ISSN 0111-1760

University of Otago
Economics Discussion Papers
No. 0511

August 2005

Are football referees really biased and inconsistent?

Evidence from the English Premier League

Peter Dawson[1], Stephen Dobson[2], John Goddard[3+], John Wilson[4]

1. Department of Economics and International Development, University of Bath, Bath BA2 7AY, UK
tel: +44 (0)1225 383074
email: p.m.dawson@bath.ac.uk

2. Department of Economics, University of Otago, PO Box 56, Dunedin, New Zealand
tel: +64 3 479 5296
email: s.dobson@business.otago.ac.nz

3. School of Business and Regional Development, University of Wales, Bangor, Gwynedd, LL57 2DG, UK
tel: +44 (0)1248 383221
email: j.goddard@bangor.ac.uk

4. Department of Management, University of St Andrews, The Gateway, North Haugh, St Andrews, Fife, KY16 9SS, UK
tel: +44 (0)1334 462803
email: jsw7@st-and.ac.uk

\+ Corresponding author.

The authors are grateful to participants at a Sports Economics Workshop held at the University of Groningen in March 2005 for comments on an earlier draft of this paper. We are grateful to Rob Simmons and Tunde Buraimo for providing access to their list of televised Premier League matches.

Abstract

This paper presents a statistical analysis of patterns in the incidence of disciplinary sanction (yellow and red cards) taken against players in the English Premier League over the period 1996-2003, using bivariate negative binomial and bivariate Poisson regressions. Several questions concerning sources of inconsistency and bias in refereeing standards are examined. Evidence is found to support a time consistency hypothesis, that the average incidence of disciplinary sanction is predominantly stable over time. However, a refereeing consistency hypothesis, that the incidence of disciplinary sanction does not vary between referees, is rejected. The tendency for away teams to incur more disciplinary points than home teams cannot be attributed to the home advantage effect on match results, and appears to be due to a refereeing bias favouring the home team.

Keywords

refereeing bias and inconsistency, English Premier League football, bivariate Poisson regression, bivariate negative binomial regression

1. Introduction

In professional team sports with a high public profile, including association football (soccer), disciplinary transgressions by players and sanctions taken by referees provide a rich source of subject matter for debate among pundits, journalists and the general public. Although newspaper and television pundits routinely and piously deplore incidents involving foul play or physical confrontation, there is no doubt that a violent incident, immediately followed by the referee's theatrical action of brandishing a yellow or red card in the direction of the miscreant, makes an important contribution to the popular appeal of the football match as spectacle or drama.

Due to the ever-increasing scope of television coverage of football especially at the highest level, together with improvements in video technology, the actions of players and referees have never been more keenly and intensely scrutinised than they are in the modern-day game. In sporting terms, the margins separating success from failure can be slender, and often depend ultimately on split-second decisions taken by referees and players in the heat of battle. Yet the financial implications of success or failure for individual football clubs and their players can be vast. In England, for example, the financial reward (in terms of increased TV revenue, gate revenue, advertising, sponsorship and merchandising) associated with winning the annual end-of-season promotion play-off fixture (which elevates the winning team to the Premier League) has been estimated as £35 million (Deloitte and Touche, 2004). Therefore the football authorities are under intense pressure from all sides to take steps to ensure that refereeing decisions are as fair, consistent and accurate as is humanly possible.

Bearing all of these considerations in mind, it is perhaps surprising that academic research on the incidence of disciplinary sanction in professional sports is relatively sparse. This paper seeks to fill this gap, by presenting a statistical analysis of patterns in the incidence of disciplinary sanction taken against players in English professional football's highest division, the Premier League, over a seven-year period from 1996 to 2003.

The empirical analysis is used to address several questions concerning possible sources of inconsistency and home team bias in refereeing standards. Among the hypotheses investigated are: a home advantage hypothesis, that the tendency for away teams to incur more disciplinary points than home teams is solely a corollary of home advantage, or the tendency for home teams to win more often than away teams; a refereeing consistency hypothesis, that the propensity to take disciplinary action does not vary between referees; and a time consistency hypothesis, that the overall incidence of disciplinary sanction is stable over time and unaffected by changes to the content or interpretation of the rules. The following questions are also examined in the course of the empirical analysis: Does the average rate of disciplinary sanction against each team depend upon which team is the favourite and which is the underdog? Does it depend upon whether the match is competitive (between two evenly balanced teams) or uncompetitive? Does it depend upon whether end-of-season outcomes are at stake for either team? Finally, is it affected by the stadium audience, and does it depend upon whether the match is broadcast live on TV? By providing answers to these questions based on statistical analysis, this study aims to provide the football authorities and other interested parties with a firmer factual basis than has previously been available for policy decisions and debate concerning the interpretation and implementation of the rules governing disciplinary sanction in football.

The structure of the paper is as follows. Section 2 reviews the previous academic literature on the topic of disciplinary sanction in professional team sports. Section 3 discusses statistical distributions that may be regarded as candidates for modelling the incidence of disciplinary sanction in football: specifically, the univariate and bivariate Poisson and negative binomial distributions. Section 4 compares the performance of these distributions when they are used to model the unconditional expectations of the incidence of disciplinary sanction against the home and away teams. Section 5 develops a theoretical analysis of the relationship between team quality and the incidence of disciplinary sanction. Section 6 reports estimations for the expectations of the incidence of disciplinary sanction conditional on a number of covariates, and reports a series of hypothesis tests concerning the sources of refereeing bias and inconsistency. Finally, Section 7 offers some conclusions and policy implications for the football authorities.

2. Literature review

Previous academic scrutiny of the topic of disciplinary sanction in professional team sports has focused mainly on the impact of dismissals on match results; and on issues of incentives, monitoring and detection, which arise in the economics literature on crime and punishment.

Ridder et al. (1994) model the effect of a dismissal on football match results, using Dutch professional football data from the period 1989-92. Probabilities are estimated for the match result conditional on the stage of the match at which a dismissal occurs. A method is developed for estimating the earliest stage of the match at which it is optimal for a defender to resort to foul play punishable by dismissal in order to deny an opposing forward a goal scoring opportunity (conditional on the probability that the opportunity would be converted), assuming the defender's objective is to minimise the probability of his team losing the match. In a multivariate analysis of the determinants of match results from the 2002 FIFA World Cup, Torgler (2004) also finds a significant association between player dismissals and match results.

In the literature on the economics of crime and punishment, rule changes in professional sports have occasionally created opportunities for empirical scrutiny of the question whether increasing the resources assigned to monitoring or policing leads to an increase or a decrease in the incidence of crimes being detected. This incidence increases if the monitoring effect (more monitoring increases detection rates) exceeds the deterrent effect (the tendency for criminals to be deterred from offending because monitoring has increased).

In North American college basketball's Atlantic Coast Conference, an increase in the number of referees from two to three per match was implemented in 1979. McCormick and Tollison (1984) find that the number of fouls called per game fell sharply. If refereeing competence improved with the increase in the number of officials (with fewer fouls being missed), the actual crime rate must have decreased by even more than is suggested by the fall in the number of fouls called.

In the North American National Hockey League (NHL), an increase from one to two referees per match was phased in during the 1998-9 and 1999-2000 seasons. Heckelman and Yates (2002) note that fouls detected are observed but fouls committed are unobserved. The difference between the two enters the error term of a regression for fouls detected when the latter is used as a proxy for fouls committed. Because this difference is likely to be correlated with the number of referees, instrumental variables are used to model the latter in the regression for fouls detected. Although more fouls were detected in NHL matches with two referees than in the matches with one, this appears to have been due solely to a monitoring effect. The incidence of fouls being committed was the same under both refereeing regimes. Distinguishing between violent and non-violent offences, Allen (2002) finds detection of the latter was significantly higher with two referees than with one. Again, this suggests the monitoring effect outweighs the deterrent effect.

As part of a wide-ranging investigation of the impact of changes in reward structures on effort using Spanish football data, Garciano and Palacios-Huerta (2000) draw comparisons between the numbers of yellow and red cards incurred before and after the introduction (in the 1995-6 season) of the award of three league points for a win and one for a draw. Previously two points had been awarded for a win and one for a draw. More yellow cards were awarded after the reward differential between winning and drawing was increased. This finding is consistent with theoretical models of tournaments, in which players can engage in sabotage activity. Following a rule change implemented at the start of the 1998-9 season requiring an automatic red card punishment for the tackle from behind, Witt (2005) finds evidence of an increase in the incidence of yellow cards (awarded for lesser offences), but no increase in the incidence of red cards. This finding suggests a deterrent effect was operative: football teams modified their behaviour in response to the rule change.

There are some limitations to the statistical models and methods that have been employed in this literature. While McCormick and Tollison (1984) and Garciano and Palacios-Huerta (2000) estimate separate equations for the winning and losing teams, the other papers cited above all report equations for the total number of offences called against both teams combined, factoring out many of the team-specific determinants of the incidence of disciplinary sanction. Despite the discrete structure of a 'fouls' or 'cards' dependent variable, McCormick and Tollison (1984) and Heckelman and Yates (2002) report ordinary least squares (OLS) regressions. Garciano and Palacios-Huerta (2000) discard univariate Poisson regressions in favour of OLS, because the former could not be estimated using fixed effects for teams. Witt (2005) reports both OLS and univariate Poisson regressions, while Allen (2002) uses a univariate negative binomial regression.

Alleged refereeing bias in favour of the home team is a frequently-aired grievance on the part of managers, players and spectators, which has also received some attention in the academic literature. Garicano et al. (2001) and Sutter and Kocher (2004) find a tendency for referees to add more time at the end of matches when the home team is trailing by one goal than when the home team is leading. Nevill et al. (2002) played videotapes of tackles to referees who, having been told the identities of the home and away teams, were asked to classify the tackles as legal or illegal. One group of referees viewed the tape with the soundtrack (including crowd's reaction) switched on, while a second group viewed silently. The first group were more likely to rule in favour of the home team, and the first group's rulings were more in line with those of the original match referee. Using German Bundesliga data, Sutter and Kocher (2004) analyse reports on the referee's performance, which comment on the legitimacy of penalties awarded and on cases of failure to award a legitimate penalty. There is evidence of home team bias in such decisions.

3. Modelling the incidence of disciplinary sanction in English Premier League football

Tables 1 and 2 show the frequency distributions for the numbers of yellow cards and red cards incurred by the home and away teams in the N=2,660 Premier League matches played during the seven English football seasons from 1996-7 to 2002-3 inclusive. A yellow card, also known as a booking or caution, is awarded for less serious disciplinary offences. There is no further punishment within the match, unless a player previously cautioned commits a second cautionable offence, in which case a red card is awarded and the player is expelled for the remainder of the match (with no replacement permitted, implying that the team completes the match one player short). A red card, also known as a sending-off or dismissal, is awarded directly for more serious offences, and results in immediate expulsion (again, with no replacement permitted). The Football Association (FA) uses a formula to determine when a cumulative number of disciplinary offences committed by a player within a season triggers a suspension, preventing the player from appearing in a number of his team's next scheduled matches: this number (usually one, two or three) depends partly upon the nature of the offences committed.

The dependent variables in the estimations reported below are the total numbers of disciplinary ‘points’ incurred by the home (i=1) and away (i=2) teams in match j for j=1...N, denoted $Z_{i,j}$ and calculated by awarding one point for a yellow card and two for a red card. Only two points (not three) are awarded when a player is dismissed having committed two cautionable offences in the same match. This metric is slightly simpler than the one used by the FA to determine suspensions, for which complete data are unpublished; but it accurately reflects the popular notion that a red card is in some sense equivalent to two yellow cards. In fact, this notion was literally true of just under one-half of the red cards awarded during the observation period (227 out of 462 dismissals in total), which resulted from two cautionable offences having been committed by one player in the same match. Attempts to estimate versions of the models that are reported below using yellow cards and red cards dependent variables separately were successful for the former but unsuccessful for the latter, presumably because the incidence of red cards is too sparse for reliable estimation. The results for the estimations using a dependent variable based solely on the number of yellow cards were very similar to those that are reported below, and are available from the authors on request.

In the applied statistics literature, several methods have been used to model professional team sports bivariate count data, where each match yields two values of a discrete dependent variable (one for each team: commonly the number of goals or points scored, but the disciplinary points dependent variable in the present case has the same structure). Maher (1982), Dixon and Coles (1997), Karlis and Ntzoufras (2003), Dixon and Pope (2004) and Goddard (2005) use the bivariate Poisson distribution to model English football goal scoring data, while Cain et al. (2000) use the univariate negative binomial distribution. Lee (1999) models Australian rugby league scores data using a bivariate negative binomial distribution. For the first time in this literature, this paper draws direct comparisons between the performance in modelling professional team sports bivariate count data of the univariate and bivariate Poisson and negative binomial distributions.

Specifically, to model the disciplinary points dependent variable $Z_{i,j}$ as defined above, four specifications are investigated. In the first case, $Z_{i,j}$ for i=1,2 are modelled using two univariate

Poisson probability functions. The joint probability is the product of the two univariate Poisson probabilities:

$$P(Z_{1,j}=z_1, Z_{2,j}=z_2) = [\exp(-\lambda_{1,j})\,\lambda_{1,j}^{z_1}/z_1!][\exp(-\lambda_{2,j})\,\lambda_{2,j}^{z_2}/z_2!] \quad \text{for } z_1,z_2=0,1,2,\ldots;\ j=1\ldots N \qquad (1)$$

In (1), $\lambda_{i,j}$ represent the expectations of $Z_{i,j}$. In the case of the Poisson distribution, $\lambda_{i,j}$ is also the variance of $Z_{i,j}$.

In the second case, $Z_{i,j}$ are modelled using a bivariate Poisson probability function, allowing for non-zero covariance or correlation between $Z_{1,j}$ and $Z_{2,j}$. The joint probability function is:

$$P(Z_{1,j}=z_1, Z_{2,j}=z_2) = \exp(-\lambda_{1,j}-\lambda_{2,j}+\xi_j)\sum_{k=0}^{\min(z_1,z_2)}(\lambda_{1,j}-\xi_j)^{z_1-k}(\lambda_{2,j}-\xi_j)^{z_2-k}\xi_j^k/\{(z_1-k)!(z_2-k)!k!\}$$

$$\text{for } z_1,z_2=0,1,2,\ldots;\ j=1\ldots N \qquad (2)$$

Equation (2) can be interpreted as the product of three univariate Poisson probability functions with means $\lambda_{1,j}-\xi_j$, $\lambda_{2,j}-\xi_j$ and ξ_j, respectively. In (2), ξ_j represents the covariance between $Z_{1,j}$ and $Z_{2,j}$. $\xi_j = \eta\sqrt{\lambda_{1,j}\lambda_{2,j}}$, where η is the correlation coefficient between $Z_{1,j}$ and $Z_{2,j}$. From casual observation, we anticipate $\xi_j>0$ and $\eta>0$, or a positive association between $Z_{1,j}$ and $Z_{2,j}$. If either team is guilty of an unexpectedly high level of foul play in any match, it appears common for the opposing team to retaliate in kind, resulting in a tendency for both teams to incur relatively high numbers of yellow and red cards. Furthermore, a common opinion among pundits and supporters is that some referees, having issued a card to a player from one team, tend to look for an opportunity to penalise a player from the opposing team soon afterwards, in an effort to pre-empt the formation on the part of managers, players or spectators of any perceptions of refereeing bias.

In the third case, $Z_{i,j}$ are modelled using univariate negative binomial probability functions:

$$P(Z_{1,j}=z_1, Z_{2,j}=z_2) = f_1(z_1)f_2(z_2)$$

where $f_i(z_i) = \dfrac{\Gamma(\rho_i + z_i)}{z_i!\Gamma(\rho_i)}\left(\dfrac{\rho_i}{\lambda_{i,j}+\rho_i}\right)^{\rho_i}\left(\dfrac{\lambda_{i,j}}{\lambda_{i,j}+\rho_i}\right)^{z_i}$ for z_i=0,1,2,...; i=1,2; j=1...N (3)

In (3), $f_i(z_i)$ are the univariate marginal probability functions for $Z_{i,j}$, and Γ is the gamma function. The expectations of $Z_{i,j}$ are $\lambda_{i,j}$, as before. $\rho_i>0$ are ancillary parameters, which determine the dispersion of the distributions of $Z_{i,j}$. The variance of $Z_{i,j}$ is $\lambda_{i,j}(1+\kappa_i\lambda_{i,j})$, where $\kappa_i=1/\rho_i$. The negative binomial distribution allows for greater heterogeneity or dispersion (overdispersion) in $Z_{i,j}$ than is permitted by the Poisson distribution.

In the fourth case, $Z_{i,j}$ are modelled using a bivariate negative binomial probability function. It is assumed that the marginal distributions of $Z_{1,j}$ and $Z_{2,j}$ are univariate negative binomial. Let $F_i(z_i)$ denote the univariate marginal distribution functions for $Z_{i,j}$ corresponding to $f_i(z_i)$. As demonstrated by Lee (1999), a bivariate joint distribution function can be defined as follows:

$$
\begin{aligned}
G[F_1(z_1),F_2(z_2)] &= \frac{1}{\varphi}\ln\left(1+\frac{\{\exp[\varphi F_1(z_1)]-1\}\{\exp[\varphi F_2(z_2)]-1\}}{\exp(\varphi)-1}\right) && \text{for } \varphi \neq 0 \\
&= F_1(z_1)F_2(z_2) && \text{for } \varphi = 0 \qquad (4)
\end{aligned}
$$

In (4), φ is an ancillary parameter, which determines the direction of any relationship between $Z_{1,j}$ and $Z_{2,j}$. For $\varphi<0$ the correlation between $Z_{1,j}$ and $Z_{2,j}$ is positive, and for $\varphi>0$ the correlation is negative. The parameter φ plays a similar role in (4) to the parameters ξ_j and η in (2). The bivariate joint probability function is obtained iteratively, as follows:

$$
\begin{aligned}
&P(Z_{1,j}=0, Z_{2,j}=0) = G[F_1(0),F_2(0)] \\
&P(Z_{1,j}=z_1, Z_{2,j}=0) = G[F_1(z_1),F_2(0)] - G[F_1(z_1-1),F_2(0)] && \text{for } z_1=1,2,...;\ j=1...N \\
&P(Z_{1,j}=0, Z_{2,j}=z_2) = G[F_1(0),F_2(z_2)] - G[F_1(0),F_2(z_2-1)] && \text{for } z_2=1,2,...;\ j=1...N \\
&P(Z_{1,j}=z_1, Z_{2,j}=z_2) = G[F_1(z_1),F_2(z_2)] - G[F_1(z_1-1),F_2(z_2)] \\
&\qquad - G[F_1(z_1),F_2(z_2-1)] + G[F_1(z_1-1),F_2(z_2-1)] && \text{for } z_1,z_2=1,2,...;\ j=1...N
\end{aligned}
$$

There are several nesting relationships between these four specifications. First, the bivariate Poisson distribution reduces to two univariate Poisson distributions if the restriction $\eta=0$ is imposed on (2). Similarly, the bivariate negative binomial distribution reduces to two univariate negative binomial distributions if the restriction $\varphi=0$ is imposed on (4). Second, in (3) and (4), if $\rho_i\rightarrow\infty$ and $\kappa_i\rightarrow 0$, $\mathrm{var}(Z_{i,j})\rightarrow\lambda_{i,j}$. In this case, the negative binomial distribution converges to the Poisson distribution. Therefore (3) reduces to (1) and (4) reduces to (2) under the restriction $\kappa_1=\kappa_2=0$. Likelihood ratio (LR) tests, based on maximum likelihood (ML) estimation of (1) to (4), can be used to test the validity of these restrictions, and to select the most appropriate of the four specifications.

4. The unconditional model for the incidence of disciplinary sanction

In this section, equations (1) to (4) are fitted to the disciplinary points data using ML, under the assumption that the expected value of $Z_{i,j}$ (for i=1,2) is the same for all matches. Therefore (1) to (4) are simplified by assuming $\lambda_{i,j}=\lambda_i$ for i=1,2; j=1...N (so $\xi_j=\xi$ in (2)). λ_1 and λ_2 are interpreted as the unconditional expectations of the disciplinary points incurred by the home and away teams. The resulting estimations are labelled Models U1 to U4. In all four cases, the estimated values of λ_1 and λ_2 are the sample means of $Z_{i,j}$, $\hat{\lambda}_1=1.4650$ and $\hat{\lambda}_2=2.0451$. The first four columns of Table 3 report the ML estimates of the other (ancillary) parameters. The estimated values $\hat{\eta}=0.2655$ in Model U2 and $\hat{\varphi}=-1.8915$ in Model U4 indicate a positive correlation between $Z_{1,j}$ and $Z_{2,j}$, as expected.

Table 3 also reports the maximised values of the log-likelihood function, denoted ln(L), and the pseudo-likelihood statistic, denoted Ψ, equivalent to the geometric mean of the joint probabilities assigned by the fitted model to the observed values of $\{Z_{1,j}, Z_{2,j}\}$ across j=1...N. An LR test for $H_0:\eta=0$ in Model U2 yields $\chi^2(1)=224.6$. An LR test for $H_0:\varphi=0$ in Model U4 yields $\chi^2(1)=226.7$. LR tests for $H_0:\kappa_1=\kappa_2=0$ in Models U3 and U4 yield $\chi^2(2)=50.2$ and $\chi^2(2)=52.2$, respectively. The p-values for all of these tests are zero. The results suggest the bivariate specification is preferred to the univariate specification, and the negative binomial distribution is preferred to the Poisson distribution. Overall Model U4 (bivariate negative binomial) is the preferred specification.

In order to assess whether Model U4 (or any of the other unconditional models) provides a satisfactory representation of the data, Table 4 compares the joint distribution of observed values of $\{Z_{1,j}, Z_{2,j}\}$ in the sample of N=2,660 matches with the distributions of expected values implied by the estimated versions of Models U1 to U4. A chi-square goodness-of-fit test is used to evaluate the null hypothesis that the observed data are drawn from the assumed probability distribution in each case. The test results are reported at the foot of Table 3. The null favouring the relevant distribution is rejected at the 0.01 significance level in all cases. However, Model U4 appears to outperform each of the other three models, with the rejection of the null borderline at the 0.01 level in this case.

Comparisons between the observed and expected values reported in the cells of Table 4 identify the sources of the superior performance of Model U4. First, the Poisson Models U1/2 understate the marginal probabilities for small and large values of $Z_{1,j}$ and $Z_{2,j}$, and overstate those for intermediate values. The 'flatter' negative binomial Models U3/4 are more successful in reproducing the shapes of the marginal distributions. Second, the univariate Models U1/3 understate the probabilities of both teams simultaneously obtaining either small or large values of $Z_{i,j}$ ($\{Z_{1,j}=0, Z_{2,j}=0\}$, $\{Z_{1,j}\geq 5, Z_{2,j}\geq 5\}$); and overstate the probabilities for large differences ($\{Z_{1,j}=0, Z_{2,j} \geq 5\}$, $\{Z_{1,j}\geq 5, Z_{2,j}=0\}$). By controlling for the correlation between $Z_{1,j}$ and $Z_{2,j}$, the bivariate Models U2/4 are more effective in reproducing the shape of the joint empirical distribution of $Z_{1,j}$ and $Z_{2,j}$ in such cases.

5. Team quality and the incidence of disciplinary sanction

In Sections 5 and 6, we develop an empirical model for the determinants of $\lambda_{i,j}$ interpreted as the conditional expectations of the disciplinary points incurred by the home (i=1) and away (i=2) teams in match j. In Section 5, we begin by investigating the theoretical relationship between team quality and the incidence of disciplinary sanction. In the theoretical analysis, it is assumed that the win probabilities for the two teams depend on: the talent differential between the teams; a home advantage effect; and a tactical decision variable representing the level of 'aggression' contributed by each team. For simplicity, in the theoretical analysis we ignore draws and consider win probabilities for the two teams that sum to one. Let $t_{i,j}$ and $a_{i,j}$ denote the playing talent and aggression level of team i in match j, respectively. It is assumed that $t_{i,j}$ and $a_{i,j}$ are scaled such that team 1's win probability against team 2 can be represented as $\Phi[x_j+\theta(a_{1,j})–\theta(a_{2,j})]$, where Φ denotes the standard Normal distribution function, $x_j=t_{1,j}–t_{2,j}+h$, h is a scalar that allows for home advantage, and θ is a function (see below). It is also assumed that aggressive play by either team imposes a cost, represented by the function $\nu(a_{i,j})$, reflecting the deleterious effect on future match results of player suspensions resulting from disciplinary points incurred in the current match.

We assume the functions θ and ν are continuous and twice-differentiable, with $\theta'(0)>0$, $\theta''(a_{i,j})<0$, $\nu'(a_{i,j})>0$ and $\nu''(a_{i,j})>0$ for all $a_{i,j}$. At low levels, more aggression enhances a team's win probability; but this relationship is subject to diminishing returns. Beyond a certain point further aggression becomes counterproductive: as noted above, Ridder et al. (1994) find that a team's win probability is reduced in matches in which it incurs one or more red cards. The cost function ν is increasing in aggression, at an increasing rate as the level of aggression increases.

We assume both teams decide their own aggression levels independently and before the start of the match, and we derive a Nash equilibrium for the aggression levels contributed by both teams. A distinction is drawn between this tactical decision concerning aggression, and any tendency to engage in foul play on a retaliatory or tit-for-tat basis once the match is underway. As discussed in Section 3, in the bivariate models this retaliatory effect is represented by the parameters ξ_j or η in (2) or φ in (4).

At the Nash equilibrium, each team selects its own aggression level, conditional on the other team's aggression level being taken as fixed at its current value. For example, consider team 1's choice of $a_{1,j}$, conditional on $a_{2,j}$ (and x_j). Team 1 selects $a_{1,j}$ to maximise the objective function $\pi_1(a_{1,j}; a_{2,j}, x_j)$, representing the net benefit to team 1 of an aggression level of $a_{1,j}$, rather than zero aggression:

$$\pi(a_{1,j}; a_{2,j}, x_j) = \Phi[x_j+\theta(a_{1,j})–\theta(a_{2,j})] – \Phi[x_j–\theta(a_{2,j})] – \nu(a_{1,j})$$

The maximisation of this objective function yields team 1's reaction function, $a_{1,j} = r_1(a_{2,j}, x_j)$. A similar optimisation procedure yields team 2's reaction function, $a_{2,j} = r_2(a_{1,j}, x_j)$. The Nash equilibrium $\{a^*_{1,j}, a^*_{2,j}\}$ is located at the intersection of the two reaction functions, at which point $a^*_{1,j} = r_1(x_j, a^*_{2,j})$ and $a^*_{2,j} = r_2(x_j, a^*_{1,j})$.

Figure 1 illustrates the Nash equilibrium for the three cases x_j=0, 0.5, 1. The quadratic functional forms $\theta(a_{i,j})=a_{i,j}-0.2a_{i,j}^2$ and $\nu(a_{i,j})=0.2a_{i,j}+0.1a_{i,j}^2$ are used for illustrative purposes. The model is symmetric, so $a^*_{1,j} = a^*_{2,j}$ for any x_j, and the Nash equilibrium values for x_j =–0.5, –1 are the same as those for x_j=0.5, 1, respectively. The maximum numerical values for $\{a^*_{1,j}, a^*_{2,j}\}$ occur in the case x_j=0 where, taking account of talent and home advantage, the teams are equally balanced and have identical win probabilities. Figure 1 illustrates the following general property of the theoretical model: as the degree of competitive imbalance increases, the Nash equilibrium values $\{a^*_{1,j}, a^*_{2,j}\}$ decrease. If the match is evenly balanced, a little extra aggression by either team has a large effect on the win probabilities, and the aggression levels of both teams are high at the Nash equilibrium. Conversely, if the match is unbalanced, a little extra aggression by either team has little effect on the win probabilities, and the aggression levels are low at the Nash equilibrium.

In the empirical model, a proxy for x_j is obtained using the ordered probit match results forecasting model developed by Goddard (2005). This model generates match result forecasts in the form of probabilities for home win, draw and away win outcomes, based solely on historical data that is available prior to the match in question. The forecasting model's covariates are: the win ratios of both teams over the 24 months prior to the current match; both teams' recent home and away match results; dummy variables indicating the significance of the match for end-of-season outcomes (championship, European qualification and relegation); dummy variables indicating current involvement in the FA Cup; both teams' recent average home attendances; and the geographic distance separating the teams' home towns. To generate match result probabilities for each season (1996-7 to 2002-3 inclusive), seven versions of the forecasting model are estimated, using data for the preceding 15 seasons in each case. Full details of the forecasting model are reported in Goddard

(2005) and are not repeated here; see also Dobson and Goddard (2001), Goddard and Asimakopoulos (2004) and Forrest et al. (2005). Using the generated match result probabilities for each of the N=2,660 sample matches, we define the covariate q_j = P(home win in match j) + 0.5P(draw). We note that q_j is a monotonic increasing function of x_j: when x_j is large and negative, $q_j \cong 0$; when $x_j=0$, $q_j=0.5$; and when x_j is large and positive, $q_j \cong 1$.

The empirical model allows for two forms of relationship between q_j and the incidence of disciplinary sanction. First, a weaker team that is forced to defend for long periods can be expected to commit more fouls than a stronger team that spends more time attacking. This suggests a negative (positive) linear relationship between q_j and the disciplinary points incurred by the home (away) team. Second, the theoretical analysis developed in Section 5 suggests there is also a non-linear dimension to the relationship between q_j and the incidence of disciplinary sanction. In the empirical model, this is represented by the quadratic covariate $q_j(1-q_j)$. A positive relationship is expected between this covariate and the incidence of disciplinary sanction against both teams.

6. The conditional model, and tests for refereeing bias and inconsistency

In Section 6, we report estimations for the conditional expectations of the numbers of disciplinary points incurred by the home and away teams. Semi-logarithmic functional forms are used in order to ensure the non-negativity of the dependent variable $\ln(\lambda_{i,j})$, which in the conditional model is assumed to depend on covariates that vary from match to match. The team quality covariates q_j and $q_j(1-q_j)$ have been described in Section 5. The remaining covariate definitions are as follows:

$sig_{i,j}$ = 0-1 dummy variable, coded 1 if match j is significant for end-of-season championship, European qualification or relegation outcomes, for the home (i=1) or away (i=2) team.

$DM_{i,m,j}$ = 1 if match j falls within managerial spell m for the home (i=1) or away (i=2) team; 0 otherwise (m=1...56 represents managerial spells that contained at least 30 Premier League matches within the observation period; the matches in 24 other spells that contained fewer than 30 matches in total form the reference category).

$DR_{r,j}$ = 1 if match j is officiated by referee r; 0 otherwise (r=1...28 represents referees who officiated at least 30 Premier League matches within the observation period; nine other referees who officiated fewer than 30 matches each form the reference category).

$DS_{s,j}$ = 1 if match j is played in season s; 0 otherwise (s represents seasons 1997-8 to 2002-3 inclusive; 1996-7 is the reference category).

att_j = reported attendance at match j.

sky_j = 1 if match j was televised live by BSkyB; 0 otherwise.

The model specification allows for tests of several hypotheses concerning patterns in the incidence of disciplinary sanction. The principal hypotheses of interest are as follows:

H1: The *home advantage hypothesis*. The tendency for away teams to incur more disciplinary points than home teams is solely a corollary of home advantage: the tendency for home teams to win more frequently than away teams.

H2: The *refereeing consistency hypothesis*. The average incidence of disciplinary sanction does not vary between referees.

H3: The *consistent home team bias hypothesis*. The degree to which away teams incur more disciplinary points than home teams on average (after controlling for home advantage) does not vary between referees.

H4: The *time consistency hypothesis*. The average incidence of disciplinary sanction is stable over time.

H5: The *audience neutrality hypothesis*. The incidence of disciplinary sanction is invariant to the size of the crowd inside the stadium, and is the same notwithstanding whether the match is broadcast live on TV.

The models for $\ln(\lambda_{i,j})$ are labelled Model C2 (bivariate Poisson specification) and Model C4 (bivariate negative binomial). Summary estimation results for these models are reported in the final two columns of Table 3. Because the bivariate distributions are preferred to the univariate distributions, the equivalent univariate specifications (Models C1 and C3) are not reported. Before assessing the contributions of the covariates and other effects to the fitted models, we reconsider the issue of the choice between the bivariate Poisson and bivariate negative binomial specifications, in respect of the conditional model. The LR test for H_0:κ_1=κ_2=0 in Model C4 is shown in the final column of Table 3. In the unconditional Model U4, H_0:κ_1=κ_2=0 was rejected; but in the conditional Model C4 the LR test fails to reject H_0 at any reasonable significance level. Models C2 and C4 both achieve Ψ=.0419. Therefore in this case, the more parsimonious bivariate Poisson specification is preferred to the bivariate negative binomial specification.

In general, the use of the negative binomial distribution is necessary in cases where population heterogeneity produces an observed distribution with more dispersion than is permitted by the Poisson distribution. However, if sufficient covariates and other effects are included in the conditional expectation equations to identify the sources of heterogeneity explicitly, the ability of the negative binomial distribution to allow for overdispersion becomes superfluous. A general implication for modelling bivariate count data is that if an unconditional specification is used, the choice between the Poisson and negative binomial distributions is important; but if covariates can be found that represent adequately the sources of heterogeneity in the equation for the conditional expectation, this choice becomes either less important or unimportant.

The estimated home team and away team equations in Model C2 are reported below as equations (5) and (6), with z-statistics for the significance of the estimated coefficients shown in parentheses (intercept and dummy variable coefficients are not reported). The estimation results are interpreted and discussed throughout the rest of Section 5.

$$\begin{aligned}\ln(\hat{\lambda}_{1,j}) = \hat{\alpha}_{1,0} &- \underset{(-2.67)}{0.5469}\, q_j + \underset{(5.35)}{5.0749}\, q_j(1-q_j) + \underset{(0.03)}{0.0012}\, sig_{1,j} + \underset{(3.27)}{14.1675}\, att_j + \underset{(0.83)}{0.0369}\, sky_j \\ &+ \sum_{s=1}^{6}\hat{\beta}_{1,s} DS_{s,j} + \sum_{m=1}^{56}\hat{\delta}_{1,m} DM_{i,m,j} + \sum_{r=1}^{28}\hat{\gamma}_{1,r} DR_{r,j}\end{aligned} \tag{5}$$

$$\begin{aligned}\ln(\hat{\lambda}_{2,j}) = \hat{\alpha}_{2,0} &+ \underset{(-3.56)}{0.8600}\, q_j + \underset{(4.02)}{3.1873}\, q_j(1-q_j) + \underset{(3.02)}{0.1269}\, sig_{2,j} + \underset{(1.32)}{2.2336}\, att_j + \underset{(0.29)}{0.0110}\, sky_j \\ &+ \sum_{s=1}^{6}\hat{\beta}_{2,s} DS_{s,j} + \sum_{m=1}^{56}\hat{\delta}_{2,m} DM_{i,m,j} + \sum_{r=1}^{28}\hat{\gamma}_{2,r} DR_{r,j}\end{aligned} \tag{6}$$

Relative team quality and home advantage

In Model C2, the coefficients on q_j and $q_j(1-q_j)$ are all correctly signed and significantly different from zero at the 0.01 level. Therefore the estimations support the hypotheses advanced in Section 5, that the incidence of disciplinary sanction tends to be greater for the team with the lower win probability, and greater in matches between evenly balanced teams.

In general, home teams tend to attack more than away teams, and away teams defend more than home teams. In the models for the unconditional expectations of $\lambda_{1,j}$ and $\lambda_{2,j}$ presented in Section 4, the difference between the mean disciplinary points per match incurred by the home and away teams is highly significant: a Wald test of H_0:λ_1=λ_2 yields $\chi^2(1)$=289.9 (p-value= .0000) in Model U4. We are now able to examine the empirical validity of H1, the *home advantage hypothesis*. H1 asserts that the propensity for away teams to collect more disciplinary points on average than home teams is solely a corollary of the home advantage effect on match results. If so, the expected incidence of disciplinary sanction for a (relatively strong) away team should be the same as that for a (relatively weak) home team if the two teams' win probabilities (allowing for home advantage) are the same. Under H1, all coefficients in (5) should be identical to their counterparts in (6) (except the coefficients on q_j, which should be equal and opposite in sign). In this case, the difference between the average disciplinary points incurred by the home and away teams would be attributable solely to the home advantage effect (and its implications for competitive balance), measured by the extent to which the sample mean value of the covariate q_j (=0.6025) exceeds 0.5.

A Wald test of a null hypothesis specifying the appropriate cross-equation restrictions on the coefficients of Model C2 yields $\chi^2(96)$=170.07 (p-value=.0000). Therefore H1 is rejected decisively, and we infer that there are significant differences between the coefficients of the two equations. Rejection of H1 implies the tendency for away teams to endure a higher incidence of disciplinary sanction cannot be explained solely by the home advantage effect; although this effect does contribute towards the observed pattern.

Other controls for team behaviour

In order to isolate the contribution of referees to the variation in the incidence of disciplinary sanction, the conditional model includes a number of covariates (in addition to the relative team quality measures q_j and $q_j(1–q_j)$) that control for the effects of team behaviour. The contribution to the model of these controls is examined in this subsection.

The incidence of disciplinary sanction for either team might be affected by the importance of the match for end-of-season championship, European qualification or relegation outcomes. A team that still has end-of-season issues at stake might be expected to be more determined than a team with nothing at stake, and might therefore be more liable to commit disciplinary transgressions. In the definitions of the dummy variables $sig_{i,j}$, the algorithm that determines whether a match is significant for either team assesses whether it is arithmetically possible (before the match is played) for the team to win the championship, qualify for European competition or be relegated, if all other teams currently in contention for the same outcome take one point on average from each of their remaining fixtures.

In Model C2, the coefficient on $sig_{1,j}$ in (5) is insignificant, but the coefficient on $sig_{2,j}$ in (6) is positively signed and significant at the 0.01 level. A possible interpretation is that away teams feel able to 'ease off' in unimportant end-of-season matches; but home teams, perhaps conscious of their own crowd's critical scrutiny, feel obliged to demonstrate maximum commitment at all times, even when no end-of-season issues are at stake.

Differences between football teams in playing personnel, styles of play and tactics represent a further possible source of variation in the incidence of disciplinary sanction. With 22 players (plus substitutes) participating in every match, in an empirical analysis at match level it is impossible to control for every change of playing personnel. In preliminary experiments with the model specification, we encountered a tendency for estimations including separate dummy variables for each team in each season to fail to converge, due to the excessive number of coefficients. Therefore we have chosen to use managerial spells as a proxy for football team-related factors that might produce differences in the incidence of disciplinary sanction. This can be justified on the grounds that managers are primarily responsible for tactics and playing styles. Casual observation suggests managerial change is a good proxy for turnover of playing personnel: the removal of a manager is often followed by high player turnover, as the new incumbent seeks to reshape his squad in accordance with his own preferences. A Wald test of $H_0{:}\delta_{i,m}=0$ for i=1,2 and m=1...56 in Model C2 yields $\chi^2(112)=252.7$ (p-value=.0000), suggesting that choices of personnel and tactics made by managers do have a highly significant effect on the incidence of disciplinary sanction.

Individual referee effects

Inconsistency in the standards applied by different referees is among the most frequent causes of complaint from football managers, players, supporters and media pundits. Table 5 summarises the average numbers of disciplinary points per match awarded against the home and away teams and against both teams combined, by each of the 28 referees who officiated at least 30 Premier League matches during the observation period. (The data for a further nine referees who each officiated fewer than 30 matches are excluded.) There appears to be considerable variation between the propensities for individual referees to take disciplinary action. For example, the most lenient referee (Keith Burge) averaged 2.526 disciplinary points per match over 57 matches, and the most prolific (Mike Reed) averaged 4.541 points over 85 matches.

Does this degree of variation in the incidence of disciplinary sanction per referee constitute statistical evidence of inconsistency in refereeing standards? H2, the *refereeing consistency hypothesis*, imposes zero restrictions on the coefficients on the individual referee dummy variables $DR_{r,j}$, which identify matches officiated by the 28 referees listed in Table 5. In Model C2, a Wald test of $H_0{:}\gamma_{i,r}=0$ for i=1,2 and r=1...28 yields $\chi^2(56)=174.0$ (p-value=.0000). Therefore H2 is rejected, suggesting there was significant variation in standards between referees. Since the conditional model includes controls for team quality and other potential influences on the incidence of disciplinary sanction, the rejection of H2 should not be attributable to any non-randomness in the assignment of referees to matches: for example, the tendency for referees with a reputation for toughness to be assigned to matches at which disciplinary issues are anticipated by the authorities.

The earlier rejection of H1, the *home advantage hypothesis*, suggests there is a bias favouring the home team in the incidence of disciplinary sanction, even after controlling for home advantage in match results. With H2 also having been rejected, it is relevant to examine whether there are significant differences between referees in the degree of home team bias. In other words, do variations in the degree of home team bias on the part of different officials contribute to the observed pattern of refereeing inconsistency? H3, the *consistent home team bias hypothesis*, imposes the restriction that the corresponding coefficients on the individual referee dummy variables in the home and away team equations are the same. H3 would imply that the rate at which away teams tend to incur more disciplinary points than home teams does not vary between referees. In Model C2, a Wald test of $H_0:\gamma_{1,r}=\gamma_{2,r}$ for r=1...28 yields $\chi^2(28)=53.29$ (p-value=.0027). Therefore H3 is rejected at a significance level of 0.01 (although not at a significance level of 0.001). This constitutes evidence that there are significant differences between referees in terms of the degree of home team bias.

Season effects

The individual football season dummy variables $DS_{s,j}$ are included in the conditional model primarily as a control for changes over time in the content and interpretation of the rules relating to the award of yellow and red cards. The key changes during the observation period are detailed in Table 6. Most of the changes have increased the range of offences that are subject to disciplinary sanction, although there has occasionally been movement in the opposite direction.

Table 7 reports the average numbers of yellow and red cards awarded against the home and away teams per match by season, and the average disciplinary points compiled on the same basis. There appears to be little or no trend in the overall incidence of disciplinary sanction, despite the increase in the range of sanctionable offences. Two possible explanations are as follows. First, when there is an addition to the list of sanctionable offences, players may tend to modify their behaviour so that the numbers of cautions and dismissals remain approximately constant (Witt, 2005). Second, referees may tend to modify their interpretation of the boundaries separating non-sanctionable from sanctionable offences, and those separating cautionable from dismissable offences, so as to maintain an approximately constant rate of disciplinary sanction.

The directive issued at the start of the 1998-9 season making the tackle from behind punishable by automatic dismissal is the only rule change that appears to have had a discernible impact on the data reported in Table 7. The mean incidence of disciplinary sanction is higher for 1998-9 than for any of the other six seasons in the observation period. Within the 1998-9 season as well, the process of adjustment to the new disciplinary regime is visible in the data: during the first three months of this season the average disciplinary points incurred by both teams per match was 4.336, while the average for the rest of the season was 3.883 (see also Witt, 2005). In subsequent seasons, although this directive remained in force, the incidence of disciplinary sanction returned to levels similar to those experienced before the directive came into effect.

In order to test H4, the *time consistency hypothesis* that the average incidence of disciplinary sanction is stable over time, the null hypothesis (expressed in terms of the coefficients of the conditional model) is $H_0:\beta_{i,s}=0$ for i=1,2 and s=1997-8 to 2002-3 (inclusive). A Wald test yields $\chi^2(12)=37.95$ (p-value=.0002), suggesting there was significant season-to-season variation in the incidence of disciplinary sanction. However, if the zero restrictions on the coefficients for 1998-9 are excluded from the null ($H_0:\beta_{i,s}=0$ for i=1,2 and s=1997-8 and 1999-2000 to 2002-3 inclusive), the Wald test yields $\chi^2(10)=13.53$ (p-value=.1963). This suggests that with the (temporary) exception of the 1998-9 season, there was no other significant season-to-season variation in the incidence of disciplinary sanction. Therefore H4 receives qualified support from the estimation results.

Match attendance and live TV broadcast

Under H5, the *audience neutrality hypothesis*, the incidence of disciplinary sanction is unaffected by the crowd inside the stadium, and is also the same notwithstanding whether the match is being broadcast live on TV. To control for crowd effects, the covariate att_j, defined as the reported attendance at match j, is included in the regressions for $\ln(\lambda_{i,j})$. If H5 is not supported in respect of the stadium audience, more than one prior concerning the direction of any effect is possible. A large attendance might be expected to add to the intensity or excitement of the occasion, resulting in more determined play by either or both teams, and therefore an increased incidence of disciplinary sanction. Alternatively, a large attendance, presumably dominated by supporters of the home team, might put pressure on the referee to treat disciplinary transgressions by the home team more leniently, and those by the away team more severely. This would produce coefficients on att_j in the home and away team equations opposite in sign.

In Model C2, the coefficient on att_j in (5) is positively signed and significant at the 0.01 level. The equivalent coefficient in (6) is also positively signed, but insignificant. With respect to the stadium audience H5 is rejected, but there is no evidence of any tendency for referees to treat the home team more leniently when the crowd size is larger; if anything, the opposite seems to apply. This might suggest a tendency for the home team's level of determination to increase with the size of the home crowd.

During the entire observation period, the satellite broadcaster BSkyB held the Premier League's live TV broadcasting rights. These rights permitted BSkyB to screen a specified number of matches per season: this number varied from 60 matches at start of the observation period to around 100 by the end (the total number of Premier League matches per season is 380). If H5 is not supported, a tendency for players or referees to 'play to the camera' might be discernible in a different incidence of disciplinary sanction between televised and non-televised matches. In the estimations, the coefficients on sky_j in both the home team and away team equations are positively signed, but neither is significant. Therefore in respect of the live TV audience, H5 is supported. There is no evidence that the behaviour of players or referees is affected when the match is broadcast live on TV.

7. Conclusion

In this paper, we have reported estimations for the unconditional and conditional expectations of the incidence of disciplinary sanction against footballers in English Premier League matches. By providing a comprehensive statistical analysis of patterns in the award of yellow and red cards over a seven-year period, this study aims to provide the football authorities and other interested parties with a firmer factual basis than has been available previously for policy decisions and debate concerning the interpretation and implementation by referees of the rules governing disciplinary sanction in professional football.

In the estimations of the conditional expectations of the numbers of disciplinary points incurred by the home and away teams, it is found that relative team strengths matter: underdogs tend to incur a higher rate of disciplinary sanction than favourites. The incidence of disciplinary sanction tends to be higher in matches between evenly balanced teams. It also tends to be higher in matches with end-of-season outcomes at stake and in matches that attract high attendances. Home teams appear to play more aggressively in front of larger crowds, but perhaps surprisingly the crowd size does not influence the incidence of disciplinary sanction against the away team. There is no evidence that the behaviour of players or referees is any different in matches that are broadcast live on TV.

Despite an increase over time in the number of offences subject to disciplinary sanction, there was no consistent time-trend in the yellow and red cards data: players and officials appear to have adjusted to changes in the rules so that in the long run the rate of disciplinary sanction remained approximately constant. However, individual referee effects make a significant contribution to the explanatory power of the conditional model, indicating that there are inconsistencies between referees in the interpretation or application of the rules. An obvious but important policy implication for the football authorities is that measures need to be implemented in order to improve refereeing consistency.

The empirical analysis suggests that the tendency for away teams to incur more disciplinary points than home teams cannot be explained solely by the home advantage effect on match results. Even after controlling for team quality, a (relatively strong) away team can expect to collect more disciplinary points than a (relatively weak) home team with the same win probability. Therefore the statistical evidence seems to point in the direction of a home team bias in the incidence of disciplinary sanction. This interpretation is consistent with evidence of home team bias in several other recent studies, which find that the home team is favoured in the calling of fouls, or in the addition of stoppage time at the end of matches. Finally, evidence is found of variation between referees in the degree of home team bias; and this variation contributes to the overall pattern of refereeing inconsistency. These findings suggest that while all referees should be counselled and encouraged to avoid (presumably unintentional) home team bias in their decision-making, the extent to which corrective action is required is also likely to vary between officials.

References

Allen, W.D. (2002). Crime, punishment and recidivism: lessons from the National Hockey League. *Journal of Sports Economics* 3, 39-60.

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.

Deloitte and Touche (2004). *Annual Review of Football Finance*. Manchester: Deloitte.

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 Pope, P.F. (2004). The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting* 20, 697-711.

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

Forrest, D., Goddard, J. and Simmons, R. (2005). Odds setters as forecasters: the case of English football. *International Journal of Forecasting*, forthcoming.

Garicano, L. and Palacios-Huerta, I. (2000). An empirical examination of multidimensional effort in tournaments. Graduate School of Business, University of Chicago, mimeo.

Garicano, L., Palacios-Huerta, I. and Prendergast, C. (2001). *Favoritism under social pressure*. NBER Working Paper 8376.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting* (forthcoming).

Goddard, J. and Asimakopoulos, I. (2004). Forecasting football match results and the efficiency of fixed-odds betting. *Journal of Forecasting* 23, 51-66.

Heckelman, J.C. and Yates, A.C. (2002). And a hockey game broke out: crime and punishment in the NHL. *Economic Inquiry* 41, 705-712.

Karlis, D. and Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. *The Statistician* 52, 381-393.

Lee, A. (1999). Modelling rugby league data via bivariate negative binomial regression. *Australian and New Zealand Journal of Statistics* 41, 153-171.

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

McCormick, R.E. and Tollison, R.D. (1984). Crime on the court. *Journal of Political Economy* 92, 223-235.

Nevill, A.M., Balmer, N.J. and Williams, A.M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise* 3, 261-272.

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

Sutter, M. and Kocher, M.G. (2004). Favoritism of agents – the case of referees' home bias. *Journal of Economic Psychology* 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 anticipated sanction changes? Evidence from the English Premier League. *Scottish Journal of Political Economy*, forthcoming.

Tables

Table 1 Observed numbers of yellow cards incurred by the home and away teams, English Premier League, seasons 1996-7 to 2002-3

	Away team								
Home team	0	1	2	3	4	5	6	7	Total
0	189	254	158	86	35	9	1	0	732
1	110	260	264	147	66	23	6	1	877
2	64	162	158	126	47	25	6	1	589
3	18	77	96	72	39	14	3	4	323
4	3	13	29	32	16	8	2	0	103
5	1	3	12	11	2	1	0	1	31
6	0	0	1	2	1	1	0	0	5
Total	385	769	718	476	206	81	18	7	2660

Source: **www.premierleague.com**

Table 2 Observed numbers of red cards incurred by the home and away teams, English Premier League, seasons 1996-7 to 2002-3

	Away team			
Home team	0	1	2	Total
0	2258	231	8	2497
1	119	34	3	156
2	2	3	0	5
3	2	0	0	2
Total	2381	268	11	2660

Source: **www.premierleague.com**

Table 3 Unconditional and conditional models: estimation results

	Model U1 univariate Poisson	Model U2 bivariate Poisson	Model U3 univariate neg. bin.	Model U4 bivariate neg. bin.	Model C2 bivariate Poisson	Model C4 bivariate. neg. bin.
Ancillary parameters						
η	-	0.2655	-	-	.2712	-
		(15.99)			(14.37)	
κ_1	-	-	.1231	.1259	-	.0286
			(5.24)	(5.34)		(1.47)
κ_2	-	-	.0523	.0548	-	.0027
			(3.45)	(3.59)		(0.20)
φ	-	-	-	-1.8915	-	-1.79
				(-14.70)		(-12.90)
Maximised log-likelihood function:						
ln(L)	-8866.3	-8754.0	-8841.2	-8727.9	-8441.3	-8440.3
Ψ	.0357	.0372	.0360	.0376	.0419	.0419
LR test for H_0:κ_1=κ_2 in Models U4 and C4 (bivariate negative binomial vs. bivariate Poisson):						
$\chi^2(2)$	-	-	50.2	52.2	-	2.0
p-value	-	-	.0000	.0000	-	.3734
Chi-square goodness of fit test (see also Table 4):						
$\chi^2(25)$	367.6	102.0	271.0	44.6	-	-
p-value	.0000	.0000	.0000	.0092	-	-

Note: z-statistics for the significance of ancillary parameters are reported in parentheses.

Ψ = Pseudo-likelihood statistic = exp[(1/N)ln(L)].

Table 4 Observed and expected numbers of occurrences of combinations of disciplinary points for home and away teams

	Away team												**Home total**	
	0		**1**		**2**		**3**		**4**		**5+**			
Home team ↓	O	E	O	E	O	E	O	E	O	E	O	E	O	E
		79.4		162.5		166.2		113.3		58.0		34.9		614.3
0	**182**	125.7	**235**	199.4	**141**	158.2	**84**	83.7	**45**	33.2	**18**	14.1	**705**	614.3
		98.9		182.9		177.9		121.1		64.7		45.7		691.0
		163.1		235.8		160.4		78.5		33.8		21.0		692.4
		116.4		238.1		243.6		166.1		85.0		51.2		900.3
1	**104**	126.5	**244**	258.4	**238**	250.8	**137**	156.9	**66**	71.8	**49**	35.9	**838**	900.3
		122.8		227.2		221.0		150.4		80.4		56.7		858.5
		130.0		247.5		229.4		139.3		67.1		44.2		857.5
		85.3		174.5		178.5		121.7		62.3		37.5		659.8
2	**65**	63.7	**150**	159.1	**140**	185.5	**121**	136.5	**51**	72.2	**57**	42.8	**584**	659.8
		85.7		158.4		154.1		104.9		56.0		39.6		598.7
		57.3		133.6		162.9		124.8		69.3		49.5		597.5
		41.7		85.2		87.2		59.5		30.4		18.3		322.3
3	**17**	21.4	**71**	63.1	**92**	86.6	**73**	74.2	**48**	45.1	**34**	31.9	**335**	322.3
		44.2		81.7		79.5		54.1		28.9		20.4		308.7
		22.0		56.3		80.1		71.7		44.4		34.0		308.5
		15.3		31.2		31.9		21.8		11.1		6.7		118.1
4	**4**	5.4	**18**	8.3	**35**	29.0	**42**	28.6	**18**	19.9	**16**	16.9	**133**	118.1
		18.8		34.7		33.8		23.0		12.3		8.7		131.1
		8.1		21.5		32.8		31.6		20.8		16.6		131.4
		5.8		12.0		12.2		8.3		4.3		2.6		45.2
5+	**2**	1.3	**7**	5.1	**17**	9.4	**20**	10.9	**10**	8.8	**9**	9.6	**65**	45.2
		10.3		19.0		18.5		12.6		6.7		4.7		71.9
		4.2		11.2		17.6		17.6		12.0		9.7		72.6
Away total		343.8		703.5		719.6		490.8		251.0		151.3		2660
	374	343.8	**725**	703.5	**663**	719.6	**477**	490.8	**238**	251.0	**183**	151.3	**2660**	2660
		380.6		703.9		684.7		465.9		249.0		176.0		2660
		384.6		706.0		683.3		463.5		247.4		175.2		2660

Note: O is the number of matches (out of 2660) in which each combination of disciplinary points for home and away teams was observed (figures in **bold**). E is the expected number of occurrences for each combination of disciplinary points, generated from Models U1, U2, U3 and U4 respectively (figures displayed vertically).

Table 5 Average total disciplinary points awarded per match, by referee

	Referee	Matches	Disciplinary points awarded				Referee	Matches	Disciplinary points awarded		
			Home team	Away team	Total				Home team	Away team	Total
1	Reed	85	1.788	2.753	4.541	15	Bennett	68	1.603	1.853	3.456
2	Willard	60	1.900	2.350	4.250	16	Barry	117	1.385	2.060	3.444
3	Barber	147	1.728	2.463	4.190	17	Jones	112	1.411	1.991	3.402
4	Riley	131	1.626	2.511	4.137	18	Ashby	33	1.212	2.152	3.364
5	Harris	52	1.750	2.327	4.077	19	Wilkie	81	1.358	1.975	3.333
6	Knight	41	1.829	2.171	4.000	20	Dunn	136	1.368	1.956	3.324
7	Styles	56	1.929	2.018	3.946	21	Elleray	129	1.295	1.984	3.279
8	Rennie	94	1.819	2.096	3.915	22	Winter	143	1.231	1.979	3.210
9	Dean	54	1.685	2.111	3.796	23	Gallagher	122	1.262	1.918	3.180
10	Wilkes	30	1.400	2.333	3.733	24	Halsey	74	1.338	1.730	3.068
11	D'urso	85	1.624	2.094	3.718	25	Alcock	78	1.000	2.026	3.026
12	Poll	160	1.619	2.069	3.688	26	Wiley	90	1.433	1.578	3.011
13	Bodenham	44	1.455	2.045	3.500	27	Durkin	145	1.248	1.469	2.717
14	Lodge	102	1.392	2.108	3.500	28	Burge	57	0.877	1.649	2.526

Source: **www.premierleague.com**

Note: Referees who officiated at fewer than 30 Premier League matches between the 1996-7 and 2002-3 seasons (inclusive) are not shown in Table 5.

Table 6 Rule changes and changes of interpretation, by season

SEASON	RULE CHANGES/CHANGES OF INTERPRETATION
1996-7	Referees are reminded to severely punish the tackle from behind.
1997-8	Failure to retreat the required distance at free kicks and delaying the restart of play are to be interpreted as yellow card offences.
1998-9	The tackle from behind which endangers the safety of an opponent is to be interpreted as a red card offence. The red card offence of denying an opponent a goal scoring opportunity is changed to denying an opposing team a goal scoring opportunity (widening the scope of this offence).
1999-2000	Simulation (diving, feigning injury or pretending that an offence has been committed) is to be punishable with a yellow card. Referees are reminded to punish racist remarks with a red card. Swearing is also an offence warranting a red card.
2000-1	Offensive gestures are to be punishable with a red card.
2001-2	Some relaxation of the rule requiring referees to issue a yellow card if a player celebrates a goal by removing his shirt. However, celebrations that are provocative, inciting, ridiculing of opponents or spectators or time wasting remain punishable with a yellow card. Referees are reminded to punish intentional holding or pulling offences with a yellow card.
2002-3	Referees are reminded to be strict in punishing simulation and the delaying of restarts, especially if players remove shirts for any length of time celebrating a goal.

Source: *Rothmans Football Yearbook* (various editions).

Table 7 Average numbers of yellow and red cards and total disciplinary points awarded per match, by season

Season	Home team			Away team			Both teams		
	Yellow	Red	Total	Yellow	Red	Total	Yellow	Red	Total
1996-7	1.305	0.026	1.350	1.808	0.084	1.934	3.113	0.111	3.284
1997-8	1.303	0.058	1.405	2.016	0.124	2.189	3.318	0.182	3.595
1998-9	1.582	0.074	1.695	2.147	0.116	2.316	3.729	0.189	4.011
1999-2000	1.411	0.055	1.497	1.932	0.129	2.118	3.342	0.184	3.616
2000-1	1.355	0.084	1.487	1.800	0.084	1.921	3.155	0.168	3.408
2001-2	1.247	0.084	1.389	1.803	0.103	1.955	3.050	0.187	3.345
2002-3	1.326	0.071	1.432	1.703	0.124	1.882	3.029	0.195	3.313

Source: **www.premierleague.com**

Figure 1 Nash equilibria with x_j = 0, 0.5, 1