By contrast, a consideration of the combined linear and quadratic effects from an increase in the difference in bookmakers probability shows an increased likelihood that an away team will receive a yellow card. For example, the turning point for the *Bundesliga* is where

the difference in bookmakers probability is approximately 0.36. This is beyond the maximum value that is observed in the sample. Thus, within the sample range, the association between the probability of an away yellow and the difference in bookmaker win probability is positive. Given efficiency in the bookmakers' betting market for match results, these results suggest that *ex ante* relative team strength is a significant predictor of the likelihood of cautions for the competing teams in a match. The greater the likely superiority of the home team, as signalled by betting odds, the fewer yellow cards the home team is predicted to receive and the more yellow cards the away team is predicted to receive.

We can test for refereeing bias in the *Bundesliga* through the exploitation of information about the presence of a running track in stadia (Dohmen, 2008). The inclusion of the dummy variable track accounts for whether the crowd was or was not separated from the field of play by a running track. This is an irrelevant consideration for the Premier League where no stadium has a track. The results in Table 2 show that, controlling for both within-game and pre-game effects, home teams face an increased probability of receiving a yellow card where there is a track, with a coefficient that is significant beyond the 5% level. We also see from Table 3 that the track has a positive effect on the home team's probability of receiving a red card. However, owing to the much lower frequency of home red cards relative to home yellow cards and consequent low precision of the parameter estimate, the effect of track is only marginally significantly different from 0 ($t = 1.72$).

That whether there is a running track has a statistically significant influence on the probability that a home team receives a yellow card could, in principle, follow from modification of either referee or player behaviour. For referees, the presence of a track creates less intimidation and noise from the crowd. If referees react to this change in atmosphere, it would be expected to be in the direction of issuing *more* cards to the home team. For home players, a running track also changes the atmosphere. They will be less subject to being enthused by their supporters; less 'stirred up'. If home players react to this change, it would be expected to be in the direction of playing less intensely and less aggressively and, as a consequence, they should receive *fewer* cards. The two effects, through referee and player responses, work in opposite directions. That the net effect of a running track is to increase cards issued to home players suggests that the result is being driven by the referee's response to the proximity of the crowd and this is consistent with referees typically being biased towards the home team because of the presence of partisan spectators. It is an example of 'favouritism under social pressure' (Garicano *et al.*, 2005).

The findings on refereeing bias in the *Bundesliga* here are consistent with those of Dohmen (2008) who examined data from independent consultants, who were appointed by the German Football Federation, on correctness of referees' decisions to award penalty kicks. More penalties were awarded to home teams in stadia without a running track. This suggests, reinforced by our result, that, given equal revenue-generating potential, the removal of a running track was a rational decision by the three *Bundesliga 1* clubs that did so in our sample period.

Behind all these findings may be that pressure on the referee is exerted through the volume of noise. Nevill *et al.* (2002) performed an experiment in which two sets of referees viewed a videotape of a Premier League match under different conditions and were asked to nominate the award of free kicks. One set viewed the replay with the sound of fan noise eliminated while the other group watched with sound retained. The latter group offered more decisions in favour of the home team.

We turn next to results based on estimating the models for the probability that cards will be issued to favourite and underdog teams as specified in Section 2. The covariates are similar to those before except that we now exclude those (track and log attendance) that are associated with the identity of the home team. Here, teams are defined according to whether they were the favourite or the underdog in a match. Each category includes a mix of home and away teams.

Table 4. Bivariate probit model of yellow card with favourites and underdogs†

Covariate coefficient	Results for Bundesliga model		Results for Premier League model	
	Favourite team	*Underdog team*	*Favourite team*	*Underdog team*
Within game				
Minute	0.014	0.016	0.015	0.015
	(11.30)	(12.97)	(11.44)	(12.50)
Minute squared	−0.000085	−0.000102	−0.000100	−0.000098
	(6.46)	(8.23)	(7.48)	(7.89)
45th minute	0.336	0.285	0.457	0.521
	(6.21)	(5.42)	(9.47)	(11.95)
90th minute	0.510	0.558	0.637	0.731
	(10.03)	(11.49)	(13.61)	(17.44)
Favourite yellow last 3 minutes	−0.196	0.089	−0.022	0.174
	(6.02)	(3.15)	(0.66)	(6.11)
Underdog yellow last 3 minutes	0.119	−0.196	0.166	−0.090
	(4.37)	(6.40)	(5.79)	(3.02)
Favourite yellow prior	−0.081	0.007	−0.035	0.046
	(7.74)	(0.82)	(3.54)	(5.06)
Underdog yellow prior	0.022	−0.082	0.042	−0.063
	(2.50)	(9.10)	(4.51)	(6.58)
Favourite 2nd yellow	−0.072	0.141	−0.004	−0.024
	(0.78)	(2.04)	(0.06)	(0.30)
Underdog 2nd yellow	0.024	−0.057	0.029	−0.003
	(0.48)	(1.09)	(0.48)	(0.06)
Favourite red	0.119	0.111	0.047	0.026
	(2.31)	(1.98)	(0.82)	(0.57)
Underdog red	0.011	−0.044	0.007	−0.025
	(0.21)	(0.93)	(0.15)	(0.53)
Goal difference	−0.011	0.004	−0.017	0.019
	(1.47)	(0.65)	(2.35)	(2.77)
Goal difference squared	−0.018	−0.010	−0.014	−0.016
	(5.77)	(3.74)	(3.75)	(5.26)
Pre-game				
Home underdog match	−0.010	−0.127	0.090	−0.134
	(0.42)	(5.49)	(3.90)	(5.89)
Derby	0.030	0.022	0.065	0.116
	(0.64)	(0.50)	(2.53)	(4.86)
Difference in bookmaker probability	−0.229	0.002	−0.236	−0.013
	(3.84)	(0.05)	(4.22)	(0.27)
Constant	−2.301	−2.323	−2.356	−2.360
	(27.48)	(29.46)	(34.02)	(33.75)
Pseudo-R^2	0.026	0.021	0.036	0.036
ρ		0.185		0.124
		(11.37)		(6.53)
Log-likelihood		−31968		−31816
Number of observations (minutes)		154980		198180

†The footnote for Table 2 applies here also.

Consequently, any effects from track and log attendance could be captured only by employing a set of interaction terms, which would make an unwieldy model. We focus instead on the very clear question of whether it makes a difference to the number of cards according to whether a team is playing at home or away. This can be inferred from the coefficients on the variable home underdog match.

Table 5. Bivariate probit model of straight red card with favourites and underdogs†

Covariate coefficient	Results for Bundesliga model		Results for Premier League model	
	Favourite team	*Underdog team*	*Favourite team*	*Underdog team*
Within game				
Minute	0.016	0.013	0.017	0.005
	(5.93)	(2.35)	(7.30)	(1.05)
Minute squared	−0.000087	−0.000065	−0.000118	−0.000010
	(3.08)	(1.23)	(4.94)	(0.22)
45th minute	0.196	0.173	0.468	0.269
	(1.62)	(0.85)	(5.99)	(1.53)
90th minute	0.628	0.512	0.665	0.457
	(6.54)	(3.94)	(8.62)	(3.37)
Favourite yellow last 3 minutes	−0.128	0.147	−0.097	−0.073
	(2.09)	(1.59)	(1.85)	(0.56)
Underdog yellow last 3 minutes	0.157	0.080	0.175	0.330
	(2.78)	(0.85)	(3.51)	(3.82)
Favourite yellow prior	−0.129	0.036	−0.072	−0.007
	(6.48)	(1.15)	(3.95)	(0.22)
Underdog yellow prior	0.050	0.021	0.057	0.050
	(2.53)	(0.80)	(3.33)	(1.48)
Favourite 2nd yellow	0.045	0.177	−0.066	0.295
	(0.37)	(0.76)	(0.46)	(1.67)
Underdog 2nd yellow	0.072	0.027	0.047	−4.839
	(0.99)	(0.17)	(0.43)	(26.99)
Favourite red	0.116	−0.143	0.108	0.311
	(1.28)	(0.70)	(0.98)	(2.70)
Underdog red	0.140	−0.181	0.149	−0.178
	(1.10)	(1.19)	(1.96)	(1.44)
Goal difference	0.040	0.058	0.023	0.018
	(2.64)	(2.34)	(1.86)	(0.58)
Goal difference squared	−0.038	−0.015	−0.021	−0.016
	(4.86)	(1.31)	(3.47)	(1.02)
Pre-game				
Home underdog match	1.384	−0.039	1.230	−0.030
	(22.25)	(0.45)	(22.20)	(0.33)
Derby	0.001	−0.044	0.143	−0.038
	(0.01)	(0.24)	(3.32)	(0.40)
Difference in bookmaker probability	−0.365	−0.269	0.228	−0.535
	(2.81)	(1.03)	(0.71)	(1.00)
Constant	−3.776	−3.913	−3.753	−3.512
	(20.05)	(10.21)	(22.21)	(12.21)
Pseudo-R^2	0.234	0.077	0.197	0.061
ρ		0.540		0.306
		(7.34)		(3.55)
Log-likelihood		−4180		−5746
Number of observations (minutes)		154980		198180

†The footnote for Table 2 applies here also.

The variable, difference in bookmaker probability, is retained. Of course, this variable is now constrained to the positive range, since the favourite, by definition, has the higher win probability. Continuing to control for win probabilities is key here. As in Dawson *et al.* (2007), it should be recognized that the reason for a greater incidence of yellow cards for visiting teams might be that *typically* they are underdogs and may therefore attempt to employ more foul play as an

Table 6. Bivariate probit model of red card (including 2nd yellow) with favourites and underdogs†

Covariate coefficient	Results for Bundesliga model		Results for Premier League model	
	Favourite team	*Underdog team*	*Favourite team*	*Underdog team*
Within game				
Minute	0.016	0.018	0.017	0.006
	(6.04)	(4.31)	(7.25)	(1.52)
Minute squared	−0.000082	−0.000108	−0.000112	−0.000018
	(3.08)	(2.70)	(4.83)	(0.47)
45th minute	0.165	0.171	0.473	0.414
	(1.41)	(1.07)	(6.28)	(3.20)
90th minute	0.595	0.515	0.686	0.477
	(6.61)	(4.85)	(9.39)	(4.76)
Favourite yellow last 3 minutes	−0.097	0.149	−0.039	0.022
	(1.68)	(2.06)	(0.79)	(0.26)
Underdog yellow last 3 minutes	0.128	0.156	0.166	0.397
	(2.37)	(2.30)	(3.45)	(6.17)
Favourite yellow prior	−0.103	0.030	−0.030	−0.000
	(5.49)	(1.32)	(1.82)	(0.01)
Underdog yellow prior	0.043	0.014	0.046	0.153
	(2.46)	(0.70)	(2.89)	(7.48)
Favourite red	0.116	0.153	0.151	0.227
	(1.36)	(1.24)	(1.52)	(2.74)
Underdog red	0.118	−0.152	0.096	−0.187
	(1.02)	(1.32)	(1.37)	(1.59)
Goal difference	0.034	0.046	0.018	0.012
	(2.37)	(2.50)	(1.55)	(0.59)
Goal difference squared	−0.037	−0.006	−0.022	−0.005
	(5.02)	(0.87)	(3.94)	(0.53)
Pre-game				
Home underdog match	1.242	−0.183	1.099	−0.065
	(24.45)	(2.63)	(23.74)	(0.89)
Derby	−0.012	−0.017	0.160	−0.103
	(0.15)	(0.14)	(3.96)	(1.31)
Difference in bookmaker probability	−0.218	−0.442	−0.234	−0.213
	(1.80)	(2.39)	(2.61)	(1.40)
Constant	−3.637	−3.571	−3.546	−3.190
	(21.29)	(13.60)	(24.74)	(15.76)
Pseudo-R^2	0.217	0.073	0.184	0.084
ρ		0.451		0.274
		(7.61)		(4.16)
Log-likelihood		−5201		−6730
Number of observations (minutes)		154980		198180

†The footnote for Table 2 applies here also.

extra input in the absence of sufficient talent. If this is the reason for differential yellow card totals for home and away teams, then a visiting team that has similar prospects of victory to those of a typical home team should not face a different expected number of yellow cards from that typical home team. So long as we control for win probabilities in the equations for team yellow cards, the inclusion of a dummy variable to indicate which team is at home should not add significant explanatory power to the model. If the coefficient on the home team dummy were significant, this would be evidence that there is differential incidence of sanctions incurred by home and away teams that could not be explained away by the correlation between home–away status and the teams' prospects of winning.

In Tables 4–6, the variable home underdog match distinguishes matches where the underdog is the team playing at home and the favourite is the team playing away. We estimate bivariate probit models of the probability of receiving a yellow card, straight red card and any red card (including second yellow). In Table 4, the coefficient estimate on home underdog match is negative and statistically significant in the underdog equation for yellow cards for each league. Hence, controlling for the size of the difference in win probabilities as well as for within-game factors, underdog teams are less likely to receive yellow cards when playing at home rather than away. Similarly, for the Premier League, the chance of receiving a yellow card is higher for the favourite playing away rather than at home. However, for the *Bundesliga* this effect is small and negative but not significantly different from 0. Therefore, this pattern of results clearly supports the finding in Dawson *et al.* (2007) that the tendency of away teams to receive relatively more yellow cards cannot be fully explained by the fact that, because of home advantage, they are more often than not the team that is more likely to win the match.

From the results that are shown in Table 5, we observe a large, positive and highly statistically significant effect of home underdog match on the probability that the favourite team receives a red card in each league, i.e. controlling for the degree of favourite status in the odds, match score and within-game dynamics of yellow and red cards, we find that favourites face a higher probability of receiving a straight red card when playing away rather than at home. This is consistent with referees generally exhibiting bias in their treatment of red cards in favour of home teams, and not just in their behaviour with respect to less severe cautionable offences. This finding is reinforced by the results that are shown in Table 6 for all red cards, including second yellow cards and not just straight red cards. In each league, teams that are favourites have a significantly higher probability of receiving a red card, of either type, when they are playing away. For red cards that are awarded to underdogs we find that the lower probability when they are playing at home is statistically significant only for the *Bundesliga* and that only for red cards including second yellow cards.

We have, then, found several indications that are consistent with home team bias on the part of referees in both Germany and England. On their own, these indications are not conclusive evidence of referee bias since it cannot be ruled out that players also react to the presence of a crowd, whether friendly or unfriendly. For this latter possibility to account for the pattern of results in Tables 4–6, it would be necessary for away teams to commit more offences, not only because they are typically underdogs and often trailing in the match, but also because an opposition crowd generates from them a response, in terms of more aggressive play and committing more offences, that is stronger than any increase in intensity of play that is triggered among home players. This is a possibility and cannot be ruled out as a possible explanation of the results in Tables 4–6. However, attributing our results to referee bias is more plausible to the extent that this interpretation is consistent with the results from Tables 2 and 3, which, we argued, could not plausibly be attributed to player behaviour, and with findings from other types of evidence that are reported in the literature. For example, added time is typically greater when it suits the interests of the home team. In this case, the finding can be interpreted only as a referee-induced effect because added time is at his sole discretion and not something that players determine. Further, Dohmen's (2008) analysis of 'right' and 'wrong' referee calls, as adjudged by experts studying film evidence after the event, also shows a tendency for decisions to be biased towards home teams. It would therefore not be surprising if similar bias were exercised in the specific case of the award of red and yellow cards.

Our view is therefore that our findings should be a source of concern to the respective governing bodies that are responsible for assignment, training and monitoring of referees. However, it cannot be assumed that neutrality would be socially preferable to any given degree of home

team bias. If referees indeed exhibit systematic bias towards home teams, then this may help to reinforce home advantage and to help smaller teams to win more home games. This may in turn raise competitive balance which is alleged by many sports economists to raise audience interest in the competition. Assuming that reluctance to issue cautions to home teams has a bearing on the outcome of matches, then small home teams are more likely to win games than they would otherwise.

4. Conclusions

We have presented a novel disaggregated analysis of potential bias and favouritism displayed by soccer referees, in which the unit of observation is the minute of play in a match. This disaggregated approach allows us to control for within-game factors of fluctuating scores and dynamics of the award of cards in a match. With these additional controls in place, inferences on referee bias are more robust and more compelling than those derived from analyses that use the match as the unit of observation.

We have two primary exhibits for referee bias in favour of home teams in the award of yellow and red cards.

(a) From the *Bundesliga* we find evidence that is consistent with referees being influenced by home support: home teams playing in stadia without running tracks, and therefore with the crowd close to the action, have lower probabilities of receiving yellow and red cards than home teams playing in grounds with running tracks. This is observed even though *a priori* we might suspect that allowing the crowd closer to the pitch would raise home players' level of aggression rather than lower it. It is therefore indicative of a successful influence of fans' 'social pressure' on referees, as proposed by Dohmen (2008). The fact that three teams switched stadium design with removal of a track in our sample period suggests that this is more than just a specific team effect.

(b) When matches in our two sample leagues are respecified as favourite *versus* underdog, rather than home *versus* away, we obtain evidence, via significant coefficients on the home underdog match dummy variable, that is consistent with biased treatment of teams in both the Premier League and the *Bundesliga*, in the award of both yellow and red cards.

Clearly, it would be desirable to assess whether our results are supported by analysis from other European football leagues, and other sports leagues where sanction design and refereeing technology may differ. Most importantly, further research is needed to assess the implications of referee bias for league design and corporate governance of sports leagues. Is the bias that we have detected harmful to stakeholders in the sport? Are audiences both at the stadium and in front of television sets deterred both by increased sanctions applied by referees and referee bias in these sanctions? And what exactly are the effects of measures to reduce referee bias on the uncertainty of outcome of a match and competitive balance in football leagues? These are deeper research questions that demand attention. For now, our findings, which are strongly suggestive of favouritism of referees towards home teams, are an essential precursor for these investigations.

References

Benz, M.-A., Brandes, L. and Franck, E. (2009) Do soccer associations really spend on a good thing?: empirical evidence on heterogeneity in the consumer response to match uncertainty of outcome. *Contemp. Econ. Poly*, **27**, 211–235.

Boyko, R., Boyko, A. and Boyko, M. (2007) Referee bias contributes to home advantage in English Premiership football. *J. Sports Sci.*, **25**, 1185–1194.

Caliendo, M. and Radic, D. (2006) Ten do it better, do they?: an empirical analysis of an old football myth. *Discussion Paper 2158*. Institute for the Study of Labor, Bonn.

Dawson, P., Dobson, S., Goddard, J. and Wilson, J. (2007) Are football referees really biased and inconsistent?: evidence on the incidence of disciplinary sanction in the English Premier League. *J. R. Statist. Soc.* A, **170**, 231–250.

Dohmen, T. (2008) The influence of social forces: evidence from the behavior of football referees. *Econ. Inq.*, **46**, 411–424.

Forrest, D., Goddard, J. and Simmons, R. (2005) Odds setters as forecasters. *Int. J. Forecast.*, **16**, 551–564.

Forrest, D. and Simmons, R. (2002) Outcome uncertainty and attendance demand in sport: the case of English soccer. *Statistician*, **51**, 229–241.

Frick, B., Gürtler, O. and Prinz, J. (2008) Men in black: monitoring and performance of German soccer referees. In *Soccer: Economics of a Passion* (eds H. Dietl, E. Franck and H. Kempf). Schorndorf: Hofmann.

Garicano, L. and Palacios-Huerta, I. (2006) Sabotage in tournaments: making the beautiful game less beautiful. *Mimeo*. University of Chicago, Chicago.

Garicano, L., Palacios-Huerta, I. and Prendergast, C. (2005) Favoritism under social pressure. *Rev. Econ. Statist.*, **87**, 208–216.

Greene, W. H. (2003) *Econometric Analysis*, 5th edn. Upper Saddle River: Prentice Hall.

Johnston, R. (2008) On referee bias, crowd size and home advantage in the English soccer Premiership. *J. Sports Sci.*, **26**, 563–568.

Lucey, B. and Power, D. (2004) Do soccer referees display home team favouritism? *Mimeo*. Trinity College Dublin, Dublin.

Nevill, A., Balmer, N. and Williams, A. (2002) The influence of crowd noise and experience upon refereeing decisions in football. *Psychol. Sport Exercs.*, **3**, 261–272.

Pettersson-Lidbom, P. and Priks, M. (2007) Behavior under social pressure: empty Italian stadiums and referee bias. *Working Paper 1960*. Centre for Economic Studies and Ifo Institute for Economic Research, Munich.

Price, J. and Wolfers, J. (2007) Racial discrimination among NBA referees. *Working Paper 13206*. National Bureau of Economic Research, Cambridge.

Rickman, N. and Witt, R. (2008) Favouritism and financial incentives: a natural experiment. *Economica*, **75**, 296–309.

Ridder, G., Cramer, J. and Hopstaken, P. (1994) Down to ten: estimating the effect of a red card in soccer. *J. Am. Statist. Ass.*, **89**, 1124–1127.

Scoppa, V. (2008) Are subjective evaluations biased by social factors or connections?: an econometric analysis of soccer referee decisions. *Empir. Econ.*, **35**, 123–140.

Stata (2007) *Stata 10 Base Reference Manual*, vol. 1. College Station: Stata Press.

Sutter, M. and Kocher, M. (2004) Favouritism of agents—the case of referees' home bias. *J. Econ. Psychol.*, **25**, 461–469.

Torgler, B. (2004) The economics of the FIFA football World Cup. *Kyklos*, **57**, 287–300.

Witt, R. (2005) Do players react to sanction changes?: evidence from the English Premier League. *Scot. J. Polit. Econ.*, **52**, 623–640.